THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Alexandre GAILLARD

Le 7 juin 2022

Essays in Macroeconomics with Heterogeneous Agents

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
TSE-R - Toulouse School of Economics - Recherche

Thèse dirigée par
Patrick FEVE

Jury

M. Sang Yoon (Tim) LEE, Rapporteur
Mme Axelle FERRIèRE, Rapporteure
M. Christian HELLWIG, Examinateur
M. Patrick FEVE, Président

Essays in Macroeconomics with Heterogeneous Agents

PhD. Thesis

Alexandre Gaillard[1]
Toulouse School of Economics

July 26, 2022

[1]Alexandre Gaillard. Address: 1, Esplanade de l'Université, 31080 Toulouse, France; email: alexandre.gaillard@tse-fr.eu.

Acknowledgment

First of all, I want to thank Patrick Fv̀e for accepting to supervise my thesis. His advice has been invaluable and I will fondly remember our many interactions, about research or life in general. I wish to thank Christian Hellwig for the inestimable help he provided and the time he took to read and discuss my research. His guidance has been enlightening, especially when I had to write my first papers. I also thank Nicolas Werquin, Tim Lee, Fabrice Collard, Martial Dupaigne, Eugenia Gonzalez-Aguado, Tiziana Assenza and Nour Meddahi for all their support, especially during the job market. All of them made my life very easy and interesting at TSE.

In particular, I am very grateful to Sumudu Kankanamge. Not only is he a great co-author and friend but I do not think anybody else could have found better words to help me put things into perspective over the last few years. Together with Tim Lee, they have sparked my interest in studying inequality through the lens of modern tools and theories. I am particularly grateful to François de Soyres with whom I collaborated on a captivating research agenda in international trade. I am very fortunate for being one of his co-authors. No words can describe how prodigious and excellent Christophe Bruneel and Philipp Wangner are as coauthors, colleagues, and friends. I will never forget, nor regret a single moment of my years in Toulouse thanks to them.

I am also grateful to Miguel Zerecerro, Stefan Pollinger, Jonas Gathen, Tuuli Vanhapelto, Thierry Magnac, Friedrich Lucke, Fernando Stipanicic, various speakers at the TSE Macro seminar, participants to various workshop and conferences, in particular participants to the TSE Macro group, as well as many fellow PhD students in Toulouse for making this journey joyful while discussing and questioning my work along the way.

Last but particularly important, I have no words to thank Adeline Audrerie who has been the best partner, especially during the most important moments of my PhD. I also deeply thank Lola Ballard, Eric Bouiller, and many friends, who pushed me so much during my PhD and with whom I had lived incredible moments outside research. I finally warmly thank my parents Myriam and Bernard Gaillard and my brothers Cyril and Yann. It is entirely thanks to their curiosity for politics and sciences that I got interested in economics in the first place. I dedicate this thesis to all of them because they helped me so much in finding the right balance between my research and the many other important things.

Summary

This thesis is divided into four essays on the broad implication of heterogeneity in macroeconomics. It puts a special emphasis on (i) the relation between the sources of wealth inequality and the effects of wealth redistribution, (ii) the effects of unemployment insurance on the selection into self-employed, (iii) the role of the market for businesses in transmitting the intangible value of firms across generations of entrepreneurs, and (iv) the role of international trade linkages with many interconnected countries in generating cross-country GDP correlations.

In my first chapter, together with Philipp Wangner, we show that the aggregate and welfare implications of redistributing wealth depends on the underlying forces behind capital investment and wealth inequality. In the data, rich households invest a higher fraction of their wealth into risky assets. This is the result of two distinct channels. Wealthy households may invest differently due to heterogeneity in specific skill or risk tolerance (type-dependence) or because wealth itself induces wealthy households to undertake riskier investments (scale-dependence). We first clarify the role of type and scale dependence and we argue that a number of existing frameworks studying the macro consequences of micro investment heterogeneity rest on a particular combination of type and scale dependence. Second, we show that their distinction is crucial for assessing the effects of wealth taxation. In an incomplete markets quantitative model calibrated to the US using micro datasets, both channels are found to lead to opposite predictions regarding wealth taxation. Under type-dependence, rich individuals with low capital returns dissave faster than those with high capital returns. By taxing the stock of wealth of the richest households, *only the fittest survive at the top* which reinforces the selection of agents with high investment skills among the rich. When returns to wealth reflect capital productivity, taxing wealth at a high rate is optimal because it raises productivity. Under scale-dependence, a wealth tax reduces productive investments, such that subsidizing wealth becomes optimal. In a benchmark model calibrated to take into account both channels, it is optimal to tax wealth at 0.8% above an exemption threshold of $550K with little effects on overall productivity. This tax rate remains robust when returns reflect rents instead of productivity. However, in such a case, the optimal wealth tax increases under scale dependence but decreases under type dependence such that both effects on the size of extracted rents almost cancel each other in the benchmark economy.

In the second chapter, Sumudu Kankanamge and I study how entrepreneurship contributes to the micro and macro-level patterns of gross labor market flows, and explains why the selection into that occupation is highly responsive to unemployment insurance (UI) changes. Our framework merges search models in the spirit of Mortensen and Pissarides (1994) with an occupational choice models with entrepreneurship along the lines of Quadrini (2000a) and Cagetti and De Nardi (2006a). We show that our model is able to replicate key facts regarding occupational flows at the aggregate level and along the ability and wealth distributions. Higher UI is associated with

strong disincentive to start businesses out of unemployment and employment, consistent with CPS data. Intuitively, higher UI generosity changes the riskiness of self-employment relative to paid-employment. In turn, this has important aggregate implications on occupational masses. Surprisingly, we find that an increase in UI generosity leads to an increases in the unemployment and employment rates, but substantially decreases the self-employment rate.

In the third chapter, Sumudu Kankanamge and I study the role of the market for small and medium-sized enterprises (SME) for the transmission of the intangible value of businesses within and across generations of entrepreneurs. In the data, a large fraction of entrepreneurs enter this occupation by acquiring an existing business instead of founding a new one. In contrast, the share of inherited businesses accounts only for a small fraction of business transfers. Using a new dataset, we also document that the market for SME is subject to important selling frictions. Motivated by these facts, a large-scale life-cycle model with entrepreneurs is developed to understand and quantify the role of the market for SME. A key attribute of the model is that newly established businesses are, on average, less productive and face higher failure risks than older ones, consistent with the selection of the best firms over time. The market for SME lets those businesses be transferred across individuals. Entrepreneurs want to sell their businesses upon retirement or because of exogenous decisions to exit. Younger potential entrepreneurs would like to buy an existing business, but selling frictions and borrowing constraints prevent them from doing so. Shutting the market down leads to a substantial drop in aggregate productivity and output, and alters the pool of firms, incentives to enter and exit, and the wealth distribution.

In the fourth chapter, François de Soyres and I study the role of international trade linkages in increasing GDP synchronization. Empirically, trade linkages are associated with higher cross-country GDP correlations. However, international real business cycle models fail to generate the magnitude of this relationship. This puzzle is known as the Trade Comovement Puzzle (TCP) since Kose and Yi (2006). We argue that the way GDP is measured by statistical agencies, using double deflation, may largely account for this puzzle. When base period prices are used in real GDP construction, any distortion between the price of imported input and their marginal revenue product generates a link between input usage and movements in real GDP, which in turn increases the strength of propagation of shocks across countries. Focusing on two common sources of such a distortion, markups and love of variety, we construct a many country international business cycle model with imperfect competition and an extensive margin of imported goods. The model is shown to quantitatively replicate the strong relationship between trade linkages and cross country GDP correlation when GDP is measured using the same method as in statistical agencies. Each component, markups and love of variety, accounts for a significant fraction of the relationship. The results highlight that, when comparing a macroeconomic model to the data, it is key to define aggregate variables in a way that is consistent with statistical agencies' procedures.

Résumé

Cette thèse est divisée en quatre essais sur le rôle de l'hétérogénéité en macroéconomie. Spécifiquement, j'étudie comment : (i) les effets de la taxation de la richesse dépendent des sources de l'inégalité de richesse, (ii) l'assurance chômage affecte la sélection dans l'entrepreneuriat, (iii) le marché des entreprises permet de transmettre la valeur intangible des entreprises entre les générations et les entrepreneurs, et (iv) le commerce international entre une multitude de pays participe au co-mouvement des PIBs entre ces pays.

Dans mon premier chapitre, nous montrons avec mon coauteur Philipp Wangner que les implications de la redistribution de la richesse sur les agrégats macroéconomique et le bien-être dépendent des forces sous-jacentes aux décisions d'investissement et à l'inégalité de richesse. Dans les données, les ménages riches investissent une plus grande proportion de leur richesse dans des actifs risqués. Cette observation est le résultat de deux mécanismes. Les ménages riches peuvent être hétérogène en termes de compétence et de tolérance au risque (dépendance au type) ou parce que la richesse elle-même conduit à des comportements plus risqués (dépendance à la richesse). Premièrement, nous clarifions le rôle de la dépendance au type et à la richesse et montrons qu'une classe de modèle utilisé pour étudier les conséquences de l'hétérogeneité micro de l'investissement s'appuient sur une combinaison particulière de dépendance au type et à la richesse. Deuxièmement, nous montrons que leur distinction est cruciale pour évaluer les effets de la taxation sur la richesse. A l'aide d'un modèle quantitative en marchés incomplets calibrés sur des données microéconomiques aux USA, nous trouvons que les deux dépendances génèrent des résultats opposés sur l'effet de la taxation de la richesse. Dans un modèle de type-dependence, les individus riches avec des retours sur l'investissement faibles desépargnent plus rapidement que ceux qui obtiennent des retours sur l'investissement plus élevés. Par conséquent, en taxant le stock de richesse des plus riches, *seuls les plus riches dans l'économie survivent au top* de la distribution de la richesse, renforçant la sélection des agents avec des compétences d'investisseurs parmi les plus riches. Lorsque les retours sur l'investissement reflètent la productivité, taxer la richesse à un taux plus élevé est optimal parce qu'il accroît la productivité. Dans un modèle de dépendance à la richesse, une taxation sur la richesse réduit l'investissement productif, tel que subventionner la richesse devient optimal. Dans un modèle calibré pour prendre en compte les deux mécanismes, taxer positivement la richesse à 0.8% au delà d'un seuil d'exonération de 550K% est optimal avec peu d'effets sur la productivité. Ce résultat est robuste si les retours élevés sur l'investissement représentent des rentes plutôt qu'une différence de productivité. Dans ce cas, l'argument inverse se produit. Il est optimal d'accroître la taxation sur la richesses lorsqu'il y a dépendance à la richesse, mais de décroître la taxe sur la richesse lorsqu'il y a dépendance au type, tel que les deux effets sur l'importance des rentes dans l'économie se neutralisent.

Dans le second chapitre, nous étudions avec mon coauteur Sumudu Kankanamge comment

l'entrepreneuriat contribuent aux flux du marché du travail au niveau macro et micro. Nous expliquons pourquoi la sélection dans ce type d'occupation est fortement sensible aux variations de l'assurance chômage. Notre cadre théorique combine les modèles de *search* dans l'esprit de Mortensen and Pissarides (1994) avec les modèles d'entrepreneuriat dans la lignée de Quadrini (2000a) et Cagetti and De Nardi (2006a). Nous montrons que notre modèle réplique les faits stylisés principaux de flux entre les différentes occupations au niveau agrégé mais aussi le long de la distribution des revenus et de la richesse. Une assurance chômage plus élevée est associée avec un effet dissuasif fort sur la création des entrepreneurs des personnes au chômage et dans l'emploi, confirmant nos résultats empiriques obtenus dans les CPS. Intuitivement, une assurance chômage plus élevée change le risque de l'occupation "entrepreneur" relativement à l'occupation "en emploi". Cet effet a des conséquences agrégées sur les masses d'occupation dans l'économie. De façon surprenante, nous trouvons qu'un accroissement de l'assurance chômage accroît le chômage et l'emploi dans l'économie mais diminue le nombre de personnes dans l'entrepreneuriat.

Dans le troisième chapitre, nous étudions avec mon coauteur Sumudu Kankanamge le role du marché des petites et moyennes entreprises (PME) dans la transmission de la valeur intangible des entreprises au sein et entre les générations d'entrepreneurs. Dans les données, une large proportion des entrepreneurs ont commencé leur entreprise en acquérant une entreprise existante au lieu d'en créer une nouvelle. Au contraire, la fraction d'entreprises cédées ou héritées ne compte que pour une faible proportion des acquisitions. En utilisant de nouvelles données, nous documentons également que le marché des PME est sujet à d'importantes frictions de vente. Motivé par ces faits, nous construisons un modèle à grande échelle de cycle de vie avec des entrepreneurs pour comprendre et quantifier le role du marché des PMEs. Un élément clé du modèle est que les nouvelles entreprises établies sont, en moyenne, moins productives et font faces à un risque de faillite plus élevé que les entreprises plus agées, en cohérence avec un effet de sélection des meilleures entreprises au cours du temps. Le marché des PMEs permet la transmission de ces entreprises entre les individus. Les entrepreneurs veulent vendre leur entreprise soit à cause d'un choc exogène ou parce qu'ils souhaitent partir à la retraite. Les jeunes entrepreneurs voudraient acheter une entreprise existante mais des frictions de vente et financière les empêchent de le faire. Nous trouvons que la fermeture du marché des PMEs génère une diminution de la productivité agrégée et de la production, altère la distribution des entreprises, les incitations d'entrer et de sortir de l'entrepreneuriat et la distribution de la richesse.

Dans un quatrième chapitre, nous étudions avec François de Soyres le lien entre les échanges internationaux de biens et le co-mouvement entre les pays. Empiriquement, deux pays qui échangent davantage de biens sont, en moyenne, davantage synchronisés. Cependant, les modèles internationaux de cycles d'affaires réels ne permettent pas de générer la magnitude de cette relation empirique. Ce résultat est à l'origine du Trade Comovement Puzzle (TCP) introduit par Kose and Yi (2006). Nous montrons que la façon dont le PIB est mesuré par les agences statistiques, en utilisant la méthode de double déflation, peut largement résoudre ce puzzle. Lorsque les prix constants sont utilisés dans la construction de PIB réel, une distorsion entre le prix des biens importés et le revenu marginal produit par leur utilisation génère un lien mécanique entre l'usage d'inputs importés et les mouvements de PIB, augmentant l'importance de la propagation des chocs entre pays. En se focalisant sur deux sources de distorsion, nous construisons un modèle quantitatif de cycles d'affaires réels internationaux avec une multitude de pays en compétition imparfaite et qui introduit une marge extensive des variétés de biens importés. Le modèle réplique la relation entre les échanges de biens et la correlation des PIBs lorsque ceux-ci sont mesurés en utilisant la même méthode que les agences statistiques. Chaque component, la présence de markups et la préférence pour la variété, comptent pour une fraction importante de la relation. Le résultat montre qu'il est crucial de bien définir les agrégats au sein d'un modèle macroéconomique de la même façon que ce qui est mesuré par les agences statistiques.

Contents

1 Wealth, Returns, and Taxation: A Tale of Two Dependencies **14**
1.1 Introduction . . . 15
1.2 An Analytical Two-Period Model . . . 21
1.2.1 Environment . . . 21
1.2.2 Efficiency Gains and Redistribution . . . 23
1.2.3 From Efficiency to Welfare . . . 31
1.2.4 Generalization . . . 32
1.3 A Dynamic Quantitative Model with Investment Heterogeneity . . . 34
1.3.1 Environment . . . 34
1.3.2 Equilibrium . . . 37
1.4 Taking the Model to the Data . . . 39
1.4.1 Capital Heterogeneity: A First Glance into the Data . . . 39
1.4.2 Functional Forms and Calibration . . . 43
1.5 Properties of the Model . . . 50
1.5.1 Wealth Inequality and Returns to Wealth . . . 51
1.5.2 Wealth Inequality – Output Relationship . . . 53
1.6 Wealth Taxation: the Role of Type and Scale Dependence . . . 55
1.6.1 Results . . . 57
1.6.2 Dissecting the Effects from Type and Scale Dependence . . . 59
1.6.3 The Interaction between Type and Scale Dependence and Rents . . . 62
1.6.4 Further Robustness . . . 64
1.7 Conclusion . . . 64

Appendices **66**
1.A Theoretical appendix . . . 66
1.A.1 Proofs for Section 1.2 . . . 66
1.A.2 Fit of the Static Model under pure type/scale dependence . . . 81
1.B Empirical Appendix . . . 82
1.B.1 Adjusted Survey of Consumer Finance . . . 82
1.B.2 Measuring Returns to Wealth in the PSID . . . 84
1.B.3 PSID participation and diversification . . . 89

1.C Quantitative Appendix 90
1.C.1 Calibration 90
1.C.2 Computational appendix 92
1.C.3 Wealth Mobility 94

2 Gross Labor Market Flows, Entrepreneurship, and the Role of Unemployment Insurance 95
2.1 Introduction 96
2.2 Labor Market Flows and Entrepreneurship: the Data 100
2.2.1 Labor Market Mobility and Entrepreneurship 101
2.2.2 Responsiveness to UI Generosity: Empirical Analysis 102
2.3 A Model of Gross Labor Market Flows with Entrepreneurship 107
2.3.1 Households 107
2.3.2 Corporate sector 112
2.3.3 Government 112
2.3.4 Equilibrium 112
2.3.5 Parameterization 113
2.4 Gross Flows: Aggregate Characteristics and Micro Level Behaviors 115
2.4.1 Optimal search efforts 116
2.4.2 Resulting gross flows 117
2.4.3 Additional validation 121
2.5 Impact of UI on Gross Flows: Quantitative Evaluation 122
2.5.1 A sample of counterfactual experiments 122
2.5.2 Effects of UI generosity on Occupational Choices 122
2.5.3 Long Run Occupational Masses 127
2.5.4 UI Generosity and Long Run Aggregate Outcomes 128
2.6 Model Implications and Robustness 129
2.6.1 Policy Experiment: *on-the-business* Unemployment Insurance 129
2.6.2 UI extensions during the Great Recession 131
2.6.3 Robustness 131
2.7 Conclusion 133

3 Buying and Selling Entrepreneurial Assets 134
3.1 Introduction 135
3.2 Business Transfers and Dynamics: Stylized Facts 139
3.2.1 Business Acquisition and Exit 139
3.2.2 Sources of Heterogeneity and Maturity of a Business 141
3.3 Model 145

3.3.1 Agents . . . 145
3.3.2 Dynamic Problem . . . 146
3.3.3 Corporate Sector . . . 149
3.3.4 The Small and Medium Sized Enterprises for Sale Market (SMESM) . . . 150
3.3.5 Demography and Bequest . . . 151
3.3.6 Government . . . 151
3.3.7 Equilibrium . . . 152
3.4 Parameterization . . . 152
3.4.1 Fixed Parameters . . . 153
3.4.2 Joint Parameterization . . . 156
3.5 Properties of the Baseline Model . . . 157
3.5.1 Model Validation . . . 157
3.5.2 The Decision to Enter Entrepreneurship . . . 158
3.5.3 The Decision to Exit Entrepreneurship . . . 160
3.6 Businesses for Sale Market: Quantitative Analysis . . . 161
3.6.1 Assessing the Importance of Business Transfers . . . 162
3.6.2 Decomposing Maturity Effects on the Businesses for Sale Market . . . 164
3.6.3 Businesses for Sale Market and Matching Efficiency . . . 165
3.6.4 Wealth Concentration and Inequality . . . 167
3.6.5 Sensitivity Analyses . . . 168
3.7 Conclusion . . . 169

Appendices **170**
3.A Empirical appendix . . . 170
3.A.1 Acquisition type . . . 170
3.A.2 Business performances and owner characteristics . . . 170
3.A.3 Exit rate and type of exit . . . 172
3.A.4 Small and medium-sized enterprises for sale market . . . 173
3.B Model: further details . . . 174
3.B.1 Numerical solution method: discrete-continuous model with the endogenous grid method (DC-EGM) under taste shocks . . . 174
3.B.2 Calibration details . . . 175
3.C Sensitivity Analyses . . . 175

4 Value-Added and Productivity Linkages Across Countries **177**
4.1 Introduction . . . 178
4.2 On the definition and measurement of real GDP . . . 181

4.2.1 A Simple Accounting Framework . . . 181
4.2.2 Productivity and technology . . . 186
4.3 A Model of International Trade with Cross-Border Input Linkages . . . 188
4.3.1 Consumption and Labor Supply . . . 188
4.3.2 Production side . . . 189
4.3.3 Equilibrium . . . 190
4.3.4 Real Value Added definitions . . . 192
4.3.5 Calibration of the Model Parameters . . . 194
4.3.6 Quantification of the Empirical Trade-Comovement Slope . . . 196
4.4 Results . . . 199
4.4.1 A Quantitative Resolution of the Trade Comovement Puzzle . . . 199
4.4.2 The Importance of Measurement . . . 202
4.5 Further Investigations . . . 202
4.5.1 Measured Productivity, Profits and Trade in the Data . . . 202
4.5.2 Measured productivity in the Model . . . 203
4.5.3 A Focus on the Role of Markups and Profits in the Model . . . 205
4.5.4 Extensive Margin fluctuations and Real GDP Comovement in the Data . . . 205
4.5.5 Robustness of the Quantitative Results. . . . 207
4.6 Conclusion . . . 208

Appendices **209**
4.A Empirical Appendix: Markup Measures . . . 209
4.B Theoretical and Quantitative appendix . . . 209

List of Figures

1.1 The inequality–efficiency diagram. . . . 28

1.2 The inequality–efficiency diagram. . . . 30

1.3 Average portfolio share of gross assets by wealth percentile. . . . 40

1.4 Average return on wealth by gross wealth quantile. . . . 42

1.5 Average share held in equity (top panel) and average number of private equity business investments (bottom panel), conditioning on investors. . . . 48

1.6 Consumption equivalent variation Δ^{CEV} across wealth distribution. . . . 58

1.7 Fraction of high investor type ($\vartheta = 1$) relative to the status quo per wealth decile. . . 61

1.8 CEV welfare as function of τ_a under type and wealth dependence. . . . 62

1.9 CEV welfare as function of τ_a under type and *wealth* dependence: sensitivity rent-extraction. . . . 63

1.B.1 Pareto shape estimation for the United States using the SCF. . . . 83

1.B.2 Portfolio composition in the PSID. . . . 84

1.B.3 Private equity (PE) investment participation and diversification in the PSID. . . . 90

1.B.4 *Scale* dependence in investment participation. . . . 91

1.B.5 *Scale* dependence private equity business investment. . . . 92

2.2.1 Average quarterly gross flow toward entrepreneurship among the unemployed individuals . . . 104

2.4.1 Implied optimal probability to find a job for unemployed individuals (top left) and entrepreneurs (top right). Implied optimal probability to find a business idea for unemployed individuals (bottom left) and workers (bottom right). . . . 117

2.4.2 Gross labor market flows by ability level ϑ (model in dashed red) and educational attainment (CPS data in straight black). . . . 119

2.5.1 UI generosity and model average flows from the insured (left panel) and uninsured (right panel) unemployed pools. . . . 124

2.5.2 UI generosity and model average flows between entrepreneurship and employment. 125

2.5.3 UI generosity and occupational masses. . . . 128

2.6.1 On-the-business UI and flow from insured unemployment. . . . 130

2.6.2 Percent deviation from the mean sample occupational mass for alternative specifications. . . . 132

3.2.1 Entrepreneurial life cycle, acquisition and business selling 141
3.2.2 Failure rate by acquisition type and for all businesses. 142
3.3.1 Timing. 146
3.5.1 Life-cycle pattern of entrepreneurship 159
3.5.2 Occupational choice and start-up capital as a function of wealth for new entrants. . . 160
3.5.3 Density of start-up capital by type of acquisition. 161
3.6.1 Entrepreneurial entry decision 164
3.6.2 The importance of the matching efficiency on the SME for sale market 166
3.A.1 Failure rate by acquisition type and for all businesses, employer group 171
3.A.2 Type of exit and establishment data 172
3.A.3 Age profile of businesses for sale 173
3.A.4 Probability to sell with listing price as a proxy for size. 174

4.4.1 Model-based association between RGDP correlation and Trade intensities. Left chart shows Intermediate Inputs trade, right chart shows Final goods trade. 200
4.4.2 Decomposition of the Input Trade-comovement slope using alternative model specifications. Left chart shows a model without Extensive Margin adjustments, right chart shows a model with neither extensive margin nor markups. 201

List of Tables

1.1 Returns to wealth in the PSID. 41
1.2 Calibrated parameters. 44
1.3 Life cycle earning profile, mortality and transition probabilities. 45
1.4 Wealth distribution in the data and models.[a] 51
1.5 Mean returns to wealth (in %) along the wealth distribution: data and model. 52
1.6 Wealth distribution under alternative model counterfactuals.[a] 53
1.7 Responses to a permanent 1% wealth tax levied on the top 1% wealthiest households. 54
1.8 Aggregate variables after optimal tax reform. 57
1.9 Aggregate variables after the optimal tax reform derived under the benchmark economy. 60
1.B.1 Scale dependence regression 89
1.C.1 Resulting scale dependence in risky portfolio share: benchmark model and data . . 91
1.C.2 Resulting scale dependence in the risky investment participation: benchmark model and data 91
1.C.3 Wealth mobility: data and model 94

2.2.1 Aggregate quarterly occupational gross flows rates. 101
2.2.2 UI generosity and exit probability toward an occupation. 106
2.3.1 Calibrated parameters and fit. 116
2.4.1 Bimonthly gross flow between occupations in the data and the model. 118
2.4.2 Occupational flow rates by wealth quantiles in the SIPP and the model. 120
2.5.1 Elasticity of unemployment flows to UI generosity: model and data 123
2.5.2 Model-based elasticity of insured unemployment to UI generosity by ability and wealth. 125
2.6.1 UI duration versus UI benefits: data and model 132

3.2.1 Business acquisition by type [a] 140
3.2.2 Key Heterogeneity Components by Acquisition Type 143
3.4.1 Fixed parameters 156
3.4.2 Model parameters calibrated within the model [a] 157
3.5.1 Survival rate: model versus data 162
3.6.1 Aggregate outcomes with and without the SMESM 163

3.6.2 Occupational choice and distribution of firms with and without a SMESM 164
3.6.3 SME for sale market, maturity effects and intangible value decomposition 165
3.6.4 Wealth concentration and inequality 168
3.A.1 Business acquisition by type in U.S. surveys 170
3.A.2 Characteristics of firms within 3 years of acquisition by type 171
3.A.3 Loan acceptance rate for firms within 4 years of acquisition by type (%) 172
3.A.4 Main reasons why credit was denied (%) 172
3.A.5 Descriptive statistics: BBS data on businesses for sale (2018-2020) 173
3.B.1 Life cycle earning profile 175
3.C.1 Sensitivity analyses on the SME market and aggregate outcomes. 175

4.3.1 Fixed parameters of the model. 195
4.3.2 Calibrated parameters of the model. 196
4.3.3 Panel estimation: Trade proximity and GDP correlation [a] 198
4.4.1 Quantitative assessment of the Trade Comovement Slope: Data versus Model. 201
4.5.1 Trade, Measured Productivity and Net Operating Surplus 203
4.5.2 Association between trade and SR comovement: Data versus Model. 204
4.5.3 The role of price distorsions and heterogenous markups [a] 206
4.5.4 Real GDP correlations and the margins of intermediate inputs trade 207
4.B.1 Model-based simulations: sensitivity analysis [a] 211

Chapter 1

Wealth, Returns, and Taxation: A Tale of Two Dependencies

Alexandre Gaillard[1,2]   Philipp Wangner[3]

Abstract

We study wealth redistribution in a framework where individual portfolio choices and associated returns are correlated with wealth through: (i) *type* dependence, which reflects that investment skills drive return differences, and (ii) *scale* dependence, which captures that wealth itself triggers returns. Using an analytical framework, we argue that several common heterogeneous agent models can be understood through the lens of a type and scale dependence representation. We show that four key statistics characterize the macroeconomic and welfare implications of wealth taxation: the right tail of the wealth distribution, the degree of scale and type dependence, and the extent to which returns reflect investment productivity. We then build a quantitative model calibrated using micro US data and find an optimal marginal wealth tax rate of 0.8 percent above an exemption level of $550K. The result is driven by two opposing forces. Under scale dependence, productivity and wealth accumulation decrease with the tax, as risk-taking depends on wealth. Under type dependence, a higher wealth tax reinforces the selection of skilled investors at the top and improves productivity. Finally, the marginal wealth tax only slightly increases when returns partially reflect rent motives, as both forces almost quantitatively offset each other.

Keywords: Wealth taxation, Return heterogeneity, Type and scale dependence, Inequality.

---

[1]A previous version was entitled "*The Growth Irrelevance Frontier of Wealth Inequality*". We are extremely grateful to Christian Hellwig, Sumudu Kankanamge, Patrick Fève, Nicolas Werquin, Fabrice Collard and Tim Lee for helpful discussions and invaluable guidance. We also thank Andrew Atkeson, Adrien Auclert, Marianne Andries, George-Marios Angeletos, Tiziana Assenza, Saki Bigio, Timo Boppart, Pierre Boyer, Christophe Bruneel-Zupanc, Harold Cole, Mariacristina De Nardi, François de Soyres, Martial Dupaigne, Xavier Gabaix, Cécile Gaubert, Christian Gollier, Eugenia Gonzalez-Aguado, Alexander Guembel, Ulrich Hege, Nir Jaimovich, Thierry Magnac, Andreas Schaab, Ludwig Straub, Stéphane Villeneuve, Miguel Zerecero and participants at the 2021 SED, the 2021 EEA and various seminars for their constructive comments and discussions. The authors gratefully acknowledge funding from Banque de France. All remaining errors are our own.

[2]Toulouse School of Economics, University of Toulouse Capitole, 31080 Toulouse, France. Email: alexandre.gaillard@tse-fr.eu.

[3]Toulouse School of Economics, University of Toulouse Capitole, 31080 Toulouse, France. Email: philipp.wangner@tse-fr.eu.

1.1 Introduction

Wealth is highly concentrated at the top. In the US, for instance, Saez and Zucman (2016) report that the wealth share of the richest 1% households has risen from 25% of the total wealth in 1980 to 40% in 2012. This trend has recently renewed academic and political interests on whether, and how, economies should redistribute the wealth of their richest individuals (see e.g. Piketty et al. (2014), Saez and Zucman (2019)). On the practical side, many of the richest OECD countries have considered a wealth tax in the last decades; some have implemented it, while others have withdrawn it.[4] In 2021, the Ultra-Millionaire Tax Act was proposed at the US Congress to introduce a tax on the wealth of the top 0.05% households.

In this paper, we investigate the macroeconomic and welfare implications of wealth taxation. To assess these implications, we build a general equilibrium model that accounts for key determinants behind the wealth accumulation of the richest individuals. We follow the influential work by Benhabib et al. (2011, 2019) and introduce heterogeneity in returns to wealth which constitutes, to date, one of the most compelling factors in explaining the high wealth concentration (Smith et al. (2019a), Hubmer et al. (2020), Xavier (2020)).[5] To do this, our model features two important departures from existing frameworks. First, in the spirit of Gabaix et al. (2016) we explicitly introduce two channels through which individuals differ in their returns to wealth: *type* and *scale* dependence. Type dependence reflects the fact that wealthy individuals obtain high returns because they differ in their innate or persistent characteristics, e.g. outstanding investment skills or high risk tolerance. Instead, scale dependence captures the fact that wealthier agents generate higher returns, regardless of their specific type, e.g. due to costly access to high-yield investments or decreasing relative risk aversion. This is especially relevant since recent contributions by Bach et al. (2020) and Fagereng et al. (2020) show that both concepts explain a substantial part of the cross-sectional correlation between returns to wealth and wealth. Second, we allow for the idea that private returns to wealth may only partially reflect differences in investment productivity due to some forms of rent-extraction. This may arise due to bargaining and market power, elite connections, or an unequal opportunity to access certain investments or markets (Piketty et al., 2014; Rothschild and Scheuer, 2016; Lockwood et al., 2017; Smith et al., 2019a).

We find that the optimal wealth tax rate in our benchmark model calibrated to the US economy is positive and large, at a rate of 0.8 percent above an exemption level of $550K. Our first key result is to show that this tax rate can be traced back to the underlying forces behind return heterogeneity and thus behind wealth accumulation. Specifically, the degree of type and scale dependence

[4] Abstracting from estate taxation and property taxes, twelve European countries levied an annual tax on net wealth in 1990. By 2018, only France, Norway, Spain, and Switzerland still imposed such a tax (Scheuer and Slemrod, 2021).

[5] Heterogeneity in the portfolio allocation of households is a commonly used factor to explain heterogeneity in returns to wealth (Calvet et al. (2019), Smith et al. (2019b), Meeuwis (2019), Xavier (2020)). Another factors that explain wealth concentration include, for instance, differential saving rates (Straub, 2018; Hubmer et al., 2020).

determines the sign and magnitude of the wealth tax. Under type dependence, a non-trivial selection effect of high-skilled investors at the top of the wealth distribution rationalizes a high wealth tax rate. In such a world, a wealth tax has the potential to increase overall productivity. In contrast, under scale dependence, a wealth tax substantially decreases wealth accumulation and productivity and provides a rationale for low wealth taxes or even a subsidy. Our second result shows that the optimal tax rate is surprisingly almost unresponsive to the extent to which returns to wealth reflect differences in investment productivity. When the strength of rent-extraction increases, the opposing forces arising from type and scale dependence almost exactly offset each other in our preferred benchmark calibration. We substantiate our quantitative results in two steps.

In a first step, we lay out the main concepts behind our results within an analytical two-period model. In this model, households' risk aversion is correlated with their initial wealth and their innate type, which determines their willingness to invest in risky but more productive assets. With a perfectly elastic supply of capital, aggregate productivity and output are determined in equilibrium by aggregating risky and riskless capital investments. We isolate and clarify the key parameters that characterize the macroeconomic and welfare implications of a change in top wealth inequality, due to, for instance, a wealth tax. We show that these implications depend on four statistics; (i) the Pareto tail of the wealth distribution, (ii) the elasticity of risk-taking with respect to wealth, i.e. scale dependence, (iii) the sorting of individuals with different investment skill-types along the wealth distribution captured by the correlation between investor's types and wealth, i.e. type-dependence, and (iv) the extent to which returns to wealth reflect differences in productivity of investments rather than rents. While (i) – (iii) can in principle be measured empirically (Vermeulen, 2016; Bach et al., 2020; Fagereng et al., 2020), there is only little recent evidence concerning (iv) (cf. Lockwood et al. (2017) and Smith et al. (2019a)).

Despite its simplicity, the analytical model provides key insights regarding the role of inequality on aggregate output. We derive an intuitive diagram which captures all the possible relationships between changes in inequality and aggregate output or welfare resulting from the signs and the magnitude of scale and type dependence. We view this representation as a compelling device that can unify the existing literature studying, and disagreeing about, the relationship between inequality and output growth: a specific model can be classified in a particular region of our diagram given the underlying – implicit or explicit – assumptions regarding the above key parameters. For instance, models that incorporate mechanisms related to saving and investment decisions with a type and/or scale dependence representation (see among others Galor and Zeira (1993), Angeletos (2007), Cagetti and De Nardi (2006a), Moll (2014), Gomez et al. (2016), Kaplan et al. (2018), Guvenen et al. (2019), Hubmer et al. (2020)) fall in a particular decomposition of our setup. This is especially important as scale dependence implies a strong behavioral response to a change in household wealth and thus makes aggregate responses considerably more sensitive to wealth inequality changes.

In our framework, the welfare-maximizing top wealth tax, based on a utilitarian consumption-equivalent variation welfare criterion, balances three effects. First, a marginal decrease in top wealth inequality through wealth redistribution affects productivity, output, and equilibrium wage rate, as it reallocates wealth among households who differ in their intrinsic investment skill or risk tolerance type and wealth. Second, whenever returns to wealth imperfectly reflect the productivity of investments, a change in inequality generates a change in the size of rents in the economy. That is, some individuals benefit from extra-returns, without actually affecting production. Therefore, aggregate returns adjust in equilibrium to ensure that the total capital income received by households equalizes the total product of capital redistributed in the economy. Third, redistribution from the top to the bottom induces a standard equity motive as wealth-rich and wealth-poor households differ in their marginal utility of consumption.

In a second step, we extend the framework to a full-blown quantitative model to carefully evaluate the implications of a wealth tax. The model is a variant of the standard incomplete-markets model with heterogeneous agents facing uninsurable labor income risk pioneered by Bewley (1986) – Huggett (1993) – Aiyagari (1994). Like in our simple model, heterogeneity in investment decisions and associated returns to wealth is introduced through type and scale dependence. In the spirit of Cagetti and De Nardi (2006a), Moll (2014), and Benhabib et al. (2019), type dependence arises as households differ in their intrinsic ability to undertake risky productive investments. Importantly, this ability evolves stochastically but is highly persistent. The higher the persistence, the more likely high-skilled investors generate high returns during many periods, and the more frequently they are represented at the top of the wealth distribution. Furthermore, we incorporate two empirically relevant forms of scale dependence. First, conditional on being investors, richer agents invest a larger fraction of their wealth (intensive margin). Second, following Hurst and Lusardi (2004) or Fagereng et al. (2017), richer agents are more likely to become investors (extensive margin). Finally, on top of the possibility that rent-seeking may explain heterogeneity in returns to wealth (Rothschild and Scheuer, 2016), we also introduce elements that have been previously identified in the literature as having potential large implications for optimal capital taxation, i.e. we add a life-cycle structure with endogenous labor supply (Conesa et al., 2009; Kindermann and Krueger, 2014).

The model is calibrated to replicate the empirical labor income and wealth distributions and moments regarding the observed heterogeneity in portfolio choices across households. We separate risky but potentially more productive assets, such as private equity and public equity, from safe assets using estimates of returns from the PSID. Like Cagetti and De Nardi (2006a) and Guvenen et al. (2019), heterogeneity in returns to wealth reflects investment productivity differences in our benchmark. We distinguish two types: highly skilled investors who manage a significant amount of risky equity assets, and non-investors who do not invest and constitute the vast majority of households in the SCF. As types are persistent, this approach generates an endogenous type dependence within the model. Second, following Hurst and Lusardi (2004), we exploit the panel dimension of the PSID to pin down scale dependence in the risky investment participation with respect to wealth such that it aligns with its empirical counterpart. Third, the share of wealth invested in risky equity, conditional on being an investor, is increasing along the wealth distribution. To distinguish scale dependence in the share arising from net risky investments only, we use detailed information from the SCF on the timing and the allocation of private equity business investments. Specifically, a large proportion of the increase in equity investments at the top is driven by recent additional private equity investments. To remain conservative, we only attribute this margin to scale dependence in the risky share invested, with the underlying assumption being that those additional investments are unlikely to drive the fortune of already rich households.

The benchmark model replicates the high concentration of returns at the top from both the type and scale dependence channels. To further investigate the properties of the model, we study alternative specifications with type or scale dependence only. We find that these alternatives are almost observationally equivalent to the benchmark model regarding the distributions of returns and wealth, i.e. both are able to generate high wealth concentration from persistent return heterogeneity. However, the aggregate responses to a wealth tax differ: the response is substantially amplified under a high degree of scale dependence.

Within a restricted class of wealth tax functions, we use our benchmark model to compute the long-run optimal one-time wealth tax reform, which we redistribute by lowering labor income taxes to obtain revenue neutrality. We jointly determine the marginal wealth tax rate and the exemption level above which it applies, which induces a common form of tax progressivity. Our result of an optimal tax of 0.8 percent above an exemption level of 550K is the first quantitative outcome of this setup. This reform generates a welfare gain equivalent to 0.14% of yearly consumption with large heterogeneity: they are high below the 70^{th} wealth percentile and negative at the very top. We then extend the model to account for the presence of rents in returns. Following evidence in Lockwood et al. (2017) and Rothschild and Scheuer (2016), we attribute the excess returns to wealth extracted from law and finance sectors to rent-seeking motives. Under this calibration and fixing the exemption at its benchmark level, the optimal marginal tax rate only slightly increases with the size of rents, to a rate of 0.92%. Dissecting our results, we find that they depend critically on whether top wealth inequality is driven by type or scale dependence.

If we first suppose the absence of rent extraction motives and scale-dependence is the dominant source of wealth concentration, then a higher wealth tax, by discouraging capital accumulation, causes a *snowball effect*: as agents become less wealthy, their rate of return falls, which further discourages productive investments. These self-enforcing effects imply that a wealth tax generates a large adverse behavioral response which decreases aggregate productivity. Instead, if we suppose that type-dependence is the dominant force at play, the wealth tax has a disproportionately large adverse effect on the investment of agents with high wealth but low returns, i.e. they dissave at a higher rate. Thus, the wealth tax creates an environment where *only the fittest survives at the top*, i.e. a selection effect whereby the top of the wealth distribution ends up being composed of the most productive investors. In this last case, although households accumulate less, the wealth tax has the property to raise aggregate productivity. In such a scenario, a wealth tax becomes a powerful instrument and, as shown by Guvenen et al. (2019), even superior to a capital income tax. In more concrete terms, fixing the exemption level at $550K, it is optimal to subsidy wealth with a negative tax rate of -0.8 percent in a model featuring scale dependence only. In contrast, it is optimal to heavily tax wealth at a rate of 2.4 percent in a model featuring type dependence only. In a world where both dependencies coexist, such as in our benchmark economy, the optimal tax rate falls in between those two bounds.

Now consider the case where high returns on wealth reflect rent extraction instead of more productive investment opportunities. In that case, the above conclusions are reversed. Under scale dependence, a wealth tax is desirable because it discourages inefficient rent-seeking behavior. Under type dependence, by contrast, the endogenous selection mechanism from the implementation of the wealth tax previously described still implies that agents with higher returns will be more concentrated at the top; but those higher returns are now a reflection of higher rents rather than higher productivity, thus making the wealth tax relatively undesirable. These two opposing forces rationalize a low response of the wealth tax rate to the size of the rent in the benchmark economy, as the two forces almost quantitatively offset each other.

Related literature Our work is related to a number of papers studying the relation between the distribution of wealth and its strong interplay with macroeconomic aggregates. Many macroeconomic models incorporate mechanisms related to saving and investment decisions with a type and/or scale dependence representation to generate realistic wealth distributions. For instance, Benhabib et al. (2019) construct a quantitative model designed to identify the determinants of wealth inequality and wealth mobility in the US. Their baseline model features wealth return heterogeneity due to type dependence only. Relatedly, Hubmer, Krusell and Smith Jr (2020) study how several determinants, comprising heterogeneity in wealth returns, account for the recent rise of wealth inequality in the US. They use estimates of returns from Bach et al. (2020) and interpret the observed heterogeneity in wealth portfolio and returns as scale dependence only. Other relevant examples include, among others, Cagetti and De Nardi (2006a), Moll (2014), Kaplan et al. (2018), and Guvenen et al. (2019).[6] Yet, surprisingly, their systematic distinction has

[6] Among many others, Kaldor (1956), Stiglitz (1969), or Bourguignon (1981) study the role of wealth inequality in a neoclassical economy with convex scale dependence in saving behaviors. Other mechanisms include risk-taking behavior (Peress (2004), Brunnermeier and Nagel (2008), Calvet et al. (2009), Robinson (2012), and Meeuwis (2019)) non-convex investment cost and DRS (Banerjee and Newman (1993), Galor and Zeira (1993)), economies of scale in wealth management (Kacperczyk et al., 2019), social status derived from wealth holdings (Roussanov, 2010), investment in

been neglected thus far. This paper fills this gap. We show that while both mechanisms are independently capable of generating large wealth inequality through return heterogeneity, they imply distinct macroeconomic and welfare implications from wealth redistribution.

Generally, our paper is related to a large literature quantifying optimal taxation in general equilibrium models with heterogeneity in household capital investments (see among others, Aiyagari (1995); İmrohoroğlu (1998); Kitao (2008a); Conesa et al. (2009); Kindermann and Krueger (2014); Brüggemann (2020); Moll and Itskhoki (2019); Boar and Midrigan (2020)). Considering a tax on the stock of wealth, Shourideh et al. (2012) shows that a positive progressive tax on savings is optimal. Cagetti and De Nardi (2009a) and De Nardi and Yang (2016) show that it is not welfare improving to abolish the estate tax in the US. Such taxes can be reinterpreted as intergenerational wealth taxes. Closer to our work is the recent contribution by Guvenen et al. (2019), which shows that heterogeneity in returns has the property to break the equivalence result between taxing capital flow relative to taxing the stock of capital under homogeneous returns. They find that replacing the capital income tax with a wealth tax reduces misallocation and increases overall welfare in a model in which heterogeneity in returns comes mainly from differential entrepreneurial skill-types. Our results point to the key role of type and scale dependence, together with whether returns reflect the productivity of capital, in deriving the welfare implications of a wealth tax. Empirically, Jakobsen et al. (2020) find a strong role of top wealth taxation on wealth accumulation. To our knowledge, two papers discuss the role of type and scale dependence for capital taxation in the spirit of Diamond (1998) and Saez (2001). Gerritsen et al. (2020) find that capital income tax is positive with returns heterogeneity under both dependencies. In complementary work, Schulz (2021) shows that the degree of scale dependence in returns significantly affects the capital income tax. Relative to them, we rather follow a different approach by studying optimal wealth taxation in a general equilibrium incomplete markets economy, and quantitatively show that it is critical to model the endogenous selection of investment skill-types along the wealth distribution.

Layout In section 1.2, we construct an analytical two-period version of our model to lay out the main concepts and forces at play. Section 1.3 sets out the quantitative dynamic model. Section 1.4 discusses the model's calibration and section 1.5 investigate its properties. In section 1.6, we use our model to study the welfare-maximizing wealth tax, and section 4.6 concludes the paper. The appendix contains all proofs, empirical analyses, and computational details.

financial sophistication (Lusardi et al., 2017), wealth-dependent offshore investments (Alstadsæter et al., 2018), or non-convex investment costs in high return assets (Kaplan et al., 2018). In contrast, Kihlstrom and Laffont (1979), Moll (2014), Herranz et al. (2015a), or Moll and Itskhoki (2019) introduce type dependence in which the distribution of types, and their persistence, is crucial to deriving aggregate efficiency. Combined dependencies arise in Quadrini (2000a) and Cagetti and De Nardi (2006a) through type dependence because entrepreneurs self-select at the top of the wealth distribution and through scale dependence due to wealth-driven occupational choices and a DRS technology.

1.2 An Analytical Two-Period Model

We begin with an analytical two-period framework to illustrate the conceptual distinction between type and scale dependence and the main trade-offs. The purpose of this section is to provide simple insights, and to introduce the notations used throughout the dynamic quantitative model.

1.2.1 Environment

Households A unit mass $i \in [0,1]$ of heterogeneous households lives for two periods, $t \in \{1,2\}$, with initial wealth a_0^i and innate risk-taking type ϑ^i drawn from the joint distribution $\mathcal{G}_0(\vartheta, a_0)$, with marginal distributions $g_\vartheta(\vartheta)$ and $g_{a_0}(a_0)$ defined over the support $\Theta \subset \mathbb{R}_+$ and $\mathcal{A}_0 \subset \mathbb{R}_+$. Households have CARA preferences over consumption c^i, i.e. $(1/\alpha^i)\left(1 - e^{-\alpha^i c^i}\right)$, where the absolute risk aversion α^i correlates with their initial wealth and risk-taking type, such that

$$\alpha^i \equiv \overline{\vartheta} \cdot \left[\vartheta^i (a_0^i)^\gamma\right]^{-1} , \tag{1.1}$$

where $\overline{\vartheta} \equiv \mathbb{E}[\vartheta]$ scales the average economy-wide risk tolerance. The parameter $\gamma \geq 0$ governs the shape of the household's risk tolerance in initial wealth. This preference specification captures in a reduced form various mechanisms driving type and scale dependence in portfolio choices and capital returns mentioned in the related literature.[7] In period $t = 1$, households invest optimally a share ω_1^i of their beginning of period wealth $a_1^i = a_0^i - t_a(a_0^i)$ into a *risky* innovative asset with stochastic gross return R_r^i, and the complementary share $(1 - \omega_1^i)$ into a *risk-free* asset with certain gross return R_f. The function $t_a(\cdot)$ defines a wealth tax on initial wealth and T is a second period lump-sum transfer. Agents inelastically supply one labor unit and obtain a wage w. In $t = 2$, returns and wage realize and households consume c_2^i. The objective of household i is given by

$$\max_{\{\omega_1^i\}} \left(1/\alpha^i\right) \left(1 - \mathbb{E}_1\left[e^{-\alpha^i c_2^i}\right]\right) \qquad \text{s.t.} \qquad c_2^i \leq \left(\underline{r} + R_f(1-\omega_1^i) + R_r^i \omega_1^i\right) a_1^i + w + T , \tag{1.2}$$

where $\underline{r}$ is an aggregate return component, which is determined in equilibrium and common to all households.

Production In period $t = 2$, a competitive final good producer uses aggregate labor n and a continuum $j \in [0,1]$ of intermediate good projects x_s^j from two technologies $s \in \{N, I\}$, an innovative technology I with risky returns and a safe non-innovative technology N. The aggregate produc-

[7] This special form of CARA preferences extends the utility functions used in Alpanda and Woglom (2007) or Makarov and Schornick (2010) by specifying the wealth normalization with a power function. This form is also ultimately linked to Guiso and Paiella (2000) and Gollier (2001), who specify the shape of risk tolerance in terms of consumption rather than wealth.

tion function is given by $Y = Xn^{\varphi}$, where $X = \left(\sum_s \int_j x_s^j \, dj\right)$ and $\varphi \in [0,1)$.[8] Profit maximization follows

$$\max_{\{n, \{x_s^j\}_{j,s}\}} Xn^{\varphi} - wn - \sum_s \int_j p_s^j \, x_s^j \, dj \,, \tag{1.3}$$

where p_s^j denotes the price of an intermediate good j in sector s.

An intermediate good producer uses risk-free capital k_N^j and risky capital k_I^j to run a project j with linear technologies. *Innovative* projects produce $x_I^j = (\phi\mu + A(1-\mu))k_I^j$, where $\phi > A$ denotes the expected innovate asset net return and $\mu \in [0,1]$. *Traditional* projects operate with technology $x_N^j = Ak_N^j$. The revenue generated by a traditional safe project is $p_N^j x_N^j$ and the revenue from an innovative project is $p_I^j x_I^j$.

Market clearing The first order condition with respect to aggregate labor yields $w = \varphi X n^{\varphi - 1}$, with $n = 1$. Substituting for labor demand, the objective of the profit maximization (1.3) can be rewritten as $(1-\varphi)Xn^{\varphi} - \sum_s \int_j p_s^j \, x_s^j \, dj$. As intermediate goods are perfect substitutes, their prices are identical, and each unit is sold at a price $p_s^j = (1-\varphi)$.

The intermediate project i uses capital invested by household i to a run a project, such that $k_N^i = a_1^i(1-\omega_1^i)$ and $k_I^i = a_1^i \omega_1^i$. It redistributes revenues to household i as follows. Revenues from riskless assets are redistributed such that their returns equal the marginal product of capital net of wage payments, i.e. $R_f = (1-\varphi)A$. In contrast, the returns to innovative investments are given by $R_r^i = (1-\varphi)\kappa^i$, and may deviate from the net marginal product for two reasons. First, there is an idiosyncratic luck component $\kappa^i \sim \mathcal{N}(\phi, \sigma_\kappa^2)$ that introduces return risk on the household side. Second, we assume that there exists a return wedge between the expected risky return to wealth, $(1-\varphi)\phi$, and the net marginal product of innovative capital $(1-\varphi)\left(\phi\mu + A(1-\mu)\right)$, henceforth MPK_r, on the production side.[9] When $\mu = 1$, expected returns to innovative capital investments equal the MPK_r. Whenever $\mu < 1$, expected risky returns are higher than the MPK_r. In the extreme case where $\mu = 0$, the risk premium $(1-\varphi)(\phi - A)$ observed on the household side does not arise from productivity differences across asset classes.

A rationale for $\mu < 1$ comes from the presence of rent-extraction motives due to some forms of bargaining, market power or political connections of investors, i.e. in the words of Rothschild and Scheuer (2016), "*the pursuit of personal enrichment by extracting a slice of the existing economic*

[8] In Appendix OA 1.5 we derive the case with aggregate decreasing returns to scale in X. In this case the portfolio choice is increasing in X, as a higher X tends to depress the dispersion of returns. Therefore, the risky capital supply in this alternative model is upward-sloping.

[9] In our specification, idiosyncratic risk materializes as return risk on the investor side rather than idiosyncratic production risk. In this respect, our framework deviates from the seminal incomplete market growth economies of Angeletos and Calvet (2006); Angeletos (2007). We impose this assumption out of tractability. If the idiosyncratic capital income risk is modeled as an idiosyncratic productivity shock, one needs to integrate over the joint distribution of wealth and types to obtain aggregate output and productivity, similar to Gabaix (2011). In Appendix OA 1.2, we show that the policy functions are isomorphic in both cases. Aggregation follows under the additional assumption that there is a sub-continuum of agents in each state (ϑ, a_0).

pie rather than by increasing the size of that pie".[10] We view this return wedge as a stylized way to reconcile two approaches by acknowledging that empirically measured returns to wealth can not easily be partitioned into a rent component and the marginal product of capital. On the one hand, some work disentangles returns to wealth from MPK, either because of their partial equilibrium structure (Benhabib et al., 2019) or because of implicit full rent extraction (Hubmer et al., 2020). On the other hand, models with capitalists often assume a perfect pass-through between MPK and returns (see among others Cagetti and De Nardi (2006a, 2009a) or Guvenen et al. (2019)). Instead, we derive results for a range of values for the return wedge μ.

Under this structure, whenever $\mu < 1$, the aggregate return component $\underline{r}$ adjusts in equilibrium to ensure that the total product of capital generated on the production side coincides with the total capital income redistributed to the households by the intermediate good producer, such that

$$\underbrace{\int_i \left(\underline{r} + R_f(1-\omega_1^i) + R_r^i \omega_1^i\right) a_1^i \, di}_{\text{returns}} = \underbrace{\int_i \left(A(1-\varphi)k_I^i + (\phi\mu + A(1-\mu))\,(1-\varphi)k_N^i\right) di}_{\text{capital product}} \,. \tag{1.4}$$

Distributional assumption Individual terminal wealth is affine in κ, i.e. its distribution is Gaussian $c_2^i \sim \mathcal{N}\left(\mu_{c_2}^i, \sigma_{c_2}^i\right)$ with mean $\mu_{c_2}^i = \varphi Y + T + (A(1-\varphi) + \underline{r})a_1^i + (1-\varphi)\omega_1^i a_1^i(\phi - A)$ and variance $\sigma_{c_2}^i = \left((1-\varphi)\omega_1^i a_1^i\right)^2 \sigma_\kappa^2$. Together with CARA preferences, this property ensures tractability of the equilibrium allocation. Finally, the initial wealth is assumed to be Pareto distributed.

Assumption 1 (Initial Wealth Distribution). *Initial wealth is drawn from a Pareto law with scale $\underline{a}$ and shape $\eta > \max\{\gamma, 1\}$, such that $A_0 \sim \mathcal{P}a\left(\underline{a}, \eta\right)$ with $\mathbb{P}(A_0 \geq a_0) = \left(\underline{a}/a_0\right)^\eta$, $\forall a_0 \geq \underline{a}$.*

While theoretically convenient, this assumption is consistent with well-known empirical evidence documenting that the wealth distribution is right-skewed and displays an heavy upper tail (Vermeulen, 2016; Klass et al., 2006). The shape parameter η is inversely related to wealth inequality.[11] As changing η leads *ceteris paribus* to a change in aggregate wealth, we sometimes study the effect of varying η (*redistribution* effect) while preserving the same aggregate wealth level by adjusting the scale $\underline{a}$ (*level* effect).

1.2.2 Efficiency Gains and Redistribution

We begin by characterizing the inequality–efficiency trade-off in this economy. We focus on the aggregate allocation when investment decisions are driven by type and scale dependence and discuss wealth taxation at the end of this section. As such, we solve first for the case $t_a(a_0) = 0$.

[10] See notably the discussion in Scheuer and Slemrod (2020) and the work of Piketty et al. (2014) and Rothschild and Scheuer (2016). In a related paper, Boar and Midrigan (2019) study a setup with entrepreneurs and workers in which entrepreneurial returns to capital investment reflect partially market power.

[11] The Pareto tail is also inversely related to the wealth share $q(p)$ of the p wealthiest households by $q(p) = p^{1-1/\eta}$.

Lemma 1 (POLICY FUNCTIONS)**.** *Let us define* $\widetilde{\omega} \equiv \frac{\phi - A}{(1-\varphi)\sigma_\kappa^2}$ *such that the standard CARA risky asset share is* $(\widetilde{\omega}/\overline{\vartheta})$*. Under our extended preference form, household* i*'s risky asset share is given by*

$$\omega_1^i = \widetilde{\omega} \cdot \left(\vartheta^i / \overline{\vartheta}\right) \cdot (a_0^i)^{\gamma - 1} . \tag{1.5}$$

If $\gamma = 1$ and $\vartheta^i = 1\, \forall i$, Lemma 1 provides the well-known result of Merton (1969) and Samuelson (1969), i.e. the share of risky asset holdings equals the baseline CARA solution captured by the risk premium over the variance of the risky return times the risk tolerance. Conditional on type ϑ^i, our CARA specification mimics IRRA (respectively DRRA) behavior if $\gamma < 1$ (respectively $\gamma > 1$). When $\gamma = 1$, our CARA specification nests CRRA behavior with constant risky asset share, while $\gamma = 0$ implies CARA behavior with constant risky asset holdings.[12] Therefore, the parameter γ pins down the elasticity of risky investments to initial wealth.

Using Lemma 1, we derive equilibrium quantities and prices.

Lemma 2 (AGGREGATE QUANTITIES)**.** *Given the joint distribution of types and wealth* $\mathcal{G}_0(\vartheta, a_0)$*, aggregate risky capital* K_I*, output* Y*, productivity* Z *and the wage rate* w *satisfy*[13]

$$K_I = \int_{\Theta \times \mathcal{A}_0} \omega_1(\vartheta, a_0) a_0 \, d\mathcal{G}_0(\vartheta, a_0) = (\widetilde{\omega}/\overline{\vartheta}) \Big(Cov(\vartheta, a_0^\gamma) - \mathbb{E}[\vartheta]\, \mathbb{E}\left[a_0^\gamma\right] \Big) , \tag{1.6}$$

$$Y = Z\mathbb{E}[a_0] , \quad \textit{with} \quad Z = \mu(\phi - A)\frac{K_I}{\mathbb{E}[a_0]} + A , \tag{1.7}$$

$$\textit{and} \quad w = \varphi Y , \quad \underline{r} = (\mu - 1)(\phi - A)(1 - \varphi)\frac{K_I}{\mathbb{E}[a_0]} . \tag{1.8}$$

Due to the CRS structure of final good production, demand for intermediate goods is perfectly elastic. Yet, its supply is bounded as households are risk-averse. Consequently, the risky portfolio shares of households, together with the joint distribution of wealth and types $\mathcal{G}_0(\vartheta, a_0)$, determine aggregate productivity Z and output Y. Therefore, wealth redistribution impacts productivity to the extent that it alters household i's investment in risky assets. The second condition of (1.8) states that $\mu < 1$ implies $\underline{r} < 0$, i.e. rent-extraction from risky investments induces a general equilibrium effect that decreases the common component of wealth returns, $\underline{r}a_1^i$, for all households.

We now formalize the effect of wealth redistribution on aggregate risky capital investment.

Proposition 1 (DISTRIBUTIONAL RELEVANCE)**.** *Consider without loss of generality a small mean pre-*

[12] Recall that one can always approximate the demand for risky assets for an arbitrary von Neumann-Morgenstern utility function as being proportional to risk tolerance, i.e. $\omega_1^i a_0^i \approx \frac{\mu_a^p}{var_a}\mathcal{T}(a_0^i)$. Contrary, under *generalized* CARA preferences risky investment is *exactly* proportional to our generalized risk tolerance $\mathcal{T}^i(\vartheta^i, a_0^i) = (\vartheta^i/\overline{\vartheta})(a_0^i)^\gamma$.

[13] The second condition in (1.8) follows by assuming that there is a sub-continuum of households in each state (ϑ, a_0^i).

serving change in the Pareto tail $\eta' > \eta$. Its effect on aggregate risky capital K_I can be decomposed into

$$\Delta K_I(\eta',\eta) = \underbrace{\Delta^a K_I(\eta',\eta)}_{\substack{\text{scale dependence in} \\ \text{portfolio holdings}}} + \underbrace{\Delta^\vartheta K_I(\eta',\eta)}_{\substack{\text{type heterogeneity and} \\ \text{selection}}} ,$$

where $\Delta^a K_I(\eta',\eta)$ is zero if $\gamma \in \{0,1\}$, increasing in η' if $\gamma \in (0,1)$ and decreasing in η' if $\gamma > 1$. A sufficient condition for $\Delta^\vartheta K_I(\eta',\eta)$ to decrease in η' is $\left(\frac{\partial corr(\vartheta,a_0^\gamma)}{\partial \eta} + \frac{\partial corr(\vartheta,a_0^\gamma)}{\partial \underline{a}} \frac{\underline{a}}{\eta(\eta-1)}\right) \frac{1}{corr(\vartheta,a_0^\gamma)} \leq 0$.

Proposition 1 establishes general conditions under which the wealth distribution is a *relevant* equilibrium object by decomposing the effect of a change in the tail of the wealth distribution, η, on aggregate risky capital K_I into two terms: (i) a scale dependence term $\Delta^a K_I(\eta',\eta)$, which hinges on the risk taking elasticity γ, and a (ii) type dependence term $\Delta^\vartheta K_I(\eta',\eta)$, which encapsulates the selection of ϑ-types across the wealth distribution. A change in the Pareto tail is called *distributional relevant* if both effects do not offset each other.

In the absence of type dependence, e.g. $\vartheta^i = \overline{\vartheta}\,\forall i$, wealth redistribution from the top to the bottom decreases (respectively increases) K_I if $\gamma > 1$ (respectively $0 < \gamma < 1$). When $\gamma = \{0,1\}$, distributional *irrelevance* arises as aggregate variables do not depend on the distribution of wealth, either because risky investments are constant ($\gamma = 0$), or because the share invested is constant ($\gamma = 1$). For the sake of clarity, we refer to scale dependence as a situation in which $\Delta^a K_I(\eta',\eta) \neq 0$, while a negative (respectively positive) scale dependence corresponds to the case where $\Delta^a K_I(\eta',\eta) > 0$ (respectively $\Delta^a K_I(\eta',\eta) < 0$). Our notion of scale-dependence therefore corresponds to cases in which scale effects, through γ, generate a *distributional relevant* link between wealth redistribution and aggregate quantities. This arises when the individual portfolio policy function ω_1^i depends non-linearly on initial wealth.

In the presence of type selection, the sufficient condition in Proposition 1 provides the bound on the change in the correlation between innate types and initial wealth such that $\Delta^\vartheta K_I(\eta',\eta) < 0$. Therefore, even if $\gamma = 1$, the distribution of wealth may be *relevant* through type dependence.

The Efficiency-Inequality Decomposition: A Closed-form Representation

Although Proposition 1 is general, we subsequently study a tractable representation of the equilibrium and the effects of wealth redistribution by putting a structural assumption on $Cov(\vartheta, a_0^\gamma)$.

Assumption 2 (Joint Distribution). *Let $\vartheta \sim \mathcal{P}a(\underline{\vartheta},\epsilon)$ such that $\overline{\vartheta} = \frac{\underline{\vartheta}\epsilon}{(\epsilon-1)}$. The joint cdf $\mathcal{G}_0(\vartheta, a_0)$ is constructed based on the Farlie-Gumbel Morgenstern copula with dependence parameter $\varrho \in [-1,1]$.*

Under Assumption 2, when $\varrho > 0$ (respectively $\varrho < 0$), there is a positive (respectively negative) correlation of types and wealth, while $\varrho = 0$ induces no correlation.[14] The level of ϱ translates

[14] The dependence parameter ϱ and the Spearman's correlation, ϱ^s, are under the Farlie-Gumbel Morgenstern copula related by $\varrho^s = \varrho/3$.

into the degree of selection, which is, for simplicity, exogenous in this section.

The following result decomposes the trade-off between inequality and efficiency into four terms which capture type dependence, scale dependence, an interaction term and the extent to which the excess return to wealth reflects productivity differences.

Proposition 2 (EFFICIENCY-INEQUALITY RELATION)**.** *Given Assumptions 1-2 and an aggregate positive riskless capital supply in equilibrium, i.e.* $\frac{K_I}{\mathbb{E}[a_0]} < 1$*, wealth-normalized output is given by* $\widetilde{Y}(\eta) = A + \mu(\phi - A)\widetilde{\omega}\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{\eta-1}{\eta-\gamma}\underline{a}^{\gamma-1}$*.*[15] *The marginal effect of wealth redistribution on* $\widetilde{Y}(\eta)$ *is*

$$\frac{\partial\widetilde{Y}(\eta;\gamma,\varrho)}{\partial\eta} \propto -\mu(\phi - A)\Big(\underbrace{\Omega^{\gamma}(\eta,\gamma)\cdot(\gamma-1)}_{\text{scale dependence}} + \underbrace{\Omega^{\varrho}(\eta,\gamma)\cdot\varrho}_{\text{type dependence}} + \underbrace{\Omega^{\varrho\gamma}(\eta,\gamma)\cdot\rho(\gamma-1)}_{\text{interaction term}}\Big), \tag{1.9}$$

where $\Omega^{\gamma}(\eta,\gamma)$*,* $\Omega^{\varrho}(\eta,\gamma)$ *and* $\Omega^{\varrho\gamma}(\eta,\gamma)$ *are strictly positive inequality multipliers.*

A key property of Proposition 2 is the ambiguous effect of rising wealth inequality on normalized output that depends on the relative strength of scale dependence, type dependence and their interaction, respectively captured by the terms $\gamma - 1$, ϱ and $\varrho(\gamma - 1)$. If preferences mimic DRRA behavior ($\gamma > 1$) and there is a positive selection of types ($\varrho > 0$), then output unambiguously rises in response to higher wealth inequality, since it reallocates wealth to agents investing in riskier and more productive assets ($\mu > 0$). This effect is scaled to the degree to which returns to investment reflect differential capital productivity, captured by the term $\mu(\phi - A)$.

Importantly, variations of the Pareto tail exhibit highly nonlinear effects on output captured by the inequality multipliers $\Omega^{\gamma}(\eta,\gamma)$, $\Omega^{\varrho}(\eta,\gamma)$ and $\Omega^{\varrho\gamma}(\eta,\gamma)$. In practice, the precise decomposition and the associated inequality multipliers are model-specific; however, as discussed below, the general idea and mechanisms unify a number of frameworks. In complete unequal economies, i.e. $\eta \to \max\{\gamma, 1\}$, small variations of the Pareto tail result in rather large output variations. In contrast, in a complete egalitarian societies, i.e. $\eta \to \infty$, small variations in η result in small output variations. Intuitively, in more unequal economies, scale dependence and selection effects are stronger in magnitude such that even small variations of η lead to strong investment reallocations.

Equation (1.9) also implies that, for a given level of inequality η, there exists an infinite number of possible combinations of type and scale dependence on a bounded two-dimensional set consistent with a given marginal effect of wealth redistribution on output. In Definition 1, we specify the notion of *iso-growth* (a kind of isoquant) which describes all parameter pairs (γ, ϱ) for which a marginal variation of the Pareto shape η generates a given output response $\overline{g}$.

Definition 1 (ISO-GROWTH OF INEQUALITY)**.** *For a given wealth Pareto tail* η *and* $\mu > 0$*, the iso-growth at level* $\overline{g}$ *is defined by the pair* (γ, ϱ) *that satisfies* $isoG(\eta, \overline{g}) \equiv \left\{(\gamma,\varrho) \in \Gamma \times [-1,1] : -\frac{\partial\widetilde{Y}(\eta;\gamma,\varrho)}{\partial\eta} = \overline{g}\right\}$.

[15]The condition $\frac{K_I}{\mathbb{E}[a_0]} < 1$ is satisfied for a given $\underline{a}$ if $\frac{\eta-\gamma}{\eta-1} \geq \widetilde{\omega}\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\underline{a}^{\gamma-1}$.

A special case ensues for $\overline{g} = 0$ for which the *iso-growth* curve separates the *growth enhancing region*, i.e. $-\frac{\partial \widetilde{Y}(\eta;\gamma,\varrho)}{\partial \eta} > 0$, from the *growth dampening region*, i.e. $-\frac{\partial \widetilde{Y}(\eta;\gamma,\varrho)}{\partial \eta} < 0$. For this reason, we label this special *iso-growth* curve the *Growth Irrelevance Frontier* of wealth inequality. Lemma 6 in Appendix 1.A.1 provides conditions for its existence based on the strength of type and scale dependence and the wealth Pareto tail η. If these parameter restrictions do not apply, an infeasible pair $(\gamma, \varrho) \notin \Gamma \times [-1, 1]$ would be required to obtain *growth neutrality*.[16] On the *Growth Irrelevance Frontier (GIF)* of wealth inequality, type and scale dependence exactly offset each other or are absent. In Lemma 3, we provide key properties of the *iso-growth*; it is decreasing in the space (γ, ϱ) and rotates clockwise in the wealth Pareto shape η.

Lemma 3 (PROPERTIES OF THE GIF AND ISO-GROWTH)**.** *The GIF is strictly decreasing on the defined set of Lemma 6, i.e.* $\frac{\partial \gamma}{\partial \varrho}|_{d\eta=0} < 0$. *A higher tail* η *rotates the GIF such that* $\frac{d\gamma}{d\eta}|_{d\varrho=0} > 0$ *for* $\gamma > 1$ *and* $\frac{d\gamma}{d\eta}|_{d\varrho=0} \leq 0$ *for* $\gamma \leq 1$*. Also, for a higher level* $\overline{g}$*, isoG*$(\eta, \overline{g})$ *is the translation of the GIF and* $\frac{d\gamma}{d\overline{g}}|_{d\varrho=0} > 0$.

Figure 1.1 illustrates the *iso-growth* for two different Pareto tails η and output levels $\overline{g}$. There are four regions which are delimited by the sign of type and scale dependence. In the top-right and bottom-left regions, the wealth-dependent risk-taking and the selection effect move in the same direction. An increase in inequality therefore unambiguously induces more (respectively less) economy-wide risk-taking and higher (respectively lower) productivity. In the top-left and bottom-right regions, the wealth-dependent risk-taking and the selection effect move in opposite directions. In these regions, there exist infinite combinations of type and scale dependence such that both effects offset each other giving rise to the GIF. In the top-left region characterized by $\{\varrho < 0, \gamma > 1\}$, an increase in inequality leads to higher output only if the positive scale dependence is sufficiently strong. In contrast, in the bottom-right region characterized by $\{\varrho > 0, \gamma < 1\}$ an increase in inequality decreases output if the wealth-dependent risk-taking elasticity γ is sufficiently low for a given selection $\varrho > 0$.

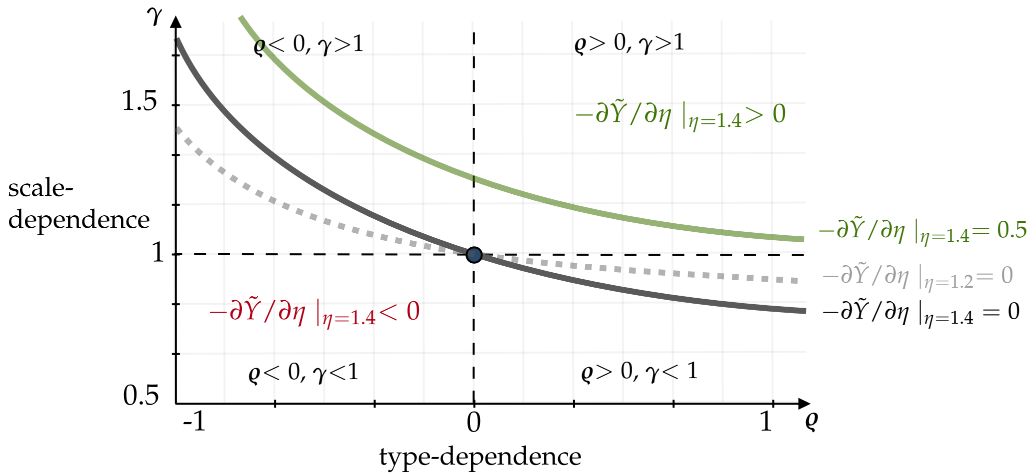
For a higher effect of wealth inequality on the level of output, i.e. an increase in $\overline{g}$, the iso-growth moves upward in the (γ, ϱ) diagram; the higher effect of greater wealth inequality on output can only be rationalized with higher degrees of positive type and/or scale dependence. Finally, the level of wealth inequality changes the relative strength of type and scale dependence effects; a higher inequality (lower η) reinforces the scale dependence effect relative to the selection effect and, as a result, the GIF (and the translated *iso-growth* curve) flattens. Finally, for $\overline{g} > 0$ (respectively $\overline{g} < 0$), a given *isoG*$(\overline{g}, \eta)$ shifts downward (respectively upward) with an increase in the productivity gap $\mu(\phi - A)$ as less reallocation between the two productive sectors is needed to achieve a certain level $\overline{g}$.

Despite its simplicity, the analytical decomposition derived above allows to captures key mech-

[16] Notice that the *isoG*$(\eta, 0)$ does not exist in a complete unegalitarian economy ($\eta = \max\{1, \gamma\}$), since in this case the absolute strength of scale dependence dominates the selection effect. However, in relatively egalitarian economies, both effects are small in magnitude such that the *isoG*$(\eta, 0)$ exists on a bounded set.

Figure 1.1. The inequality–efficiency diagram.

Note: numerical parameter values are $\varepsilon = 2.0, \underline{a} = \underline{\vartheta} = 1.0, A_r = 1.1, \sigma_\kappa = 0.2, A = 1.0$.



Legend: The solid black line is the GIF with a Pareto tail $\eta = 1.4$, and the dotted grey line is the GIF with $\eta = 1.2$. The solid green line is an iso-growth curve corresponding to a growth level of 0.5 percent.

anisms driving real-world household behavior, and carries over different quantitative models with heterogeneity in household investments such as the one studied in section 1.3. It shows that, in order to understand the effects of wealth redistribution on aggregates in a framework that accounts for heterogeneous capital investments, there are four key parameters required: (i) the Pareto tail of the wealth distribution, η, (ii) the elasticity of risk-taking with respect to wealth, γ, and (iii) the selection or sorting of types along the wealth distribution, as captured by the dependence, ϱ, and (iv) the excess wealth return augmented by the extent to which higher returns reflect higher capital investment productivity, $\mu(\phi - A)$.

Discussion and practical implications A number of frameworks with type and scale dependent mechanisms can be unified within the representation of the inequality-efficiency diagram of Figure 1.1 and equation (1.9).[17] In the basic Aiyagari (1994) economy, the distribution of wealth is almost growth–irrelevant due to the quasi linearity of the decision to save of the wealth-rich households. Angeletos (2007) studies an economy with linear portfolio policy functions and no type heterogeneity, thereby implying distributional *irrelevance*. Those models are located respectively at, and close to, the anchor point of the inequality-efficiency diagram ($\gamma = 1$ and $\varrho = 0$). Models with capitalists/entrepreneurs (among others Cagetti and De Nardi (2006a, 2009a), Guvenen et al. (2019), Brüggemann (2021)) often display positive type dependence at the stationary equilibrium, as entrepreneurs who invest in high-return private equity investments self-select at the top of the wealth distribution ($\varrho > 0$). However, they feature decreasing marginal product on those investments which can be reinterpreted as negative scale dependence effects ($\gamma < 1$). In

[17] In the Online Appendix OA 1.1, we hypothetically locate various theoretical and quantitative incomplete markets models relative to the GIF.

those models, wealth redistribution from the top to the bottom has thus an ambiguous effect on output. Therefore, they are positioned in the bottom-right region. Similarly, the seminal paper by Galor and Zeira (1993) with non-convex human capital investment costs can be reinterpreted as a form of DRS ($\gamma < 1$) but without type heterogeneity ($\varrho = 0$). Finally, the model of Moll (2014) displays only type dependence in households capital productivity with linear investment policy functions. Such models therefore locate on the locus with $\gamma = 1$.[18]

From the above diagram, it is interesting to see that type and scale dependence cannot be simply identified by using information on the effect of inequality on growth, as an infinite combination of pairs (γ, ϱ) may rationalize the relationship.[19] To identify both dependencies, it is ideal to have access to detailed panel data comprising portfolio decisions and associated returns to wealth. This may for example allow an econometrician to infer individual types through fixed effects, and scale dependence by estimating the effects of wealth variations on household behavior. Recent papers, such as Fagereng et al. (2020) and Bach et al. (2020), pave the way for such an empirical analysis. Without access to panel data and by relying only on cross-sectional data it is however difficult to identify the distinction as both channels may in principle generate consistent patterns regarding the portfolio allocation and associated returns to investment along the wealth distribution.[20] This is striking given that both effects result in substantially different elasticities of macroeconomic aggregates to wealth redistribution and, as shown in the dynamic model of section 1.3, to distinct implications for optimal wealth taxation.

A numerical example Consider two distinct models, i.e. the first with scale dependence only, and the second with type dependence only. We normalize the non-innovative productivity $A = 1$ and assume $\mu = 1$. We set the labor share $\varphi = 0.67$, the shape $\eta = 1.4$ consistent with estimates for the US (Vermeulen, 2016), and the risk premium such that $(1 - \varphi)(\phi - A) = 15\%$. The variance σ_κ^2 is set to 0.16, which is consistent with estimates from the PSID discussed in section 1.3. We calibrate type and scale dependence to generate a consistent cross-sectional pattern of risky equity shares, thus targeting the equity share of the top 1% wealthiest households of 65% as observed in

[18]In Appendix OA 3.1, we show how a number of additional frameworks can be understood within this representation. Moreover, the representation above provides a rationale for the lack of clear empirical evidence on the role of wealth inequality on growth (see for instance Perotti (1996); Forbes (2000); Barro (2008)). Recent panel data estimations tend to find a weak *positive* relationship in developed countries (Forbes, 2000; Voitchovsky, 2005; Barro, 2008; Frank, 2009) and a *negative* one in developing economies. The disparity of the estimates can be reconciled within our framework as the relation crucially depends on the underlying selection of agents, as well as the direction of scale dependence, which may of course be country-specific. In the Online Appendix OA 3.1, we investigate the relationship between top wealth inequality and GDP growth using new estimates regarding wealth concentration and find a positive link.

[19]They are, however, identified under particular conditions. Figure 1.1 shows that γ and ϱ are identified with *two* different couples of observation $(\eta_1, \overline{g}_1)$ and $(\eta_2, \overline{g}_2)$, but this requires that type and wealth dependence are constant over time. Moreover, estimating this relationship is somewhat complex as shown by the variety of empirical results in this related literature. See among others Forbes (2000), Voitchovsky (2005) or Barro (2008).

[20]See Section 1.3 for a detailed discussion regarding the identification of type and scale dependence in micro datasets.

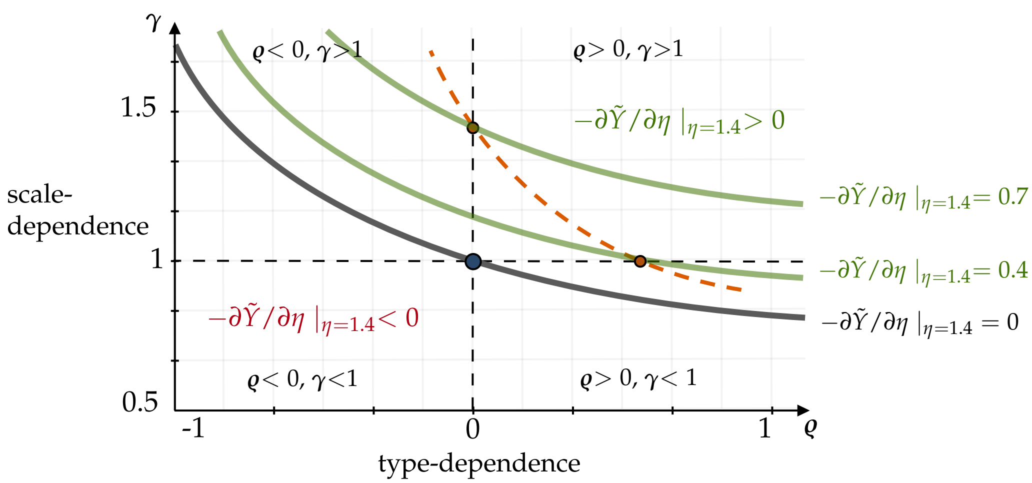
the 2010 SCF. Note that because the average risky equity share is increasing in wealth in the cross-section, the two models are located respectively on the $\{\varrho = 0, \gamma > 1\}$ and $\{\varrho > 0, \gamma = 1\}$ *loci*.

In the type dependence model, we set the Pareto shape of types $\varepsilon = 2$ and vary the correlation between types and wealth to match the top 1% risky share, such that $corr(\vartheta, a_0) = 0.65$.[21] In the scale dependence model, we match the same target by varying the wealth-dependent risk-taking elasticity and obtain $\gamma = 1.39$. In Appendix 1.A.2, we show that both models reproduce well the overall cross-sectional pattern of portfolio shares, at the bottom and at the upper end of the wealth distribution. However, the responses of output to a proportional top marginal wealth tax of 1% on the top 1% wealthiest households, which is redistributed through lump-sum transfers, differ substantially. We find that output drops by 0.43% under type dependence, which is in effect substantially lower than the 0.70% reduction found under scale dependence. In Figure 1.2, we report their respective iso-growth location. The difference in output responses originates from the behavioral response triggered by scale dependence as wealth varies (cf. Lemma 1).

This simple numerical example illustrates the importance of unraveling the economic forces behind capital investment heterogeneity. Notice that in a dynamic model, scale effects will be further amplified, as future wealth is a function of current investment itself, and the selection of skill-types along the distribution will be endogenous to inequality changes.

Figure 1.2. The inequality–efficiency diagram.

Note: numerical parameter values are $\varepsilon = 2.0, \underline{a} = \underline{\vartheta} = 1.0, A_r = 1.1, \sigma_\kappa = 0.2, A = 1.0$.



Legend: the solid black line is the GIF with an inequality level $\eta = 1.4$. The solid green lines are iso-growth curves corresponding to the two orange dots illustrating the numerical example described in the main text. The orange dashed line represents all combinations (ϱ,γ) such that the model replicates the cross-sectional portfolio shares and returns in the data, for a given $\eta = 1.4$.

[21] Instead, it is also possible to fix the correlation between types and wealth but vary the extent to which individuals are different by varying the shape ϵ. There is marginal difference in considering one over the other alternative, as long as the cross-sectional distribution of portfolio shares and wealth are well matched.

1.2.3 From Efficiency to Welfare

We now shift our focus to a welfare analysis. For tractability, we consider the case of a constant rate of progressivity (CRP) tax (Feldstein, 1969; Heathcote et al., 2017) on initial wealth such that $t_a(a_0^i) = a_0^i - (a_0^i)^{1-p_a}$, where $p_a \in (-\infty, 1)$ captures the progressivity of the wealth tax schedule. We denote with "$\sim$" *post-tax* variables. Under this assumption, individual choices are isomorphic, replacing initial wealth a_0^i with $\widetilde{a}_1^i = a_0^i - t_a(a_0^i)$, which implies an updated first period wealth Pareto tail $\widetilde{\eta} = \frac{\eta}{1-p_a}$ and scale $\underline{\widetilde{a}} = \underline{a}^{1-p_a}$, where $p_a \to 1$ implies a complete egalitarian economy. We measure welfare in terms of consumption equivalents, defined as the amount $\Delta^{CE,i}$ that makes household i in the reformed economy as well off as in the initial status quo economy, such that $\mathbb{E}[u(\widetilde{c}_2^i - \Delta^{CE,i})] = \mathbb{E}[u(c_2^i)]$. Under our CARA-Normal structure, this gives $\Delta^{CE,i} = \widetilde{x}_{c_2}^i - x_{c_2}^i + \frac{\Delta_c^i}{\widetilde{\alpha}_i}$, where $x_{c_2}^i$ and $\widetilde{x}_{c_2}^i$ denote certainty equivalents of the second period pre- and post-tax consumption, and the term Δ_c^i arises as the utility is a positive function of initial wealth through the risk aversion α^i. Given an utilitarian equivalent variation-based welfare measure, the planner solves

$$\mathcal{W} = \arg\max_{p_a} \int \Delta^{CE}(\vartheta, a_0)\, d\mathcal{G}_0(\vartheta, a_0) \qquad \text{s.t.} \qquad T = \frac{\eta \underline{a}}{\eta - 1} - \frac{\eta \underline{a}^{1-p_a}}{\eta - 1 + p_a}\,. \tag{1.10}$$

where the last equality balances the government budget constraint, such that $\int t_a(a_i^0) di = T$.[22]

Lemma 4 (Optimal wealth redistribution)**.** *Assume that the excess return is sufficiently large, i.e. $\phi - A - \frac{\sigma_\kappa^2}{2}(\widetilde{\omega}/\overline{\vartheta}) > 0$. The optimal progressivity p_a^* solves $\frac{\partial \mathcal{W}}{\partial p_a} = 0$, with*

$$\frac{\partial \mathcal{W}}{\partial p_a} = \Bigg(\underbrace{\varphi \frac{\partial \widetilde{Y}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}} \mathbb{E}[a_0]}_{\substack{\text{GE wage} \\ \text{efficiency if } \mu > 0}} + \underbrace{\frac{\partial \underline{r}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}} \mathbb{E}[\widetilde{a}_0]}_{\substack{\text{GE rent component} \\ \text{if } \mu < 1}} \Bigg) \frac{\partial \widetilde{\eta}}{\partial p_a} + \underbrace{\frac{\partial T}{\partial p_a}}_{\substack{\text{lump-sum} \\ \text{transfers}}} + \underbrace{\int \mathcal{R}(a_0, \vartheta) d\mathcal{G}_0(a_0, \vartheta)}_{\text{direct effects of the wealth tax}}$$

where $\mathcal{R}(\vartheta, a_0)$ captures the direct effects of the wealth tax on second period consumptions and on risk aversion $\widetilde{\alpha}_i$. The sign of $\frac{\partial \widetilde{Y}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}}$ is characterized in Proposition 2 and $sgn\left(\frac{\partial \underline{r}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}}\right) = -sgn\left(\frac{\partial \widetilde{Y}(\widetilde{\eta}; \varrho, \gamma)}{\partial \widetilde{\eta}}\right)$.

Lemma 4 characterizes the main trade-offs of wealth taxation on welfare. First, there is an equity channel as individuals differ in initial wealth, and thus in their marginal utility of consumption. Welfare increases with the lump-sum transfers T, but decreases with the efficiency losses from wealth taxation. The latter affects terminal wealth, captured by the term $\mathcal{R}$, through behavioral investment responses, i.e. changes in ω_1^i and changes in the curvature of the utility function through the risk aversion α^i. Second, a wealth tax affects efficiency, captured by $\frac{\partial \widetilde{Y}}{\partial \widetilde{\eta}}$, by reallocating wealth from the top to the bottom of the wealth distribution. Depending on the size

[22] Between periods $t = 0$ and $t = 2$, the government may in principle invest tax revenues in risky or riskless assets, K_N or K_I, and obtain returns from those investments which affect the amount of second period lump-sum transfers T. We simplify the exposition here and assume that the government does not invest.

of type and scale dependence, this affects the amount of capital that is invested in the innovative sector, and thus aggregate productivity and wage w. Third, whenever $\mu < 1$, the size of the rents in the economy responds to the inequality change induced by the wealth tax, captured by the aggregate return component $\underline{r}$. If rent-extraction increases, it lowers welfare as high capital return investors obtain a larger fraction of the overall product of capital, such that risky private returns to wealth become larger than their social value, i.e. the marginal product of capital investments. This in turns lowers the aggregate capital return component $\underline{r}$ to ensure that the total product of capital equalizes the amount of returns redistributed to households. In other words, the existence of rents widens the dispersion of returns to wealth across households, without actually reflecting dispersion in investment productivity.

The optimal proportional wealth tax thus trades-off equity and rent-extraction *versus* efficiency considerations. In this static model with exogenous type dependence, a lower μ implies a higher progressivity of the wealth tax. In the quantitative dynamic model in which the joint distribution of skill-types and wealth is endogenous, we argue that this result crucially depends on how the selection of skilled investors reacts to the implementation of a wealth tax in the long-run.

1.2.4 Generalization

The assumptions made throughout the special case allowed to isolate risk-taking decisions arising from *type* and *scale* dependence. We now briefly extend the analysis to saving decisions, alternative sources of *scale* dependence, and aggregate productivity shocks.[23]

Portfolio choice with explicit saving decision

Households now consume over the two periods, c_1^i and c_2^i, and preferences have a recursive form $u_1^i = U(c_1^i) + \beta U\Big(G^{-1}\Big(\mathbb{E}\Big[G\Big(U^{-1}(u_2^i)\Big)\Big]\Big)\Big)$, where it holds that $u_2^i = U(c_2^i)$, $U(c^i) = \frac{c^{1-1/\sigma}}{1-1/\sigma}$ and $G(c^i) = (1/\alpha^i)\left(1 - e^{-\alpha^i c^i}\right)$. As a result, the maximization objective becomes

$$\max_{\{c_1^i,\, \omega_1^i,\, a_1^i \geq 0\}} \frac{1}{1-1/\sigma}\left((c_1^i)^{1-1/\sigma} + \beta\left\{ -\frac{1}{\alpha^i}\log\left(\mathbb{E}\left[e^{-\alpha^i c_2^i}\right]\right)\right\}^{1-1/\sigma}\right), \tag{1.11}$$
$$\text{s.t.} \qquad c_1^i + a_1^i \leq a_0^i - t_a(a_0^i), \qquad c_2^i \leq \left(\underline{r} + R_f(1-\omega_1^i) + R_r^i\omega_1^i\right) a_1^i + w + T\,,$$

where $\sigma > 0$ and $\beta \in (0,1)$ define, respectively, the intertemporal elasticity of substitution (IES) and the discount rate. In this case, the joint heterogeneity in marginal propensities to save and marginal propensities to take risk (Kekre and Lenel, 2020) is key to studying aggregate allocations and gives rise to generalized *iso-growth* curves. For simplicity, we neglect wealth taxation in the

[23] Additional extensions include uninsurable labor income risk and participation decisions. We relegate those additional analyses to the online appendix OA 1.8.

following, i.e. $t_a(a) = T = 0$.

Lemma 5 (INDIVIDUAL PORTFOLIO CHOICE). *Denote $\widetilde{\beta} = (\widetilde{R}\beta)^{\sigma} + \widetilde{R}$ where $\widetilde{R} = \underline{r} + R_f$ and let us assume an interior solution to (1.11) such that $c_1^i > 0$. Individual portfolio choices of risky and riskless assets are denoted, respectively, $k_1^i = \omega_1^i a_1^i$ and $b_1^i = (1-\omega_1^i)a_1^i$, and given by*

$$k_1^i = \widetilde{\omega}\frac{\vartheta^i}{\overline{\vartheta}}(a_0^i)^{\gamma}\ , \quad b_1^i = \frac{1}{\widetilde{\beta}}\Bigg(\underbrace{(\widetilde{R}\beta)^{\sigma}a_0^i - ((\widetilde{R}\beta)^{\sigma} + \underline{r} + \phi)k_1^i - \varphi Y}_{\text{Intertemporal substitution}} + \underbrace{\frac{1}{2}\alpha^i\sigma^2_{c_2^i}}_{\text{Precautionary savings}}\Bigg).$$

Lemma 5 is a generalized counterpart to Proposition 1 in Angeletos and Calvet (2006) derived under a baseline CARA specification. The two-period structure leads to an intertemporal substitution effect due to the risky asset holdings and a precautionary savings effect that arises from uncertainty about the realization of second period consumption c_2^i. In this economy, the effects of a change in the shape of the wealth distribution on output can be characterized by means of sufficient statistics.

Corollary 1 (EFFICIENCY AND INEQUALITY). *The second period output in this economy is given by $Y = \mathbb{E}[b_1(\vartheta, a_0)]A + (\mu\phi + (1-\mu)A)\mathbb{E}[k_1(\vartheta, a_0)]$. The effect of a change in inequality on wealth-normalized output is*

$$\frac{\partial\widetilde{Y}}{\partial\eta} = A\cdot Cov\left(mps^i, \frac{da_0^i}{\mathbb{E}[a_0]}\right) + \mu(\phi - A)\cdot Cov\left(mpr^i \times mps^i, \frac{da_0^i}{\mathbb{E}[a_0]}\right),$$

where $mps^i \equiv \frac{\partial a_1^i}{\partial a_0^i}$ and $mpr^i \equiv \left(\frac{\partial k_1^i}{\partial a_1^i}\right)/\left(\frac{\partial a_1^i}{\partial a_0^i}\right)$ are the marginal propensity to save and to take risk.

Corollary 1 characterizes the overall effect of a wealth inequality change on aggregate efficiency into sufficient statistics in an economy with capital accumulation. In such an economy, type and scale dependence determine the extent to which agents invest in risky assets, captured by the distribution of mpr^i, and also how much they accumulate wealth, captured by the distribution of mps^i. As such, both covariance terms depend on γ, ρ and η in the aggregate. The quantitative setting in Section 1.3 also incorporates an explicit saving decision into a dynamic framework.

Extensions

Other sources of *scale* dependence Appendix OA 1.3.1 introduces an entrepreneurship type of model along the lines of Cagetti and De Nardi (2006a), Guvenen et al. (2019) or Brüggemann (2021). The model is shown to map into the representation of equation (1.9). In this setting, we show that wealth-normalized output depends negatively on wealth inequality due to decreasing returns to scale on private equity investments (negative wealth-dependence $\gamma < 1$), but positively with the selection of entrepreneurs at the top of the distribution (positive type-dependence $\varrho > 0$).

As stated before, such models are located in the bottom-right area of the inequality-efficiency diagram of Figure 1.1. Second, we consider the case of wealth-dependent borrowing constraint and show that it generates similar results as the one derived above under wealth-dependent risk-aversion.

Aggregate shocks Throughout the paper, we assumed that investment return risk is idiosyncratic following the findings of Bach et al. (2020) on private equity, which represents the largest share of wealth in the hands of the wealthy. We now check how our insights change under the assumption of aggregate production risk. Therefore, we assume that the productivity of innovative projects is stochastic and given by $z(\mu\phi + (1-\mu)A)$ with $z \sim \mathcal{N}(1, \sigma_z^2)$ an aggregate shock. In this case, a *growth – variance trade-off* arises as an increase in wealth inequality does not only affect expected growth, but also its volatility. A social planner seeking to redistribute wealth has an additional incentive to stabilize the wage rate, pushing towards less inequality when higher inequality is linked to higher aggregate risky investments. Under the special case of Section 1.2.2, this creates a positive link between inequality and wealth-normalized output volatility $\sigma^2_{\tilde{Y}}(\eta)$ if a certain model economy falls into the growth-enhancing region regarding type and scale dependence. Overall, the main insights and trade-offs of interest remain valid under this assumption.

1.3 A Dynamic Quantitative Model with Investment Heterogeneity

Section 1.2 derived an analytical representation of the link between wealth inequality, aggregate output and welfare. We isolated four key parameters: (i) the Pareto tail of the wealth distribution, (ii) the elasticity of risk-taking to wealth, (iii) the sorting of types along the wealth distribution, and (iv) the extent to which returns to wealth reflects higher investment productivity. Focusing on those elements, we now build a quantitative model in which the wealth distribution arises endogenously, and study the distinct effects of type and scale dependence for wealth taxation.

1.3.1 Environment

The distribution of wealth arises endogenously from two empirically relevant features: heterogeneity in labor productivity as in a standard incomplete markets model (Aiyagari, 1994) and heterogeneity in capital investment and associated returns. While Benhabib et al. (2011, 2019) and Hubmer et al. (2020) show that the latter is key to generate the right tail of the wealth distribution, we study the role of the sources of this heterogeneity, type and/or scale dependence, for wealth accumulation and redistribution. We also introduce a life-cycle structure with endogenous labor supply which has been previously demonstrated in the literature to have potential for generating positive capital taxes (Conesa et al., 2009). Apart from those elements, the rest of the model is kept deliberately parsimonious.

Demographics, preferences and endowments

Time is discrete. Households derive a per period flow of utility $u(c_t^i, \ell_t^i)$ from consumption c_t^i and labor supply ℓ_t^i. At the beginning of each period agents differ in their wealth a_t^i, their age bracket j_t^i, their permanent component of labor productivity h_t^i, and their innate risk taking or investment skill type ϑ_t^i. They discount future periods at rate $\beta \in (0,1)$ and die with probability $d_j^i \equiv d(j_t^i)$. The expected life-time utility is given by

$$\mathcal{W}^i = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left(1 - d_j^i\right)^t u(c_t^i, \ell_t^i) \right] . \tag{1.12}$$

Unless necessary, we drop the time and households indexes. Our model incorporates stochastic aging to capture the dynamics of income and wealth accumulation over the life-cycle. We thus assume that agents live through a discrete number of stages, i.e. $j \in \mathcal{J} \equiv [1, \ldots, J]$. From stage 1 to $J-1$ households participate in the labor market. Stage J comprises households beyond retirement. The probability of switching between age brackets j and $j+1$ is denoted by $\pi_j(j+1|j)$. Upon death, a household is replaced by a newborn household who inherits their wealth. There are no annuity markets such that households leave unintended bequests.

Individuals who work earn pre-tax labor income defined by $wy\mathcal{H}(h)\zeta_j\ell$, where w is the equilibrium wage rate, $y \sim F_y(y)$ and $\mathcal{H}(h)$ denote respectively the transitory and the persistent labor productivity components, while ζ_j is the age component of earnings. The evolution of the persistent component follows a first order Markov chain with transition probability $\pi_h(h'|h)$. At retirement, we assume that the working ability h stays constant over time, such that pensions are given by $w\mathcal{H}(h)\zeta_J$, with ζ_J defining the replacement rate. Upon death, a newborn imperfectly inherits the persistent component of her parents. With probability p_h she draws her parent's persistent labor productivity, and with probability $(1 - p_h)$ she draws her productivity from the invariant distribution $F_h(h)$ generated by $\pi_h(h'|h)$.

The risk-taking type $\vartheta \in \{\vartheta_1, ..., \vartheta_S\} \in \Theta$ follows a Markov chain with transition probability $\pi_\vartheta(\vartheta'|\vartheta)$. A newborn draws her parent's risk-taking type with probability p_ϑ and from the invariant distribution $F_\vartheta(\vartheta)$ otherwise.

Households are heterogeneous in their capital investments. They split their savings into safe and risky assets. An agent with risk-taking type ϑ and wealth a invests a share $\omega(a, \vartheta)$ in the risky asset. For the sake of clarity, our portfolio specification should be understood as a reduced form of a more elaborated portfolio choice. Let r_F and r_R, with $r_R > r_F$ be the safe and risky net returns determined in equilibrium, respectively. The pre-tax return on total investment is given by

$$r(a, \vartheta, \kappa) = \underline{r} + r_F \cdot (1 - \omega(a, \vartheta)) + (r_R \kappa) \cdot \omega(a, \vartheta) \,, \tag{1.13}$$

where $\underline{r}$ is an aggregate return component and $\kappa \sim F_\kappa(\kappa)$ an idiosyncratic element of *luck*. The variance of returns is therefore given by $\sigma_r^2 = (r_R\omega(a,\vartheta))^2\sigma_\kappa^2$ and implies that households with higher equity shares experience higher portfolio risk. Such a feature is supported in the PSID, and documented by Bach, Calvet and Sodini (2020) and Fagereng, Guiso, Malacrino and Pistaferri (2020). From equation (1.13), it is clear that returns are correlated over time through wealth itself, and through the process governing the evolution of investment skill-type ϑ.

Agents optimally choose their saving a', labor supply ℓ, consumption c and cannot borrow. Their recursive program is

$$v(a,\vartheta,h,j) = \mathbb{E}_{\kappa,y}\left\{\max_{c>0,\,a'\geq 0,\,\ell\geq 0}\left\{u(c,\ell) + \beta(1-d_j)\mathbb{E}_{j',h',\vartheta'|j,h,\vartheta}\left[v(a',\vartheta',h',j')\right]\right\}\right. \tag{1.14}$$

$$\text{s.t.} \quad c + t_c(c) + a' = \mathcal{Y}^{inc} - t_w(\mathcal{Y}^{inc}) + r(a,\vartheta,\kappa)a - t_r(r(a,\vartheta,\kappa)a) + a - t_a(a)\ , \tag{1.15}$$

$$\mathcal{Y}^{inc} = w\mathcal{H}(h)\left(\mathbb{1}_{\{j<J\}}\zeta_j y\ell + \mathbb{1}_{\{j=J\}}\zeta_J\right)\ , \tag{1.16}$$

where the functions $t_r(\cdot)$, $t_w(\cdot)$, $t_c(\cdot)$ and $t_a(\cdot)$ are taxes on capital income, labor income, consumption and wealth. Upon death, bequests are taxed such that $a^{child} = a^{parents} - t_b(a^{parents})$.[24]

Production, government, and equilibrium

Production The production sector is similar to the one in section 1.2. An intermediate producer operates at no cost a continuum of projects in sectors $s \in \{N, I\}$. Each project uses assets supplied by a household with wealth holdings a and skill type ϑ to produce x intermediate goods with technology

$$x^N(a,\vartheta) = A[(1-\omega(a,\vartheta))a]^{\nu_N}\ , \qquad x^I(a,\vartheta) = (\phi\mu + A(1-\mu))[\omega(a,\vartheta)a]^{\nu_I}, \tag{1.18}$$

where $\phi > A$ holds and $(\nu_N,\nu_I) \in \mathbb{R}_+^2$ are returns to scale on the technologies. Similar to section 1.2, μ is a wedge capturing the extent to which returns on risky investments reflect the associated capital productivity. The intermediate producer sells intermediate goods to a final good producer at price $p^s(a,\vartheta)$ and obtains revenues $\Pi(a,\vartheta) = \sum_s p^s \cdot x^s(a,\vartheta)$, which are redistributed to investors. Recall that intermediate producers do not face any risk, however, investors are exposed to the investment shock κ.

A competitive final good producer uses labor L and intermediate goods $X = \sum_s\left(\int_i x_i^s di\right)$ to produce with technology $Y = F(X,L)$, where $F(\cdot)$ satisfies the Inada conditions. Profit maximiza-

[24]The outer expectation comes from the fact that y and κ are *iid*. An alternative way to write this value function is

$$v(a,\vartheta,h,j,\kappa,y) = \max_{c>0,\,a'\geq 0,\ell\geq 0}\left\{u(c,1-\ell) + \beta(1-d_j)\mathbb{E}_{j',\kappa',y',h',\vartheta'|j,h,\vartheta}\left[v(a',\vartheta',h',j',\kappa',y')\right]\right\}. \tag{1.17}$$

tion, i.e. $\max_{x_i^s, L} Y - \sum_s \int_j (p_i^s + \delta) x_i^s \, di - wL$, yields the following set of prices: $p_i^s = \frac{\partial F(X,L)}{\partial X} \frac{\partial X}{\partial x_i^s} - \delta$, and $w = \frac{\partial F(X,L)}{\partial L}$, where $\delta \in (0,1)$ is the depreciation rate. As intermediate goods are perfect substitutes, it follows that $p_i^s = p \; \forall i, s$.

Given the intermediate goods equations (1.18), the return wedge and the profit maximization, the returns to wealth to safe and risky asset investments are given by

$$r_F := \frac{px^N(a,\vartheta)}{(1-\omega(a,\vartheta))a} = MPK_F = pA[(1-\omega(a,\vartheta))a]^{\nu_N - 1}, \tag{1.19}$$

$$r_R := \frac{px^I(a,\vartheta)}{\omega(a,\vartheta)a} = p\phi[\omega(a,\vartheta)a]^{\nu_I - 1} \geq MPK_R = p(\phi\mu + A(1-\mu))[\omega(a,\vartheta)a]^{\nu_I - 1}, \tag{1.20}$$

where $\mu < 1$ implies $r_R > MPK_R$. This thus describes the case in which risky returns to wealth do not only reflect investment productivity but also some form of rent-extraction, for example.

Government The government finances an exogenous expenditure level G as well as social security retirement pensions. It raises total revenues from consumption, capital income, bequest, wealth, and labor income taxes. Consumption, capital income and labor income are subject to a linear tax, respectively $t_c(\mathrm{x}) = \mathrm{x}\tau_c$, $t_r(\mathrm{x}) = \mathrm{x}\tau_r$, and $t_w(\mathrm{x}) = \mathrm{x}\tau_w$. There is no wealth tax in the baseline economy, i.e. $t_a(\mathrm{x}) = 0$. We will however study the case of a progressive wealth tax in section 1.6. The bequest tax is also linear, $t_b(\mathrm{x}) = \tau_b(\mathrm{x} - t_a(\mathrm{x}))$, where we assume that the wealth tax is paid first. Consequently, the government budget is given by

$$\begin{aligned} G + \int_{(a,\vartheta,h)} w\mathcal{H}(h)\zeta_J \, d\mathcal{G}(a,\vartheta,h,j=J) = \int_{(a,\vartheta,h,j)} \Big(\tau_w \int_y \mathbb{1}_{\{j<J\}} w\ell\mathcal{H}(h)y\zeta_j \, dF_y(y) + \\ + \tau_r \int_\kappa r(a,\vartheta,\kappa)a \, dF_\kappa(\kappa) + \tau_c c(a,\vartheta,h,j) \\ + \tau_a(a) + \tau_b d_j(a - t_a(a))\Big) d\mathcal{G}(a,\vartheta,h,j). \end{aligned} \tag{1.21}$$

1.3.2 Equilibrium

In each period t, the aggregate state of the economy is described by the joint measure $\mathcal{G}_t$ over asset positions, labor productivity, investment skill-type and age.

Definition 2. *Denote the state space by $\boldsymbol{s} = (a,\vartheta,h,j) \in \mathbb{S} \equiv \mathbb{R}_+ \times \Theta \times \mathbb{H} \times \mathcal{J}$. A steady-state equilibrium of this economy is a vector of quantities $\{Y, X, L\}$, a set of policy functions $\{c(\boldsymbol{s}), a'(\boldsymbol{s}), \ell(\boldsymbol{s})\}$, a set of prices $\{p, w, \underline{r}\}$, a set of tax functions $\{t_w, t_r, t_c, t_b, t_a\}$, and a probability distribution of households $\mathcal{G}$ defined over $\mathbb{S}$, such that*

(1) The representative final producer maximizes profits, i.e. $\max_{\{X,L\}} F(X,L) - (p+\delta)X - wL$, where p and w are given by their respective marginal products.

(2) Given prices, households solve the stationary version of their decision problem (1.14), giving rise to an invariant distribution $\mathcal{G}(\boldsymbol{s})$.[25]

[25] At each state (a,ϑ,h,j), there is a continuum of individuals experiencing the *iid* shocks y and κ. The stationary

(3) The government budget constraint (1.21) is satisfied.

(4) Labor and intermediate goods markets clear, i.e.

$$L = \int_y \int_{(a,\vartheta,h,j)} \mathbb{1}_{\{j<J\}} w\ell\mathcal{H}(h)y\zeta_j \, d\mathcal{G}(a,\vartheta,h,j) \; dF_y(y) \;, \tag{1.22}$$

$$X = \int_{(a,\vartheta,h,j)} \Big(A[(1-\omega(a,\vartheta))a]^{\nu_N} + (\phi\mu + A(1-\mu))[\omega(a,\vartheta)a]^{\nu_I} \Big) \, d\mathcal{G}(a,\vartheta,h,j) \;. \tag{1.23}$$

(5) The total capital product distributed by the intermediate producer for each project $\Pi(a,\vartheta)$ is consistent with the total capital income received by households, i.e.

$$pX = \int_\kappa \int_{(a,\vartheta,h,j)} \Big(\underline{r} + r_F(a,\vartheta)(1-\omega(a,\vartheta)) + r_R(a,\vartheta)\kappa\omega(a,\vartheta) \Big) a \; d\mathcal{G}(a,\vartheta,h,j) \; dF_\kappa(\kappa). \tag{1.24}$$

Finally, by Walras Law, the good market clears. Moreover, our measure of total wealth in this economy is given by $K = \int_{(a,\vartheta,h,j)} a \, d\mathcal{G}(a,\vartheta,h,j)$.

Condition (1.23) states that the total efficiency units of capital used in the production sector must correspond to the total capital supplied by households given their investment choice between risky and safe assets. When $\mu < 1$, each unit of risky investment yields returns to wealth higher than its corresponding marginal product. As such, the equilibrium base return $\underline{r}$ must be negative to satisfy condition (1.24). Therefore, besides X/L that pins down p and w, the aggregate return component $\underline{r}$ in equation (1.13) is the second object that adjusts in equilibrium.

Numerical solution

The model admits no analytical solution. We solve it numerically using a version of the endogenous grid method (Carroll, 2006). Appendix 1.C.2 describes the algorithm. Under certain calibrations, e.g. the one of the benchmark economy, the model induces a Pareto distribution of wealth at the top. As a consequence, the support of the stationary distribution of wealth is unbounded from above. To circumvent this issue in practice, we use a large value for the upper bound on wealth in our numerical implementation.[26]

distribution is obtained using non-stochastic simulations.

[26] To check the size of the error implied by this truncation, we estimate the Pareto tail, $\widehat{\eta}_\xi$, of the model-generated wealth distribution at the top above the a_ξ wealth threshold, which is chosen large enough such that the upper part of the distribution is indeed Pareto. The details of the estimation procedure are similar to the one implemented in Appendix 1.B.1. In the truncated model, for $a \geq a_\xi$, there is a mass $\mathcal{G}_\xi$ of households at the top with total wealth K_ξ who thus hold a fraction $\theta_K = K_\xi/K$ of total wealth in the economy. We then reconstruct a Pareto distribution with shape $\widehat{\eta}_\xi$ and scale a_ξ, and compute the implied fraction $\widehat{\theta}_K$ from this theoretical distribution. We increase the upper bound of the grid on wealth until $|\widehat{\theta}_K - \theta_K|$ is negligible.

1.4 Taking the Model to the Data

Before studying wealth taxation, we first choose a parameterization that allows the model to closely replicate key features of the US economy. We begin with a discussion of the data used.

1.4.1 Capital Heterogeneity: A First Glance into the Data

We use the SCF and the PSID micro datasets. Our concept of wealth is *net worth*, which is defined as the market value of all assets minus total liabilities. Assets comprise riskless assets (deposits, savings, cash), direct and indirectly (mutual funds, individual retirement accounts) held public equities, net private equity business investments, primary and secondary residences, and other non-financial assets. Liabilities comprise student loans, mortgages, consumer credits and other loans. We restrict the sample population to those aged 20–70, essentially to focus on decision makers. We define private and public equity as productive risky assets while assets such as riskless assets, residential properties and other non-financial assets are considered as safe productive assets.[27] This classification is in line with the literature; for example, Cagetti and De Nardi (2006a) separate private equity assets from other investments into corporate firms, while Kaplan et al. (2018) consider only equity and commercial or business real estate as productive assets.

Our SCF waves span from 1989 to 2019 and include detailed information on households' portfolio composition, comprising a number of very wealth-rich households. When computing moments related to wealth inequality, we will sometimes refer to the *adjusted SCF*. The adjustments are made using the method of Vermeulen (2016) for *all* periods. First, we correct for underreporting of assets by adjusting survey estimates of real assets, financial assets, and liabilities such that they align with aggregate national balance sheets. Second, we adjust for under-representation at the very top by merging the SCF with households from the Forbes World's Billionaires lists and extrapolate wealth shares based on estimating a Pareto Law. Appendix 1.B.1 details the procedure.

Additionally, we use PSID waves from 1998 to 2018 to compute empirical moments of investment returns and extract complementary information on investors. The PSID constitutes a large and representative biennial survey with a long panel dimension that contains information by broad asset classes on capital income and costs, asset prices, inflows and outflows.[28]

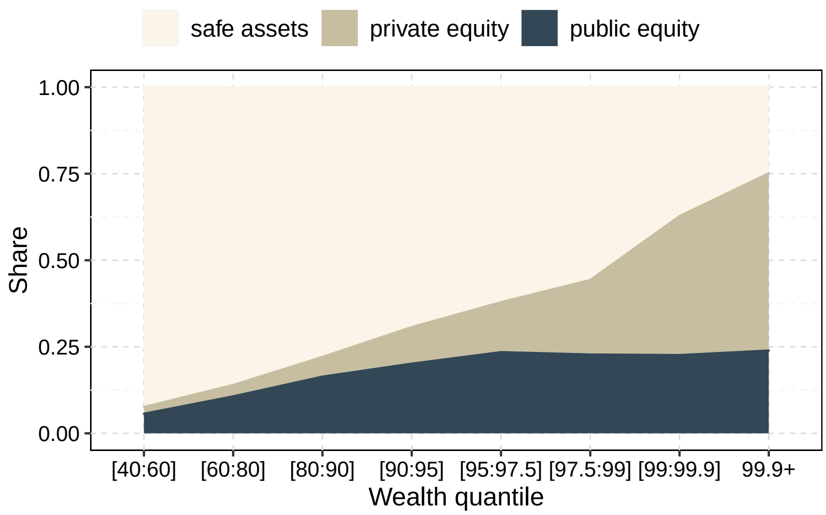
Portfolio composition. In Figure 1.3 we first use the SCF to get a full *cross-sectional* picture of the average household's portfolio composition across the US wealth distribution. As already documented in the literature, private and public equity investments correlate strongly and positively with wealth in the cross-section. The top 1% in the US hold on average 65% of wealth in risky

[27] Our motivation for classifying housing in safe assets comes from the fact that returns from housing assets are less volatile than other assets. Moreover, their volatility relates mostly to differential mortgage interest rates.

[28] See Pfeffer et al. (2016) for an excellent comparison between both surveys and Flavin and Yamashita (2002) for a discussion about returns estimated from the PSID. Appendix 1.B.2 provides further details. A drawback of the PSID is the presence of only a small number of households at the very top (within the top 1%) of the wealth distribution.

equity, while the corresponding share for the median household is 7%.[29]. As it can be seen, the noticeable increase in risky equity share at the very top of the wealth distribution (above the 95^{th} percentile) is mostly driven by assets held in private equity business investments.

Figure 1.3. Average portfolio share of gross assets by wealth percentile.



Source: adjusted SCF from 1989 to 2019, averaged from different SCF imputations.

Returns to wealth. We measure returns to wealth in the US using the PSID. We compute returns for each broad asset class $l \in \{riskfree, home, secondary, priv, public, other\}$, where $riskfree$ assets bundle savings, bonds and checking accounts, and $home$ denotes assets linked to the primary residence. As stated before, risky assets comprise public equities and non-financial private equity business investments. The return on asset class l for household i in year t is given by

$$r_{i,l,t} = \frac{R^K_{i,l,t} + R^I_{i,l,t} - R^D_{i,l,t}}{a^g_{i,l,t-1} + F_{i,l,t}/2} \, , \tag{1.25}$$

where $a^g_{i,l,t-1}$ is the beginning-of-period amount of asset class l held and $F_{i,l,t}$ are inflows minus outflows (i.e. net investment), that we divide by two assuming that they occur in mid-year.[30] The values $R^K_{i,l,t}$ and $R^I_{i,l,t}$ and $R^D_{i,l,t}$ correspond respectively to capital gains, asset income such as dividends, interests and other payments and to the cost of debts (if any). Returns to total net worth are similarly computed as $r_{i,net\,worth,t} = \frac{\sum_l R^K_{i,l,t} + R^I_{i,l,t} - R^D_{i,l,t}}{\sum_l (a^g_{i,l,t-1} + F_{i,l,t}/2)}$. Finally, nominal returns are converted to real returns using the consumer price index for each year. The complete and detailed procedure regarding the construction of returns is provided in Appendix 1.B.2.

Table 1.1 provides selected descriptive statistics on the returns to net worth (before-tax), to riskless asset, to public equity and to private equity. Notably, returns to private equity display the

[29]These numbers are comparable to estimates from detailed Swedish administrative data as in Bach et al. (2020), who use a similar definition of risky assets

[30]This is mostly due to reduce the bias due the acquisition and sale that generate large returns (i.e. division by zero).

highest expected returns (15.6%) with substantial heterogeneity and skewness to the right. To a lower extend, public equity generates also substantial returns (5.8%). Despite the absence of very wealthy households in the PSID, our results are comparable to estimates in Fagereng et al. (2020) in Norway and Bach et al. (2020) in Sweden using administrative data. As a direct comparison to US estimates, Xavier (2020) evaluates, using cross-sectional information from the SCF, that aggregate returns are 13.6% for private equity, 6.4% for public equity, and between 0.4%-2.1% on the different safe assets.[31] A noticeable difference, however, is that our aggregate estimate of returns to net worth (before-tax) is 3.3%, which is substantially lower than the one evaluated by Xavier (2020) (6.8%), but closer to the ones estimated using a quantitative structural model of inequality by Benhabib et al. (2019) (3.1%) and to the empirical estimates in Fagereng et al. (2020) for Norway (3.8%). In comparison to Sweden, Bach et al. (2020) find a median return to net worth of 4.5% with a standard deviation of 13% per year. We attribute this discrepancy to the fact that the PSID does not account for very wealthy households and to using different methodologies.

Table 1.1. Returns to wealth in the PSID.

Wealth component	Descriptive Statistics					
	Mean	St.Dev.	Skewness	Kurtosis	20^{th} perc.	80^{th} perc.
Net worth (before-tax)	0.033	0.158	0.897	6.243	-0.035	0.089
Private equity	0.156	0.614	2.071	10.967	-0.225	0.500
Public equity	0.058	0.417	-0.122	0.085	-0.248	0.385
Riskless assets	0.004	0.009	3.234	10.267	0.000	0.003

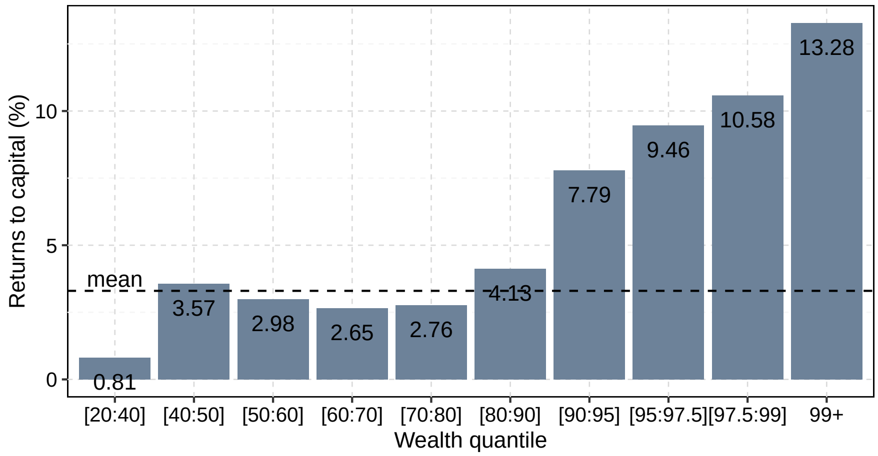
Note: we apply a trimming of 0.5% at the top and the bottom for each asset class.

Figure 1.4 documents the before-tax returns to wealth by wealth quantile in the PSID. Returns to wealth positively correlate with wealth in the cross-section, which is likely to be driven by heterogeneity in household's portfolio composition, as documented above.[32] These findings are consistent with existing work establishing a positive correlation between private equity ownership and wealth (Quadrini, 2000a; Cagetti and De Nardi, 2006a). Of course, those numbers are not informative on whether the correlation is driven by type or scale dependence. In practice, there is no obvious way to disentangle the two as both channels are likely to drive the observed cross-sectional relationship, as we will demonstrate subsequently.

[31] Her methodology is very different from ours. We use the panel dimension of the PSID to compute returns of a given household over time, while she computes returns from the SCF waves using cross-sectional information on capital income and stocks, averaged by wealth percentile. A drawback of her analysis is sample selection, as individuals may in principle move in and out of a given wealth percentile over time, as shown in Gomez et al. (2018). Contrary, a drawback of our analysis is the under-representation of very wealthy households in the PSID.

[32] This inference is shared by Bach et al. (2020) while Fagereng et al. (2020) find that there remains substantial heterogeneity within a broad asset class, which may reflect an important role for heterogeneity in skills. In fact, conditional on a broad asset class, it is difficult to disentangle whether higher returns are due to specific skills or due to higher risk-taking. For private equity investments, the latter may arise due to diversification motives (Penciakova, 2018) or due to the interaction between business risk-taking and borrowing limits (Robinson, 2012). In Appendix 1.B.2, we decompose the *cross-sectional* relation and find that this correlation is not observed *within* asset classes.

Figure 1.4. Average return on wealth by gross wealth quantile.



To sum up, the increasing share of risky assets at the top of the wealth distribution is substantially driven by private and public equity holdings displaying the highest expected returns. Among other possible determinants, the positive correlation between returns to wealth and wealth *itself* is thus likely to be driven by differing portfolio composition.

Type and scale dependence in capital investments in the US economy

Exploiting the panel dimension of the PSID by controlling for individual characteristics, Hurst and Lusardi (2004) find evidence for scale dependence in the propensity to select into private equity business ownership among the top 5% wealthiest households. Their estimates show that the average probability is flat at a rate of 3 percent for the bottom 80%. It increases to 4 percent for the 95^{th} percentile and reaches 7% for the 98^{th} percentile.

Using information on returns to capital endowments of US universities, Piketty (2018) (Chapter 12) finds that returns substantially increase with wealth and argue that this may arise from economies of scale in portfolio management.[33] However, US universities may not be representative of the US population as their investment strategies may substantially differ.

Bach et al. (2020) use Swedish administrative panel data and test for scale dependence in returns to wealth. According to them, even within a sample of twins and controlling for twin-pair fixed effects, there is strong evidence of scale dependence, especially at the top of the wealth distribution (Table 9, p. 2738). They argue that scale dependence is likely driven by changes in the individual portfolio composition, e.g. due to returns to scale on management costs or DRRA behavior. Controlling for individual and year fixed effects, Fagereng et al. (2020) use a Norwegian administrative panel on wealth tax records and regress the average return to wealth on the individual's wealth percentile at the beginning of the period. Both wealth (scale) and individual fixed

[33] In Piketty's words, "*The most obvious one is that a person with 10 million euros rather than 100,000, or 1 billion euros rather than 10 million, has greater means to employ wealth management consultants and financial advisors.*".

effects are found to be statistically significant. Their estimates imply that scale dependence alone explains 48% of the 18 percentage point return difference between the 10^{th} and 90^{th} net worth percentiles.

Finally, Robinson (2012) shows that wealthier private equity business owners tend to start relatively riskier business investments, with higher expected profitability. Penciakova (2018) confirms a similar pattern with data on US firms using the Census Bureau's Longitudinal Business Database, patenting data, and Compustat. She documents that private equity investors who diversify start riskier additional private equity investments. As we will show, this diversification among private equity owners occurs mostly among the wealthiest households.

All in all, this leads us to conclude that both type and scale dependencies are likely to drive the observed correlation between returns to wealth and wealth.

1.4.2 Functional Forms and Calibration

The benchmark model aims to capture features of the SCF and PSID described above. Apart from this, the model is parameterized to account for realistic government policy, demographics, labor income process and production technology.

We map the stationary equilibrium of the quantitative model to US data in two steps. We first fix some parameters based on model-exogenous information. In a second step, we calibrate the remaining parameters endogenously by numerically simulating the model to match particular empirical moments. Table 1.2 summarizes all parameters.

Exogenously set common parameters

Preferences and technology The model is calibrated to the US economy and the period is one year. Preferences over consumption c and labor supply ℓ are represented by a standard time-separable utility function of the form

$$u(c,\ell) = \frac{c^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \chi \frac{\ell^{1+\frac{1}{\lambda}}}{1+\frac{1}{\lambda}} \,. \tag{1.26}$$

We set the IES $\sigma = 0.5$, and the Frisch labor elasticity $\lambda = 0.6$ following Brüggemann (2021) and Kindermann and Krueger (2014). The disutility cost χ is chosen so that households spend, on average, 1/3 of disposable time on market work. The discount factor β matches a capital-output ratio $\frac{K}{Y}$ of 2.6, which is consistent with Kitao (2008a).

Demographics We model four age brackets. The first three brackets span from age 20 to age 65, each with a length of 15 years. The probability of switching from age bracket j to the next one $j+1$ equals accordingly $\frac{1}{15}$ for agents in the first three age brackets, while retired households stay

Table 1.2. Calibrated parameters.

PARAMETER	SYMBOL	VALUE	SOURCE/TARGET
Demographics and Preferences			
Survival probability	d_j	in text	Data (US social security statistics) [a]
Discount factor	β	0.925	Capital-output ratio of 2.6
Risk aversion	σ^{-1}	2.0	Conesa et al. (2009)
Frisch elasticity	λ	0.6	Kindermann and Krueger (2014)
Disutility of labor	χ	18	1/3 of time on market work
Labor productivity process			
Age component	ζ_j	in text	Guvenen et al. (2021)
Permanent component	$\{\rho_h, \sigma_h\}$	$\{0.95, 0.2\}$	Storesletten et al. (2004)
Transitory component	$\{\sigma_y\}$	0.15	Heathcote et al. (2010)
Pareto tail	$\{\eta_h, q_h\}$	$\{1.9, 0.9\}$	Piketty and Saez (2003)
Intergenerational corr. h	p_h	0.35	Chetty et al. (2014)
Low income state	y_0	0.08	12% households with zero wealth (SCF)
Investment			
Intergenerational corr. ϑ	p_ϑ	0.15	Fagereng et al. (2020)
Risky share param. – scale	in text	in text	Shape of portfolio distribution (SCF)
Risky share param. – type	$\underline{\omega}$	0.4	Conditional risky share (SCF)
Transition between types	in text	in text	Data (PSID)
Excess wealth return	ϕ	6.2	Top 1% wealth share of 0.36
Technology			
Labor share	$1-\alpha$	0.67	Data
Depreciation	δ	0.045	Assumption
Safe technology constant	A	1.0	Normalization
Returns to scale	$\{\nu_N, \nu_I\}$	1.0	Assumption
Government policy			
Consumption tax	τ_c	0.05	Conesa et al. (2009)
Labor income tax	τ_w	0.225	Guvenen et al. (2019)
Capital income tax	τ_r	0.25	Guvenen et al. (2019)
Bequest tax	τ_b	0.4	Data, statutory tax rate

[a] See https://www.ssa.gov/oact/STATS/table4c6.html.

in the terminal bracket J until death. The death probability d_j for each age bracket j is taken from the US social security statistics and is reported in Table 3.B.1.

Technology We specify $F(X,L) = X^\alpha L^{1-\alpha}$ with $\alpha = 0.33$ and set the depreciation rate to 4.5%. We normalize $A = 1$. In the baseline, we set $\mu = 1$ such that $r_F = MPK_r$, and consider later cases with $\mu < 1$ when analyzing the aggregate effects of wealth taxation.

As already stressed above, micro returns to scale on investment (ν_N, ν_I) constitute another important form of scale dependence on returns. Existing data from Bach et al. (2020) find no evidence on the presence of increasing or decreasing returns on risky investments, especially on private equity returns.[34] We thus set $\nu_N = \nu_I = 1$. However, it contrasts with the variety of entrepreneurship

[34] Furthermore, we use our PSID sample and estimate, fixing the broad asset class l, the effect of asset holdings $a_{i,l,t}$ on asset returns $r_{i,l,t}$ according to $\log(r_{i,l,t}) = \beta_l \log(a_{i,l,t}) + FE_i + FE_t + \epsilon_{i,l,t}$, where β_l reflects the returns to scale, while

Table 1.3. Life cycle earning profile, mortality and transition probabilities.

	ζ_j	d_j	$\pi_j(j+1\|j)$ conditional on survival			
			$[20,35)$	$[35,50)$	$[50,65)$	≥ 65
$[20,35)$	0.81	0.15%	0.933	0.067	0.0	0.0
$[35,50)$	1.00	0.4%	0.0	0.933	0.067	0.0
$[50,65)$	1.35	1.1%	0.0	0.0	0.933	0.067
≥ 65	0.40	9.9%	0.0	0.0	0.0	1.0

models assuming a DRS technology on risky private equity businesses investment (Cagetti and De Nardi, 2006a; Brüggemann, 2021; Guvenen et al., 2019). Therefore, as CRS is not a standard assumption, we will also investigate the case where $\nu_I < 1$.

Government policies The government levies consumption, labor income, capital income and bequest taxes which approximately equal the current rates of the US economy. Labor income and capital income tax rates are set to $\tau_w = 22.5\%$ and $\tau_r = 25\%$, which is consistent with Guvenen et al. (2019). The bequest tax rate is fixed to $\tau_b = 40\%$ according to the corresponding statutory tax rates. Finally, the consumption tax is set to 5%, which is consistent with the value used in Conesa et al. (2009). Given the tax rates in place, the share of total government expenditure to GDP is equal to 0.24 in the stationary equilibrium, respectively 0.17 without social security payments.

Labor income process As stated above, a household's labor productivity depends on three components: an age-dependent component ζ_j, a persistent component h and a transitory component y. The natural logarithm of the individual hourly wage of a household writes

$$log(w) + log(\mathcal{H}(h)) + log(y) + log(\zeta_j) \ . \tag{1.27}$$

The life-cycle average earning profile ζ_j is taken from Guvenen et al. (2021) (Supplementary Appendix, Figure C.36). Table 3.B.1 provides the parameter values used throughout this paper for each age bracket, including the transition probability matrix across age brackets.

The processes for labor productivity aim to generate a realistic earning distribution and contribute to the overall wealth inequality. We follow Hubmer et al. (2020) and improve the fit of the earnings distribution by assuming that the persistent component $\mathcal{H}(h)$ follows a lognormal AR(1) process with persistence ρ_h and variance σ_h^2. However, at the top of the income distribution, h is drawn from a Pareto distribution with shape $\eta_h > 1$,

$$\mathcal{H}(h) = \begin{cases} e^h & \text{if} \quad F_h(h) \leq q_h \ , \\ F^{-1}_{Pareto(\eta^h)}\left(\frac{F_h(h)-q_h}{1-q_h}\right) & \text{otherwise} \ , \end{cases} \tag{1.28}$$

FE_i and FE_t stand for household and year fixed effects. If β_l does not statistically differ from zero, the CRS hypothesis on returns cannot be rejected. Our estimates suggest that there is no statistical evidence for either IRS or DRS, even among private equity business holdings.

where $F_h(h)$ is the CDF of h and $F^{-1}_{Pareto(\eta^h)}(\cdot)$ the inverse CDF for a Pareto distribution with lower bound $F_h^{-1}(q_h)$ with $q_h \in [0,1]$. The persistent component is discretized into nine bins $h \in \mathbb{H} \equiv \{h_1, ..., h_9\}$. We set the threshold of the Pareto distribution to $q_h = 0.9$ and the shape to $\eta_h = 1.9$, consistent with 1990-2010 estimates for the US (Piketty and Saez, 2003). In line with Storesletten, Telmer and Yaron (2004), parameters of the persistent labor productivity component are set to $\rho_h = 0.95$ and $\sigma_h = 0.2$.[35] The correlation between parents' labor productivity with the one of their heir is $p_h = 0.35$, which is consistent with Chetty et al. (2014).

Following Heathcote et al. (2010), the transitory process follows a log-normal distribution, i.e. $y \sim \mathcal{LN}(0, \sigma_y^2)$, where $\sigma_y = 0.15$. The process y is discretized into three states using Gauss-Hermite quadratures. We further add a low income state y_0 that occurs with probability $\pi_y(y_0) = 7.5\%$ and reflects for instance involuntary unemployment or part-time work, independently of (y, h) and over time. We choose y_0 to generate a realistic fraction of individuals with zero wealth.

Calibrating capital heterogeneity and wealth returns

We now calibrate variables associated with returns to wealth. The standard deviation of risky returns σ_κ is set to 0.45; a value consistent with our PSID estimates for public and private equity (Table 1.1) and, for instance, Fagereng et al. (2020). The excess return parameter $\phi = 6.2$ generates a top 1% wealth share of 36% by controlling the dispersion of returns to wealth between households. As we will show later on, this value also produces a realistic distribution of returns to wealth.

To carefully pin down the type and scale dependence, one would need sufficiently detailed panel data on investment decisions to test for a statistical relation between portfolio composition, investment returns and wealth while controlling for household characteristics. Fagereng et al. (2020) and Bach et al. (2020) follow this strategy using administrative data, and find strong support that returns feature both type and scale dependence, acknowledging a crucial role for portfolio composition.[36] However, the results are conditioned by the statistical model used to estimate type and scale dependence, e.g. a linear model with fixed effects (type) and wealth percentiles (scale).

We pursue another strategy by recognizing that there are two common ways to generate scale dependence in models featuring private or public equity investments. On one hand, the extensive margin decision to invest might be wealth-dependent. This is the case in many occupational choice models in the presence of borrowing constraint (Cagetti and De Nardi, 2006a; Brüggemann, 2021) or models with fixed participation cost (Fagereng et al., 2017). On the other hand, conditional on

[35] For the sake of transparency, we reduce the computational burden by using a reduced transition matrix $\hat{\Pi}_h(h'|h)$ such that: $\hat{\Pi}_h(h'|h) = \begin{cases} \Pi_h(h'|h) & \text{if} \quad \Pi_h(h'|h) \geq \epsilon, \\ 0 & \text{otherwise.} \end{cases}$ with $\epsilon = 10e^{-6}$ and normalizing the matrix $\sum_{h'} \hat{\Pi}_h(h'|h) = 1$. This allows us to exploit the sparsity of the transition matrix.

[36] Moreover, notice that the role of private equity is substantial in Fagereng et al. (2020): *"All in all, heterogeneity in our most comprehensive measure of returns to wealth can be traced in the first place to heterogeneity in returns to private equity and the cost of debt and only partially to heterogeneity in returns to financial wealth"*.

being an equity investor, there might be an intensive margin effect such that the portfolio share of investors is *itself* a function of wealth.

To capture both margins, we first assume that there are two investment skill-types, $\vartheta \in \{\vartheta_1 = 0, \vartheta_2 = 1\}$, i.e. those with investment skills, and those without, respectively. The probability of switching from an unskilled to a skilled investor type is a function of wealth, such that

$$\pi_\vartheta(\vartheta'|\vartheta, a) = \begin{bmatrix} 1 - \overline{\pi}_\vartheta - \lambda(a) & \overline{\pi}_\vartheta + \lambda(a) \\ \underline{\pi}_\vartheta & 1 - \underline{\pi}_\vartheta \end{bmatrix}, \tag{1.29}$$

where the components $\underline{\pi}_\vartheta$ and $\overline{\pi}_\vartheta$ capture switching probabilities that are unrelated to wealth, e.g. time-variations in skills or preferences. In the data, the fraction of public equity investors is substantially larger than the fraction of private equity holders. Moreover, the adjustment margin of portfolio allocation at the top of the wealth distribution comes mainly from private equity business investments (cf. Figure 1.3). As such, households investing in public equity and who do not invest in private equity are counted as investor only when they hold more than 50 percent of their wealth in public equity. This assumption means that 15% of households are investors. We set the exit probability $\underline{\pi}_\vartheta = 0.10$ consistent with the PSID and choose $\overline{\pi}_\vartheta = 0.018$ to match the fraction of investors. We calibrate $\lambda(a)$ using the following parametric form:

$$\lambda(a) = \min\{\lambda_1(\max\{a - \underline{a}_\lambda, 0\})^{\gamma_\lambda}, \lambda_2\}. \tag{1.30}$$

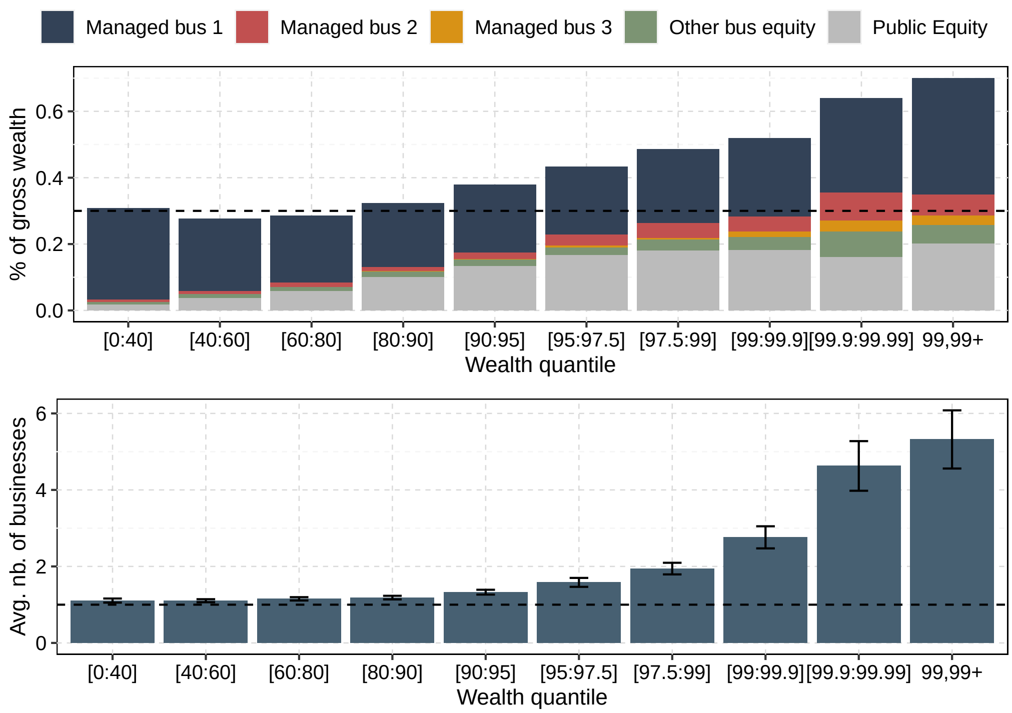
Using the results in Hurst and Lusardi (2004) derived from the PSID, parameters $\{\underline{a}_\lambda, \lambda_1, \lambda_2, \gamma_\lambda\}$ are chosen to generate the increasing probability to become an investor as a function of wealth, conditional on household's characteristics.[37] Consistent with the data, $\underline{a}_\lambda$ is set such that the probability to become an investor starts to be a function of wealth only above the wealth level corresponding to the 80^{th} wealth percentile. Below this wealth percentile, the participation rate is only generated through the process governing the ϑ-type, i.e. through $\overline{\pi}_\vartheta$. The level and the shape parameters of the transition probability with respect to wealth are endogenously set to $\lambda_1 = 0.071$ and $\gamma_\lambda = 0.30$ in order to replicate the average transition rate of 3.2% for households within the [95-97.5] wealth quantile, and of 6.1% for households within the [99-99.9] wealth quantile. A maximal value of $\lambda_2 = 0.045$ matches a transition rate of 7% for the top 1% wealthiest households.

We calibrate the intensive margin of risky investment as follows. In the SCF, conditioning on being private equity investor, the share of risky equity increases with wealth in the cross-section. One practical issue, however, is to distinguish whether the share of private equity held by those owners increases because of systematic capital gains or whether it is the result of net investments. Even in the latter case, it is difficult to disentangle whether private equity owners who obtain

[37] To save on space, we relegate the full empirical analysis of this observed relationship to Appendix 1.B.3.

higher returns are more likely to end up at the top, or whether wealth itself induces households to undertake riskier investments. To circumvent those issues, we use the SCF and exploit detailed information on the biggest three household's private equity business investments and a bundle of the remaining ones, including their acquisition date and the share of wealth in each private equity.

Figure 1.5. Average share held in equity (top panel) and average number of private equity business investments (bottom panel), conditioning on investors.



Source: SCF (1989-2019). The dashed line represents the median.

The top panel of Figure 1.5 shows the average share of private equity investments per net worth quantile, decomposed in different business investments. The average share of private equity investment over total gross wealth appears to be strongly correlated with wealth, especially above the top 5%. This relation is driven by diversification at the top. From our standpoint, this pattern cannot be solely driven by type dependence. First, borrowing constraints are likely to prevent relatively wealth-poor households to invest in multiple businesses, limiting the concern regarding the possible reverse causality that an owner of multiple businesses is more likely to select over time at the top of the wealth distribution. Second, those additional businesses are generally newly founded; 85% of the second businesses were created in the past ten years relative to the survey date, and 67% were created within the past five years. As a comparison, 47% of the first main business of those private equity owners were created within the past ten years. Therefore, given that wealth accumulation is a slow process, it seems unlikely that multiple business owners at the top became rich due to multiple private equity investments. Instead, the decision to open additional private equity investments is likely, among other determinants, to be wealth-driven.[38]

We further elaborate on this point using two additional pieces of evidence. First, as shown in the bottom panel of Figure 1.5, the average *number* of private equity business investments substantially rises at the upper end of the wealth distribution. Again, looking at the timing of those additional businesses, the last acquired business is particularly recent relative to the main business. Finally, despite the lack of observations at the very top in the PSID, we use its panel dimension and confirm that investment in additional private equity business investments, conditional on being a business owner and controlling for individual characteristics, is statistically positively correlated with net worth. To save on space, we defer this additional evidence to Appendix 1.B.3.

We attribute the part of the observed increase in equity investments due to diversification in private equity investments in the top panel of Figure 1.5 to scale dependence in the model. We view this choice as conservative as it constitutes a *lower bound* on the effect of wealth on the equity share of investors. To match this increase within the model, we specify the portfolio share as:

$$\omega(a,\vartheta) = \vartheta\Big(\underline{\omega} + \varpi(a)\Big), \qquad \varpi(a) = \min\Big\{\omega_1\left(\max\left\{a - \underline{a}_{\omega}, 0\right\}\right)^{\gamma_\omega}, \omega_2\Big\}, \tag{1.31}$$

where $\underline{\omega} = 0.4$ is the average share invested in equity, conditional on being an investor.[39] From Figure 1.5, $\underline{a}_{\omega}$ is chosen to correspond to the wealth level of the 70^{th} wealth percentile in the model, such that there is no scale dependence in risky portfolio observed below this percentile. The level and the shape parameters are endogenously set to $\omega_1 = 0.072$ and $\gamma_\omega = 0.30$ to replicate the average share invested in risky equity (through additional investments) of 11% for households within the [95-97.5] wealth quantile, and of 20% for households within the [99-99.9] wealth quantile. A maximal value of $\omega_2 = 0.20$ matches the average risky portfolio share above the top 0.1%.

In Appendix 1.C.1, we report the model fit regarding the scale and type dependence parameters.

[38] This result is not due to composition effects. Even when focusing on single households, private equity investments increase with wealth through diversification. Moreover, additional businesses are different than the first established business: 70% of additional private equity investments are made in a different sector. Additionally, they are only slightly more represented in finance-related industries, thereby limiting the concern that it may constitute a financial affiliate company. Those facts challenge the widely held view that business owners are poorly diversified (Moskowitz and Vissing-Jørgensen, 2002) and are consistent with Penciakova (2018). Diversification occurs, but only at the very top. In numbers, 12% of business owners own multiple managed businesses in the US. Among the top 1%, however, this number increases to 40%.

[39] In an alternative calibration strategy, we used longitudinal information of returns to wealth in the PSID to calibrate portfolio shares $\omega(a,\vartheta)$ such that they are consistent with the shape of returns to net worth. However, it is difficult, given the small number of observations in the PSID, to test for non-linearity of scale-dependence at the top of the distribution. Moreover, empirical evidence suggests that scale-dependence occurs in both the extensive and the intensive margins of equity investments, thus introducing important interactions which are captured within our specification.

1.5 Properties of the Model

Before turning to the main wealth taxation experiment, we first discuss key properties of the calibrated model regarding wealth inequality and returns to wealth. To isolate the driving forces behind the results of our benchmark model, we describe versions of our model with various combinations of type and scale dependence. In these versions, parameters are *always* recalibrated to match the same targets as in the benchmark economy. We find that several combinations of both dependencies deliver close to observationally equivalent cross-sectional moments with respect to wealth and return distributions, but they imply distinct aggregate responses to a wealth tax.

Comparison with alternative models We will subsequently compare our benchmark model (denoted M1) with the following model versions:

[label=(M0)]A pure scale dependence model (scale-model) in line with the information acquisition model of Peress (2004) and the incomplete markets models of Meeuwis (2019) and Hubmer et al. (2020). In this version, we shut down type-dependence. All households in the economy now invest in risky assets (i.e. $\vartheta = 1$), yet the amount depends on wealth only. Parameters ω_1, ω_2 and γ_ω match the average portfolio of risky equity observed in the SCF data (cf. Figure 1.3). A pure type dependence model (type-model) that resembles a stylized version of the capitalist/entrepreneur framework along the lines of Moll (2014) or Gomez et al. (2016). There are no scale dependence effects, i.e. the wealth-dependent propensity to select as an investor is set to $\lambda(a) = 0$, the wealth-dependent portfolio component to $\varpi(a) = 0$, and parameters $\overline{\pi}_\vartheta$ and $\underline{\omega}$ are recalibrated to match the fraction of investors and the average portfolio share. Despite our focus on type versus scale dependence, one may ask how a version of the widely used capitalist/entrepreneur framework (Cagetti and De Nardi, 2006a; Kitao, 2008a; Guvenen et al., 2019) with heterogeneity in household investments and returns compares to our benchmark model and the data. The key difference relative to our benchmark stems from the assumption of decreasing returns to scale (DRS) on private equity investments. For this reason, we denote this alternative version the type–DRS entrepreneur model. Apart from this, there is limited scale dependence in equity investment, as those investments are often tight to a borrowing constraint proportional to wealth.[40] To closely replicate such a model version, we impose a risky asset investment share $\underline{\omega} = 1$, no return risk $\sigma_\kappa = 0$, DRS on the entrepreneurial technology $\nu_I = 0.9$, a transition matrix of entrepreneur skill-type ϑ taken from Cagetti and De Nardi (2006a) to obtain a fraction of entrepreneurs of 8%.[41] A no return heterogeneity model that is characterized by $\omega(a, \vartheta) = 0$.

[40]To be precise, those frameworks may feature additional scale dependence in the selection into private equity business investments through an occupational choice (worker versus entrepreneur). A given entrepreneur, however, can invest at most a fixed fraction of her wealth, $k \leq \lambda a$, where λ captures the tightness of the borrowing constraint. In Cagetti and De Nardi (2006a), this borrowing limit is endogenous, but turns out to be quasi-linear in wealth.

[41]In our view, this specification is close to the traditional entrepreneur/capitalist type of model where entrepreneurs

In this version, it is not possible to replicate the observed high concentration of wealth at the very top.

In each version, ϕ and β are recalibrated to match the $\frac{K}{Y}$ ratio and the top 1% wealth share.

1.5.1 Wealth Inequality and Returns to Wealth

As demonstrated in the analytical framework of section 1.2, it is important that the quantitative model captures well the shape of the wealth distribution, especially at the very top, as this conditions the relative strength of type and scale dependence effects. In Table 3.6.4, we first assess the models' accuracy in generating a realistic wealth distribution relative to its empirical counterpart. A striking result is that, beyond the targeted top 1% wealth share, the benchmark model and the alternatives (i.e. models M1 to M4) account remarkably well for the empirical top wealth shares. We do not consider the ability of our model to reproduce the wealth distribution as a success *per se*, since return heterogeneity has been shown to generate high concentration of wealth (Benhabib et al., 2011). However, the fact that our benchmark economy indeed successfully replicates wealth inequality allows us to appropriately study tax experiments that are highly redistributive across the wealth distribution. Moreover, our results point to the observation that type and scale dependence may not be distinguishable based on their ability to generate high inequality, as both mechanisms actually deliver a good fit of top wealth shares.[42]

Table 1.4. Wealth distribution in the data and models.[a]

		Gini [c]	Share of wealth (in %) held by the top x%						
			40	20	10	5	1	0.1	0.01
US data (World Inequality Database)		0.82	97.5	85.1	70.6	57.7	35.5	18.0	9.0
US data (adjusted SCF)[b]			97.2	86.4	72.7	59.7	37.2	17.8	7.3
M1	benchmark model	0.80	93.4	84.2	71.9	59.3	35.4	18.2	8.9
M2	scale model	0.82	94.6	85.7	73.6	60.3	35.2	20.7	11.7
M3	type model	0.78	92.5	82.0	67.1	56.2	35.7	20.2	10.9
M4	type–DRS entrepreneur model	0.78	93.3	82.0	68.9	56.6	35.9	14.7	4.9

[a] The top one percent wealth share is targeted.
[b] Adjusted for under-representation and underreporting using the procedure in Vermeulen (2016).
[c] The wealth Gini is based on the average estimated from the SCF waves from 1989 to 2019.

This high concentration of wealth can be traced back to substantial heterogeneity in wealth returns implied by the different equity portfolio allocations among households. Table 1.5 shows

invest the total amount of their assets in private equity investments subject to a DRS technology. It induces an optimal amount of equity that a household would like to invest. This maximum is never reached under our parameterization. Notice that DRS is often imposed as a relevant assumption in the firm dynamics literature (Lucas Jr, 1978).

[42]This observation goes back to Benhabib, Bisin and Luo (2019) (Table 9). They find that including scale dependence in returns to wealth in a model with type dependence does not provide further explanatory power on the model ability to match top wealth inequality.

that consistent with estimates from various data sources, the benchmark model and the alternatives produce average returns to wealth which increase along the wealth distribution.

In the type-model (M3), the increase in returns to wealth is driven by selection only. As investment skill-types are persistent, households with a high propensity to invest in equity have higher expected returns for several periods and are thus more likely to be represented at the top of the distribution, hence driving the observed cross-sectional relationship.[43] In the scale-model (M2), the relationship is generated intuitively, as higher levels of wealth are associated with higher risk-taking and higher expected returns. In the benchmark model (M1), both scale and type dependence drive the observed pattern. To see this, we report the average returns to wealth across the wealth distribution assuming that the *pure orthogonal* type component in (1.43) is zero, i.e. $\underline{\omega} = 0$, and assuming no scale dependence on the intensive margin, i.e. $\varpi(a) = 0$. In line with the aforementioned empirical evidence in Fagereng et al. (2020) and Bach et al. (2020), we find that both forces shape average returns along the wealth distribution. In contrast, in the type–DRS model (M3) the scale dependence shifts sign due to the DRS specification, and the overall shape of returns across the wealth distribution is now hump-shaped, i.e. it decreases at the very top. Such a negative dependence in returns is not observed in data.[44] Finally, it should be noted that the idiosyncratic *luck* component κ contributes to the overall wealth inequality (Benhabib et al., 2011). In the benchmark model, the standard deviation of returns to wealth is 12.5%, compared to 15% in the PSID. Finally, in Appendix 1.C.3 we show that the 5 years wealth mobility matrices from the models M1 – M3 are comparable to the one obtained from the PSID.

Table 1.5. Mean returns to wealth (in %) along the wealth distribution: data and model.

	Data[a]				Model						
	PSID	SCF	Norway	Sweden	benchmark[b]			scale	type	type–DRS	
Wealth group					M1	type $\varpi(a)$=0	scale $\underline{\omega}$=0	M2	M3	M4	scale effect
P40-P50	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF	REF
P50-P60	−0.6	|	∼ 1.0	0.2	0.0	0.0	0.3	0.0	0.1	0.1	−0.2
P60-P70	−0.9	−0.4		0.3	0.3	0.3	0.6	0.0	0.6	0.7	−0.5
P70-P80	−0.8	0.0	∼ 2.5	0.3	0.9	0.9	1.0	0.3	1.3	1.6	−0.7
P80-P90	0.5	0.2		0.5	1.7	1.7	1.6	1.4	2.1	2.7	−1.0
P90-P95	3.8	1.4	∼ 4.0	0.8	3.6	1.5	2.0	3.8	2.7	3.4	−1.4
P95-P97.5	5.8	2.6	∼ 6.0	1.1	4.6	1.3	3.3	5.9	2.9	5.6	−1.6
P97.5-P99	6.9	3.8		1.5	7.9	3.9	4.0	7.8	7.4	9.9	−1.9
Top 1%	9.6	4.6	∼10.0	2.5	12.5	7.0	5.5	12.2	9.8	7.0	−2.4

Note: "REF" stands for reference wealth bracket, i.e. returns are computed as the difference to the REF.
[a] Estimates are our own for the PSID. They are taken from Xavier (2020) for the SCF, from Bach et al. (2020) for Sweden and from Fagereng et al. (2020) and Halvorsen et al. (2021) for Norway.
[b] The returns are computed assuming $\underline{\omega} = 0$ in the no type model and $\varpi(a) = 0$ in the no scale model.

[43]See Benhabib et al. (2011) and Moll (2014) for a theoretical illumination on the role of persistence in capital returns.

[44]In Bach et al. (2020), there is a slight decrease in private equity returns with respect to wealth due to leverage.

Decomposition In Table 3.6.4, we then ask how much type and scale dependence contribute to the observed wealth inequality in the calibrated benchmark model. Specifically, we shut down one component at a time and recompute the stationary distribution without this component, keeping everything else unchanged. Counterfactual (D1) isolates the orthogonal type dependent component by assuming $\underline{\omega} = 0$. Counterfactual (D2) isolates the effects of scale dependence by shutting down both intensive and extensive margin scale effects. Lastly, type and scale dependence in portfolio choices are jointly shut down in counterfactual (D3). Both type and scale dependence are important drivers of wealth inequality, as evidenced by the lower top wealth shares under those counterfactuals. As expected, a model without heterogeneity in capital investments fails to account for the high wealth concentration observed in the data. In such cases, the tail of the wealth distribution inherits the (lower) tail of the labor income distribution. In the model, the income Pareto tail is 1.9 against 1.4 for the empirical wealth distribution from the adjusted SCF.

Table 1.6. Wealth distribution under alternative model counterfactuals.[a]

		Gini	Share of wealth (in %) held by the top x%						
			40	20	10	5	1	0.1	0.01
	benchmark model	0.80	93.4	84.2	71.9	59.3	35.4	18.2	8.9
D1	no pure type dependence, $\underline{w} = 0$	0.76	93.5	81.3	68.0	48.5	17.0	3.3	0.1
D2	no scale, $\lambda(a) = \varpi(a) = 0$	0.75	91.1	77.8	63.9	51.5	31.7	16.3	8.0
D3	no portfolio heterogeneity	0.63	86.4	68.2	48.4	31.3	8.7	1.1	0.0

1.5.2 Wealth Inequality – Output Relationship

We now analyze the response to a *permanent* wealth redistribution. Conditional on the strength of type and scale dependence, our goal is to give a sense of how the alternative model versions locate relative to a situation in which inequality is neutral (cf. the GIF in Figure 1.1). To do so, we compute the long-run effects of a 1 percent tax levied on the wealth of the top 1% wealthiest households. We report the responses in Table 1.7.

While models M1–M4 produce close to observationally equivalent wealth distributions, they substantially differ in terms of output response. Moving from the scale (M2) to the type (M3) model reduces output losses from -1.19 percent to -0.64 percent. Under scale-dependence, risky asset holdings and thus future wealth are a function of current wealth level itself. This generates a quantitatively strong behavioral dynamic self-enforcing multiplier, which leads to lower returns of the richest households and thus lowers top inequality. In the Type model, the responses are an order of magnitude lower as individuals continue to invest a given share of their wealth into equity, and thus experience high capital returns. The benchmark model (M1), a composite of positive type and scale dependence, falls in between the previous two alternatives. Interestingly, the type–DRS entrepreneur model (M4) generates a lower response, as the negative scale dependence coming from the DRS assumption counterbalances the positive type dependence arising from the sorting of skilled entrepreneurs at the top of the distribution. Finally, a model without portfolio heterogeneity (M5) (a standard Aiyagari (1994) type of model) produces a slight reduction in output of -0.10 percent and is much closer to growth neutrality. In that case, the response comes from a reduction of capital accumulation (K) and is small due to the quasi-linearity of saving decisions for individuals sufficiently far away from the borrowing constraint.

Table 1.7. Responses to a permanent 1% wealth tax levied on the top 1% wealthiest households.

		ΔGDP (in %)	Δ Top 1% (in pp deviation)	Semi elasticity
M1	benchmark	-0.77	-1.82	0.42
M2	scale	-1.19	-3.50	0.34
M3	type	-0.64	-1.77	0.36
M4	type–DRS	-0.36	-1.48	0.24
M5	no portfolio heterogeneity	-0.10	-0.80	0.13

Additional validation from cross-country evidence In the Online Appendix OA 3.1, we revisit the empirical cross-country relationship between inequality and GDP growth. We extend previous results relying mostly on the Gini coefficient and top income shares to a sample of 29 developed countries by complementing existing estimates of wealth concentration measures with own estimates constructed based on survey data.[45] An increase by 1 percentage point of the top 1% wealth share is associated with a 0.27 percent increase in the subsequent five years average GDP growth.[46] Using data from the most recent Penn World Table, a decomposition of the inequality-growth relationship shows that GDP growth responses to changes in the top 1% wealth shares are mainly reflected in changes of the Solow residual and physical capital accumulation, while the relation is not significant regarding human capital. These findings are in line with the reallocation channel between productive capital investment outlined in this paper.

Using estimates in Table 1.7, we find that in the type (M3) and scale (M2) dependence models, a 1 percentage point decrease in the top 1% wealth share is respectively associated with a decrease in long-run GDP of 0.36 and 0.34 percent. In contrast, in absence of portfolio heterogeneity (M5), the relation substantially falls, to 0.13 percent. It is worth noting that the benchmark model (M1) produces a slightly stronger association. Among other things, this may be due to the fact that return heterogeneity may imperfectly reflect productivity differences due to some forms of rent-extraction, i.e. $\mu < 1$. In the next section, we study optimal wealth taxation while we allow returns

[45]A wide range of empirical papers studies the link between inequality and growth. For instance, Forbes (2000), Barro (2000) and Halter et al. (2014) use the income Gini coefficient as measure of inequality, while Barro (2008) and Voitchovsky (2005) use quintile and decile income shares. None of the previous papers explore the *Inequality-Growth-slope* using wealth concentration measures. To the best of our knowledge, only Voitchovsky (2005) and Frank (2009) look at the impact of income concentration at different quantiles.

[46]Note that we do not claim any causal relationship in here. Even in the model, a negative wealth shock at the top of the distribution affects inequality and GDP, and their combined change in turn feeds back into inequality and GDP. Therefore, it is hard to identify causality even based on our simulated results.

to imperfectly reflect investment productivity.

Implications for the dynamics of wealth inequality While Gabaix et al. (2016) show that type and scale dependence account for the recent rise of income/wealth inequality, their analysis is uninformative on whether the two are equivalent when both channels are calibrated to match the same empirical moments. Our results substantiate the idea that various degrees of type and scale dependence that are consistent with portfolio and return heterogeneity produce large difference when it comes to understanding the dynamics of wealth inequality. We find that the response of the top 1% wealth share to a wealth tax in the scale model (M2) is twice as large as the one found under the type model (M3). This may have large consequences, as Hubmer et al. (2020) find that changes in top capital income taxes are a key driver of the recent rise in wealth inequality observed in the US. In their words, *"the marked decrease in tax progressivity is by far the most powerful force for the cumulative increase in wealth inequality"*. However, this result is derived based on a model that features scale dependence in portfolio choices only. In the Online Appendix OA 2.1, we support our results in table 1.7 and show that changes in capital income taxes in the US since 1980 have substantially large differences on the dynamics of wealth inequality depending on whether returns are driven by scale or type dependence.

1.6 Wealth Taxation: the Role of Type and Scale Dependence

We now proceed to our main experiment and assess the quantitative implications of taxing household wealth at the steady state. We conduct our experiment in three steps. First, we compute the optimal wealth tax in our benchmark economy and decompose the resulting welfare gains. We find that a positive wealth tax above the top 80^{th} percentile is optimal. Second, we deviate from the benchmark calibration and study cases for which returns to wealth imperfectly reflect differences in capital productivity. We numerically argue that the existence of rents does not substantially change the welfare-maximizing tax rate relative to the one obtained under the benchmark calibration. Third, we dissect the key underlying forces behind our results and unravel the distinct effects arising from type and scale dependence.

Thought experiment We assume that tax rates cannot be personalized. The government uses a restricted class of wealth tax functions described by

$$t_a(a;\tau_a,\underline{a}_{max}) = \mathbb{1}_{a\geq \underline{a}_{max}}\tau_a(a-\underline{a}_{max})\,, \tag{1.32}$$

and optimizes over the exemption level $\underline{a}_{max}$ and the marginal wealth tax rate τ_a. Imposing restrictions on the class of wealth tax functions the government can choose from is necessary to ensure that the maximization objective is computationally feasible.[47] In equilibrium, the labor income tax τ_w adjusts to ensure that the government budget is balanced. In the Online Appendix OA 1.6, we provide further results when using alternative tax instruments to balance the government budget.

The main trade-offs shaping optimal redistribution are those identified in section 1.2. The wealth tax balances: (i) equity by reallocating wealth from households with low to high marginal utilities of consumption, (ii) efficiency as wealth-rich households trigger general equilibrium effects through asset reallocation and thus affect productivity and wages, and (iii) rent-extraction, as whenever $\mu < 1$, higher risky asset investments lead to an overall downward adjustment in the aggregate component $\underline{r}$ of returns to wealth.

The criterion to rank different wealth tax functions is based on a consumption-equivalent variation (henceforth CEV) approach. After solving for the stationary equilibrium of a specific tax reform $(\tau_a, \underline{a}_{max})$, we compute the variation Δ^{CEV} of consumption that makes every household in the post-reform economy on average as well-off as in the pre-reform economy. Under this *utilitarian* criterion, aggregate welfare in the *post*-reform economy, $\mathcal{W}^{post}(\tau_a, \underline{a}_{max})$, has to be equal to aggregate welfare in the status quo economy without wealth tax $\mathcal{W}^{pre}(\Delta^{CEV})$, where optimal consumption has been changed by Δ^{CEV} percent, such that

$$\mathcal{W}^{post}(\tau_a, \underline{a}_{max}) = \mathcal{W}^{pre}(\Delta^{CEV}) \, ,$$
$$\int_{\mathbf{s}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \widetilde{\beta}^t u\Big(c_t^{post}(\mathbf{s}), \ell_t^{post}(\mathbf{s})\Big) \right] d\mathcal{G}^{post}(\mathbf{s}) = \int_{\mathbf{s}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \widetilde{\beta}^t u\Big((1+\Delta^{CEV}) c_t^{pre}(\mathbf{s}), \ell_t^{pre}(\mathbf{s})\Big) \right] d\mathcal{G}^{pre}(\mathbf{s}) \, ,$$

where $\widetilde{\beta} = \beta(1-d_j)$. Given our time-separable utility function, it is straightforward to show that the government problem can be stated as follows

$$\arg \max_{\{\tau_a, \underline{a}_{max}\}} \Delta^{CEV}(\tau_a, \underline{a}_{max}) = \left[\frac{\mathcal{W}^{post}(\tau_a, \underline{a}_{max}) - \mathcal{W}^{pre}(0)}{\int_{\mathbf{s}} \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t (1-d_j)^t \frac{c_t^{pre}(\mathbf{s})^{1-\sigma}}{1-\sigma} \right] d\mathcal{G}^{pre}(\mathbf{s})} + 1 \right]^{\frac{1}{1-\sigma}} - 1 \, .$$

This welfare criterion is widely used in the quantitative macroeconomics literature (among many others Conesa et al. (2009), Guvenen et al. (2019) or Brüggemann (2021)). In our setting, this criterion relies, however, on a "representative utility" that ranks people according to the same preferences whatever the underlying scale or type dependent mechanism is. As such, we do not claim that this criterion captures all relevant welfare effects associated with a wealth tax, which may arise in reality due to a particular scale or type dependent mechanism. Our criterion implicitly assumes, for instance, that type and scale dependence do not alter household preferences. It should thus be understood as a transparent way to compare welfare consequences among vari-

[47] Furthermore, many developed countries adopted this wealth tax schedule. A typical wealth tax features an exemption level ranging from the wealth level corresponding to the top 50% (in Switzerland) to the top 5 to 1% wealthiest households (in France).

ous economies with different degrees of type and scale dependence without actually changing the underlying planner objective.

1.6.1 Results

The optimal wealth tax system in the benchmark economy is given by a positive marginal tax rate of 0.82 percent with an exemption level of $550K. To put these numbers into context, the reform is equivalent to imposing a wealth tax on the top 20 percent wealthiest households, that currently hold approximately 85 percent of total wealth in the US economy.

In Table 1.8, we report that under the optimal tax reform capital and productivity fall below the level of the benchmark economy. Consequently, aggregate output falls as well. This is an immediate implication of the wealth tax, which disproportionately concerns individuals contributing to risky and more productive assets. Additionally, their saving rate drops substantially. The tax reform also induces adjustments in aggregate labor supply arising from two opposite forces. First, wealth-rich households become richer over time which induces a negative wealth effect on labor supply. Second, wealth-rich households become poorer, which induces them to work more. Finally, notice that under the wealth tax, the top 0.1% wealth share increases.

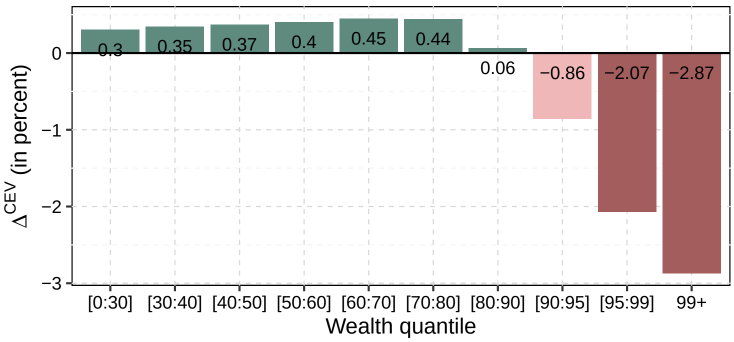
Table 1.8. Aggregate variables after optimal tax reform.

	Percent change relative to status quo
Total labor supply L	0.26
Capital stock X	−6.19
Productivity X/K	−0.08
Output Y	−2.10
Top 0.1% wealth share (in pp)	0.50
Δ^{CEV}	0.14

Decomposing the welfare gains

The consumption equivalent variation in response to the optimal tax reform is modest with an average of 0.14 percent of yearly consumption. This number, however, conceals important heterogeneous effects across the wealth distribution. In Figure 1.6, we document that the largest drop in terms of CEV occurs within the top 1% wealth bracket, in which wealth is largely above the exemption level. In contrast, welfare gains of wealth brackets below the 70^{th} wealth percentile are on average positive, in between 0.3 to 0.45 percent of yearly consumption. Furthermore, the CEV is slightly hump-shaped across the wealth redistribution, as revenues that are raised from the wealth tax are balanced with a lower labor tax which disproportionately benefits households with relatively high labor incomes. Finally, the reform is also politically feasible as the majority of households benefits in terms of CEV.

Figure 1.6. Consumption equivalent variation Δ^{CEV} across wealth distribution.



4.

The effects of rent-extraction

To what extent do differential returns reflect the productivity of capital investments? So far, our baseline results abstract from rents in private returns to wealth. However, private returns may reflect both the productivity of capital investments (MPK) and some form of rents. For example, Smith et al. (2019a) observe for private businesses that the share of value added allocated to owners has increased over time, irrespective of productivity gains. Incorporating this feature into a quantitative exercise is challenging as it requires detailed data allowing to link household investments to their marginal productivity, which must then be compared to private returns of the corresponding household.

Despite this limitation, we show how the presence of a realistic amount of rent extraction would alter the top marginal wealth tax rate. Specifically, we distinguish two types of risky assets. On the one hand, rent-seeking investments, and on the other hand, investments linked to higher productivity and output.[48] Lockwood et al. (2017) report that an increase in the aggregate income share of finance service and law sectors is associated with a decrease in aggregate income. They interpret this finding as indirect evidence for a negative externality from those sectors on aggregate income. Using their estimates, Rothschild and Scheuer (2016) derive an optimal labor income tax taking into account rent-seeking. Following their lead and for the sake of illustration, we make the assumption that returns to wealth obtained from investments in finance and law sectors are associated to rent-seeking activities.

To calibrate the degree of returns associated to rent extraction motives, we use our SCF sample and compute the average share of household equity investments into law and finance sectors between 1998 and 2019. Those equity investments account for roughly 20% of total equity, which justifies our choice to set the return wedge to $\mu = 0.8$.[49] Again, the aggregate return component

[48] It should be noticed that both Piketty et al. (2014), Rothschild and Scheuer (2016) and Lockwood et al. (2017) focus mainly on rent-seeking from labor income. Moreover, the recent studies by Piketty et al. (2014) and Lockwood et al. (2017) focus on the case where externalities from rent-seeking reduce everyone else's income in a lump-sum fashion rather than the proportional reduction that we consider here through $\underline{r}$. Relatedly, Scheuer and Slemrod (2021) discuss the role of rent-extraction in capital returns.

[49] The share of equity invested into both sectors displays substantial heterogeneity across the wealth distribution. Notably, it is particularly important at the top of the wealth distribution, from 15% at the bottom 99% to 22% for the

$\underline{r}$ adjusts in equilibrium to ensure that total revenues obtained by households coincide with total revenues distributed by the intermediate good producer. As a higher return wedge, i.e. a lower value of μ, increases the dispersion of returns between households, we recalibrate the values for ϕ and β such that the model matches the top 1% wealth share and a capital-output ratio of 2.6.

To have a direct comparison to our baseline results without rent extraction, i.e. the case of $\mu = 1$, we preserve the wealth exemption threshold to the wealth level corresponding to the 80^{th} wealth percentile, hereafter referred to as $\underline{a}_{max} = F_a^{-1}(0.80)$, and optimize our welfare criterion over the marginal tax rate τ_a. We find that the optimal top marginal wealth tax rate increases slightly relative to the case without rent-extraction, to a rate of 0.92 percent. Implementing this tax reform leads to overall welfare gains equivalent to 0.2 percent of yearly consumption. Therefore, the marginal wealth tax rate *slightly increases* in the degree of rent extraction. In fact, simulating various model versions with different degrees for $\mu < 1$ shows that the marginal tax rate τ_a monotonously increases in the return wedge. This implies that our benchmark tax rate of 0.82 percent, derived when private returns from investment coincide with their associated marginal productivity, constitutes a conservative *lower bound*.

1.6.2 Dissecting the Effects from Type and Scale Dependence

We now isolate the driving forces behind our two main quantitative results while fixing the wealth exemption level, that is

(A) the optimal marginal wealth tax rate is positive,

(B) the optimal marginal wealth tax rate *slightly increases* in the degree of rent-extraction.

Subsequently, we demonstrate that results *(A)* and *(B)* depend on the quantitative importance of type and scale dependence, and how both mechanisms interact with the extent to which returns to wealth reflect the productivity of capital investments. To unravel the driving forces, we compare the aggregate equilibrium statistics relative to the status quo along several model alternatives where various combinations of type and scale dependence and the return wedge μ are adopted. Under all alternatives, the parameters ϕ and β are *always* recalibrated to match the same targets regarding the top 1% wealth share and the capital-output ratio.

The role of type and scale dependence

We first shed light on the role of type and scale dependence in driving our quantitative results. We compute the aggregate implications of taxing wealth at the optimum of the benchmark economy, characterized by $\mu = 1$, $\tau_a = 0.0082$ and $\underline{a}_{max} = F_a^{-1}(0.80)$, in the scale model (M2) and in the type

top 1% wealthiest households. In an alternative experiment, we capture this non-linearity by assuming that the return wedge is wealth dependent, i.e $\mu(a) = \mu_1 a^{\mu_2}$, for some positive parameters μ_1 and μ_2. We find similar results.

model (M3). Table 1.9 displays the aggregate responses. A striking result is that scale and type models exhibit opposite responses to a positive wealth tax with respect to aggregate labor supply and productivity. Moreover, the welfare gains turn out to be significant and positive under the type model, and negative under the scale model.

Table 1.9. Aggregate variables after the optimal tax reform derived under the benchmark economy.

	Percent changes relative to status quo		
	scale model (M2)	benchmark model (M1)	type model (M3)
Total labor supply L	0.50	0.26	−0.08
Capital stock X	−7.33	−6.19	−4.84
Productivity X/K	−1.55	−0.08	1.27
Output Y	−2.39	−2.10	−1.82
Top 0.1% wealth share (in pp)	−0.15	0.50	0.16
Δ^{CEV}	−0.36	0.14	0.52

Two forces rationalize the above findings. First, the scale model (M2) triggers important general equilibrium effects in response to wealth taxation. As shown in Table 1.9, this manifests in relatively large output losses relative to the type model (M3). The reason is that risky investment behavior is itself a function of wealth, and thus any change in wealth is accompanied by strong contemporaneous behavioral responses in terms of portfolio allocation and savings which transmit across periods. This *snowball* effect induced by the wealth tax reduces wealth accumulation and the amount of risky asset investments undertaken by wealth-rich households, translating into permanent lower productivity and equilibrium wage w. Nevertheless, households work more in the post reform equilibrium, as they face lower marginal labor income taxes and the income effect is sufficiently strong compared to the substitution effect. Finally, these efficiency losses generated by the aforementioned general equilibrium effects outweigh the equity gains from redistribution such that the average consumption equivalent variation turns out to be negative.

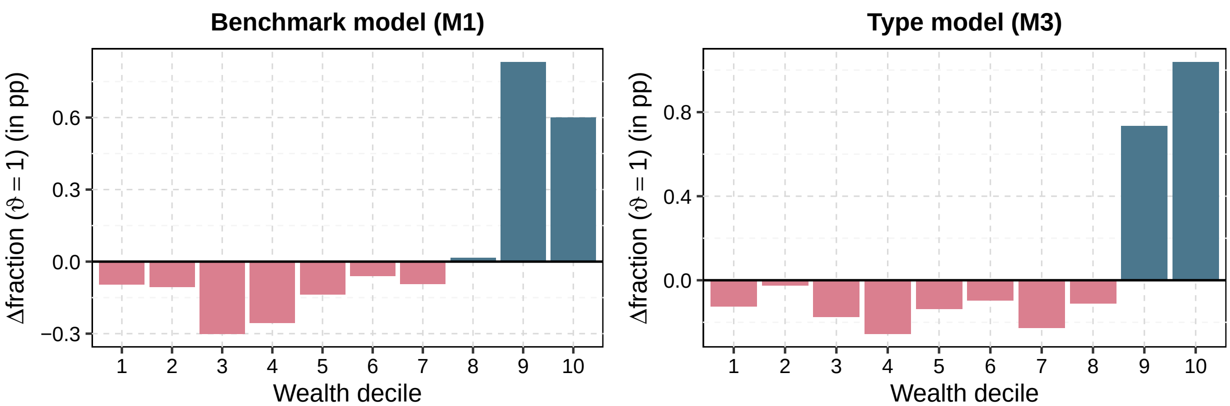
Second, under the type model (M3), we find opposite effects on labor, productivity and welfare gains. To understand the underlying mechanisms, it is important to bear in mind that the persistence of types is crucial for engendering a selection of households with high returns more frequently into the top of the wealth distribution, as they experience high returns to their wealth during several consecutive periods. When wealth-rich households face a tax on the stock of their wealth, those with high returns to wealth are relatively less affected by the wealth tax. As a result, they dissave at a lower rate relative to wealth-rich households who invest in less risky assets and experience lower returns to wealth.[50] Therefore, by taxing the stock of wealth of the richest households, the government creates an environment where *only the fittest survive at the top*.

[50] The fact that higher returns lead individuals to save at a higher rate is a reminiscence of previous findings in the literature, in which entrepreneurs with high returns on their capital save at a higher rate relative to workers with low returns (see, for instance, Cagetti and De Nardi (2006a)). This point is also formalized in Fagereng et al. (2019) in the absence of return risk.

Wealth taxation thus reinforces the selection of agents with higher capital income – associated with higher capital productivity – even further. In Figure 1.7, we show that the wealth tax indeed selects a higher fraction of investors at the top, both in the benchmark model (M1) (left panel) and the type model (M3) (right panel). Note that the hump-shaped pattern at the top in the benchmark economy comes from the scale dependence effects on the risky investment participation margin. In the end, highly skilled investors hold a higher fraction of total wealth, which raises productivity. This effect outweighs the negative effects of a lower capital accumulation on GDP such that welfare gains are large relative to the ones obtained in the benchmark economy.

Due to those two opposing forces from scale and type dependence, aggregate productivity is approximately irresponsive to the implementation of a wealth tax in the benchmark economy. In the next section, we show that these distinct implications on the aggregates also shape the optimal marginal wealth tax rate.

Figure 1.7. Fraction of high investor type ($\vartheta = 1$) relative to the status quo per wealth decile.



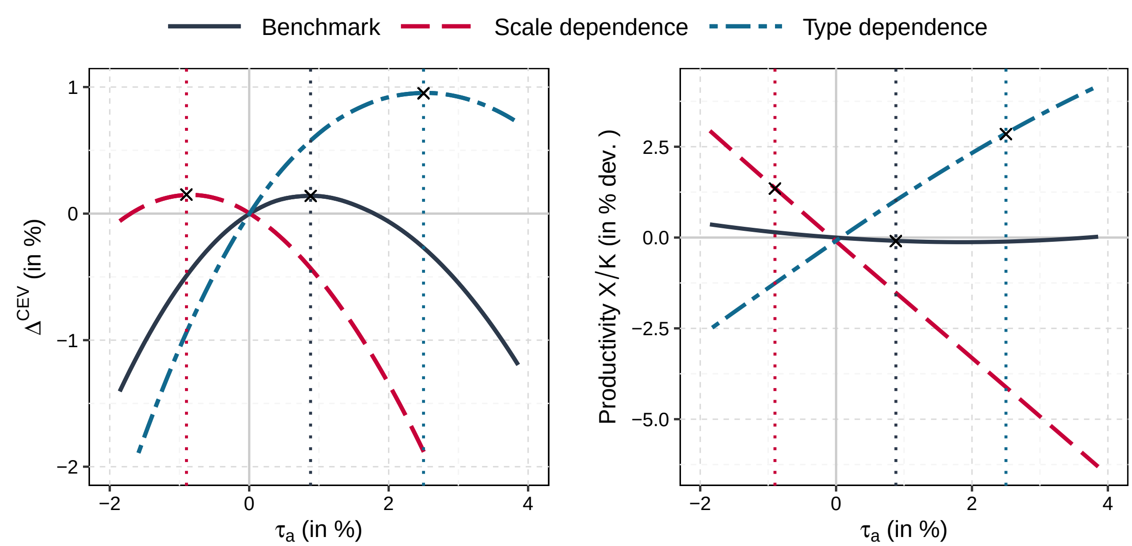
Optimal wealth tax under type and scale dependence

In our second experiment we optimize over the marginal wealth tax rate under the scale model (M2) and under the type model (M3), keeping the wealth exemption level constant such that it corresponds to the 80^{th} percentile. As shown in the left panel of Figure 1.9, the optimal marginal wealth tax rate under the scale model is negative, at a rate of -0.88 percent, while it is substantially positive under the type model, at a rate of 2.41 percent. Which mechanisms explain these different implications for optimal redistribution under the two polar cases?

In the right panel of Figure 1.8, we plot the capital productivity, i.e. the ratio X/K, as a function of the marginal wealth tax rate τ_a. Under type dependence, the change in the selection of skilled types along the wealth distribution leads to an increase in productivity despite the negative effects on capital accumulation. The welfare maximizing wealth tax rate is large with sizable welfare gains equivalent to 0.96 percent of yearly consumption. The reasoning behind this finding is in line with the optimal wealth tax result stated in Guvenen et al. (2019), in which return heterogeneity stems mostly from heterogeneous entrepreneurial skills among households, i.e. from type-dependence. In contrast, scale dependence triggers a strong behavioral response on risky asset investments, such that productivity decreases in the marginal wealth tax, and so do welfare gains as well.

Those distinct forces on productivity rationalize Result *(A)* under our benchmark economy: the welfare-maximizing wealth tax rate is positive. In this economy, type and scale dependence outweigh each other and the productivity becomes roughly irresponsive to the implementation of a wealth tax. As such, the optimal welfare-maximizing wealth tax rate is positive but far below the one implied under pure type dependence.

Figure 1.8. CEV welfare as function of τ_a under type and wealth dependence.



Remark: results are derived by comparing the welfare measure within different long-run stationary economy in which the marginal tax rate τ_a varies.

1.6.3 The Interaction between Type and Scale Dependence and Rents

We now show how the share of rents in private returns affects the welfare-maximizing wealth tax rate under the scale, the type and the benchmark models (M1–M3). Again, we set the wealth exemption level to the one implied by the optimal exemption level under the benchmark model assuming $\mu = 1$. We then compute the optimal marginal tax rate τ_a in the presence of rent-extraction by fixing $\mu = 0.8$ and recalibrate ϕ and β such that the alternative model versions match a top 1% wealth share of 0.36 and a capital-output ratio of 2.6.

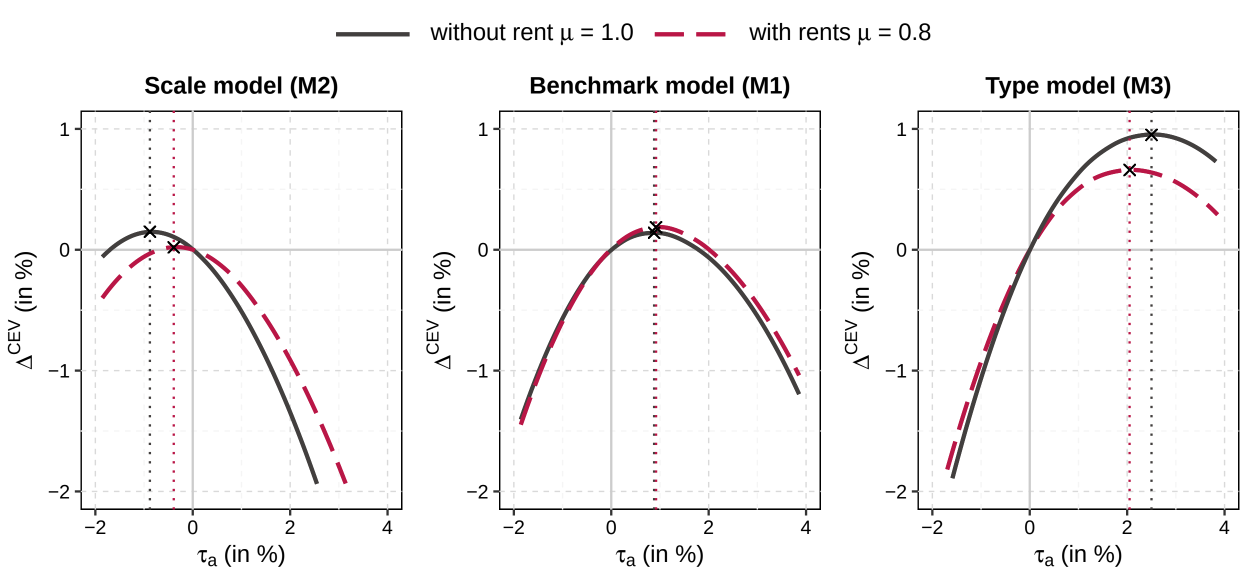
In Figure 1.9 we show that the optimal wealth tax rate is *increasing* in the presence of pure rents in returns to wealth under the scale model (M2). In contrast, the optimal wealth tax rate is *decreasing* in the presence of pure rents in returns to wealth under the type model (M3). This finding mirrors our productivity result, but goes in the opposite direction concerning welfare.

In the Scale model M2, wealth-rich households invest in risky assets, but only a share of those assets are associated with higher productivity, while the remainder is associated with higher rent-extraction. As discussed earlier, in that case, it is optimal to lower the subsidy relative to the case without rent-extraction. The optimal tax rate τ_a thus rises from -0.88 percent to -0.36 percent.

In the type model (M3), highly skilled investors obtain high returns because they systematically invest a higher share of their wealth in riskier investments which now partly reflects productivity *and* rents. As before, taxing wealth leads high return investors to be even more represented at the top, raising in this case both productivity and the size of rents in the economy. Productivity gains are, however, lower than in the baseline case with $\mu = 1$. Therefore, everything else equal, a wealth tax in this economy is less powerful relative to the case of a type model (M3) without rents. On top of this, due to rent-extraction, the stronger selection of high return investors induces a general equilibrium adjustment of the aggregate return $\underline{r}$ for all individuals in the economy, which pushes towards a lower wealth tax. Those two forces lead the optimal wealth tax rate to *decrease* with the size of rents, from a rate of 2.41 percent without rent-extraction to a rate of 2.06 percent when $\mu = 0.8$.

Quite surprisingly, we find that, again, type and scale dependence effects outweigh each other in the benchmark economy, such that the marginal wealth tax rate is almost not responding to the presence of rent-extraction. This leads to Result *(B)*: the welfare-maximizing wealth tax rate is slightly *increasing* in the share of rents in returns.

Figure 1.9. CEV welfare as function of τ_a under type and *wealth* dependence: sensitivity rent-extraction.



Remark: results are derived by comparing the welfare measure within different long-run stationary economies in which the marginal tax rate τ_a varies.

Finally, notice that our results under pure scale or type dependence with $\mu = 1$, i.e. in the absence of a return wedge, can be viewed as lower and upper bounds on the optimal marginal wealth tax. Our results point out that depending on the relative strength of scale and type dependence consistent with the observed return heterogeneity in the data, the welfare-maximizing tax rate lies in between $[-0.8, 2.4]$ percent.

To summarize, a key take away is that in order to understand the consequences of wealth taxation in an economy where returns to wealth may or may not coincide with capital productivity, it is essential to take into account the relative importance of type and scale dependence in household investments and associated returns to wealth.

1.6.4 Further Robustness

In addition to the previous analysis, we have also explored the role of particular elements of our quantitative model including (i) the life-cycle structure with the mortality rate $d_j > 0$, the age-dependent earnings component $\zeta_j \neq 1$ and the social security pension ζ_J , (ii) the endogenous labor choice, (iii) the tax instrument used to balance the government budget, (iv) the presence of idiosyncratic return risk. We do so by reassessing the optimal wealth tax in versions of our model where various combinations of these elements are shut down. We found that (i) and (ii) have little influence on our main qualitative message regarding the distinction between type and scale dependence. Concerning (iii), we find that redistribution through a lump-sum tax provides similar results, while redistribution through capital income tax reinforces even further the selection of high types at the top. Removing the risky component κ in returns lowers wealth inequality in the stationary equilibrium. This is compensated by increasing the excess wealth return ϕ to match the top 1% wealth share. More importantly, under all previously discussed model alternatives, we find that our results do not hinge on particular elements. A high positive wealth tax is optimal under type dependence and a low or negative wealth tax is optimal under scale dependence.

1.7 Conclusion

In this paper, we first develop a conceptual framework to study the macroeconomic and welfare implications of wealth redistribution, unraveling and clarifying the key economic forces behind many heterogeneous agents incomplete markets models. Despite its stylized nature, our analytical two-period model identifies four statistics that are crucial for understanding these implications: (i) the Pareto tail of the wealth distribution, (ii) the elasticity of risk-taking to wealth (scale dependence), (iii) the sorting of types along the wealth distribution (type dependence), and (iv) the extent to which returns to wealth reflects investment productivity.

In a second step, we construct a full-blown quantitative model and analyze the key elements that we identified as particularly relevant within our theoretical framework. The model accounts for the highly concentrated wealth and returns distributions through type and scale dependence in portfolio choices. Our model is consistent with empirical evidence from the SCF and PSID. We show that the underlying force behind wealth accumulation and inequality, i.e. type or scale dependence, shapes distinct aggregate responses to a top wealth tax, both qualitatively and quantitatively. The aggregate responses of productivity and inequality are large under pure scale dependence. Under type dependence, however, the joint distribution of investor-types and wealth non-trivially reacts to the implementation of a wealth tax, as a top marginal wealth tax selects high capital income households more effectively into the top of the wealth distribution.

The welfare implications of a wealth tax depend on the degree of scale and type dependence together with the extent to which returns to wealth reflect the productivity of investments. In our benchmark economy, the optimal top marginal wealth tax rate is 0.8 percent above an exemption level of $550K, as long as the model features scale and type dependence consistent with data.

Future research is needed to understand the relationship between the joint distribution of wealth, returns, and portfolio allocation. Specifically, empirical studies are helpful in disentangling scale and type dependence in the decision of agents, and, in turn, to determine the optimal taxation of wealth-rich households. Moreover, future research should attempt to empirically evaluate the pass-through between the productivity of investments and returns to wealth along the lines of Lockwood et al. (2017) or Smith et al. (2019a). Finally, we abstracted from tax avoidance motives. Substantial amounts of wealth are held abroad (Alstadsæter et al., 2018), and a wealth tax may foster incentives for tax avoidance. Such considerations induce a form of negative scale dependence that our quantitative model may additionally consider. These are important and, to a large extent, unexplored issues that we leave for future work.

Appendix

We organize the appendix as follows. Section 1.A contains appendix of the simple analytical model. Section 1.B contains details regarding our empirical work. Section 1.C contains the computational appendix and calibration details.

1.A Theoretical appendix

1.A.1 Proofs for Section 1.2

Final producer maximisation

For clarity, we detail the steps of the final good producer who maximizes the use of labor n and intermediate goods x_s^j, with

$$\max_{\{n,x_s^j\}} \left(\sum_s \int_j x_s^j \, dj \right) n^{\varphi} - wn - \sum_s \int_j p_s^j \, x_s^j \, dj \, .$$

Taking the first order condition with respect to labor gives $w = \varphi \left(\sum_s \int_j x_s^j \, dj \right) n^{\varphi - 1}$. Plugging this condition together with the assumption that $n = \int_i h^i di = 1$ into the profit function yields

$$\Pi^f = (1-\varphi) \left(\sum_s \int_j x_s^j \, dj \right) n^{\varphi} - \sum_s \int_j p_s^j \, x_s^j dj = (1-\varphi) \left(\sum_s \int_j x_s^j \, dj \right) - \sum_s \int_j p_s^j \, x_s^j dj \, .$$

Terminal Wealth Distribution

Proof. Because of $u'(c_2^i) > 0$, we know that the second period budget constraint holds in equilibrium with equality. Thus, second period consumption is given by $c_2^i = \underline{r} a_1^i + w + T + R_f a_1^i + \omega_1^i a_1^i (R_f - R_r^i)$. Substituting from the wage rate $w = \varphi Y$ and returns R_f and R_r^i, we get

$$c_2^i = \underline{r} a_1^i + \varphi Y + A(1-\varphi)(1-\omega_1^i) a_1^i + \kappa^i (1-\varphi) \omega_1^i a_1^i + T \, .$$

We obtain that $c_2^i \sim \mathcal{N} \left(\mu_{c_2}^i, \sigma_{c_2}^i \right)$, with

$$\mu_{c_2}^i = \underline{r} a_1^i + \varphi Y + T + A(1-\varphi)(1-\omega_1^i) a_1^i + \phi(1-\varphi) \omega_1^i a_1^i ,$$
$$\sigma_{c_2}^i = \sigma_\kappa^2 (1-\varphi)^2 (\omega_1^i a_0^i)^2 \, .$$

In the Online Appendix OA 1.8, we extend the results with a case in which we introduce labor income risk. □

Proof of Lemma 1

Proof. We first need to derive $\mathbb{E}\left[u(c_2^i)|\mathcal{I}_1\right]$ analytically. To do so, we use an arbitrary Gaussian distribution with mean $\mu^i_{c_2}$ and variance $(\sigma^i_{c_2})^2$. Using terminal wealth distribution, we obtain

$$\begin{aligned}\mathbb{E}\left[u(c_2^i)|\mathcal{I}_1\right] &= \frac{1}{\alpha_i}\int\left(1-e^{-\alpha_i c_2^i}\right)\times\frac{1}{\sqrt{2\pi(\sigma^i_{c_2})^2}}e^{-\frac{1}{2(\sigma^i_{c_2})^2}\left(c_2^i-\mu^i_{c_2}\right)^2}dc_2^i\\ &= \frac{1}{\alpha_i}-\frac{1}{\alpha_i}\int\frac{1}{\sqrt{2\pi(\sigma^i_{c_2})^2}}e^{-\frac{1}{2(\sigma^i_{c_2})^2}\left[\left(c_2^i-\mu^i_{c_2}\right)^2+2\alpha_i(\sigma^i_{c_2})^2c_2^i\right]}dc_2^i\\ &= \frac{1}{\alpha_i}-\frac{1}{\alpha_i}e^{-\alpha_i\mu^i_{c_2}+\frac{1}{2}\alpha_i^2(\sigma^i_{c_2})^2}\int\frac{1}{\sqrt{2\pi(\sigma^i_{c_2})^2}}e^{-\frac{1}{2(\sigma^i_{c_2})^2}\left(c_2^i-(\mu^i_{c_2}-\alpha_i(\sigma^i_{c_2})^2)\right)^2}dc_2^i\end{aligned}$$

Recognizing that the term in the integral is the pdf of a normally distributed random variable with mean $\mu^i_{c_2}-\alpha_i(\sigma^i_{c_2})^2$ and variance $(\sigma^i_{c_2})^2$, we finally obtain

$$\mathbb{E}\left[u(c_2^i)|\mathcal{I}_1\right] = \frac{1-e^{-\alpha_i\mu^i_{c_2}+\frac{1}{2}\alpha_i^2(\sigma^i_{c_2})^2}}{\alpha_i}\,.$$

Under the additional set of assumptions within the special case section, we have $a_1^i \equiv a_0^i$ and $c_2^i = \varphi Y + T + R_f(1-\omega_1^i)a_0^i + R_r^i\omega_1^i a_0^i$ such that $\mu^i_{c_2} = \varphi Y + T + A(1-\varphi)(1-\omega_1^i)a_0^i + \phi(1-\varphi)\omega_1^i a_0^i$ and $\sigma^i_{c_2} = \omega_1^i a_0^i\sigma_\kappa(1-\varphi)$. We solve for the maximization problem given by

$$\max_{\{\omega_1^i\}}\left[1-exp\Big\{-\alpha_i\Big(\mu^i_{c_2}-\frac{\alpha_i}{2}\sigma^i_{c_2}\Big)\Big\}\right]\alpha_i^{-1}\,,$$

Denoting $\mathbb{V} = exp\Big\{-\alpha_i\Big(\mu^i_{c_2}-\frac{\alpha_i}{2}\sigma^i_{c_2}\Big)\Big\}$, the corresponding first order condition is given by

$$-\frac{1}{\alpha_i}\mathbb{V}\left[-\alpha_i(\phi-A)(1-\varphi)a_0^i+\alpha_i^2(a_0^i)^2\omega_1^i\sigma_\kappa^2(1-\varphi)^2\right]=0\,,$$

which results after rearranging

$$\omega_1^i = \frac{\phi-A}{(1-\varphi)\alpha_i\sigma_\kappa^2}(a_0^i)^{-1} = \frac{\phi-A}{(1-\varphi)\sigma_\kappa^2}\frac{\vartheta^i}{\vartheta}(a_0^i)^{\gamma-1}\,.$$

To ensure that the solution is indeed a maximum, we derive the second order condition as

$$-\frac{1}{\alpha_i}\mathbb{V}\left[-\alpha_i(\phi-A)(1-\varphi)a_0^i+\alpha_i^2(a_0^i)^2\omega_1^i\sigma_\kappa^2(1-\varphi)^2\right]^2-\frac{1}{\alpha_i}\mathbb{V}\alpha_i^2\sigma_\kappa^2(1-\varphi)^2(a_0^i)^2<0\,.$$

which completes the proof. □

Proof of Proposition 2

Proof. The expression for aggregate innovative asset holdings follows straightforward from integrating over household dynamics while applying the covariance formula

$$\begin{aligned}K_I=(\widetilde{\omega}/\overline{\vartheta})\mathbb{E}\left[\vartheta a_0^\gamma\right]&=\frac{\phi-A}{(1-\varphi)\overline{\vartheta}\sigma_\kappa^2}\left(cov(\vartheta,a_0^\gamma)+\mathbb{E}\left[\vartheta\right]\mathbb{E}\left[a_0^\gamma\right]\right)\\&=\frac{\phi-A}{(1-\varphi)\overline{\vartheta}\sigma_\kappa^2}\left(\rho_{\vartheta,a_0^\gamma}\sigma_\vartheta\sigma_{a_0^\gamma}+\mu_\vartheta\mu_{a_0^\gamma}\right)\,.\end{aligned}$$

Aggregate output is given by $Y=\left(\sum_s\int_j x_s^j\,dj\right)n^\varphi$, where $n=\int_i e^i\,di=1$. Given that $K_N+K_I=\mathbb{E}[a_0]$, this can be rewritten after integrating over intermediate goods x_s^j as

$$\begin{aligned}Y&=\int_j x_I^j\,dj+\int_j x_N^j\,dj\\&=\left[\phi\mu+A(1-\mu)\right]\int_i\omega_1^i a_0^i\,di+A\int_i(1-\omega_1^i)a_0^i\,di\\&=\mu(\phi-A)K_I+A\mathbb{E}[a_0]=\underbrace{\left[\mu(\phi-A)(K_I/\mathbb{E}[a_0])+A\right]}_{:=Z}\mathbb{E}[a_0]\end{aligned}$$

The price $\underline{r}$ ensures that total capital distributed to households coincide with the total revenue distributed by the intermediate producer, such that

$$\begin{aligned}\int_i\left(R_f(1-\omega_1^i)+R_r^i\omega_1^i+\underline{r}\right)a_1^i di&=\int_i\left(A(1-\varphi)(1-\omega_1^i)+(\phi\mu+A(1-\mu))(1-\varphi)\omega_1^i\right)a_1^i di\,,\\\phi(1-\varphi)K_I+\underline{r}\mathbb{E}[a_0]&=(\phi\mu+A(1-\mu))(1-\varphi)K_I\,,\\\underline{r}&=(\mu-1)(\phi-A)(1-\varphi)(K_I/\mathbb{E}[a_0])\,.\end{aligned}$$

where the last equality, regarding the integration of $\int_i R_r^i a_1^i\omega_1^i di$, follows from the simplifying assumption that there is a sub-continuum of households in each state (ϑ^i,a_0^i). □

Proof of Proposition 1

Proof. To prove the result regarding the effect of a mean preserving change in wealth inequality on K_I, we compare the aggregate innovate asset holdings for two economies with different Pareto tails $\eta'\neq\eta$ while keeping aggregate wealth $\mu_{a_0}=\mathbb{E}[a_0]$ constant. We then proceed by case distinction.

CASE 1: $\rho_{\vartheta,a_0^\gamma}=0$

The difference in aggregate innovative asset holdings between the two economies is written as

$$\begin{aligned}\Delta^a K_I(\eta',\eta) &= (\widetilde{\omega}/\overline{\vartheta})\mu_\vartheta\left(\frac{\eta'}{\eta'-\gamma}(\underline{a}')^\gamma-\frac{\eta}{\eta-\gamma}\underline{a}^\gamma\right)\\ &= \widetilde{\omega}\frac{\eta}{\eta-\gamma}\underline{a}^\gamma\left(\frac{\eta'}{\eta'-\gamma}\frac{\eta-\gamma}{\eta}\left(\frac{\underline{a}'}{\underline{a}}\right)^\gamma-1\right)\ .\end{aligned}$$

Making use of the mean preserving assumption, i.e. $\underline{a}\frac{\eta}{\eta-1}=\underline{a}'\frac{\eta'}{\eta'-1}$, we obtain

$$\Delta^a K_I(\eta',\eta)=\widetilde{\omega}\frac{\eta}{\eta-\gamma}\underline{a}^\gamma\left(\frac{\eta'}{\eta'-\gamma}\frac{\eta-\gamma}{\eta}\left(\frac{\eta'-1}{\eta'}\frac{\eta}{\eta-1}\right)^\gamma-1\right)\ .$$

Defining $\chi(\eta,\gamma)\equiv\frac{\eta-\gamma}{\eta}\left(\frac{\eta}{\eta-1}\right)^\gamma$ and taking the derivative of the inner expression w.r.t. η', we get

$$\chi(\eta,\gamma)\left(\frac{\eta'-1}{\eta'}\right)^\gamma\left[-\frac{\gamma}{(\eta'-\gamma)^2}+\frac{\gamma}{(\eta'-\gamma)(\eta'-1)}\right]=\chi(\eta,\gamma)\left(\frac{\eta'-1}{\eta'}\right)^\gamma\frac{\gamma(1-\gamma)}{(\eta'-\gamma)^2(\eta'-1)}\ .$$

As a result, we obtain finally

$$\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'}=\widetilde{\omega}\mu_{a_0^\gamma}\chi(\eta,\gamma)\left(\frac{\eta'-1}{\eta'}\right)^\gamma\frac{\gamma(1-\gamma)}{(\eta'-\gamma)^2(\eta'-1)}\ .$$

The Lemma then follows by recognizing that $\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'}=0$ if $\gamma\in\{0,1\}$. Similarly, we obtain $\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'}>0$ if $\gamma\in(0,1)$ and $\frac{\partial\Delta^a K_I(\eta',\eta)}{\partial\eta'}<0$ if $\gamma>1$.

CASE 2: $\rho_{\vartheta,a_0^\gamma}\neq 0$

In the case of an arbitrary correlation between innate risk aversion types and wealth, we obtain

$$\Delta K_I(\eta',\eta)=\Delta^\vartheta K_I(\eta',\eta)+\Delta^a K_I(\eta',\eta)\ ,$$

where the first term denotes distributional relevance arising from the selection effect, whereas the second term resembles distributional relevance arising from wealth dependent risk taking. Notice that the latter is equivalent to Case 1. Contrary, the first effect can be written as

$$\Delta^\vartheta K_I(\eta',\eta)=(\widetilde{\omega}/\overline{\vartheta})\rho_{\vartheta,a_0^\gamma}\sigma_\vartheta\sigma_{a_0^\gamma}\left(\frac{\rho_{\vartheta,a_0^\gamma}(\eta',\underline{a}',\cdot)}{\rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}\frac{\sigma_{a_0^\gamma}(\eta',\underline{a}')}{\sigma_{a_0^\gamma}(\eta,\underline{a})}-1\right)\ .$$

Using the relation $\underline{a}\frac{\eta}{\eta-1}=\underline{a}'\frac{\eta'}{\eta'-1}$, a change in the Pareto tail η' preserves the mean wealth if

$\underline{a}'(\eta') = \underline{a}\left(\frac{\eta}{\eta-1}\right)\left(\frac{\eta'-1}{\eta'}\right)$. Using a first order Taylor approximation of $\rho_{\vartheta,a_0^\gamma}(\eta', \underline{a}'(\eta'), \cdot)$ around η, we obtain:

$$\rho_{\vartheta,a_0^\gamma}(\eta', \underline{a}'(\eta'), \cdot) \approx \rho_{\vartheta,a_0^\gamma}(\eta, \underline{a}, \cdot) + \left|\frac{\partial \rho^1_{\vartheta,a_0^\gamma}(\eta', \underline{a}'(\eta'), \cdot)}{\partial \eta} + \frac{\partial \rho^2_{\vartheta,a_0^\gamma}(\eta', \underline{a}'(\eta'), \cdot)}{\partial \underline{a}'(\eta')}\frac{\partial \underline{a}'(\eta')}{\partial \eta'}\right|_{\eta'=\eta} (\eta' - \eta)$$

$$\approx \rho_{\vartheta,a_0^\gamma}(\eta, \underline{a}, \cdot) + \left(\frac{\partial \rho_{\vartheta,a_0^\gamma}(\eta, \underline{a}, \cdot)}{\partial \eta} + \frac{\partial \rho_{\vartheta,a_0^\gamma}(\eta, \underline{a}, \cdot)}{\partial \underline{a}}\frac{\underline{a}}{\eta(\eta-1)}\right)(\eta' - \eta) \ .$$

Substituting the previous expression into the one for $\Delta^\vartheta K_I(\eta', \eta)$ we arrive at

$$\Delta^\vartheta K_I(\eta', \eta) \approx (\widetilde{\omega}/\overline{\vartheta})\rho_{\vartheta,a_0^\gamma}\sigma_\vartheta\sigma_{a_0^\gamma}\left(\left[1 + \frac{\left(\frac{\partial \rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}{\partial \eta} + \frac{\partial \rho_{\vartheta,a_0^\gamma}(\eta,\underline{a},\cdot)}{\partial \underline{a}}\frac{\underline{a}}{\eta(\eta-1)}\right)}{\rho_{\vartheta,a_0^\gamma}(\eta, \underline{a}, \cdot)}(\eta' - \eta)\right]\frac{\sigma_{a_0^\gamma}(\eta', \underline{a}')}{\sigma_{a_0^\gamma}(\eta, \underline{a})} - 1\right) \ .$$

With a slight abuse of notation, we can take the derivative w.r.t. η' to obtain

$$\frac{\partial \Delta^\vartheta K_I(\eta', \eta)}{\partial \eta'} \approx (\widetilde{\omega}/\overline{\vartheta})\rho_{\vartheta,a_0^\gamma}\sigma_\vartheta\sigma_{a_0^\gamma}\left(\frac{1}{\sigma_{a_0^\gamma}}\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} + \frac{\frac{\partial \rho_{\vartheta,a_0^\gamma}}{\partial \eta} + \frac{\partial \rho_{\vartheta,a_0^\gamma}}{\partial \underline{a}}\frac{\underline{a}}{\eta(\eta-1)}}{\rho_{\vartheta,a_0^\gamma}}\left(\frac{\sigma'_{a_0^\gamma}}{\sigma_{a_0^\gamma}} + (\eta' - \eta)\frac{1}{\sigma_{a_0^\gamma}}\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'}\right)\right) \ .$$

To get an impression about the sign of the previous derivative, let us first analyze the sign of $\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'}$. Notice that it is straightforward to show that a_0^γ follows a $\mathcal{P}a(\underline{a}^\gamma, \frac{\eta}{\gamma})$ distribution. As a result, the variance is given by

$$\sigma'_{a_0^\gamma} = (\underline{a}')^{2\gamma}\frac{\frac{\eta'}{\gamma}}{\left(\frac{\eta'}{\gamma} - 1\right)^2\left(\frac{\eta'}{\gamma} - 2\right)} = \left(\underline{a}\frac{\eta}{\eta - 1}\right)^{2\gamma}\left(\frac{\eta' - 1}{\eta'}\right)^{2\gamma}\frac{\frac{\eta'}{\gamma}}{\left(\frac{\eta'}{\gamma} - 1\right)^2\left(\frac{\eta'}{\gamma} - 2\right)} \ ,$$

where again the last equality follows from using $\underline{a}\frac{\eta}{\eta-1} = \underline{a}'\frac{\eta'}{\eta'-1}$. Defining the auxiliary variable $\tilde{\chi}(\eta, \gamma, \underline{a}) \equiv \left(\underline{a}\frac{\eta}{\eta-1}\right)^{2\gamma}$, we obtain:

$$\begin{aligned}\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} =& \tilde{\chi}(\eta, \gamma, \underline{a})\left(2\gamma\left(\frac{\eta' - 1}{\eta'}\right)^{2\gamma-1}\frac{1}{(\eta')^2}\frac{\frac{\eta'}{\gamma}}{\left(\frac{\eta'}{\gamma} - 1\right)^2\left(\frac{\eta'}{\gamma} - 2\right)}\right) \\ &+ \tilde{\chi}(\eta, \gamma, \underline{a})\left(\frac{\eta' - 1}{\eta'}\right)^{2\gamma}\left(\frac{\frac{1}{\gamma}\left(\frac{\eta'}{\gamma} - 1\right)^2\left(\frac{\eta'}{\gamma} - 2\right) - \frac{\eta'}{\gamma}\left[\frac{2}{\gamma}\left(\frac{\eta'}{\gamma} - 1\right)\left(\frac{\eta'}{\gamma} - 2\right) + \frac{1}{\gamma}\left(\frac{\eta'}{\gamma} - 1\right)^2\right]}{\left(\frac{\eta'}{\gamma} - 1\right)^4\left(\frac{\eta'}{\gamma} - 2\right)^2}\right) \ .\end{aligned}$$

Collecting terms leads to

$$\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} = \tilde{\chi}(\eta,\gamma,\underline{a}) \left(\frac{\eta'-1}{\eta'}\right)^{2\gamma} \frac{1}{\left(\frac{\eta'}{\gamma}-1\right)^2\left(\frac{\eta'}{\gamma}-2\right)} \left[\frac{2}{\eta'-1}+\frac{1}{\gamma}-\frac{2\eta'}{\gamma(\eta'-\gamma)}-\frac{\eta'}{\gamma(\eta'-2\gamma)}\right]$$
$$= \frac{1}{\eta'}\sigma'_{a_0^\gamma}\left[1+\frac{2\gamma}{\eta'-1}-\frac{2\eta'}{(\eta'-\gamma)}-\frac{\eta'}{(\eta'-2\gamma)}\right] .$$

In order to determine the sign of the bracket term, one can simplify to

$$\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} = \frac{2}{\eta'}\sigma'_{a_0^\gamma}\left[\frac{\gamma(1-2\gamma)(\eta'-\gamma)-\eta'(\eta'-1)(\eta'-2\gamma)}{(\eta'-1)(\eta'-\gamma)(\eta'-2\gamma)}\right] .$$

It is straightforward to show that the previous term is (weakly) negative if

$$\eta' \geq 2\gamma + \gamma\frac{(1-2\gamma)(\eta'-\gamma)}{\eta'(\eta'-1)} .$$

The left hand side of this expression is increasing in η', whereas the right hand side is decreasing if $\gamma \leq \frac{1}{2}$ and increasing if $\gamma > \frac{1}{2}$. Hence, for the case of $\gamma \leq \frac{1}{2}$, we obtain after substituting $\eta' = 2\gamma$ an upper limit of the right hand side given by $\overline{\eta} = \frac{3}{2}\gamma$. Contrary, in the case of $\gamma > \frac{1}{2}$ a straightforward application of L'Hopital's rule results in $\overline{\eta} = 2\gamma$. As a result, we obtain that $\frac{\partial \sigma'_{a_0^\gamma}}{\partial \eta'} \leq 0 \; \forall \eta' \geq \overline{\eta} = \max\{\frac{3}{2}\gamma, 2\gamma\} = 2\gamma$, which trivially holds due to the implicit assumed finite variance of the $\mathcal{P}a(\underline{a}^\gamma, \frac{\eta}{\gamma})$ distribution. Consequently, the result of proposition 1 follows (given a small change in the wealth Pareto tail). □

Proof of Proposition 1

Proof. The Farlie-Gumbel-Morgenstern (FGM) copula can be written for two arbitrary cumulative distribution functions $\{F(x_1), F(x_2)\}$ as

$$F(x_1,x_2) = C^{FGM}(F(x_1),F(x_2)) = F(x_1)F(x_2) + \varrho F(x_1)F(x_2)(1-F(x_1))(1-F(x_2)) ,$$

where $\varrho \in [-1,1]$. The joint probability density function of $f(x_1,x_2)$ is the obtained by

$$f(x_1,x_2) = (1+\varrho\,(1-2F(x_1))\,(1-2F(x_2)))\,f(x_1)f(x_2)$$
$$= (1+\varrho+2\varrho\,(2F(x_1)F(x_2)-F(x_1)-F(x_2)))\,f(x_1)f(x_2) .$$

Under this assumption that $\vartheta \sim \mathcal{P}a(\underline{\vartheta},\epsilon)$ and $a_0 \sim \mathcal{P}a(\underline{a},\eta)$ this provides us with

$$f(\vartheta,a_0) = (1+\varrho)f(\vartheta)f(a_0) + 2\varrho\left[2\left(\frac{\underline{\vartheta}}{\vartheta}\right)^{\epsilon}\left(\frac{\underline{a}}{a_0}\right)^{\eta} - \left(\frac{\underline{\vartheta}}{\vartheta}\right)^{\epsilon} - \left(\frac{\underline{a}}{a_0}\right)^{\eta}\right] f(\vartheta)f(a_0) .$$

where the marginals are given by $f(\vartheta)$ and $f(a_0)$. Given the Pareto assumptions, we have $\mu_\vartheta \equiv \overline{\vartheta} = \underline{\vartheta}\frac{\epsilon}{\epsilon-1}$ and $\mu_{a_0^\gamma} = \underline{a}^\gamma \frac{\eta}{\eta-\gamma}$. In order to derive $cov(\vartheta, a_0^\gamma)$, we need to compute $\mathbb{E}\left[\vartheta a_0^\gamma\right]$:

$$\mathbb{E}\left[\vartheta a_0^\gamma\right] = \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta a_0^\gamma f(\vartheta, a_0) d\vartheta da_0 \,.$$

Using the FGM copula, we proceed in four steps:

$$\begin{aligned}
(1+\varrho)\underline{\vartheta}^\epsilon \underline{a}^\eta \epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-\epsilon} a_0^{\gamma-\eta-1} d\vartheta da_0 &= (1+\varrho)\underline{\vartheta}\frac{\epsilon}{\epsilon-1}\underline{a}^\gamma \frac{\eta}{\eta-\gamma}\,, \\
4\varrho\underline{\vartheta}^{2\epsilon} \underline{a}^{2\eta} \epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-2\epsilon} a_0^{\gamma-2\eta-1} d\vartheta da_0 &= 4\varrho\underline{\vartheta}\frac{\epsilon}{2\epsilon-1}\underline{a}^\gamma \frac{\eta}{2\eta-\gamma}\,, \\
-2\varrho\underline{\vartheta}^{2\epsilon} \underline{a}^{\eta} \epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-2\epsilon} a_0^{\gamma-\eta-1} d\vartheta da_0 &= -2\varrho\underline{\vartheta}\frac{\epsilon}{2\epsilon-1}\underline{a}^\gamma \frac{\eta}{\eta-\gamma}\,, \\
-2\varrho\underline{\vartheta}^{\epsilon} \underline{a}^{2\eta} \epsilon\eta \int_{\underline{\vartheta}}^{\infty}\int_{\underline{a}}^{\infty} \vartheta^{-\epsilon} a_0^{\gamma-2\eta-1} d\vartheta da_0 &= -2\varrho\underline{\vartheta}\frac{\epsilon}{\epsilon-1}\underline{a}^\gamma \frac{\eta}{2\eta-\gamma}\,.
\end{aligned}$$

Combining the previous four equations gives

$$\begin{aligned}
cov(\vartheta, a_0^\gamma) &= \mathbb{E}\left[\vartheta a_0^\gamma\right] - \mathbb{E}\left[\vartheta\right]\mathbb{E}\left[a_0^\gamma\right] \\
&= \underline{\vartheta}\underline{a}^\gamma\left[\varrho\frac{\epsilon}{\epsilon-1}\frac{\eta}{\eta-\gamma} + 4\varrho\frac{\epsilon}{2\epsilon-1}\frac{\eta}{2\eta-\gamma} - 2\varrho\frac{\epsilon}{2\epsilon-1}\frac{\eta}{\eta-\gamma} - 2\varrho\frac{\epsilon}{\epsilon-1}\frac{\eta}{2\eta-\gamma}\right]
\end{aligned}$$

Further simplifications result in

$$cov(\vartheta, a_0^\gamma) = \underline{\vartheta}\underline{a}^\gamma \varrho \frac{\epsilon}{(\epsilon-1)(2\epsilon-1)}\frac{\eta\gamma}{(\eta-\gamma)(2\eta-\gamma)}\,.$$

As a result, aggregate innovative asset holdings from Lemma 2 are given by

$$K_I = (\widetilde{\omega}/\overline{\vartheta})\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\underline{\vartheta}\frac{\epsilon}{\epsilon-1}\underline{a}^\gamma\frac{\eta}{\eta-\gamma}\,.$$

Aggregate risk free capital holdings from Lemma 1 are (weakly) positive if the condition $\mu_{a_0} \geq K_I$ holds, which can be rewritten as

$$\frac{\eta-\gamma}{\eta-1} \geq (\widetilde{\omega}/\overline{\vartheta})\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{\epsilon}{\epsilon-1}\underline{\vartheta}\underline{a}^{\gamma-1}\,.$$

Finally, as $\mu_\vartheta = \mathbb{E}[\vartheta] = \overline{\vartheta}$, let $\widetilde{\omega} = (\widetilde{\omega}/\overline{\vartheta})\mu_\vartheta \underline{a}^{\gamma-1} = \widetilde{\omega}\underline{a}^{\gamma-1}$, $B = -\mu\widetilde{\omega}(\phi - A)$ and $C = (2\epsilon-1)$, we derive the marginal effect of a change in the Pareto tail η on wealth-normalized output $\widetilde{Y}(\eta) \equiv$

$\frac{Y(\eta)}{\mu_{a_0}} = A + \mu\widetilde{\omega}(\phi - A)\Psi(\eta)$ with $\Psi(\eta) = \left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{\eta-1}{\eta-\gamma}$ as:

$$\begin{aligned}
\frac{\partial\widetilde{Y}(\eta)}{\partial\eta} &= -B\frac{\partial\Psi(\eta)}{\partial\eta} \\
&= -B\left[\left(\frac{1}{\eta-\gamma}\right)\left(1+\frac{\varrho\gamma}{C(2\eta-\gamma)}\right) - \left(\frac{\eta-1}{(\eta-\gamma)^2}\right)\left(1+\frac{\varrho\gamma}{C(2\eta-\gamma)}\right) - 2\left(\frac{\varrho\gamma}{C(2\eta-\gamma)^2}\right)\left(\frac{\eta-1}{\eta-\gamma}\right)\right] \\
&= B\left[(\gamma-1)\underbrace{\left(\frac{1}{(\eta-\gamma)^2}\right)}_{:=\Omega^{\gamma}} + \varrho(\gamma-1)\underbrace{\left(\frac{(\gamma(2\eta-\gamma)+2(\eta-\gamma)(\eta-1))}{C(2\eta-\gamma)^2(\eta-\gamma)^2}\right)}_{:=\Omega^{\varrho\gamma}} + \varrho\underbrace{\left(\frac{2(\eta-1)}{C(2\eta-\gamma)^2(\eta-\gamma)}\right)}_{:=\Omega^{\varrho}}\right] \\
&= B\left(\frac{1}{C(2\eta-\gamma)^2(\eta-\gamma)^2}\right)\left[(\gamma-1)C(2\eta-\gamma)^2 + \varrho(\gamma-1)2(\eta-1)(\eta-\gamma) + \varrho(2\eta(\eta-1)+\gamma)\right]
\end{aligned}$$

where the before last line follows from $\varrho\gamma = \varrho + \varrho(\gamma-1)$. In order to determine the sign of the derivatives, we need to know the sign of $\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)$. To do so, let us assume that

$$1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)} \leq 0 \Leftrightarrow 1 \leq -\frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)} \leq \frac{\gamma}{(2\epsilon-1)(2\eta-\gamma)},$$

where the last inequality follows from $\varrho \in [-1,1]$. As we have $\epsilon > 1$ and $\eta > \gamma$, an upper bound of $\frac{\gamma}{(2\epsilon-1)(2\eta-\gamma)}$ is given by $\frac{\gamma}{\epsilon\eta}$, which is strictly smaller than one. Hence, we obtain a contradiction and conclude that $1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}$ is strictly positive. This completes the proof of Proposition 1. □

Existence of the Growth Irrelevance Frontier of Wealth Inequality

Lemma 6 (EXISTENCE GIF). *For a tail of wealth η and of type ϵ, type dependence ϱ, and wealth-dependent risk taking $\gamma \in (\underline{\gamma}, \overline{\gamma})$ with $\underline{\gamma} = \frac{2(2\epsilon-1)}{1+2(2\epsilon-1)}$ and $\overline{\gamma} = 2$, there exists a η^* which lies on the GIF.*

(a) For $\gamma \in (1, \overline{\gamma})$ and $-1 \leq \varrho \leq 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma}$, the GIF exists for a unique $\eta^* \in (\gamma, \infty)$.

(b) With $\gamma = 1$ and $\varrho = 0$, any Pareto tail $\eta^* \in (\gamma, \infty)$ lies on the GIF.

(c) For $\gamma \in (\underline{\gamma}, 1)$, and $2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} \leq \varrho \leq 1$, the GIF exists for a unique $\eta^* \in (\underline{\eta}^{dc}, \infty)$, $\underline{\eta}^{dc} > 1$.

item a. proves that a negative *type* dependence ($\varrho < 0$) is needed for the existence of an economy which lies on the GIF if the *scale* dependence is positive ($\gamma > 0$). Conditions in item b. without *type* dependence requires no *scale* dependence for an economy to lie on the GIF, and conditions in item c. with positive *type* dependence shows that an economy with some $\gamma < 0$ can be located on the GIF.

Proof of Lemma 6: Existence of the $isoG(\eta, 0)$

Proof. Using Lemma 1, the *iso-growth* at level $\overline{g}$ can be implicitly defined as

$$(\phi - A)\overline{g} = -(\phi - A)\Big[(\gamma-1)\Omega^{\gamma} + \varrho(\gamma-1)\Omega^{\varrho\gamma} + \varrho\Omega^{\varrho}\Big] \tag{1.33}$$

Which, for a level $\overline{g} = 0$ can be rewritten as:

$$\left(1 + \frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)}\right)\frac{1-\gamma}{(\eta-\gamma)^2} - 2\frac{\eta-1}{\eta-\gamma}\frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)^2} = 0\,,$$

which we can rearrange to

$$\frac{\varrho\gamma}{(2\epsilon-1)(2\eta-\gamma)^2(\eta-\gamma)^2}\Big((1-\gamma)(2\eta-\gamma) - 2(\eta-1)(\eta-\gamma)\Big) = -\frac{1-\gamma}{(\eta-\gamma)^2}\,,$$

If $(1-\gamma)(2\eta-\gamma) - 2(\eta-1)(\eta-\gamma) \neq 0$ (which only occurs in the case of $\gamma < 1$) we can state the GIF algebraically as

$$\varrho(\gamma, \eta, \epsilon; \overline{g} = 0) = \frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{(2\eta-\gamma)^2}{2(\eta-1)(\eta-\gamma) - (1-\gamma)(2\eta-\gamma)}\,. \tag{1.34}$$

The sign of the derivative w.r.t. to the Pareto tail is determined by

$$\begin{aligned}\operatorname{sgn}\left(\frac{\partial\varrho}{\partial\eta}\right) &= \operatorname{sgn}\Big((1-\gamma)\Big)\operatorname{sgn}\Big(4(2\eta-\gamma)\Big(2(\eta-1)(\eta-\gamma) - (1-\gamma)(2\eta-\gamma)\Big) - (2\eta-\gamma)^2\Big(4(\eta-1)\Big)\Big)\\ &= \operatorname{sgn}\Big((1-\gamma)\Big)\operatorname{sgn}\Big((\gamma-\eta)(2-\gamma)\Big)\,.\end{aligned}$$

Hence, we obtain due to $\eta > \gamma$

$$\frac{\partial\varrho}{\partial\eta}\begin{cases} < 0 & \text{if } \gamma > 2\,,\\ = 0 & \text{if } \gamma = 2\,,\\ > 0 & \text{if } 1 < \gamma < 2\,,\\ = 0 & \text{if } \gamma = 1\,,\\ < 0 & \text{if } 0 < \gamma < 1\,.\end{cases}$$

Before turning to the existence of η^{GIF}, we first study the limits of (1.34). Thus, we obtain

$$\lim_{\eta\to\gamma^+}\varrho(\gamma, \eta, \epsilon) = -\frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{(2\eta-\gamma)}{1-\gamma} = -(2\epsilon-1)\,.$$

Similarly, we have

$$\lim_{\eta \to 1^+} \varrho(\gamma,\eta,\epsilon) = -\frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{(2-\gamma)}{1-\gamma} = -\frac{(2\epsilon-1)(2-\gamma)}{\gamma} .$$

Finally, by an application of L'Hopitals rule we derive

$$\lim_{\eta \to \infty} \varrho(\gamma,\eta,\epsilon) = \frac{(2\epsilon-1)(1-\gamma)}{\gamma}\frac{2\eta-\gamma}{\eta-1}|_{\eta=\infty} = 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} .$$

As a result, we obtain the following bounds on the copula dependence parameter

$$\begin{aligned} -(2\epsilon-1) < \varrho < 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} \qquad & \text{if } \gamma > 1 \\ \varrho = 0 \qquad & \text{if } \gamma = 1 \\ -\frac{(2\epsilon-1)(2-\gamma)}{\gamma} < \varrho < 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma} \qquad & \text{if } \gamma < 1 \end{aligned}$$

It is straightforward to see for $\gamma > 2$ that the required $\varrho \notin \mathcal{R}$ such that the GIF is empty (i.e. $\nexists\, \eta^* \in (\gamma,\infty)$ s.t. $\text{GIF}(\eta^*) = 0$). Contrary, for $1 < \gamma < 2$ there exists by an application of the intermediate value theorem a unique $\eta^* \in (\gamma,\infty)$ such that $\text{GIF}(\eta^*) = 0$ if $-1 < \varrho < \overline{\rho}^{FGM} < 0$, where $\overline{\rho}^{FGM} \equiv 2\frac{(2\epsilon-1)(1-\gamma)}{\gamma}$. Additionally, in the case of $\gamma = 1$, being on the growth irrelevance frontier requires $\varrho = 0$. As a result, for any $\eta \in (1,\infty)$ the GIF goes through the point $\{\gamma = 1, \varrho = 0\}$. Finally, in the case of $\gamma < 1$ we can show that the growth irrelevance frontier is discontinuous at the points (cf. denominator of equation (1.34))

$$\eta_{dc}^{1,2} = 1 \pm \sqrt{1 - \frac{1}{2}\left(3\gamma - \gamma^2\right)} .$$

Recognizing that $3\gamma - \gamma^2$ is strictly increasing in γ on the interval $(0,1)$, we conclude that the expression in the square brackets is strictly positive and lies on the interval $(0,1)$. Due to the imposition of $\eta > 1$, we can hence exclude the smaller solution. Let us subsequently denote by $\eta_{dc}^* \in (1,2)$ the only feasible discontinuity point. As $\frac{\partial \varrho}{\partial \eta} < 0$, the denominator of (1.34) is strictly increasing in η and $\lim_{\eta \to 1^+} < -1$ for $\gamma \in (0,1)$, we conclude that $\varrho > 0$ is a necessary condition for the existence of a GIF solution. This implies that the feasible set of Pareto tails reduces to $\eta \in (\eta_{dc}^*, \infty)$. Ensuring that $\varrho \leq 1$ gives us finally a lower bound on the wealth dependent risk taking parameter such that $\gamma \geq \underline{\gamma} = \frac{2(2\epsilon-1)}{1+2(2\epsilon-1)}$. As a result, for all $\gamma < \underline{\gamma}$ a GIF solution does not exists. Contrary, by an application of the intermediate value theorem an unique $\eta^* \in (\eta_{dc}^*, \infty)$ exists for $\underline{\gamma} \leq \gamma < 1$. This completes the proof of Lemma 6. □

Proof of Lemma 3: Properties of the GIF and iso-growth

Proof. Using equation (1.34) from Lemma 6, we obtain

$$\frac{\partial \varrho}{\partial \gamma} = -(2\epsilon - 1)\frac{1}{\gamma^2}\frac{(2\eta-\gamma)^2}{2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)} + (2\epsilon-1)\frac{1-\gamma}{\gamma}\frac{-2(2\eta-\gamma)\left[2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\right]-(2\eta-\gamma)^2\left[2(1-\gamma)+1\right]}{\Big(2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\Big)^2}\,.$$

Hence, the sign of the previous expression is determined by the sign of

$$\text{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big) = \Big(-(2\eta-\gamma)^2\Big[2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\Big]\Big) + \Big((1-\gamma)\gamma(2\eta-\gamma)\Big[-4(\eta-1)(\eta-\gamma)+2(1-\gamma)(2\eta-\gamma)-2(2\eta-\gamma)(1-\gamma)-2(\eta-\gamma)\Big]\Big)\,.$$

Simplifying terms provides us with

$$\text{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big) = \text{sgn}\Big(\underbrace{-(2\eta-\gamma)\Big[2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma)\Big]}_{\equiv\mathcal{A}} + \underbrace{(1-\gamma)\gamma\Big[-4(\eta-1)(\eta-\gamma)-2(\eta-\gamma)\Big]}_{\equiv\mathcal{B}}\Big)\,.$$

Let us begin with the case $\gamma = 1$. It is straightforward to see that

$$\text{sgn}\Big(\frac{\partial \varrho}{\partial \gamma}\Big)|_{\gamma=1} = \text{sgn}(\mathcal{A}) = \text{sgn}\Big(-2(2\eta-\gamma)(\eta-1)(\eta-\gamma)\Big) < 0\,,$$

such that the GIF is strictly decreasing in the point $\gamma = 1$. For the case $\underline{\gamma} < \gamma < 1$, the reasoning is slightly more evolved. First, notice that $\mathcal{B} < 0$ in this case. Second, requiring that the inner bracket of $\mathcal{A}$ is weakly positive is equivalent to requiring that

$$2(\eta-1)(\eta-\gamma)-(1-\gamma)(2\eta-\gamma) \geq 0 \Leftrightarrow 2\left(\eta^2-\eta\gamma-\eta+\gamma\right)-2\eta+\gamma+2\eta\gamma-\gamma^2 \geq 0$$

Collecting terms gives us the following condition

$$2\eta^2 - 4\eta + 3\gamma - \gamma^2 \geq 0\,. \tag{1.35}$$

We know from Lemma 6 that the GIF is only defined in this case if $\eta > \eta^*_{dc} = 1 + \sqrt{1 - \frac{1}{2}(3\gamma - \gamma^2)}$. Substituting this expression into the former inequality yields

$$\begin{aligned}
\text{LHS} &= 2\left(1 + \sqrt{1 - \frac{1}{2}(3\gamma - \gamma^2)}\right)^2 - 4\left(1 + \sqrt{1 - \frac{1}{2}(3\gamma - \gamma^2)}\right) + 3\gamma - \gamma^2 \\
&= 2 + 4\sqrt{1 - \frac{1}{2}(3\gamma - \gamma^2)} + 2\left(1 - \frac{1}{2}(3\gamma - \gamma^2)\right) - 4 - 4\sqrt{1 - \frac{1}{2}(3\gamma - \gamma^2)} + 3\gamma - \gamma^2 \\
&= 0\,.
\end{aligned}$$

As the left hand side of the inequality (1.35) is strictly increasing in η on the set $\eta > \eta^*_{dc} > 1$, we know that the above inequality is always strictly satisfied. Hence, we conclude that $\mathcal{A} < 0$. Finally, this provides us with

$$\text{sgn}\left(\frac{\partial \varrho}{\partial \gamma}\right)\Big|_{\underline{\gamma} \leq \gamma < 1} = \text{sgn}\left(\mathcal{A} + \mathcal{B}\right) < 0\,.$$

Let us finally consider the case $1 \leq \gamma < \overline{\gamma}$. It is evident that $\text{sgn}(\mathcal{A}) < 0$ and $\text{sgn}(\mathcal{B}) > 0$ hold in this case. Hence, the sign of the derivative is *a priori* undetermined. To show our claim, let us first rewrite the claim on the sign of our initial inequality as

$$\begin{aligned}
&-(2\eta - \gamma)\Big[2(\eta - 1)(\eta - \gamma) - (1 - \gamma)(2\eta - \gamma)\Big] + (1 - \gamma)\gamma\Big[-4(\eta - 1)(\eta - \gamma) - 2(\eta - \gamma)\Big] < 0 \\
\Leftrightarrow &(2\eta - \gamma)(\eta + \gamma - 2) + 2(\eta - 1)\gamma(1 - \gamma) > 0\,.
\end{aligned} \tag{1.36}$$

We know that $\eta > \gamma$ has to hold. Substituting $\eta = \gamma$ into the previous inequality yields

$$2\gamma(\gamma - 1) > 2\gamma(\gamma - 1)^2\,,$$

which is trivially satisfied for $\gamma < 2$. Hence, it suffices to show that the left hand side of (1.36) is increasing in η. To do so, let us take the derivative of (1.36) w.r.t. η and let us simultaneously impose that the derivative is positive:

$$\mathcal{Q}(\eta) \equiv 2(\eta + \gamma - 2) + 2\eta - \gamma + 2\gamma(1 - \gamma) = 4(\eta - 1) + 3\gamma - 2\gamma^2 > 4(\gamma - 1) + 3\gamma - 2\gamma^2 \equiv \mathcal{Q}(\gamma),$$

where the second last inequality holds due to $\eta > \gamma$. Hence, $\mathcal{Q}(\gamma) > 0$ implies also $4(\eta - 1) + 3\gamma - 2\gamma^2 > 0$. To show the validity of the previous inequality, let us compute the values of γ of the second order polynomial $\mathcal{Q}(\gamma)$ for which the function equals exactly zero: $\tilde{\gamma}^{1,2} = \frac{-7 \pm \sqrt{49 - 32}}{-4} = \frac{7}{4} \pm \frac{1}{4}\sqrt{17}$. As a consequence, we have that $\tilde{\gamma}^1 < 1$ and $\tilde{\gamma}^2 > 2$ such that $\mathcal{Q}(\gamma)$ is due to continuity strictly positive on the interval $\gamma \in [1, 2]$. As a result, we know that $\mathcal{Q}(\eta)$ is also strictly positive on the entire interval $\gamma \in [1,2]$ which proves that equation (1.36) is satisfied. This shows

$$\text{sgn}\left(\frac{\partial \varrho}{\partial \gamma}\right)|_{1<\gamma<\overline{\gamma}} = \text{sgn}\,(\mathcal{A}+\mathcal{B}) < 0\,,$$

such that we overall obtain

$$\text{sgn}\left(\frac{\partial \varrho}{\partial \gamma}\right)|_{\underline{\gamma}\leq\gamma<\overline{\gamma}} = \text{sgn}\,(\mathcal{A}+\mathcal{B}) < 0\,.$$

Let us finally define the growth irrelevance equation (1.34) by $\varrho \equiv \mathcal{G}(\gamma)$, where $\{\epsilon, \eta\}$ enter the $\mathcal{G}$ function as constants. Recognize that $\mathcal{G}$ is strictly decreasing on the interval $\underline{\gamma} \leq \gamma < \overline{\gamma}$ and thus *injective*. Additionally, it is differentiable at $\mathcal{G}^{-1}(\varrho)$ and hence continuous on the interval $\mathbb{G}$. As a result, we can define the inverse function of the growth irrelevance frontier equation (1.34) by $\gamma \equiv \mathcal{G}^{-1}(\varrho)$. Consequently, it is straightforward to obtain

$$\frac{\partial \gamma}{\partial \varrho} = \frac{\partial \mathcal{G}^{-1}(\varrho)}{\partial \varrho} = \frac{1}{\mathcal{G}'(\mathcal{G}^{-1}(\varrho))} = \frac{1}{\mathcal{G}'(\gamma)} < 0\,,$$

which completes the first part of the proof of Lemma 3.

PART 2. Let us consider now the general growth irrelevance frontier at an arbitrary growth level $\overline{g}$, possibly different from zero. Without loss of generality let us further assume that $\underline{a} = 1$. The GIF is then implicitly characterized by

$$\frac{(1-\gamma)(2\eta-\gamma)-2(\eta-1)(\eta-\gamma)}{(2\epsilon-1)(2\eta-\gamma)^2(\eta-\gamma)^2}\gamma\varrho + \frac{1-\gamma}{(\eta-\gamma)^2} = \chi_{\overline{g}}\overline{g}\,,$$

where $\chi_{\overline{g}}$ denotes a strictly positive constant which is independent of $\{\varrho, \gamma\}$. Total differentiation of the previous equation yields

$$\chi_\rho d\varrho + \chi_\gamma d\gamma = \chi_{\overline{g}} d\overline{g}\,.$$

Rearranging the previous condition results in

$$d\gamma = -\frac{\chi_\rho}{\chi_\gamma} d\varrho + \frac{\chi_{\overline{g}}}{\chi_\gamma} d\overline{g}\,.$$

On the restricted set of Lemma 6, we have $\chi_\rho < 0$. Additionally, we have that $\frac{\chi_\rho}{\chi_\gamma} = -\frac{1}{\mathcal{G}'(\gamma)}$, which implies $\chi_\gamma < 0$. Hence, an increase in $\overline{g}$ (i.e. lower growth rate) decreases γ, conditional on ϱ. As a result, the GIF shifts downwards. Similarly, if $\overline{g}$ decreases (i.e. higher growth rate), γ increases conditional on ϱ which shifts the GIF upwards. This concludes the proof. □

Proof of Lemma 4

We first solve for the consumption equivalent variation $\Delta^{CE,i}$, defined as the amount of consumption that makes an individual indifferent between the reformed economy with progressivity p_a and the initial status quo situation, such that $\mathbb{E}[u(\widetilde{c}_2^i - \Delta^{CE,i})] = \mathbb{E}[u(c_2^i)]$. We get:

$$E[(1 - exp(-\widetilde{\alpha}^i(\widetilde{c}_2^i - \Delta^{CE,i})))]/\widetilde{\alpha}^i = E[(1 - exp(-\alpha^i c_2^i))]/\alpha^i \,,$$

which is, under the generalized CARA and $\Delta^\alpha = \left(1/\widetilde{\alpha}^i - 1/\alpha^i\right)$, equivalent to

$$\begin{aligned}
\Delta^\alpha - exp\left(-\widetilde{\alpha}^i(\widetilde{x}_2^i - \Delta^{CE,i})\right)/\widetilde{\alpha}^i &= -exp\left(-\alpha^i x_2^i\right)/\alpha^i \\
exp\left(-\widetilde{\alpha}^i(x_2^i - \Delta^{CE,i}) + \alpha^i x_2^i\right) &= \left[1 + \Delta^\alpha exp\left(\alpha^i x_2^i\right)\alpha^i\right](\widetilde{\alpha}^i/\alpha^i) \\
-\widetilde{\alpha}^i(\widetilde{x}_2^i - \Delta^{CE,i}) + \alpha^i x_2^i &= \Delta^c \,,
\end{aligned}$$

where $\widetilde{x}_2^i$ and x_2^i denote the certainty equivalents. Rearranging terms, this yields

$$\Delta^{CE,i} = \widetilde{x}_2^i - x_2^i + \frac{\Delta^c}{\widetilde{\alpha}^i} \,,$$

with $\Delta^c = -\widetilde{\alpha}^i\left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right)x_2^i + \ln\left(1 + \left(\frac{\alpha^i}{\widetilde{\alpha}^i} - 1\right)exp\left(\alpha^i x_2^i\right)\right) + \ln\left(\frac{\widetilde{\alpha}^i}{\alpha^i}\right)$, a term that arises because a change in the progressivity p_a affects a_0^i which impacts the curvature of the utility function through a change in the risk aversion $\widetilde{\alpha}^i$.

We now analyze the effects of introducing a proportional tax on each component. First, let us analyze the effect on $\widetilde{x}_2^i = \widetilde{\mu}_2^i - (\widetilde{\alpha}^i/2)(\widetilde{\sigma}_2^i)^2$, with

$$\begin{aligned}
\widetilde{\mu}_2^i &= \varphi Y + T + \underline{r}\widetilde{a}_0^i + A(1-\varphi)\widetilde{a}_0^i + (\phi - A)(1-\varphi)\omega_1^i\widetilde{a}_0^i \\
((\widetilde{\alpha}^i/2)(\widetilde{\sigma}_2^i)^2) &= \frac{1}{2}\sigma_\kappa^2(\widetilde{a}_0^i)^\gamma\widetilde{\omega}^2\frac{\vartheta^i}{\overline{\overline{\vartheta}}}(1-\varphi)^2 \,,
\end{aligned}$$

we get

$$\begin{aligned}
\frac{\partial\widetilde{\mu}_2^i}{\partial p_a} &= \varphi\frac{\partial Y}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a} + \frac{\partial\underline{r}}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}\widetilde{a}_0^i + \frac{\partial T}{\partial p_a} + \frac{\partial\widetilde{a}_0{}^i}{\partial p_a}\left[A(1-\varphi) + \underline{r} + \gamma\widetilde{\omega}\frac{\vartheta^i}{\overline{\overline{\vartheta}}}(\widetilde{a}_0^i)^{\gamma-1}(1-\varphi)(\phi - A)\right] \\
\frac{\partial((\widetilde{\alpha}^i/2)(\widetilde{\sigma}_2^i)^2)}{\partial p_a} &= \frac{\partial\widetilde{a}_0{}^i}{\partial p_a}\left[\gamma(1/2)\sigma_\kappa^2(\widetilde{a}_0^i)^{\gamma-1}\widetilde{\omega}^2\frac{\vartheta^i}{\overline{\overline{\vartheta}}}(1-\varphi)^2\right] \,,
\end{aligned}$$

which yields:

$$\frac{\partial\widetilde{x}_2^i}{\partial p_a} = \varphi\frac{\partial Y}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a} + \frac{\partial\underline{r}}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}\widetilde{a}_0^i + \frac{\partial T}{\partial p_a} + \frac{\partial\widetilde{a}_0{}^i}{\partial p_a}\left[A(1-\varphi) + \underline{r} + \gamma\frac{\vartheta^i}{\overline{\overline{\vartheta}}}(\widetilde{a}_0^i)^{\gamma-1}x_r\right] \,,$$

where $x_r = \widetilde{\omega}(\phi - A)(1-\varphi) - \frac{1}{2}\sigma_\kappa^2\widetilde{\omega}^2(1-\varphi)^2$.

Concerning the term Δ^c, notice that:

$$\ln\left(1+\left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right)exp\left(\alpha^i x_2^i\right)\right) = \left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right)exp\left(\alpha^i x_2^i\right) + \mathcal{E}(a_0^i)\,,$$

where since $\left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right) < 0$ we have $\mathcal{E}(a_0^i) < 0$ an approximation error to the transformation $\ln(1+x) \approx x$. Using this, we can rewrite:

$$\begin{aligned}
\Delta^c &= -\widetilde{\alpha}^i\left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right)x_2^i + \left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right)exp\left(\alpha^i x_2^i\right) + \mathcal{E}(a_0^i) + \ln(\widetilde{\alpha}^i/\alpha^i) \\
&= \left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right)\left(exp\left(\alpha^i x_2^i\right) - \widetilde{\alpha}^i x_2^i\right) + \mathcal{E}(a_0^i) + \ln(\widetilde{\alpha}^i/\alpha^i)\,,
\end{aligned}$$

using $\ln(\widetilde{\alpha}^i/\alpha^i) \approx -\left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right)\frac{\widetilde{\alpha}^i}{\alpha^i}$, we get

$$\Delta^c \approx \left(\frac{\alpha^i}{\widetilde{\alpha}^i}-1\right)\left(exp\left(\alpha^i x_2^i\right) - \frac{\widetilde{\alpha}^i}{\alpha^i}(\alpha^i x_2^i + 1)\right) + \mathcal{E}(a_0^i)\,.$$

Using the welfare function, the optimal progressivity p_a solves:

$$\frac{\partial \mathcal{W}}{\partial p_a} = \int s(a_0,\vartheta)\frac{\partial \Delta^{CE}(a_0,\vartheta)}{\partial p_a}\mathcal{G}(a_0,\vartheta) = 0\,,$$

which can be rewritten

$$\begin{aligned}
\int s(a_0,\vartheta)\left[\varphi\frac{\partial Y}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a} + \frac{\partial \underline{r}}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}\widetilde{a}_0^i + \frac{\partial T}{\partial p_a} + \frac{\partial \widetilde{a_0}^i}{\partial p_a}\left[A(1-\varphi)+\underline{r}+\gamma\frac{\vartheta^i}{\overline{\overline{\vartheta}}}(\widetilde{a}_0^i)^{\gamma-1}x_r\right] + \frac{\partial(\Delta^c/\widetilde{\alpha}^i)}{\partial p_a}\right]\mathcal{G}(a_0,\vartheta) \\
= 0\,,
\end{aligned}$$

or equivalently using $\int s(a_0,\vartheta)\mathcal{G}(a_0,\vartheta) = 1$,

$$\begin{aligned}
&\underbrace{\varphi\frac{\partial Y}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}}_{\text{efficiency}} + \underbrace{\frac{\partial \underline{r}}{\partial\widetilde{\eta}}\frac{\partial\widetilde{\eta}}{\partial p_a}\int s(a_0,\vartheta)\widetilde{a}_0\mathcal{G}(a_0,\vartheta)}_{\text{rent extraction}} + \underbrace{\frac{\partial T}{\partial p_a}}_{\text{lump-sum transfers}} \\
&\quad + \int s(a_0,\vartheta)\left(\underbrace{\frac{\partial\widetilde{a_0}^i}{\partial p_a}\left[A(1-\varphi)+\underline{r}+\gamma\frac{\vartheta^i}{\overline{\overline{\vartheta}}}(\widetilde{a}_0^i)^{\gamma-1}x_r\right]}_{\text{direct effect on level } a_0^i} + \underbrace{\frac{\partial(\Delta^c/\widetilde{\alpha}^i)}{\partial p_a}}_{\text{direct effect on } \alpha^i}\right)\mathcal{G}(a_0,\vartheta) = 0\,.
\end{aligned}$$

Proof of Lemma 5

Proof. Let us first observed that we can rewrite the utility function as:

$$\max_{k_i^1, b_1^i} \left(\frac{1}{1-1/\sigma}\right) \left[\left(a_0^i - k_1^i - b_1^i\right)^{1-1/\sigma} + \beta \left(\mu_{c_2}^i(k_1^i, b_1^i) - \frac{\alpha_i}{2}\sigma_{c_2}^i(k_1^i, b_1^i)\right)^{1-1/\sigma} \right], \tag{1.37}$$

with $k_1^i = \omega_1^i a_1^i$ and $b_1^i = (1-\omega_1^i)a_1^i$. To do that, notice that

$$U\Big(G^{-1}\Big(\mathbb{E}\Big[G\Big(U^{-1}(u_2)\Big)\Big) = \Big(G^{-1}\left(\mathbb{E}\left[G(c_2^i)\right]\right)\Big)^{1-1/\sigma} / (1-1/\sigma),$$

using the fact that $x = -(1/\alpha^i)\ln(1-\alpha^i G(x))$ and $\mathbb{E}\left[G(c_2^i)\right] = (1/\alpha^i)(1-exp(-\alpha^i x_{c_2}^i))$ as shown above, we get:

$$G^{-1}\left(\mathbb{E}\left[G(c_2^i)\right]\right) = -(1/\alpha^i)\ln\left(1-\alpha^i(1/\alpha^i)\left(1-exp(-\alpha^i x_{c_2}^i)\right)\right) = x_{c_2}^i.$$

Hence the program of the agent can be rewritten as in (1.37).

Combining both the first order conditions with respect to k_1^i and b_1^i yields

$$k_1^i = \frac{(\phi - A)(1-\varphi)}{\alpha^i(1-\varphi)^2\sigma_\kappa^2} = \widetilde{\omega}\frac{\vartheta^i}{\overline{\vartheta}}(a_0^i)^{\gamma-1}.$$

Using the condition with respect to b_1^i, we get

$$(a_0^i - b_1^i - k_1^i) = x_{c_2}^i(\widetilde{R}\beta)^{-\sigma},$$

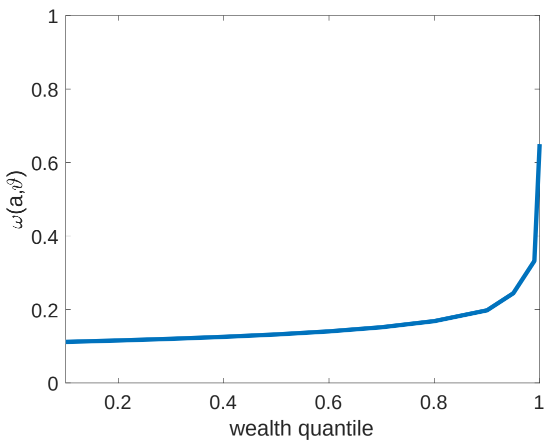
where $\widetilde{R} = \underline{r} + A(1-\varphi)$. Using the expression of $\mu_{c_2}^i = \varphi Y + T + (\underline{r} + A(1-\varphi))b_1^i + (\underline{r} + \phi(1-\varphi))k_1^i$ and $x_{c_2}^i = \mu_{c_2}^i - \frac{\alpha^i}{2}\sigma_{c_2}^i$, we obtain

$$\begin{aligned} b_1^i((\widetilde{R}\beta)^\sigma + \widetilde{R}) &= a_0^i(\widetilde{R}\beta)^\sigma - k_1^i\left[(\widetilde{R}\beta)^\sigma + \underline{r} + \phi(1-\varphi)\right] - \varphi Y - T + \frac{\alpha^i}{2}\sigma_{c_2}^i \\ b_1^i &= ((\widetilde{R}\beta)^\sigma + \widetilde{R})^{-1}\left(a_0^i(\widetilde{R}\beta)^\sigma - k_1^i\left[(\widetilde{R}\beta)^\sigma + \underline{r} + \phi(1-\varphi)\right] - \varphi Y - T + \frac{\alpha^i}{2}\sigma_{c_2}^i\right). \end{aligned}$$

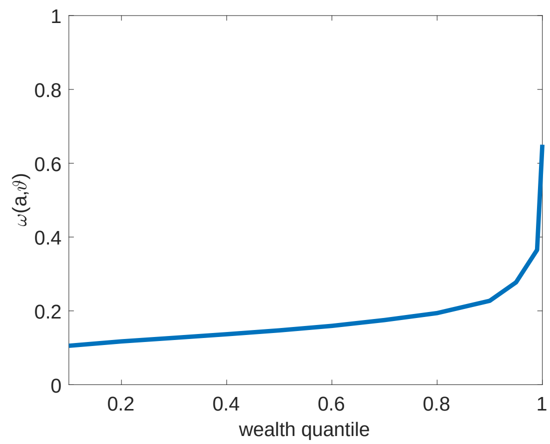
□

1.A.2 Fit of the Static Model under pure type/scale dependence

Figures 1.A.1a and 1.A.1b show the fit of the type and scale dependence models regarding the distribution of risky asset shares across the wealth distribution. To fit this shape, we fix inequality to $\eta = 1.4$. The parameter $\gamma = 1.39$ is used to match the average risky asset share of the top 1% in the pure *scale* dependence model. In the pure *type* dependence model, we use the parameter controlling the correlation (to 0.65) between the two distributions and fix the Pareto shape of types to $\varepsilon = 2$. The two models produce an extremely close fit of the observed distribution of the average portfolio allocations across the wealth distribution relative to the one observed in Figure 1.3.



(a) wealth dependence model



(b) type dependence model

1.B Empirical Appendix

1.B.1 Adjusted Survey of Consumer Finance

Throughout the paper, we use the Survey of Consumer Finance (SCF) from 1998 to 2019 (eight waves). Each wave provides cross-sectional data on U.S. households' income and wealth, including detailed information regarding portfolio allocation as well as demographic characteristics.

Details on sampling

Households in the SCF are selected from a double sampling procedure. A first sample is selected from a standard sampling procedure, providing a good representativity of the population. A second sample selects very high income families from the tax records of the Internal Revenue Service (IRS), with some that are also likely to be very wealthy. The SCF weights are used to combine individual characteristics from the two samples to make estimates for the full U.S. population.

Correcting for under-representation and under-reporting

Wealth and income concentration measures from survey data face two common issues: (i) under-representation, meaning that wealth-rich households are generally under-represented in survey data, and (ii) underreporting of assets, meaning that individuals tend to under-report wealth, especially financial wealth.

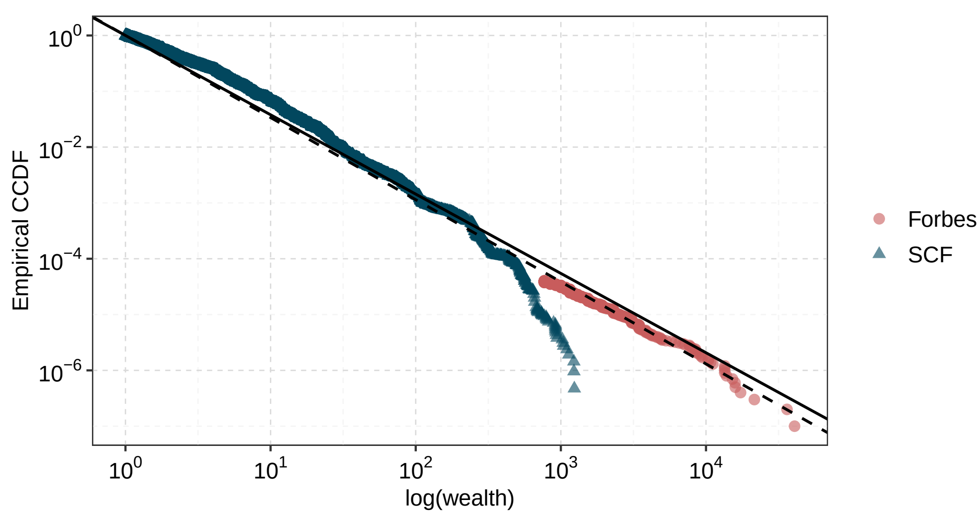
We correct for those issues using the procedure described in Vermeulen (2016). The method is iterative and proceed by assuming that the wealth distribution can be well approximated by a Pareto Law at the top. While this assumption is questionable, most countries admit a linear log-log relationship between wealth level and its empirical CCDF at the top of the distribution, indicating that a Pareto Law can well describe the distribution. The method employs this property to estimate a country-specific Pareto tail and use it to extrapolate estimates for top wealth shares.

First, observation at the top of the wealth distribution in the SCF (above a given threshold of wealth) are supplemented by an external source – such as the Forbes World's Billionaires lists – in order to estimate a Pareto Law using additional observations at the very top. It can be shown that an estimate of the Pareto tail can be obtained by regressing:

$$\ln(n(a_i)/n) = -\eta ln(a_i/a_{min}), \tag{1.38}$$

where $n(a_i)$ is the number of sample observations that have wealth at or above a_i, i.e. the rank of the observation, and a_{min} is the minimum wealth level at which we assume that the wealth distribution is Pareto (we fix it to 1 million of dollars). $\ln(n(a_i)/n)$ is the log of the relative frequency (or empirical ccdf). Figure 1.B.1 shows the resulting Pareto tail $\hat{\eta}$ estimates without and with the 2010 Forbes World's Billionaires lists using the 2010 SCF. From this, we generate corrected for missing value wealth shares by reconstructing a theoretical Pareto distribution at the top.

Figure 1.B.1. Pareto shape estimation for the United States using the SCF.



Empirical CCDF in the US combining SCF (2010) and the Forbes World's Billionaires lists. The dashed line corresponds to the estimate of the Pareto tail without the Forbes observations. The solid line includes observations from the rich list.

Second, aggregate estimates from the entire wealth distribution, below and above the threshold of wealth, are computed for total, financial and non-financial assets. Households' wealth are then adjusted such that aggregate estimates from the survey data supplemented with the Forbes's list coincide with national households balance sheet.

The procedure iterates until the distribution of wealth is invariant.[51]

1.B.2 Measuring Returns to Wealth in the PSID

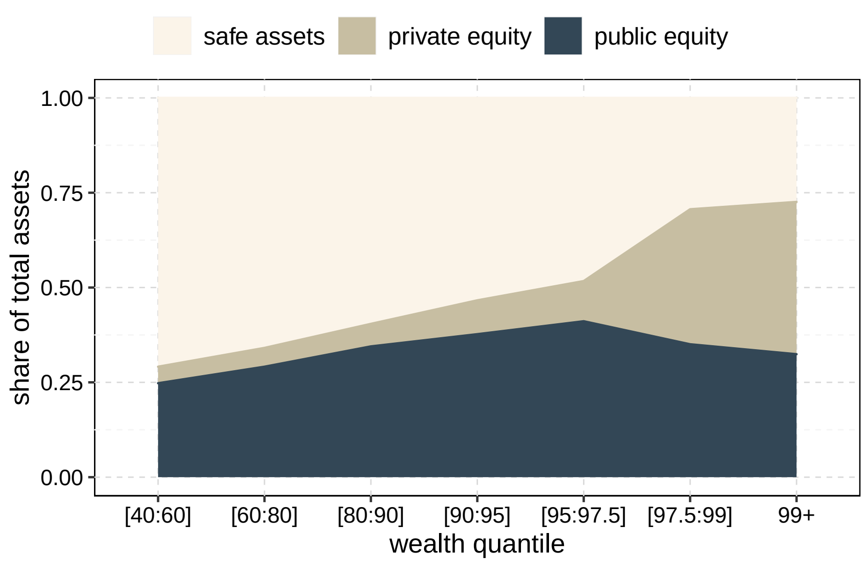
We use the PSID to compute returns to wealth along the wealth distribution. Our sample spans the period from 1998 to 2019. Every two years, there is a new wave (thus ten in total). The initial period is lost when we compute the returns.

We follow closely the procedure of Fagereng et al. (2020). Our sample considers households whose the head is aged between 20 and 70. This is to ensure that the financial decision maker is the holder of the assets. Moreover, our model abstracts from many decisions that may occur at the end of life (voluntary bequests, health expenditures etc). We restrict our attention to households with at least $1000 of net worth. This ensures that returns to wealth are finite.

Portfolio composition in the PSID

Before turning to the analysis of returns, we first display in Figure 1.B.2 the average portfolio composition across the wealth distribution in the PSID. As can be seen, and consistent with our SCF sample, the share of risky assets is increasing along the wealth distribution, reaching around 60% within the top 5% (25% in private equity and 35% in public equity).

Figure 1.B.2. Portfolio composition in the PSID.



[51] While there is no reason to think that this procedure converges to a fixed point, it appears that this is indeed the case for the US economy and all European countries from the HFCS (see the Online Appendix OA 3.1.3).

Returns to wealth

We compute returns to *net worth* and for three categories: safe assets, public equity and private equity. Following our justification in the core paper, safe assets include housing (primary and secondary residence including rental properties or cottages), riskfree assets (checking/saving accounts, money market funds, certificates of deposits, government bonds, or treasury bills) and other assets (boats, motor homes, cars, cash value in a life insurance policy, a valuable collection, or rights in a trust or estate). Private equity includes businesses and farms, and public equity includes direct holdings in publicly held corporations and indirect holdings in mutual funds, investment trusts and through employer-based pensions or IRAs.

Our definition of net worth is the amount the household would receive if they would sold their assets and paid off all debts associated with the asset. Total liabilities include loans, mortgages, consumer credits and other loans.

We define the total amount of gross assets as:

$$a^g_{i,t} = a_{i,riskfree,t} + a_{i,home,t} + a_{i,secondary,t} + a_{i,other,t} + a_{i,priv,t} + a_{i,public,t} \,. \tag{1.39}$$

The PSID reports inflows and outflows from each assets. Inflows represent all investments, additions and upgrades of assets, while outflows represent all disinvestment, liquidation, and asset sales. We denote $F_{i,l,t}$ the net inflow (total inflows minus total outflows) for asset $l \in \{riskfree, home, secondary, priv, public, other\}$.

Asset values are available for each period for holdings of public equities and for primary residences. Unfortunately, asset values for private businesses and secondary housing are only observed from 2011 onward. Prior to 2011, only the *equity* value (marketable asset value minus total debt associated with the asset $e_{i,l,t} = a_{i,j,t} - d_{j,l,t}$) is observed for secondary housing and private business assets. We impute the asset value using backward induction using $\Delta^a_{i,l,t} = a_{i,l,t} - a_{j,l,t-2} \approx e_{i,l,t} - e_{i,l,t-2}$ and net inflows $F_{i,l,t}$, such that:

$$a_{i,l,t-2} = a_{i,l,t} - \Delta^a_{i,l,t} - F_{i,l,t} \,, \qquad l \in \{priv, secondary\} \,,$$

which implies that any variation in debt $d_{i,l,t}$ translates one-to-one into variations in the value of the asset $a_{i,l,t}$. We now compute the *pre-tax* returns to wealth using our PSID sample. To do so, we make a number of steps to compute capital gains, income and costs.

Capital gains We compute (unrealized and realized) capital gains by comparing the value of each asset at two consecutive waves while taking into account inflows and outflows from this asset. Specifically, we compute capital gains of a household i, period t and asset l as follows.

- For primary residence (*home*), unrealized capital gains are defined as the difference between the current marketable value $a_{i,home,t}$ minus the past one (in the previous wave) $a_{i,home,t-2}$. Realized capital gains are defined as the selling price p_{home}^{sell} less the marketable value in the previous wave, $a_{i,home,t-2}$. To isotate the variations due to capital gains, we take into account inflows and outflows. Inflows are all additions and upgrades, denoted $I_{i,home,t}$. Outflows take into account that the stock of housing depreciate, at a rate of $\delta_h = 2.0\%$ which is the average maintenance and repair cost over the marketable asset value available from 2005 onward, such that: $F_{i,home,t} = I_{i,home,t} - \delta_h(a_{i,home,t} + a_{i,home,t-2})$. We obtain that capital gains for the primary residence are given by:

$$R_{i,home,t}^{K} = \frac{\mathbb{1}_{\{sold=1\}} p_{home}^{sell} + \mathbb{1}_{\{sold=0\}} a_{i,home,t} - a_{i,home,t-2} - F_{i,home,t}}{2},$$

where we divide by 2 to annualize the capital gains.

- For other assets (public equity, private equity and secondary housing), unrealized and realized capital gains $R_{i,l,t}^{K}$ are defined as the difference between the current marketable asset value and the one observed in the previous wave, minus net inflows $F_{i,l,t}$, such that:

$$R_{i,l,t}^{K} = \frac{a_{i,l,t} - a_{i,l,t-2} - F_{i,l,t}}{2} \qquad l \in \{priv, public, secondary\},$$

Notice that in case of new business acquisition of a value equals to A, we obtain $F_{i,priv,t} = A$ $a_{i,l,t-2} = 0$ and $a_{i,l,t} = A$. Therefore, $R_{i,priv,t}^{K} = 0$. We obtain the reverse in case of liquidation. As such, our measure of capital gains accommodates for new acquisitions and total sales.

Capital income For each category, capital income refers to the sum of capital income earned by the head and the spouse. We define capital income for each asset category l as follows.

- Capital income from primary residence, $R_{i,home,t}^{I}$, takes into account maintenance cost and the rental value in the calculation. Lacking evidence on those two components, we assume:

$$R_{i,home,t}^{I} = r_h a_{i,home,t-2} + inc_{i,home,t}^{rent} - \delta_h a_{i,home,t-2}$$

where $inc_{i,housing,t}^{rent}$ is the rental income reported to all housing assets. We attribute rental income to the primary residence when the household has no secondary residence, in such case we subtract 0.5 $utils_{i,home,t}$ from the rents, and to secondary residence otherwise. Finally, consistent with Flavin and Yamashita (2002), we assume that the housing yield have an interest component with rate $r_h = 5\%$. Again, δ_h denotes the depreciation rate. Apart from the fact that we directly observe net inflows, it should be noticed that our approach is different from Flavin and Yamashita (2002). Adopting exactly their specification would shift upward returns for primary residence, but turns to be less consistent with values reported in Fagereng et al. (2020) and Bach et al. (2020).

- Capital income from secondary residence, $R^I_{i,secondary,t}$, depends on whether the property is rented, occupied by the household, or not occupied. We assume that

$$R^I_{i,secondary,t} = \mathbb{1}_{\{occupied\}} r_h a_{i,secondary,t-2} + \mathbb{1}_{\{rented\}} inc^{rent}_{i,secondary,t} - \delta_h a_{i,secondary,t-2}\}$$

- Capital income from private equity businesses, $R^I_{i,priv,t}$ is computed as follows. Private equity income is split evenly between labor and asset income if the household actively participates in a private business and only to asset income otherwise.

- Capital income from riskfree assets, $R^I_{i,riskfree,t}$, is obtained from interest income reported in the PSID. As there is no distinction between the fraction of interest income coming from safe assets relative to public equity, we proceed as follows. We assume that interest income from riskfree assets is given by the maximum between the reported interest income $inc^{interest}_{i,t}$ and income derived from the 1-year Treasury bill secondary market rate times the value reported from bond interest, i.e. $r^{treasury}_{i,t} \times a_{i,bond,t}$, such that: $R^I_{i,riskfree,t} = \min\{r^{treasury}_{i,t} \times a_{i,bond,t}, inc^{interest}_{i,t}\}$. A positive difference $\Delta^{interest}_{i,t} = inc^{interest}_{i,t} - R^I_{i,riskfree,t}$ is then associated to public equity. Results are not very sensitive to this assumption.

- Capital income from public equity, $R^I_{i,public,t}$, is equal to the sum of dividends, income from stocks held into IRAs and pension accounts, other interest income and trusts. We obtain income from public equity as: $R^I_{i,public,t} = \Delta^{interest}_{i,t} + inc^{dividend}_{i,t} + inc^{IRA}_{i,t} + inc^{trust}_{i,t} + inc^{other\ fin}_{i,t}$.

Capital debt cost The last important component of our definition of returns are debts. For primary residence, the cost of debt, $R^D_{i,home,t}$ corresponds to the repayment of mortgages. The PSID contains information for two mortgages. We compute the average mortgage interest rate as a weighted average between them and deflate this rate using the CPI index. We follow Flavin and Yamashita (2002) and assume that interest payment are deductible, such that the household pays a real after-tax interest rate of $r^D_{i,home,t} = \frac{1+(1-\tau)r^{mortgage}_{i,t}}{1+inflation_t} - 1$, where we set $\tau = 33\%$. All other costs of debt are computed assuming an interest rate of 5% on other remaining debts.

Return measure Following Fagereng et al. (2020), our reference measure of return of an asset l is:

$$r_{i,l,t} = \frac{R^K_{i,l,t} + R^I_{i,l,t} - R^D_{i,l,t}}{a_{i,l,t-1} + F_{i,l,t}/2} \tag{1.40}$$

The numerator is the sum of income, $R^I_{i,l,t}$, capital gains, $R^K_{i,l,t}$, minus the cost of debt, $R^D_{i,l,t}$ accrued by household i on asset l in year t. The denominator is defined as the sum of beginning-of-period stock of gross wealth and net flows of gross wealth during the year. In the PSID, wealth is observed only at the time of the interview while income for each asset are observed for the past year.

Because the periodicity of the data is biennial, we need to impute the beginning-of-period asset level corresponding to the income derived from the asset. We do so by assuming that beginning-of-period asset is the interpolation between the current wealth level and the wealth level reported in the previous wave, such that: $a_{i,l,t-1} = (a_{i,l,t-2} + a_{i,l,t})/2$. The second term on the denominator, $F_{i,j,t}$, accounts for the fact that asset yields are generated not only by beginning-of-period wealth but also by additions/subtractions of assets during the year. Without this adjustment, we may bias our estimates if the beginning-of-period wealth is small but capital income is large due to positive net asset flows occurring during the period (for example, a business acquisition). As the flows occur during the year, we make the assumption that they occur on average in mid-year.

In equation (1.40), we express the dollar yield on net worth as a share of gross wealth (or total assets) to ensure that the sign of the return reflect the sign of the yield.

Notice that for *net worth*, we define:

$$r_{i,t}^{networth} = \frac{\sum_l (R_{i,l,t}^K + R_{i,l,t}^I - R_{i,l,t}^D)}{a_{i,t-1}^g + \sum_l F_{i,l,t}/2} \qquad r_{i,t}^{gross} = \frac{\sum_l (R_{i,l,t}^K + R_{i,l,t}^I)}{a_{i,t-1}^g + \sum_l F_{i,l,t}/2} \tag{1.41}$$

We convert all nominal returns to real returns using the consumer price index (CPI) from the Federal Reserve, using: $\tilde{r}_{i,l,t} = \frac{1+r_{i,l,t}}{1+inflation_t} - 1$.

Trimming Finally, we trim the distribution of returns in each year t and for each asset category l at the top and the bottom by 0.5%. This ensures that there is no outlier polluting the estimates of the mean of returns and aim to reduce measurement errors.

Scale dependence in returns

In this section, we use the PSID to evaluate the presence of scale dependence in the returns to wealth. To do so, we follow Gabaix et al. (2016) and Fagereng et al. (2020) representation and estimate the following statistical model:

$$r_{i,t}^{gross} = \theta^g P_a(a_{i,t-1}^g) + f_t + f_i + \epsilon_{i,t}^g\,, \qquad r_{i,t}^{networth} = \theta^n P_a(a_{i,t-1}^n) + f_t + f_i + \epsilon_{i,t}\,, \tag{1.42}$$

where $r_{i,t}^{gross}$ and $r_{i,t}^{networth}$ are respectively the return to gross and net wealth, $P_a(\cdot)$ is the percentile of beginning-of-period gross/net wealth (capturing scale dependence), f_i^g, f_i^n are the individual fixed effect (capturing persistent heterogeneity), f_t^g, f_t^n are time fixed effects capturing aggregate return components, and $\epsilon_{i,t}^g, \epsilon_{i,t}^n$ are error terms. *Scale* dependence is measured by the parameter θ^g for gross returns and by the parameter θ^n for net returns, while *type* dependence is captured by the individual fixed effect. Thus, similar to Fagereng et al. (2020), the *scale* dependence (comprising all sources of scale dependence, direct and indirect) parameters are identified from household-specific time variations in the wealth percentile.

Table 1.B.1. Scale dependence regression

	percentile		log specification	
	Gross Return	Net Return	Gross Return	Net Return
θ	0.152	0.121	0.034	0.024
Time FE	Yes	Yes	Yes	Yes
Observations	9286	7491	9286	7491

We find a positive and statistically significant degree of *scale* dependence in returns to wealth in the PSID, confirming the findings of Fagereng et al. (2020) and Bach et al. (2020) who use Scandinavian administrative data. Moreover, the estimates are surprisingly close to the ones obtained by Fagereng et al. (2020) using Norwegian data. Notice that we obtain a similar qualitative message regarding the relationship between returns to wealth and wealth if we regress $log(a^g_{i,t-1})$ and $log(a^n_{i,t-1})$ on gross and net returns, respectively.

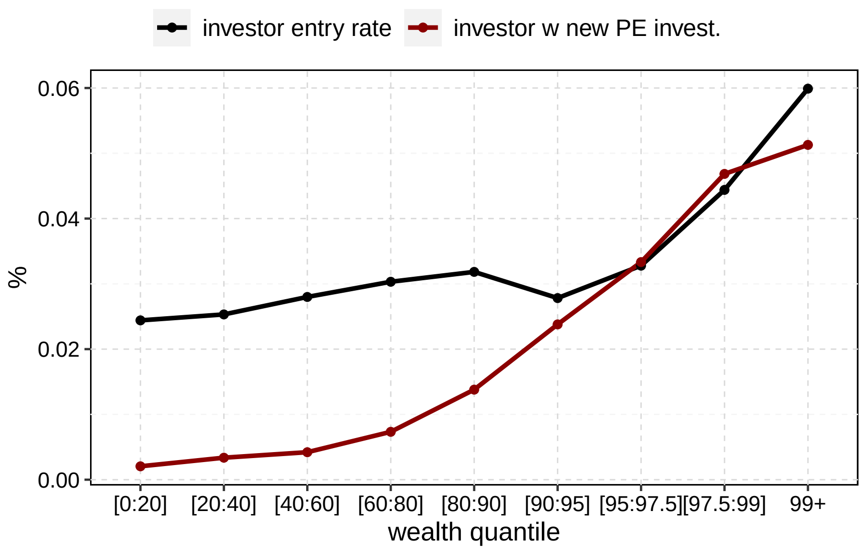
1.B.3 PSID participation and diversification

In Figure 1.B.3, we use our PSID sample to show the unconditional average probability to switch from a non private equity investor state to a private equity investor state (black line), and the unconditional fraction of private equity investors who invest in a new additional private equity investment (red line). Focusing on the black line, it is apparent that there is an increasing and convex relationship between risky investment participation and wealth, a feature which is consistent with Hurst and Lusardi (2004). Focusing on the red line, it appears that, conditional on being already private equity investor, new additional investments in private equity occur essentially at the very top. We interpret this as evidence for diversification among the wealthy households, which is consistent with the recent paper by Penciakova (2018). We now investigate these relationships while controlling for household's characteristics.

Conditional relationships In the main text, we model the fact that the entry into "investor" state increases with wealth. To this, Hurst and Lusardi (2004) estimate the probability to become a private equity investor as a function of wealth while controlling for household's characteristics. They then compute the predicted probability to participate into private equity investments along the wealth distribution. They find that only the wealthy households (above the top 95^{th} percentile) are more likely to switch to private equity business ownership when their wealth increases. Following their lead and using similar controls, we update their estimation using our sample period. In Figure 1.B.4, we report the predicted probability as estimated in Hurst and Lusardi (2004) as well as our update. Results are found to be very similar.

Finally, we report in Figure 1.B.5 the predicted probability of acquiring a new private equity business investment, conditional on being an investor. Consistent with what is observed in the

Figure 1.B.3. Private equity (PE) investment participation and diversification in the PSID.



Survey of Consumer Finance, we find that the probability to invest in multiple businesses increases at the top of the wealth distribution. While we cannot control for household's characteristics in the SCF, our results from the PSID reveal that this relationship between new acquisition of a private equity business investment and wealth is robust and do not hinge on an observed household characteristic (which may be a proxy for "type") only.

1.C Quantitative Appendix

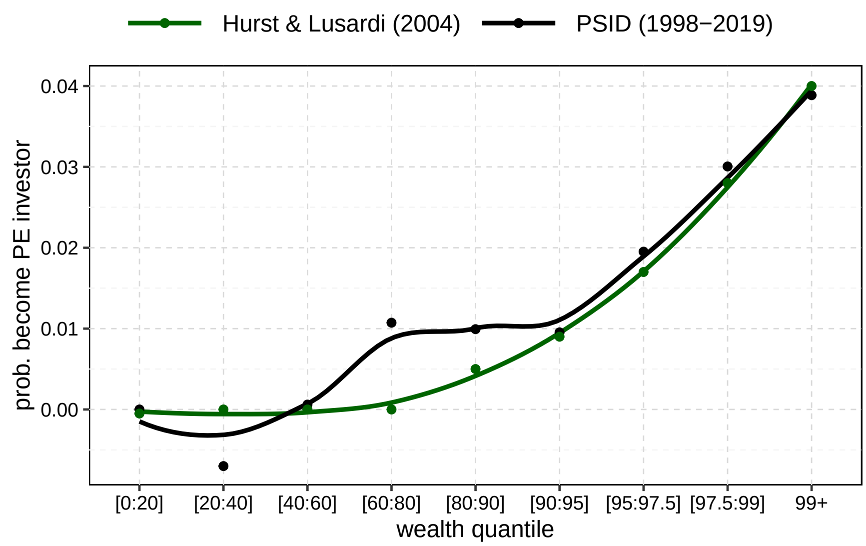
In this section, we show additional details regarding the quantitative model, the calibration, the computational algorithm and additional moments of interest.

1.C.1 Calibration

Benchmark economy (M1) Table 1.C.1 displays the scale dependence in the intensive margin of portfolio allocation, conditional on being an investor, in the data and in the benchmark economy. In the benchmark economy, the parameter $\{\underline{a}_\omega, \omega_1, \omega_2, \gamma_\omega\}$ are used to match the following moments. $\underline{a}_\omega$ is chosen to correspond to the wealth level of the 70^{th} wealth percentile in the model, such that there is no scale dependence in risky portfolio observed below this percentile. The level and the shape parameters are endogenously set to $\omega_1 = 0.072$ and $\gamma_\omega = 0.30$ in order to replicate the average share invested in risky equity (through additional investments) of 11% for households within the [95-97.5] wealth quantile, and of 20% for households within the [99-99.9] wealth quantile. A maximal value of $\omega_2 = 0.20$ is set to guarantee that the wealth distribution is stationary and to match the average risky portfolio share above the top 0.1%, as observed in Figure 1.5.

Table 1.C.2 reports the scale dependence in the risky investment participation in the bench-

Figure 1.B.4. *Scale* dependence in investment participation.



Predicted probability relative to the mean.

Table 1.C.1. Resulting scale dependence in risky portfolio share: benchmark model and data

Wealth quantile	[0–80]	[90–95]	[95–97.5]	[97.5–99]	[99–99.9]	top 0.1%
SCF average risky share	0%	6%	11%	15%	20%	20%
Benchmark model $\omega(a, 1)$	0%	6.7%	10.3%	12.6%	16.9%	20%

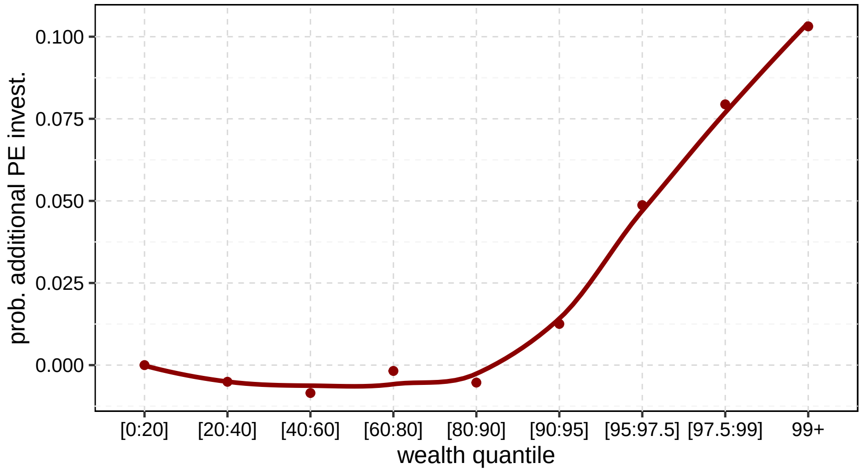
mark economy and in the PSID (see Figure 1.B.4 and Hurst and Lusardi (2004)). In the benchmark economy, the parameters $\{\underline{a}_\lambda, \lambda_1, \lambda_2, \gamma_\lambda\}$ are used to match the following moments. $\underline{a}_\lambda$ is chosen such that it corresponds to the wealth level of the 80^{th} wealth percentile in the model, such that there is no scale dependence in the risky investment participation below this percentile. In this case, the participation rate of 1.8% is only generated through the process governing ϑ-types. The level and the shape parameters of the transition probability with respect to wealth are endogenously chosen to be $\lambda_1 = 0.071$ and $\gamma_\lambda = 0.30$ to replicate the average transition rate of 3.2% for households within the [95-97.5] wealth quantile, and of 6.1% for households within the [99-99.9] wealth quantile. A maximal value of $\lambda_2 = 0.045$ is set to guarantee that the wealth distribution is stationary and to match the maximum transition rate of 7% at the very top (within the top 0.1%).

Table 1.C.2. Resulting scale dependence in the risky investment participation: benchmark model and data

Wealth quantile	[0–80]	[90–95]	[95–97.5]	[97.5–99]	[99–99.9]	top 0.1%
PSID entry rate into "investor" state	1.8%	2.1%	3.2%	4.5%	6.1%	7.0%
Benchmark model entry rate $\underline{\pi_\vartheta} + \lambda(a)$	1.8%	2.2%	3.2%	4.4%	6.3%	7.0%

Scale model (M2) In this model, we follow Hubmer et al. (2020) and match the increasing average risky asset shares along the wealth distribution through scale dependence only. In this model,

Figure 1.B.5. *Scale* dependence private equity business investment.



Predicted probability relative to the mean.

the only source of scale effects comes from the intensive margin of risky portfolio shares such that equation (1.43) in the main text becomes:

$$\omega(a,\vartheta) \equiv \omega(a) = \underline{\omega} + \varpi(a) = \underline{\omega} + \min\left\{\omega_1\left(\max\left\{a - \underline{a}_\omega, 0\right\}\right)^{\gamma_\omega}, \omega_2\right\}, \tag{1.43}$$

The parameters $\{\underline{\omega}, \underline{a}_\omega, \omega_1, \omega_2, \gamma_\omega\}$ are recalibrated to replicate the following moments. $\underline{a}_\omega$ is chosen to correspond to the wealth level of the 80^{th} wealth percentile in the model, such that there is no scale dependence in the risky portfolio share below this percentile. In this case, the average risky portfolio share is fixed to $\underline{\omega} = 10\%$, following our estimates from the SCF. The level and the shape parameters of the portfolio allocation with respect to wealth are endogenously set to $\omega_1 = 0.057$ and $\gamma_\omega = 0.57$ to replicate an average portfolio share of 37% for households within the [95-97.5] wealth quantile, and of 60% for households within the [99-99.9] wealth quantile. A maximal value of $\omega_2 = 0.75$ is set to guarantee that the wealth distribution is stationary and to match the maximum average portfolio share invested at the very top (within the top 0.1%).

Type model (M3) In the type model, there are no scale effects ($\lambda(a) = \varpi(a) = 0$). In this model, we recalibrate $\underline{\omega}$ such that the average risky share in the economy corresponds to the one observed in the SCF. The switching probability $\underline{\pi}_\vartheta$ is set to match the fraction of investors.

1.C.2 Computational appendix

State space and grid definition

In our model, an household is fully characterized by a state vector $\mathbf{s} = (a, \vartheta, h, j) \in \mathbb{S} \equiv \mathbb{R}^+ \times \Theta \times \mathbb{H} \times \mathcal{J}$. We compute the household problem using a grid of asset $\mathbf{a}$ of 350 points (adding more points only marginally increase our accuracy) spaced according to an exponential rule. We truncate the grid for asset to $A_{\max} = 1500000$ and we impose the borrowing constraint such that $A_{min} = 0$. Due to the finite upper bound on wealth and the Pareto property coming from heterogeneous returns (see Benhabib et al. (2011)), the resulting distribution of wealth is not always ergodic. In the main text, we describe a procedure to verify the size of the approximation error generated by this upper bound on wealth and conclude that it is small.

We discretize the process h with 9 grid points spaced according to the following quantiles of the persistent income component distribution: $\mathbf{q}_h = [0.01 \;\; 0.15 \;\; 0.30 \;\; 0.45 \;\; 0.60 \;\; 0.75 \;\; 0.9 \;\; 0.95 \;\; 0.99]$. Following Hubmer et al. (2020), the values h corresponding to these quantiles follow the log-normal/Pareto mixture described in the core paper. Given the calibration, h values attached to households within the top 10% highest earners are drawn from a Pareto distribution, and from a log-normal distribution otherwise.

Algorithm

We organize the algorithm in steps.

1. Initialize a full dimension grid space over asset values (a), productivity level (h), age bracket (j), and innate investment skill (ϑ).

2. Guess initial tax rates $\boldsymbol{\tau}$ and equilibrium quantities $\{\frac{X}{L}, \underline{r}\}$. Compute p and w.

3. Given prices, solve the consumption-saving-leisure problem. We use a modified version of the EGM algorithm introduced by Carroll (2006).

   Specifically, the budget constraint can be written as $c = \frac{why\zeta_j(1-\tau_w)\ell - a' + \mathcal{F}(a,\vartheta)}{1+\tau_c}$. Given the utility function $u(c,\ell)$, the optimality conditions are given by

$$\ell(c) = c^{-\sigma\lambda}\left[\frac{why\zeta_j(1-\tau_w)}{\chi(1+\tau_c)}\right]^{\lambda}, \qquad c^{-\sigma} = \beta(1-d_j)(1+\tau_c)\mathbb{E}[v_a(a',\vartheta',h)].$$

   We compute v_a numerically. To use the endogenous grid method, we invert the consumption-leisure intratemporal condition and express ℓ as a function of c. We plug the solution in the budget constraint, such that:

$$(1+\tau_c)c + a' = [why\zeta_j(1-\tau_w)]^{1+\lambda}\left(\frac{1}{(1+\tau_c)\chi}\right)c^{-\sigma\lambda} + \mathcal{F}(a,\vartheta).$$

   To gain in speed and avoid any root-finding within the policy function iteration, we pre-compute all possible realizations of (c,ℓ) on an exogenous grid of cash on hand $\mathcal{F}(a,\vartheta)$. The

EGM is performed on an endogenous grid defined as:

$$\widetilde{\mathcal{F}}(a,\vartheta) = (1+\tau_c)c(a',\vartheta,h) + a' - [why\zeta_j(1-\tau_w)]^{1+\lambda}\left(\frac{1}{(1+\tau_c)\chi}\right)c(a',\vartheta,h)^{-\sigma\lambda}.$$

4. Construct the transition matrix $\mathbf{M}$ generated by Π_h, Π_ϑ, Π_κ and Π_y, $a'(\mathbf{s})$ and $\ell(\mathbf{s},y)$. Compute the associated stationary measure of individuals $\mathcal{G}(\mathbf{s})$; by first guessing an initial distribution, and then by iterating on $\mathcal{G}'(\mathbf{s}) = \mathbf{M}\mathcal{G}(\mathbf{s})$ until convergence.

5. Compute the resulting total efficiency units of capital X, total labor supplied L, the return component $\underline{r}$ and government expenditures and revenues.

6. With a relaxation, update the vector of prices; $\{p,w\}$ are obtained using the first order conditions of the representative final good producer, $\underline{r}$ is adjusted to ensure that total returns to capital distributed in the economy is equal to total product of capital in the economy, and the tax rate τ_w is adjusted to balance the government budget if necessary.

   Back to step 2 and iterate until convergence on the equilibrium prices is reached.

1.C.3 Wealth Mobility

In this section, we investigate how the intra-generational wealth mobility matrix in our model alternatives compares with its empirical counterpart. We take as a reference the estimates by Klevmarken et al. (2003), who compute a five-state (quintiles) five-year transition matrix from the 1994–1999 PSID waves. Table 1.C.3 reports the results for the benchmark economy (M1), the type-model (M2) and the scale-model (M3). We find that the three models overstate the persistence of wealth-rank in the top quintiles while being broadly consistent with the U-shaped diagonal transition rate. Interestingly, adding portfolio heterogeneity helps in generating an empirically consistent wealth mobility. Overall, we find that it is hard to distinguish models M1, M2 and M3 based on the resulting wealth mobility matrix.

Table 1.C.3. Wealth mobility: data and model

5-years transition	Diagonal element (quintile – quintile)				
	Q1 – Q1	Q2 – Q2	Q3 – Q3	Q4 – Q4	Q5 – Q5
PSID (1994-1999), Klevmarken et al. (2003)	0.58	0.44	0.42	0.48	0.71
M1 – *Benchmark* model	0.57	0.46	0.34	0.42	0.79
M2 – *Scale*-model	0.57	0.37	0.34	0.47	0.81
M3 – *Type*-model	0.57	0.45	0.41	0.42	0.78
M5 – No portfolio heterogeneity	0.64	0.40	0.52	0.60	0.84

Chapter 2

Gross Labor Market Flows, Entrepreneurship, and the Role of Unemployment Insurance

Alexandre Gaillard[1,2] Sumudu Kankanamge[3]

Abstract

This paper investigates how the selection into entrepreneurship and its sensitivity to unemployment insurance (UI) generosity shape gross labor market flows. We empirically establish a negative relationship between higher UI provision and the selection into entrepreneurship out of unemployment. We introduce a parsimonious model of the US labor market that accounts for micro and macro-level patterns of gross flows across occupations and their responsiveness to UI generosity. We show how asymmetric occupational insurance coverage drive this responsiveness, determining gross labor market flows. Hence, beyond direct effects on unemployment, large reallocations between entrepreneurship and employment appear, shaping aggregate occupational shares.

Keywords: Entrepreneurship, Occupational Choice, Labor Market Mobility, Unemployment Insurance.

[1]We are especially grateful to Christian Hellwig for his many helpful comments. We thank in no specific order Anmol Bhandari, Mark Bils, Fabrice Collard, Patrick Fève, Eugenia Gonzalez-Aguado, Nir Jaimovich, Tim Lee, Benjamin Moll, Franck Portier, and José Víctor Ríos Rull for their useful comments and discussions. The authors are solely responsible for any remaining errors or omissions. The authors acknowledge funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

[2]Toulouse School of Economics, University of Toulouse Capitole, 31080 Toulouse, France. Email: alexandre.gaillard@tse-fr.eu.

[3]Toulouse School of Economics, University of Toulouse Capitole, 31080 Toulouse, France. Email: sumudu.kankanamge@tse-fr.eu.

2.1 Introduction

Each quarter, 3.6% of the US unemployed population become entrepreneurs, accounting for 20% of all transitions into entrepreneurship, while 0.7% of the employed population choose this occupation.[4] Conversely, 6.3% of entrepreneurs turn to employment while 1.45% become unemployed. This significant contribution of the entrepreneurial occupation to gross labor market flows has received some attention in the literature (see, among many others, Vereshchagina and Hopenhayn (2009); Dillon and Stanton (2017); Humphries (2017); Poschke (2019); Hamilton et al. (2019); Rubinstein and Levine (2020)). A number of papers point out the importance of individual labor income prospects in guiding the selection into entrepreneurship, arguing, for instance, that low-wage earners or unemployed individuals are more likely to choose this occupation (Evans and Leighton, 1989; Poschke, 2013; da Fonseca, 2021). In this regard, an important margin could influence the selection into entrepreneurship by significantly altering alternative occupations and, in turn, gross labor market flows: the design of the unemployment insurance (UI) system and, notably, its generosity.[5] While the relation between UI generosity and the incentives to search for a job has received a lot of attention, much less is known of the impact of UI generosity on the entrepreneurial occupational choice. In this paper, we show how entrepreneurship and its sensitivity to UI generosity shape gross labor market flows and masses in three steps: (1) we empirically measure the contribution of entrepreneurship to aggregate flows and its responsiveness to UI generosity; (2) we build a parsimonious occupational choice model of the US labor market and assess its ability to account for the micro and macro-level patterns of gross labor market flows across employment, unemployment, and entrepreneurship; (3) we show that our model accounts for the responsiveness of flows to UI generosity and demonstrate how it shapes occupational shares.

We start by empirically establishing a negative relationship between higher UI provision and the probability that unemployed individuals select into entrepreneurship. In terms of magnitude, a standard deviation increase in total regular UI generosity corresponds to a 7.4% decline in the likelihood of a transition from unemployment to entrepreneurship. The estimate is robust, significant, and economically large, especially when compared to that of the flow from unemployment to employment.[6] Our empirical study uses the 1994–2015 CPS micro-data and the variation in regular and extended UI benefits – from the Extended Benefits (EB) program and the successive

[4]To the extent that we consider the pool of all self-employed individuals, entrepreneurs represent 10% to 12.5% of the US labor market. Cagetti and De Nardi (2006b) provide a meticulous overview of the different definitions of entrepreneurship used in the literature. Given our focus on labor market flows and on the occupational decisions to start a business out of unemployment and employment, it is empirically relevant to consider all self-employed individuals contributing to the dynamics of the labor market. In this paper we interchangeably use the terms entrepreneurship and self-employment. The reported numbers are from the CPS (1995:2015).

[5]Generosity is defined as the product between the level of benefits and coverage duration.

[6]This result is robust to alternative periods and the type of variations used. The magnitude would correspond to an increase by about 8,000$ in total generosity. It could broadly be illustrated as the difference in total UI generosity between the states of Pennsylvania and Michigan.

Emergency Unemployment Compensation (EUC) programs – across US states and over time. We estimate the effect of a change in UI generosity on the resulting occupational choices by comparing two groups: the insured unemployed group, eligible following a layoff and the uninsured group, non-eligible for UI.[7]

We then develop a quantitative model to explain the incentives behind occupational choices and the large responsiveness of the selection into entrepreneurship to changes in the design of UI. Our framework combines two families of models widely used in the literature: search models in the spirit of Mortensen and Pissarides (1994) and occupational choice models with entrepreneurs as pioneered by Quadrini (2000a) and Cagetti and De Nardi (2006b). In the former class of models, labor market frictions are the main determinants of gross flows in and out of unemployment while self-employment is rarely distinguished from employment. Moreover, these models ignore in general the importance of wealth and associated occupational risks in guiding microeconomic occupational decisions.[8] In a recent contribution, Krusell et al. (2017) demonstrate the important effect of wealth in replicating gross flows in and out of the labor force while the liquidity effect studied in Browning and Crossley (2001) and Chetty (2008) illustrate how individual wealth determines the reactivity of flows to the UI policy. In contrast, in the occupational choice family of models with incomplete markets, wealth is a key component that conditions the selection into entrepreneurship given collateral constraints on business capital, but those models often abstract from unemployment and labor market frictions.

We build on the above two classes of models and introduce key mechanisms to better account for gross labor market flows. We use an incomplete markets setting with labor market frictions and idiosyncratic shocks where households can endogenously choose between employment, unemployment, and entrepreneurship. Employed individuals are subject to an exit risk, either due to a voluntary action or a layoff, while adverse shocks may compel entrepreneurs to cease their businesses. Unemployed agents face a standard job search friction and, similarly, entrepreneurs have to search *on-the-business* for a job. An additional search friction rationalizes the time and effort necessary to set up a new business: both unemployed individuals and workers (*on-the-job*) have to search for a business idea. Importantly, we introduce a flexible entrepreneurial production technology which let entrepreneurs produce without capital by using their own labor supply. It is motivated by the fact that in standard entrepreneurial models, only sufficiently wealthy individuals set up a firm under the typical use of a Cobb-Douglas production technology or when business capital is used as the sole production factor (Quadrini, 2000a; Cagetti and De Nardi, 2006b; Buera

[7]Our approach shares some similarities with the recent literature exploring the effects of UI generosity beyond direct effects on the decision to exit unemployment (Rothstein, 2011; Farber et al., 2015). In a different context, Hsu et al. (2018) show that UI affects the proportion of foreclosures in housing markets and Agrawal and Matsa (2013) show that it affects corporate financing decisions.

[8]See Wasmer and Weil (2004) for an extension of the Diamond-Mortensen-Pissarides framework with a role for financial funding on the entrepreneurial side and Krusell et al. (2010) for the consideration of precautionary savings in this setting.

et al., 2011). In such a world, and in line with Evans and Jovanovic (1989), capital and liquidity constraints are decisive factors for the selection into entrepreneurship. Consequently, in these models, almost no individual selects into entrepreneurship below the median wealth level, which is at odds with the empirical evidence.[9] By introducing the possibility that entrepreneurs produce with their own labor *and/or* business capital, this technology considerably improves the fitting of the data by letting wealth-poor individuals select into self-employment. By taking into account both smaller activities, using mostly self-employed labor, and larger businesses, with significant levels of entrepreneurial capital, this technology is also consistent with our broad definition of entrepreneurship. Taken together, we show that our model featuring the above frictions and this entrepreneurial technology produces an accurate characterization of the US gross labor market flows across employment, unemployment, and entrepreneurship along three dimensions: at the aggregate level with respect to the Current Population Survey (CPS) counterparts, at the micro-level by wealth quantiles with respect to the Survey of Income and Program Participation (SIPP) data, and by ability as proxied by education levels in the CPS. This is especially important given that wealth and ability are two important dimensions for the unemployment pool that may interact with the design of the UI.

We then assess the responsiveness of gross flows to UI variations within our model as well as its determinants. To this end, our setting features a rich characterization of the US UI system with a detailed accounting of benefit caps, replacement rates, and UI durations. Importantly and consistent with the UI system in place in the US and many other countries, self-employed individuals are not covered by UI.[10] We use a counterfactual experiments framework to measure the responsiveness of gross flows: in a nutshell, we run the model to generate an extensive variety of UI generosity situations that matches the empirical variety across US states and over time. This framework then lets us capture key insights about the effects of varying UI generosity and the sizable repercussions on gross occupational flows, occupational masses, and aggregate outcomes. We establish that the model produces gross flow elasticities comparable to the data. We find that the elasticity to UI generosity of the flow from the insured unemployed pool to entrepreneurship is negative and about five to six times higher than the corresponding elasticity of the flow to the employment pool. Intuitively, an increase in UI generosity improves the insurance levels of both the unemployment occupation, due to direct effects on UI benefits and duration, and the employment occupation, due to indirect effects of UI coverage in case of a layoff. In contrast, the entrepreneurial

[9] We use the panel data in the Survey of Income and Program Participation (SIPP) to document this fact. Consistent with our setup, Hurst and Lusardi (2004) estimate the role of wealth in the selection into entrepreneurship (defined as business ownership) and find a mild effect except at the top of the wealth distribution.

[10] Among advanced economies, only Finland and the Republic of Korea provide full or partial UI coverage to self-employed individuals while Spain, Germany, Denmark, and Austria provide coverage on a voluntary basis. Most of the remaining advanced economies do not consider self-employment in the scope of UI (Asenjo et al., 2019). In the US, self-employed individuals can obtain UI under the Self-Employment Assistance program. However, this policy is provided in a limited number of US states and is subject to quotas.

occupation is mostly unaffected because it is largely uncovered by the UI system. Thus, following an increase in UI generosity, the incentive to exit unemployment for entrepreneurship reduces relative to exiting for employment. The incentive to search for a self-employed activity *on-the-job* is also lowered while the incentive to search for a job *on-the business* is increased. The responsiveness also displays a large heterogeneity among households. Focusing on gross flows from insured unemployment to self-employment, wealth-poor and low-ability individuals react the most to a change in UI generosity. In the model, the profitability of many businesses is still scaled to personal wealth making the occupational decision of wealth-poor individuals more sensitive. We also show that adding even a simple form of monitoring of the job search effort of unemployed agents improves the above responsiveness with respect to the data.

Furthermore, the model establishes that flows into and out of entrepreneurship play a significant role in shaping the aggregate occupational masses. In substance, when UI generosity increases, the aggregate occupational flow from entrepreneurship to employment increases while the opposite aggregate flow decreases because of the change in the relative risk (or insurance value) between those two occupations. Because the masses of employed agents and entrepreneurs are larger than the mass of insured unemployed agents, the reallocation from entrepreneurship to employment is a key factor in counterbalancing the flow out of insured unemployment. As a result, our model generates an empirically consistent stable to slightly increasing mass of employed individuals, an increasing mass of unemployed individuals, and a decreasing mass of entrepreneurs when UI generosity increases. Our approach, thus, gives an additional perspective to the UI literature. On the one hand, UI generosity has a significant impact on optimal individual decisions and labor market flows, especially the depressing effect on the incentive to exit unemployment (Moffitt, 1985; Meyer, 1990; Chetty, 2008). On the other hand, as empirically established by Chodorow-Reich et al. (2019) and Boone et al. (Forthcoming), the effect of UI generosity on the aggregate level of employment is small or non-significant. Our framework produces a consistent gross flow mechanism to reconcile those views by accounting for the self-employment margin. Overall, an important conclusion of our responsiveness exercise is that asymmetric labor market policies, such as the coverage of UI, significantly change the trade-off between occupations and, in turn, the resulting gross flows.

Finally, we use our framework to evaluate two key predictions of our model. First, our model predicts that an asymmetric insurance coverage between occupations has large effects on selection and occupational masses. To further validate this point, we show that the implementation of an extended insurance scheme that let new entrepreneurs selected out of the insured unemployment pool keep their UI benefits *on-the-business* completely eliminates the distortive effect of the asymmetric coverage of the UI system. Under this alternative policy, a higher UI generosity lowers the employment rate at the equilibrium, while the self-employment rate is roughly constant. Second, we study changes in occupational masses during the Great Recession, taking into account the various extensions of UI duration that were implemented. Our main point is that, when separating the effects of UI generosity into duration and benefits, changes in duration have a lower impact on gross flows relative to changes in benefits, both in the data and in the model. Through the lens of the model, it is due to the discounting of future benefits associated with extended UI duration. During the Great Recession, a decomposition of the effects reveals that the impact of UI extensions on the propensity to select into entrepreneurship generated a persistent 0.4-0.45 percentage point drop in the share of entrepreneurs in the economy. Those numbers might constitute a lower bound of the effects of a temporary change in UI generosity on gross flows and masses as they are mainly the result of an adjustment in UI duration.

The remaining of the paper is organized as follows. Section 2.2 reports our main empirical findings and section 2.3 introduces our model and its parameterization. Section 2.4 evaluates the performance of our model in producing consistent gross labor market flows and masses and section 2.5 discusses the responsiveness of the model to changes in UI generosity. In section 2.6, we study model implications and robustness. Section 3.7 concludes.

2.2 Labor Market Flows and Entrepreneurship: the Data

Definition of occupations We separate individuals into three groups: employed (W), unemployed (U), and entrepreneurs (E). We use a broad definition for entrepreneurs and see them as the pool of all self-employed individuals, able to operate firms with or without capital. Some papers use a narrower definition of entrepreneurship based on business ownership with an active management role (see Cagetti and De Nardi (2006b) or Brüggemann (2020)). However, this approach ignores an important fraction of self-employment and its contribution to the gross labor market flows. Moreover, this alternative definition is relevant in the specific framework where entrepreneurs can not produce without physical capital. In contrast, our definition explicitly accounts for all entrepreneurs, comprising those who run activities with very little physical capital and use mostly their own labor instead. We will use the notions of entrepreneurship and self-employment interchangeably.

CPS Sample Construction Most of our occupational transitions are derived using the CPS data for the 1994:IV-2015-IV period. We consider all respondents between the ages of 20 and 65 and do not restrict to household heads. Importantly, our procedure to establish the flows corrects for high-frequency reversals of transitions between entrepreneurship and unemployment.[11] For instance, based on this procedure, U – – E – U transitions (from unemployment to entrepreneurship and back over the quarter) are recoded as U – – – U, with "–" representing a month. We perform similar adjustments for U – – U – E cases. As such, only U – – E – E transitions are coded as U – – – E.

[11] To follow individuals over time, we build a specific individual identifier based on the individual CPS identifier, age, sex, ethnicity, and US state. Using only the CPS identifier results in a non-negligible amount of false matches.

This restriction helps in reducing mismeasurements due to possible misreporting (see Farber et al. (2015) and Krusell et al. (2017)). All our results are qualitatively robust without this restriction. We classify as a worker an individual currently working in a paid job or declaring being temporarily absent from a paid job. We classify unemployed individuals as those who did not have a job while being in the labor force. Unemployed agents eligible for UI are job losers/on layoff and other job losers. We further condition the *layoff* category to unemployed individuals with a number of weeks in unemployment that does not exceed the maximum UI duration.[12] Entrepreneurs are all individuals that declare being currently self-employed as their main occupation.[13] We use the longitudinal weights provided by the CPS. Further details regarding data construction are available in the Online Appendix.

2.2.1 Labor Market Mobility and Entrepreneurship

In Table 2.2.1, we characterize the aggregate gross flows between occupations and associated occupational masses from our CPS sample.

Table 2.2.1. Aggregate quarterly occupational gross flows rates.

	Gross flow (%) to			**Masses** (%)
From	*Employment*	*Entrepreneurship*	*Unemployment*	
Employment	97.32 (0.45)	0.70 (0.11)	1.97 (0.43)	84.3
Entrepreneurship	6.30 (1.28)	92.26 (1.49)	1.45 (0.64)	10.3
Unemployment	44.38 (10.24)	3.56 (1.19)	52.06 (10.47)	5.4

Source: authors' computations using CPS data from 1994:IV-2015-IV. We restrict our sample to individuals between the ages 20 and 65. Gross flows are corrected for misreporting. Standard deviations between brackets.

While the existing literature has documented the flows in and out of employment and unemployment, we focus on the decomposition of the transitions in and out of entrepreneurship. We discuss two key facts that a model of the three labor market statuses should account for. First, employment is a significantly more persistent occupation than entrepreneurship. Entrepreneurs have an average quarterly exit rate of 7.75% (with 1.45% toward unemployment) compared to 2.67% for employed individuals. A possible explanation is related to the risk entrepreneurs face (Herranz et al., 2015b). Additionally, the flows out of the above two activities have differing characteristics: most of the flows out of employment are toward unemployment whereas most of those out of entrepreneurship are toward employment. Therefore, many entrepreneurs voluntarily cease their businesses for a job while not experiencing unemployment. Second, unemployed individuals are

[12] For regular UI benefits, we restrict eligibility to laid off individuals with less than 30 weeks of unemployment duration, which is the maximum regular US state UI duration. For extensions, we restrict to less than 99 weeks.

[13] In a robustness exercise, we impose that self-employed agents also own their business by using the variable HHBUS in the CPS available after 1994. Our findings are quite similar to those presented in this section.

5 times more likely to enter entrepreneurship than workers: 3.6% of the unemployed individuals and 0.7% of the workers start a business each quarter. We stress at least two explanations: (i) workers spend less time searching for business ideas and learning about potential business markets, and (ii) unemployed individuals may choose to enter entrepreneurship as a better outside opportunity or out-of-necessity. Our model will account for both margins. As a result, while representing only 5-6% of the workforce, unemployed individuals account for 20% of the individuals transiting into entrepreneurship.

2.2.2 Responsiveness to UI Generosity: Empirical Analysis

We exploit the heterogeneity in the UI system across US states and over time to study the interaction between UI generosity and occupational choice out-of-unemployment.[14] We use the CPS panel over the 1994:IV-2014:IV period. Using the procedure explained above, we distinguish recently laid-off unemployed individuals who are eligible for UI from those who either voluntarily quit their job or are not eligible for UI. Our underlying assumption is that variations in UI generosity should mostly affect the eligible group. We use quarterly frequency to obtain sufficient flows toward entrepreneurship and use the correction detailed in Section 2.2.1 to account for misreporting.[15]

We obtain data for regular UI duration and the maximum weekly benefit amount at the state level from the US Department of Labor's "significant provisions of state unemployment insurance laws". Data for UI extensions (EB and EUC) comes from Farber et al. (2015). From 1994 to 2015, variations in the generosity of regular benefits are quite large, not only in the cross-section but also over time within states. Each state applies its own benefit schedule with a typical replacement rate of 35-50% of the previous wage of an individual with the level of benefits capped at each state's inflation-adjusted maximum weekly benefit level. Moreover, each state applies a limit on the number of weeks UI benefits can be claimed. A maximum of 26 weeks has been the typical UI duration, and variations in regular UI benefits are mostly driven by changes in the maximum weekly benefit amount (WBA). Following Agrawal and Matsa (2013) and Hsu et al. (2018), we define the generosity of regular UI benefits in state s as:[16] $Max\,Regular\,UI_{st} = Max\,WBA_{st} \times$

[14] Rothstein (2011) and Farber et al. (2015) study the effect of UI extensions on unemployment exit options. They focus on the timing of the switch toward employment and the potential disincentive effect. We focus on the effect of UI on the resulting occupational choices. Xu (2019) studies the effects of regular UI benefits on the decision to exit unemployment toward employment and self-employment. Our results, however, exploit different sources of UI variations including UI extensions.

[15] We use benefit *eligibility* rather than benefit *receipt* as the latter is not available in the CPS. This is however not specific to this paper. Even in data with information on benefit receipts, Hsu et al. (2018) argue that self-reported information on UI payments is 30 to 40 percent lower than what administrative records show, and suggest *eligibility* as a proxy for actual UI receipts. Regarding our estimates, we think that they may be biased downward, as we suspect only UI beneficiaries (a subset of all eligible unemployed individuals) to respond to changes in UI generosity.

[16] Regarding the relation to actual UI benefits, Hsu et al. (2018) show that the elasticity of $Max\,Regular\,UI$ to total actual compensation payments at the state level is 1.0. Furthermore, they show that for 60% of the population, benefits are capped, and that $Max\,WBA$ captures changes in UI benefits well. To complement the analysis, we check the

$Max\ Regular\ Weeks_{st}$.

In recession periods, each state also provides extended benefits to individuals exhausting their regular benefits in the form of additional weeks. The Extended Benefits (EB) and the Emergency Unemployment Compensation (the successive names of this program having been EUC91, TEUC, and EUC08, we hereafter refer to it as simply EUC) are such regulations. During the Great Recession, heterogeneous emergency extensions (EB and EUC) of UI generosity were activated and in some states, the duration of UI benefits was extended up to 99 weeks. To represent the total UI generosity including the extensions, we define: $Max\ Extended\ UI_{st} = Max\ Regular\ UI_{st} + Max\ WBA_{st} \times Max\ EB\ EUC\ Weeks_{st}$. It is the total amount of benefits that an individual falling into unemployment in a given month could claim over the maximum number of weeks benefits could be claimed at that time, including additional variations from the activation of UI extensions.

We distinguish three data panels in which the source of variations in UI generosity differs. Panel A covers the whole sample period from 1994 to 2015. Panel B covers the 1994-2007 period and excludes the Great Recession and the significant UI benefits extensions that were then implemented. It will let us study mostly the effects of a change in regular benefits. Panel C, covering the 2008-2015 period, encompasses the impact of all benefits changes including UI extensions. It will let us verify whether UI duration adjustments observed during the Great Recession result in similar findings with respect to regular benefits. Accordingly, for a laid-off unemployed individual eligible for UI benefits, we impose that the unemployment duration does not exceed 30 weeks in Panel B and 99 weeks in Panel A/C. This corresponds to the maximum UI duration, including extensions, in the respective Panels.

UI generosity and aggregate propensity to start a business

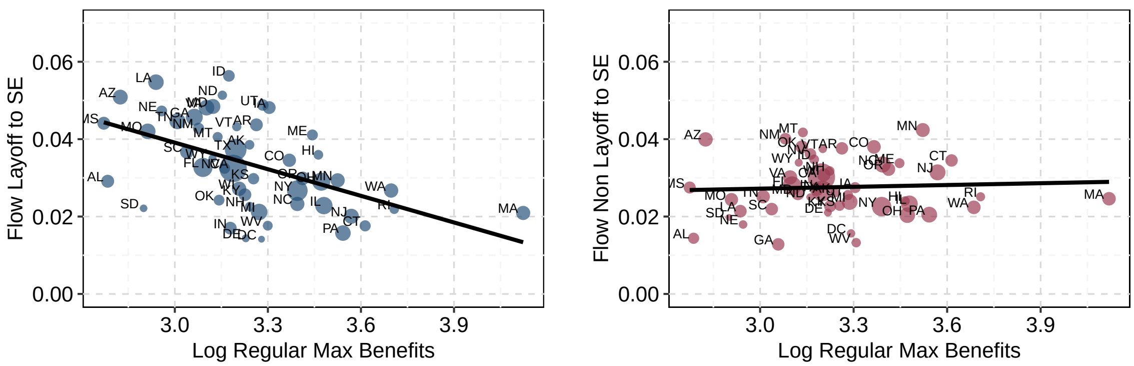
Recent papers, such as Røed and Skogstrøm (2014) and Hombert et al. (2020), show an important interaction between the selection into entrepreneurship and UI generosity. To establish whether a relationship exists between UI generosity and flows to entrepreneurship in the US, we first show in Figure 2.2.1a the flows from unemployment to entrepreneurship disaggregated by US states for Panel B. The left plot displays a downward relation between the maximum regular benefits level and flows from the pool of laid-off unemployed individuals to entrepreneurship. Namely, US states that have a higher maximum regular benefits level tend to have a smaller flow from the laid-off unemployment pool to entrepreneurship. The right plot illustrates that such a downward relation does not exist for individuals that are unemployed for reasons other than a layoff and not eligible for UI. The left plot of Figure 2.2.1b establishes the same relation for Panel C, i.e. a timeframe where most of the UI benefits extensions offered after the Great Recession had come into implementation. Accordingly, when taking into account regular benefits as well as EB and

robustness of the results against alternative UI generosity measures in the Online Appendix 1.3.

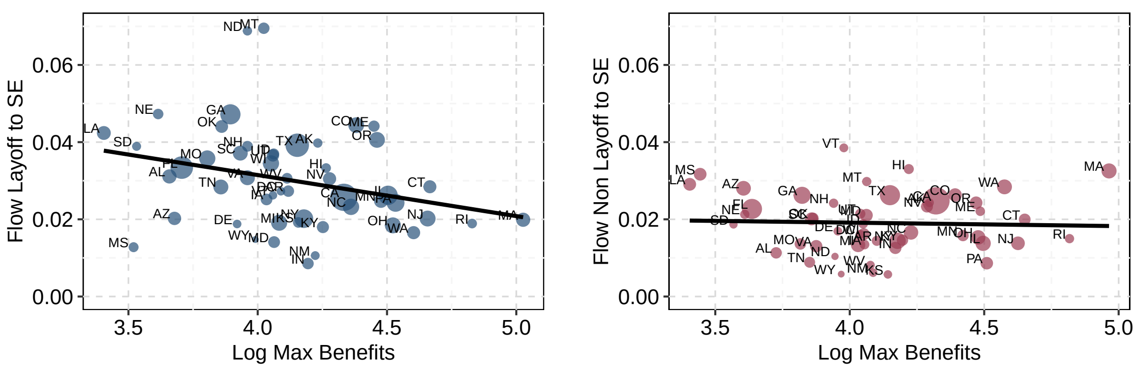
EUC extensions, a higher generosity of UI is related to lower flows from the layoff pool to entrepreneurship. The right plot shows that there is no such relation for non-eligible unemployed individuals.

Figure 2.2.1. Average quarterly gross flow toward entrepreneurship among the unemployed individuals

(a) Panel B: regular UI generosity ($Max\ Regular\ UI_{st}$) for the 1994:2007 period



(b) Panel C: regular and extended UI generosity ($Max\ Extended\ UI_{st}$) for the 2008:2015 period



Notes: aggregate gross flow rates, for individuals experiencing a layoff in the left panels and other unemployed individuals in the right panels, are calculated from the CPS. The size of the dots refers to the number of observations.

Sensitivity of occupational decisions at the micro level

To further establish the relationship between UI generosity and flows to entrepreneurship, we investigate whether it holds at the micro level when we control for characteristics of individuals and US states. Our identification rests on the assumption that state-level changes in regular UI generosity are independent of factors that might otherwise affect the propensity to select into entrepreneurship among the unemployed. Hsu et al. (2018) find that this assumption is supported in the data and we confirm this in the Online Appendix 1.2 for our sample and covariates. The maximum regular UI benefits provided by a given state are not significantly related to the unemployment rate, average wage, log real gross domestic product per capita, home price growth, or other unobservable factors captured by state and year-by-month fixed effects. Concerning the use of federal extensions of UI benefits during the Great Recession, we follow Rothstein (2011) and

Hsu et al. (2018) and control for the endogeneity of the EUC and EB activations by controlling flexibly for smooth cubic polynomial functions of the state's unemployment rate in the initial period of the transition. The aim is to control for the omitted variable bias arising from the specific activation rules of the EUC and the EB. Moreover, by taking the initial state unemployment rate, there is no endogeneity issue between the dependent variable, i.e. the occupational choice a quarter ahead, and the initial state unemployment rate. Finally, we identify the effect of UI generosity on occupational choices by comparing the impact of *within* state changes in UI benefits between the pool of eligible individuals unemployed after a layoff and the rest of the unemployed. We estimate the probability model:

$$\begin{aligned}\text{Unemp. to Occ.}_{ist} = \alpha + \beta\text{UI generosity}_{st} + \gamma\text{Layoff}_{it} + \delta\text{UI generosity}_{st} \times \text{Layoff}_{it} \\ + \boldsymbol{\zeta}\mathbf{X}_{it} + \boldsymbol{\eta}\mathbf{Z}_{st} + \boldsymbol{\lambda}_s + \boldsymbol{\mu}_t + \epsilon_{ist}\end{aligned} \tag{2.1}$$

where $\text{Unemp. to Occ.}_{ist}$ is an indicator of whether individual i in state s and quarter t is switching to the following specific occupation: Entrepreneur (E) or Worker (W). The variable $\text{UI generosity}_{st}$ depends on the specification: $\log(Max\ Regular\ UI_{st})$ for regular benefits or $\log(Max\ Extended\ UI_{st}$ for regular and extended benefits. $\mathbf{X}_{it}$ is a vector of individual characteristics that includes household income brackets, educational attainment, ethnicity, sex, age, age squared, marital status, cubic polynomial in unemployment duration, and an indicator of whether the spouse is currently employed. $\mathbf{Z}_{st}$ is a vector of time-varying US states characteristics that includes a cubic in the monthly seasonally adjusted state unemployment rate, annual state log real GDP per capita, log income per capita and a housing price index. Again, those elements aim to capture the activation threshold of the UI extensions and serve as controls as the incentive to start a business might be correlated with the economic environment. Finally, $\boldsymbol{\lambda}_s$ and $\boldsymbol{\mu}_t$ are states and year-month fixed effects and ϵ_{ist} is an error term.

The results for Panel A (1994:2015) are reported in Table 2.2.2 (top) both for OLS and multinomial Logit specifications.[17] The latter takes the remaining unemployed as a reference and also consider other flows out of the labor force. The effect of the generosity of regular UI benefits (columns (1)-(4)) and extended UI benefits (columns (5)-(8)) on the propensity to switch to entrepreneurship among the laid-off workers relative to other unemployed individuals is significant and negative. To get a sense of the magnitude of the effect, given the estimates in column (1) (resp. column (5)), a 1000$ increase of *Max Regular UI* is associated with a significant 0.021 (resp. 0.006) percentage points decline in the average probability of switching to entrepreneurship for eligible unemployed individuals.[18] Put differently, a standard deviation increase in UI benefits, i.e. a 30% increase in

[17]In the OLS specification, the $\text{Unemp. to Occ.}_{ist}$ variable is divided by the average transition rate from unemployment to the specific occupation over the sample, such that our estimates will reflect the percentage point change relative to the average probability of switching.

[18]Note that a 1000$ increase of UI corresponds to an increase of 3.2% of the average *Max Regular UI*. Therefore, in

Table 2.2.2. UI generosity and exit probability toward an occupation.

	Panel A (1994:IV-2015:IV)							
	OLS UtoE (1)	*OLS* UtoW (2)	*mLogit* UtoE (3)	*mLogit* UtoW (4)	*OLS* UtoE (5)	*OLS* UtoW (6)	*mLogit* UtoE (7)	*mLogit* UtoW (8)
layoff	5.919***	0.356	3.272**	0.846**	−1.742***	1.052***	1.824**	0.660***
	(1.140)	(0.540)	(1.499)	(0.418)	(0.579)	(0.156)	(0.770)	(0.211)
log(*Max Reg. UI*)	−0.067	0.003	−0.110	0.048				
	(0.150)	(0.060)	(0.240)	(0.066)				
layoff x log(*Max Reg. UI*)	−0.647***	−0.010	−0.363**	−0.077*				
	(0.120)	(0.058)	(0.162)	(0.045)				
log(*Max Ext. UI*)					−0.115	0.057	−0.035	0.082
					(0.155)	(0.048)	(0.193)	(0.053)
layoff x log(*Max Ext. UI*)					−0.189***	−0.085***	−0.200**	−0.056***
					(0.056)	(0.016)	(0.080)	(0.022)
Observations	140,952							
	Panel B (1994:IV-2007:IV)				Panel C (2008:I-2015:IV)			
	OLS UtoE (9)	*OLS* UtoW (10)	*mLogit* UtoE (11)	*mLogit* UtoW (12)	*OLS* UtoE (13)	*OLS* UtoW (14)	*mLogit* UtoE (15)	*mLogit* UtoW (16)
layoff	5.080**	0.460	5.280**	1.158*	2.575***	1.100***	3.122***	0.996***
	(2.347)	(0.429)	(2.214)	(0.644)	(0.738)	(0.249)	(1.192)	(0.305)
log(*Max Reg. UI*)	0.079	−0.107	0.003	−0.237*				
	(0.370)	(0.073)	(0.432)	(0.129)				
layoff x log(*Max Reg. UI*)	−0.552**	−0.024	−0.586**	−0.114				
	(0.250)	(0.046)	(0.239)	(0.070)				
log(*Max Ext. UI*)					0.018	0.034	0.096	0.033
					(0.268)	(0.040)	(0.314)	(0.079)
layoff x log(*Max Ext. UI*)					−0.256***	−0.087***	−0.319***	−0.088***
					(0.071)	(0.024)	(0.118)	(0.030)
Observations	72,112	72,112	72,112	72,112	68,840	68,840	68,840	68,840
State and year-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind. & state controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors, adjusted for clustering at the state level, are reported in parentheses. See text for the list of controls. OLS estimates are normalized by the mean transition rate of the flow, and can be interpreted as an elasticity relative to that mean transition rate.

average generosity, would lead to a 7.4% (resp. 2.2%) decline in the fraction of insured unemployed individuals moving to entrepreneurship.[19] The magnitude of the corresponding number for a flow toward employment is 3 to 6 times lower. Finally, the Logit specification corroborates the previous findings and magnitude. Table 2.2.2 (bottom) distinguishes the effects of UI generosity for Panels B (pre-recession period) and C (Great Recession period). Columns (9)-(16) show that for both periods and sources of variations, extensions, or regular benefits adjustments, UI generosity has a negative and significant effect on the propensity to select into entrepreneurship. The effect is larger during the pre-recession period, in which the source of variations is mainly due to a change in the weekly benefit amount (WBA). These results are robust to a variety of alternative specifications that we report in the Online Appendix 1.3. Notably, the results hold for an alter-

Panel A, the percentage point number corresponds to -0.65*3.2/100 (resp. -0.19*3.2/100).

[19]This magnitude would correspond to an increase by about 8,000$ in total generosity. It could broadly be illustrated as the difference in total UI generosity between the states of Pennsylvania and Michigan.

native definition of entrepreneurship based on self-employment *and* business ownership and for alternative UI generosity measures, controls, and sample periods.

The results above show that there is an economically significant and large empirical relation between UI provision and the probability that unemployed individuals select into entrepreneurship, with a significant impact on gross labor market flows. Our general equilibrium quantitative model, introduced in the next section, will provide an accurate representation of flows across all occupations and let us examine how incentives change with more generous UI provision and evaluate the macroeconomic outcomes.

2.3 A Model of Gross Labor Market Flows with Entrepreneurship

In this section, we develop an incomplete markets dynamic general equilibrium model of entrepreneurship with occupational choices and search frictions. A unit measure of *ex-post* heterogeneous agents can be either employed, entrepreneurs, or unemployed. Entrepreneurs hold small businesses and together with a representative corporate firms sector provide the production of the economy. The model parsimoniously characterizes average aggregate gross labor market flows across the above three occupations while still generating empirically consistent micro-level behaviors. We also incorporate an extensive depiction of the UI system in the US to let us measure the responsiveness of flows to UI changes.

2.3.1 Households

The economy is populated by a continuum of infinitely-lived households of measure one. Every period, each agent falls in one of three occupations $o_t \in \mathcal{O} \equiv \{e, w, u\}$: entrepreneurship ($e$), employment ($w$), or unemployment ($u$). We keep using the $\{E,\ W,\ U\}$ notations to designate respectively entrepreneurs, workers and unemployed individuals. All individuals are described, among other things, with an ability component $\vartheta \in \Theta$ and savings $a_t \in \mathcal{A}$. r_t and w_t are respectively the interest rate on savings and the wage rate in the economy.

Life-time utility is derived from consumption c_t and disutility from search:

$$\mathcal{U}_t = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t u(c_t, s_{e,t}, s_{w,t}) \right], \tag{2.2}$$

where $s_{e,t}$ and $s_{w,t}$ are the search efforts exerted to respectively start a business and find a new job. β is the discount factor. In the following, we drop the time index t unless necessary. Labor is supplied inelastically and the utility function is:

$$u(c, s_e, s_w) = \frac{c^{1-\sigma} - 1}{1 - \sigma} - \left(s_w^{\psi} + s_e^{\psi} \right)^{\phi/\psi}, \tag{2.3}$$

with ψ and ϕ the search elasticities. Note that if $\phi = \psi$, the utility is separable in s_e and s_w.

The labor income of a working household, the replacement income of an unemployed individual, and the business income of an entrepreneurial household all depend on ability ϑ. This component follows the process: $log(\vartheta) = \rho_\vartheta \log(\vartheta_{-1}) + \epsilon_\vartheta$, with $\epsilon_\vartheta \sim \mathcal{N}(0, \sigma_\vartheta)$ and the invariant distribution Π_ϑ.[20]

Workers are subject to an additional persistent idiosyncratic shock y on their labor income that we call *match-quality*.[21] It follows the process: $\log(y) = \rho_y \log(y_{-1}) + \epsilon_y$, with $\epsilon_y \sim \mathcal{N}(0, \sigma_y)$. For a new worker, this shock is drawn from the invariant distribution Π_y. Finally, entrepreneurs face a persistent idiosyncratic business shock z following the process: $log(z) = \rho_z log(z_{-1}) + \epsilon_z$, with $\epsilon_z \sim \mathcal{N}(0, \sigma_z)$. A new entrepreneur draws her initial business shock in the invariant distribution Π_z. All processes are discretized into Markov chains with support $\vartheta \in \Theta \equiv \{\vartheta_1, ..., \vartheta_T\}$, $y \in \mathrm{Y} \equiv \{y_1, ...y_Y\}$, and $z \in \mathcal{Z} \equiv \{z_1, ..., z_Z\}$.

With j the unemployment insurance status, the value function of a worker is $W(\mathbf{x}_w, j)$ with state vector $(\mathbf{x}_w, j) \in X^w \times \mathcal{J}$ and $\mathbf{x}_w \equiv (a, \vartheta, y) \in X^w \equiv \mathcal{A} \times \Theta \times \mathrm{Y}$. An entrepreneur has the value function $E(\mathbf{x}_e, j)$ with state vector $(\mathbf{x}_e, j) \in X^e \times \mathcal{J}$ and $\mathbf{x}_e \equiv (a, \vartheta, z) \in X^e \equiv \mathcal{A} \times \Theta \times \mathcal{Z}$. An unemployed individual has the value function $U(\mathbf{x}_u, j)$ with state vector $(\mathbf{x}_u, j) \in X^u \times \mathcal{J}$ and $\mathbf{x}_u \equiv (a, \vartheta) \in X^u \equiv \mathcal{A} \times \Theta$.

Workers

Workers earn labor income $h(\vartheta)yw$, where the function $h : \vartheta \mapsto R$ maps their individual ability ϑ into a working ability. They have a probability $\eta = \eta(\vartheta)$ of being fired due to no fault of their own and a probability q of voluntarily quitting their job. Only in the former case, do they face insured unemployment and can expect to get continuation value $U(\mathbf{x}'_u, J)$.[22] By providing effort s_e, workers can search for business ideas *on-the-job* and start a business in the next period with probability $\pi_e(s_e)$. Business search effort can describe market research on the feasibility of an idea, competition assessment, business education, agency costs or the time needed to fill administrative forms, validate product norms, etc. They then voluntarily change their occupation, loose their UI

[20] Ability can change over time in order to generate additional saving motives as our model abstracts from life-cycle aspects, human capital or health risks which can explain a large productivity dispersion in the data.

[21] This model does not include an explicit matching process but y can be viewed as a match-quality component because it starts and ends with a specific job while not appearing as a state for the unemployed or the entrepreneur. This process brings our generated distributions and transitional flows closer to the data.

[22] Notice that in the model, $U(\mathbf{x}_u, j) < W(\mathbf{x}_w)$, $\forall(\mathbf{x}_u, \mathbf{x}_w, j)$. Therefore, we rule out any voluntary transition to unemployment. Conversely, unemployed individuals getting a job opportunity always return to employment provided they do not get a better entrepreneurial opportunity.

rights and can expect a continuation value $E(\mathbf{x}_e', 0)$. Their recursive program is:

$$W(\mathbf{x}_w, J) = \max_{c,a',s_e} u(c,0,s_e) + \beta\mathbb{E}\Big\{(1-\eta)\Big[W(\mathbf{x}_w', J) + \pi_e(s_e)\max\left\{E(\mathbf{x}_e',0) - W(\mathbf{x}_w', J), 0\right\}\Big] \tag{2.4}$$

$$+ \eta\Big[(1-q)\left[U(\mathbf{x}_u', J) + \pi_e(s_e)\max\left\{E(\mathbf{x}_e',0) - U(\mathbf{x}_u', J), 0\right\}\right]$$

$$+ q\left[U(\mathbf{x}_u', 0) + \pi_e(s_e)\max\left\{E(\mathbf{x}_e',0) - U(\mathbf{x}_u', 0), 0\right\}\right]\Big] \Big| y, \vartheta\Big\},$$

$$\text{s.t.} \quad c + a' = (1-\tau_w)h(\vartheta)wy + (1+r)a, \tag{2.5}$$

$$c > 0,\ a' \geq 0,\ s_e \geq 0, \tag{2.6}$$

where τ_w is a flat labor income tax and equation (2.5) is the budget constraint of the worker.

Unemployed individual

We specify the UI program to capture key features of the UI system in the US while maintaining the tractability of the model. The exact number of remaining periods with UI benefits is an important component which is tracked by the state variable $j \in \{0, ..., J\} \equiv \mathcal{J}$. We note $\{b(\vartheta, j)\}_{j=\overline{J}}^{j=0}$ the path of UI benefits. Consistent with what is implemented in the US, the amount of benefits of an eligible unemployed individual, $b(\vartheta, j)$, is related to her past earnings through $h(\vartheta)w$, up to a replacement rate μ and subject to a cap defined by the maximum benefit amount $\overline{b}$. $\overline{J}$ is the exogenous regulatory maximum UI duration converted to model periods but, due to discretization, this number can fall between two model periods. To implement the exact number of regulatory UI benefits periods, we set J as the number of model periods immediately above $\overline{J}$ and then apply a linear rule to provide only partial UI benefits in the last model period before losing coverage. UI benefits in the current period are:

$$b(\vartheta, j) = \begin{cases} \widetilde{b}(\vartheta)(1-\tau_w) & \text{if } j \in [2, J] \\ \widetilde{b}(\vartheta)(1-\tau_w)\Big[1-(J-\overline{J})\Big] & \text{if } j = 1 \\ 0 & \text{if } j = 0 \end{cases}, \quad \widetilde{b}(\vartheta) = \begin{cases} w\mu h(\vartheta) & \text{if } w\mu h(\vartheta) \leq \overline{b} \\ \overline{b} & \text{otherwise} \end{cases}, \tag{2.7}$$

With the above rule, agents can be either insured ($j > 0$) or uninsured ($j = 0$). Consistent with the current US unemployment insurance scheme, laid-off workers are eligible for UI and are assumed to have maximum insurance duration (i.e. $j = J$) while non-laid-off workers and entrepreneurs are uninsured (i.e. $j = 0$).

Following the above scheme, insured unemployed individuals ($j > 0$) receive benefits $b(\vartheta, j)$, in proportion to their individual productivity ϑ. By claiming UI in the current period, they shift from j periods of remaining UI rights to $j-1$ at the end of the period. However, this shift is only allowed if individuals actively searched for a job and provided enough effort to meet the applicable UI regulations.[23] $\pi^m(s_w)$ is the probability to meet this requirement. This can be viewed as stylized imperfect monitoring of program applicants due to, for instance, asymmetric information. Non eligible individuals and those who have exhausted their rights ($j = 0$) receive no benefits. Moreover, all unemployed individuals are assumed to receive a fixed amount m from domestic production. Unemployed individuals search for both a business idea and a job opportunity with respective efforts s_e and s_w and corresponding success probabilities $\pi_e(s_e)$ and $\pi_w(s_w)$. Upon finding a job, they become workers with continuation value $W(\mathbf{x}'_w, J)$. Similarly, upon having an idea, a business can be started in the next period with continuation value $E(\mathbf{x}'_e, 0)$. Their recursive program is:

$$
\begin{aligned}
U(\mathbf{x}_u, j) = \max_{c,a',s_e,s_w} \; & u(c, s_w, s_e) + \beta\mathbb{E}\Big\{\pi_w(s_w)\Big[W(\mathbf{x}'_w, J) + \pi_e(s_e)\max\{E(\mathbf{x}'_e, 0) - W(\mathbf{x}'_w, J), 0\}\Big] \\
& + (1 - \pi_w(s_w))\Big[\pi^m(s_w)U(\mathbf{x}'_u, j-1) + (1 - \pi^m(s_w))U(\mathbf{x}'_u, 0) \\
& + \pi^m(s_w)\pi_e(s_e)\max\{E(\mathbf{x}'_e, 0) - U(\mathbf{x}'_u, j-1), 0\} \\
& + (1 - \pi^m(s_w))\pi_e(s_e)\max\{E(\mathbf{x}'_e, 0) - U(\mathbf{x}'_u, 0), 0\}\Big] \Big| \vartheta\Big\}, \\
\text{s.t.} \quad & c + a' = m + b(\vartheta, j) + (1+r)a, \qquad (2.8) \\
& c > 0,\; a' \geq 0,\; s_e \geq 0,\; s_w \geq 0,\; \textit{Equation } (2.7), \qquad (2.9)
\end{aligned}
$$

where equation (2.8) is the corresponding budget constraint.

Entrepreneurs

Entrepreneurs produce using capital k and their own labor $\underline{l}$ in their self-employed business using technology:

$$\mathcal{Y}(k, \vartheta, z) = zg(\vartheta)\left[\varpi k^p + (1-\varpi)\underline{l}^p\right]^{\nu/p} \tag{2.10}$$

with $\nu \in (0,1)$ the degree of homogeneity which characterizes returns to scale, $p > 0$ the degree of substitutability, and $\varpi \in [0,1]$ the share of each production factor. Equation (2.10) departs from the standard specification in models with entrepreneurs found in Quadrini (2000a), Cagetti and De Nardi (2006b), or Kitao (2008b) among others. This specification lets us generate self-employment income even without any business capital k. This assumption is empirically relevant given our broad definition of entrepreneurship: in the Survey of Consumer Finances (SCF), about 35-40% of the self-employed individuals are using less than 1000$ of business capital.[24] In Section 2.4, we show that this specification also lets us better capture the gross flows in and out

[23] Despite specific local UI regulations, in most states, monitoring relies on the weekly obligation to report any job search activities to the U.S. Department of Labor (Asenjo et al., 2019).

[24] In a recent contribution, Bassi et al. (2021) show that access to capital rental markets allow small firms to increase their effective scale and mechanize production, even when each individual firm would be too small to invest in expensive machines. In our context, we believe that rental market interactions would allow the selection of wealth-poor entrepreneurs into the sector, generating results similar to our approach with the flexible production technology.

of entrepreneurship estimated from the SIPP. The function $g : \vartheta \mapsto R$ maps individual ability into entrepreneurial ability. However, due to the presence of the entrepreneurial business shock z and the transitory match-quality shock y, there is an imperfect correlation between labor and entrepreneurial incomes.

To invest k, entrepreneurs can borrow from a financial intermediary funds that can only be invested in the business. Recalling that a is the current wealth of an agent, entrepreneurs choose whether to borrow ($k > a$) or save ($k < a$). If they borrow the amount $(k - a)$, we assume that it is only up to a fixed fraction λ of their total assets. This type of borrowing constraint has been widely used in the context of entrepreneurship (see Kitao (2008b); Brüggemann (2020) among others). Entrepreneurial profit $\mathcal{P}$ is defined as entrepreneurial production net of capital depreciation, any interest repayment, and the fixed cost c_f. The latter accounts for all the additional functioning costs that entrepreneurs face. By providing effort s_w, entrepreneurs can search for a job opportunity *on-the-business* and change occupation in the next period with probability $\pi_w(s_w)$ and value $W(\mathbf{x}'_w, J)$. Otherwise, if they endogenously choose to quit entrepreneurship, they can return to the uninsured unemployment pool with value $U(\mathbf{x}'_u, 0)$. Finally, an entrepreneur faces an exogenous probability to fall in an absorbing state $z_0 = 0$ with probability p_{z_0} which translates the fact that some entrepreneurs might fail independently of their ability to self-insure against bad business shocks.[25] In such a case, she will exit to either employment or unemployment depending on her job search effort. The recursive program of entrepreneurs is:

$$E(\mathbf{x}_e, 0) = \max_{c,a',k,s_w} u(c, s_w, 0) + \beta \mathbb{E}\Big\{ \pi_w(s_w) \max\{W(\mathbf{x}'_w, J), E(\mathbf{x}'_e, 0)\} \tag{2.11}$$

$$+ \big(1 - \pi_w(s_w)\big) \max\{U(\mathbf{x}'_u, 0), E(\mathbf{x}'_e, 0)\} \,\Big|\, z, \vartheta \Big\},$$

$$\text{s.t.} \quad c + a' = (1 - \tau_p)\mathcal{P}(k, \vartheta, z) + a + r(a - k)1_{\{k \le a\}}, \tag{2.12}$$

$$\mathcal{P}(k, \vartheta, z) = \mathcal{Y}(k, \vartheta, z) - \delta k - r(k - a)1_{\{k \ge a\}} - c_f, \tag{2.13}$$

$$k \le \lambda a, \tag{2.14}$$

$$c > 0, a' \ge 0, s_w \ge 0. \tag{2.15}$$

with δ the depreciation rate, equation (2.12) is the budget constraint, and equation (2.14) is the borrowing constraint.[26] τ_p is a payroll tax rate. Finally, notice that although entrepreneurs do not benefit from the UI program, they can always self-insure using their wealth.

[25] For the sake of clarity, we augment the Markov process z with an additional state $z_0 = 0$ and redefine the transition matrix with $\widetilde{\pi}_z(z'|z) = \pi_z(z'|z)(1 - p_{z_0})$ for $z, z' \in \{z_1, ..., z_Z\}$, $\widetilde{\pi}_z(z_0|z \in \{z_1, ..., z_Z\}) = p_{z_0}$ and $\widetilde{\pi}_z(z' \in \{z_1, ..., z_Z\}|z_0) = 0$, $\widetilde{\pi}_z(z_0|z_0) = 1$, with $\pi_z(z'|z)$ the discretized transition matrix of the above AR(1) process.

[26] Recall that the cash on hand of entrepreneurs in the baseline case can be written: $\mathcal{Y}(k, \vartheta, z) + (1 - \delta)k - (1 + r))(k - a)1_{\{k \ge a\}} + (1 + r)(a - k)1_{\{k \le a\}}$. Rearranging terms yield profit and budget constraint equations.

2.3.2 Corporate sector

A representative corporate firm produces Y_t using a Cobb-Douglas technology, with total factor productivity A, capital level K_t and labor L_t, such that: $Y_t = F(K_t, L_t) = AK_t^{\alpha}L_t^{1-\alpha}$, where $\alpha \in (0,1)$ is the capital share.[27] Profit maximization produces the competitive interest rate $r_t = A\alpha\left(\frac{L_t}{K_t}\right)^{1-\alpha} - \delta$ and wage rate $w_t = A(1-\alpha)\left(\frac{K_t}{L_t}\right)^{\alpha}$.

2.3.3 Government

The government runs an UI system that covers the pool of recently laid-off unemployed individuals and finances it using labor income and payroll taxes. In our benchmark economy, in order to not distort occupational choices in counterfactuals experiments, we assume that UI is equally financed by a symmetric tax scheme on entrepreneurs and workers, such that $\tau = \tau_w = \tau_p$.[28] The validity and consequences of this assumption are discussed in section 2.6.3. Notably, we explore as a robustness a case where the total tax burden rests on workers and find no reversal of our main results. Total government revenues (T) are:

$$T = \int_j \int_{\mathbf{x}_w} \tau_w h(\vartheta) wy \, d\Gamma(\mathbf{x}_w, j) + \int_{\mathbf{x}_u} \tau_w b(\vartheta, j) \, d\Gamma(\mathbf{x}_u, j) + \int_{\mathbf{x}_e} \tau_p \mathcal{P}(k, \vartheta, z) \, d\Gamma(\mathbf{x}_e, j), \tag{2.16}$$

with $\Gamma(\mathbf{x}_o, j)$ the measure of individuals in occupation o with remaining UI duration j. Total government expenditures G are equal to the allocated UI benefits: $G = \int_j \int_{\mathbf{x}_u} b(\vartheta, j) \, d\Gamma(\mathbf{x}_u, j)$.

2.3.4 Equilibrium

A stationary recursive equilibrium in this economy consists of a set of value functions $W(\mathbf{x}_w, j)$, $U(\mathbf{x}_u, j)$, $E(\mathbf{x}_e, j)$, policy rules over asset holdings $a'(\mathbf{x}_o, j)$, consumption $c(\mathbf{x}_o, j)$, job search effort $s_w(\mathbf{x}_o, j)$, business search effort $s_e(\mathbf{x}_o, j)$, business investment $k(\mathbf{x}_e, j)$, occupational choices, prices $(r, w) \in \mathbb{R}^+$, tax rate $\tau \in \mathbb{R}^+$ and a stationary measure over individuals $\Gamma(\mathbf{x}_o, j)\ \forall o, j$, such that: (i) Given prices (r, w) and tax rate τ, the policy rules and value functions solve household individual programs; (ii) The wage w and the interest rate r are equal to the marginal products of the respective production factor in the corporate sector; (iii) goods and factor markets clear: (a) capital: $\int a'(\mathbf{x}_o, j) d\Gamma(\mathbf{x}_o, j) = K + K^E$, with aggregate entrepreneurial capital $K^E = \int k(\mathbf{x}_e, j) d\Gamma(\mathbf{x}_e, j)$,

[27] Unlike Cagetti and De Nardi (2009b), we abstract from entrepreneurial labor demand. However, we believe our setup is sufficient to replicate the dynamics of the data. Indeed, with a static entrepreneurial labor demand, workers are hired in proportion to entrepreneurial capital and productivity and the wage rate. Mechanically, a higher number of entrepreneurs lead to a higher labor demand and thus a higher equilibrium wage. In our setting, as the number of entrepreneurs reduces the labor force in the corporate sector, its labor supply is lower, and therefore equilibrium wages increase.

[28] In the US, entrepreneurs pay a self-employment tax and paid-employee a labor income tax. In practice, employers pay both state and federal UI related taxes. However, Anderson and Meyer (2000) argue that average industry tax rates are largely passed on to workers through lower earnings. In the presence of entrepreneurs and given the lack of clear evidence on who eventually bears the tax burden, we choose to equally share the cost of the UI in our benchmark.

(b) the measure of corporate workers $\int d\Gamma(\mathbf{x}_w, j)$ is equal to corporate labor demand; (iv) $\Gamma(\mathbf{x}_o, j)$ is the stationary measure of individuals induced by the decision rules and the exogenous Markov processes; (v) τ balances the government budget ($T = G$). Finally, we define total output $\mathbb{Y}$ as the sum of corporate sector output Y and entrepreneurial sector output Y^E such that $\mathbb{Y} = Y + Y^E = Y + \int_{\mathbf{x}_e} \mathcal{Y}(k(\mathbf{x}_e), \vartheta, z) d\Gamma(\mathbf{x}_e, j)$.

This model has no analytical solution and must be solved numerically. We detail our numerical implementation of this problem in Online Appendix 1.C.2.

2.3.5 Parameterization

We parameterize the model to fit key features of the US gross labor market flows between employment, unemployment, and entrepreneurship as well as key feature related to the entrepreneurial sector. Our data counterparts are taken from the CPS, 1994:IV-2015-IV, the 2004 Survey of Consumer Finances (SCF), and the SIPP, 1996:2013. Some parameters are assigned using standard values or estimates in the literature while others are endogenously chosen to reproduce key moments observed in the data in the next section. A model period is set to two months.[29]

Exogenously calibrated parameters

Parameters related to the production sectors are set as follows. We normalize TFP to unity at yearly frequency, implying that $A = 1/6$ given the periodicity of the model. We set the depreciation rate δ to a standard value of 6.1% annually or 1% every two months. The capital share in the corporate sector α is set to 0.33. As in Kitao (2008b), we set the borrowing parameter λ to 1.5. The entrepreneurial labor supply $\underline{l}$ is normalized to 1. In the absence of an empirical counterpart, we set $p = 1.0$ in the baseline case and also compare to the Cobb-Douglas case ($p \to 0$). A value $p = 1.0$ implies that an entrepreneur with low capital can produce using his own labor supply which brings the model closer to the data. Consistent with estimates in Castro et al. (2015) and values used in Clementi and Palazzo (2016), the idiosyncratic business process z has a standard deviation $\sigma_z = 0.22$ and persistence $\rho_z = 0.91$, which corresponds to an annual persistence of 0.57. Moreover, the exogenous exit probability of entrepreneurs is set to 20% of the total entrepreneurial exit probability (5.8% each period), yielding $p_{z0} = 0.0116$.[30] This is in the range of the 20-30% of long-established business owners declaring ceasing their activity due to reasons not related to economic conditions in the Survey of Business Owners (2007) and the Annual Survey of Entrepreneurs (2014:2016).

We normalize the persistent individual working ability $h(\vartheta)$ to ϑ and set the persistence ρ_ϑ to

[29]We produce robustness checks for a lower and a higher frequency of the model periodicity. Results are qualitatively similar. The current periodicity was chosen because transition flows in and out of entrepreneurship in the data have more observations as compared to a monthly (or lower) frequency.

[30]The bimonthly flows used to pin down these numbers are in Table 2.4.1 while the quarterly flows are in Table 2.2.1.

0.985 corresponding to an annual persistence of about 0.91. The standard deviation σ_ϑ is set to 0.21 in order to generate an earnings Gini coefficient of 0.38. For the transitory match-quality process y, ρ_y is set to 0.85, corresponding to an annual persistence of 0.38, and σ_y to 0.175.

We set the coefficient of relative risk aversion $\sigma = 1.25$. The home production parameter m is set to 0.025, which corresponds to 11% of the average wage in the economy consistent with the fact that only 11% of total consumption spending could be replaced with home-production according to Been et al. (2020). In the benchmark case, the search elasticities ψ and ϕ are both set to 2.0 to generate separable quadratic search costs.[31]

Regarding the gross flows relative to unemployment, we use a linear relation of the $W \to U$ transition characterized by the separation rate $\eta(\vartheta)$ with respect to earnings in the CPS. We specify: $\eta(\vartheta) = \alpha_\eta + \beta_\eta wh(\vartheta)$, where α_η and β_η are estimated using earning quantiles as a proxy for $wh(\vartheta)$. We obtain $\alpha_\eta = 0.0252$ and $\beta_\eta = -0.0047$. Moreover, to better account for CPS transition flows, each period, a fraction $\zeta = 0.8\%$ of individuals retires and is replaced by young uninsured unemployed individuals that enter the workforce with zero net worth and ability ϑ drawn from the invariant distribution Π_ϑ.[32]

The benchmark UI replacement rate μ is set to 0.45, close to the US across states average replacement rate in the last decades, and the UI duration is set to $\bar{J} = 3$, which corresponds to 26 weeks of regular benefits. The UI cap $\bar{b}$ is set to represent 50% of the average wage, which corresponds approximately to the average applied across US states.

Endogenously calibrated parameters and target moments

We now describe parameters picked endogenously to match targets in the data. Although a given parameter does not affect only one particular moment due to the nonlinearity of the model, specific parameters can be somewhat tied to particular moments.

The respective probabilities of getting a business idea, a job opportunity, or meeting the active job search requirement arrive at a Poisson rate such that:

$$\pi_e(s_e) = 1 - e^{-\kappa_e s_e}, \qquad \pi_w(s_w) = 1 - e^{-\kappa_w s_w}, \qquad \pi_m(s_w) = 1 - e^{-\kappa_m s_w},$$

with κ_e and κ_w matching parameters and κ_m the elasticity of monitoring. κ_w is set to capture the 40.1% of unemployed individuals transiting toward employment as observed in the CPS and κ_e is set to match the fraction of entrepreneurs in the economy. The literature uses various definitions of entrepreneurship. As stated above, we consider self-employment in the broader sense as it is the relevant measure for our analysis.[33] In the CPS and the SCF, the mass of self-employed

[31] In section 2.6, we allow for non-separable search intensities, i.e. $\phi \neq \psi$. Our results are qualitatively similar.

[32] This assumption has negligible effects on transition flows, but let us better reproduce the margin of uninsured unemployed individuals in the economy.

[33] In an alternative setting where entrepreneurs are defined as active self-employed business owners (available upon

individuals is about 10.5% and 12.5% respectively. Therefore, we target a self-employment rate of 12%. Finally, we choose κ_m such that 4% of the insured unemployed individuals are sanctioned following detection by the UI agency in equilibrium, consistent with estimates in Grubb (2001). Importantly, none of the above parameters are chosen so as to generate the responsiveness of occupational choice to labor market policy changes (which is an outcome of the model). Rather, they are designed to accurately capture aggregate occupational masses or flows.

The estimation of $g(\vartheta)$ is challenging since the contribution of the skills of an entrepreneur to the performances of a business is generally unobservable. We indirectly infer the mapping between worker and entrepreneur individual productivities using the observed relationship in the $W \rightarrow E$ transition by earnings quantiles. We divide the labor income distribution into 3 quantiles and compute in each the ratio of workers starting a business over the average ratio of workers starting a business in the economy. Over our CPS sample period, workers in the first earnings quantile are 19-20% more likely to start a business than the average worker. In the middle quantile, they are about 14-15% less likely. In the third quantile, they are 2-3% more likely. We use those relative flows to pin down the following values: $g(\vartheta) = [0.136, 0.200, 0.289]$. The resulting transition flows relative to the average flow in the model implies that workers are 19% more likely to become entrepreneurs in the first earnings quantile, 15% less likely in the second quantile, and 1% more likely in the third quantile.

The remaining parameters are chosen as follows. The discount factor β helps to generate a realistic annual capital (excluding public capital) to output ratio of 2.6. Our model accounts for the fact that not all unemployed individuals are eligible for UI. We, therefore, set the exogenous probability for a worker to voluntarily quit her job, q, to replicate the observed insured unemployment rate (IUR) of 2.6. By targeting the IUR, the probability q also capture the margin of unemployed individuals who do not take up their UI benefits while being eligible to do so. The returns to scale parameter in the entrepreneurial sector ν lets us fit the 23% of total income received by the entrepreneurs in the SCF. The fixed cost c_f captures the 5.8% of entrepreneurs who exit each period in the CPS. The share ϖ in the entrepreneurial production function is used to generate a gross flow of unemployment transiting to entrepreneurship in the second wealth quantile relative to the average flow from unemployment to entrepreneurship of 0.90. Table 2.3.1 reports these parameters and related targets.

2.4 Gross Flows: Aggregate Characteristics and Micro Level Behaviors

In this section, we assess the ability of our benchmark economy to generate key properties of the data on labor market flows. We first discuss the gross labor market flows along the wealth and the

request), most of our results are qualitatively similar.

Table 2.3.1. Calibrated parameters and fit.

	Parameter	Value	Moment	Target	Model
Discount factor	β	0.986	K/Y (annual)	2.6	2.6
Returns to scale	ν	0.638	Entrepreneur's share of total income (%)	23.0	22.4
Monitoring	κ_m	3.770	Share of sanctioned eligible unemployed (%)	4.0	4.3
Fixed cost	c_f	0.081	Exit rate from entrepreneurship (%)	5.8	5.8
Unvoluntary exit	q	0.250	Insured unemployment rate (%)	2.6	2.6
Matching prob.	κ_e	0.274	Share of entrepreneurs (%)	12.0	12.0
Matching prob.	κ_w	0.639	$U \to W$ transition (%)	40.1	40.0
Entrep. ability	$g(\vartheta)$	*See text*	W to E by quantile/avg rate (%)	*See text*	*See text*
Entrep. capital share	ϖ	0.249	Flow $U \to E$ in $T2$ wealth relative to average	0.90	0.85

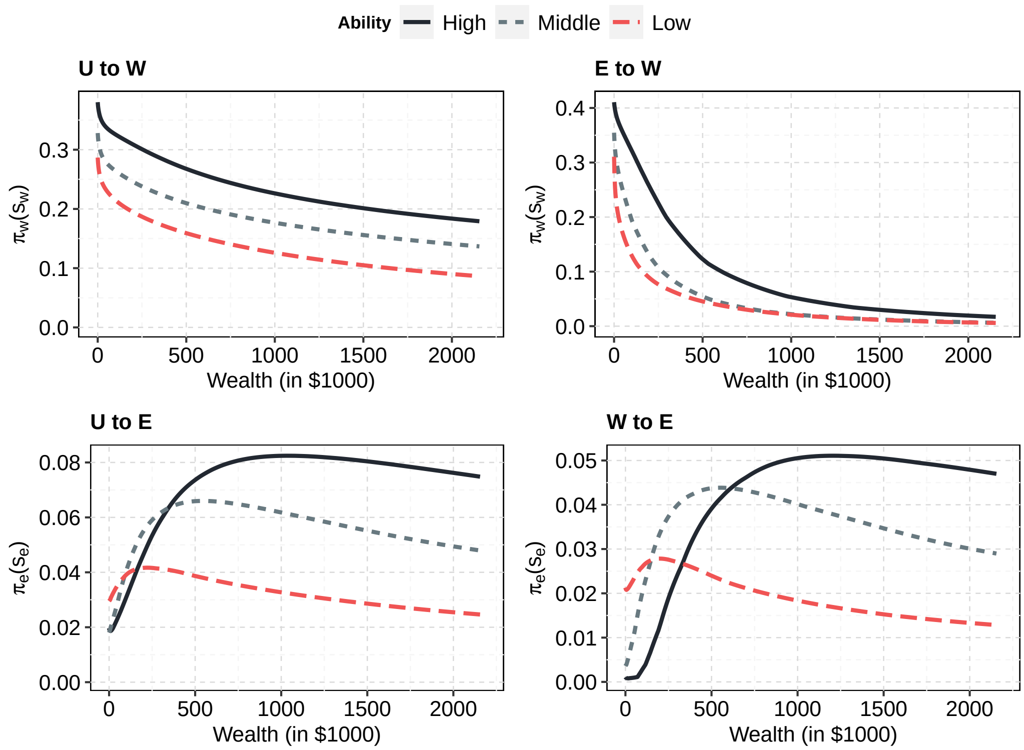
ability distributions and then verify our fit in a number of additional dimensions.

2.4.1 Optimal search efforts

With ability and wealth, the model embeds two parsimonious dimensions of heterogeneity that have an influence on occupational flows. Figure 2.4.1 (top panels) reports how those dimensions interact with optimal job and business search efforts. The model job search effort s_w is consistent with long-established results (see for example Lentz and Tranaes (2005) among others): the optimal job search effort is decreasing in wealth, for both unemployed individuals and entrepreneurs. Wealth provides a means to smooth consumption that conditions the ability to wait for a job and in turn the search effort. Moreover, wealthy individuals are able to establish bigger profitable firms that eliminate the incentive to search for a job. We also illustrate that the higher the ability, the higher the job search effort in the case of unemployed individuals. For high-ability individuals, there is a clear opportunity cost of unemployment with respect to the high wages they can earn in employment. There is a similar effect for relatively wealth-poor entrepreneurs looking for a job.

The more novel aspects here are displayed in the bottom panels: the optimal business search efforts s_e are hump-shaped with respect to wealth. This is due to two opposing effects. First, wealth-poor individuals, who are the most likely to be credit constrained, do not find it interesting to run very small firms and thus provide very small effort. This effect is mitigated, most notably for unemployed individuals, by the introduction of our CES entrepreneurial production function: the efforts to set up a business are positive at zero wealth for those individuals whereas no effort would have been provided in the standard Cobb-Douglas case. Then, as wealth increases, individuals can invest larger capital amounts in their businesses and increase their search effort. Second, beyond a certain wealth level, incentives to establish a business diminish. Similar to looking for a job opportunity, wealthy unemployed individuals face search disincentives due to their important financial wealth compared to the additional income business capital can procure. The same is true for employed individuals looking to create a business *on the job*. We also find that wealth-poor low-ability individuals search more than the corresponding high-ability individuals. For richer individuals, this ordering is reversed. Low-ability individuals have lower UI benefits which make it more advantageous for them to invest in a business earlier as their wealth increases. But they also have a low ability to run a business which makes them reach the above threshold faster. High-ability individuals receive higher UI benefits and go through the same phases but at higher levels of wealth. The same type of reasoning applies to workers searching *on the job* for a business opportunity but relative to their current wage instead of UI benefits. In the next section, we show that those policy functions generate consistent gross flows between occupations across ability levels and across the wealth distribution.

Figure 2.4.1. Implied optimal probability to find a job for unemployed individuals (top left) and entrepreneurs (top right). Implied optimal probability to find a business idea for unemployed individuals (bottom left) and workers (bottom right).



Note: the worker's match component y and the entrepreneur's business shock z are set to their average values. Policy functions out of unemployment are for $j = J$.

2.4.2 Resulting gross flows

Aggregate gross flows Our calibrated model successfully replicates a number of empirical characteristics of gross labor market flows in the US economy even outside explicitly targeted moments. Table 2.4.1 reports bimonthly aggregate gross flows between employment, unemployment, and entrepreneurship. The calibrated model is able to closely capture gross flows in the CPS, including a number of them that are endogenously generated by the model.[34]

[34]Specifically, the $W \rightarrow U$ transition is captured by the $\eta(\vartheta)$ process. As we calibrated the model to match one occupational mass and two transitions, we are left with three degrees of freedom, since targeting masses also indirectly

The model captures that unemployed individuals are more likely than workers to start a business, and replicates the high $E \rightarrow W$ transition (4.5%) and low $E \rightarrow U$ transition (1.4%). Using the 2014 CPS, the Kauffman Indicators of Entrepreneurship reports a share of new entrepreneurs out-of-unemployment of 20.5%, against 19.9% in the model. This fraction is higher for individuals with less than a high school degree (26.5%) and lower for college graduates (17.4%). In the model, the corresponding shares are consistent: 25.4% and 21.2% for low and high ability respectively. Beyond the required match of aggregate occupational flows, recent research has stressed the importance of consistent micro-level behaviors concerning gross flows (see for instance Krusell et al. (2017)). We now discuss the model ability to fit gross flows in two dimensions of interest – ability and wealth – which have been shown to play an important role in occupational decisions regarding entrepreneurship (see Quadrini (2000a) or Cagetti and De Nardi (2006b)) and for unemployed individuals (see Chetty (2008)).

Table 2.4.1. Bimonthly gross flow between occupations in the data and the model.

	Data				**Model**			
	To			Masses	To			Masses
From	W	E	U		W	E	U	
W	97.83	0.50	1.67	84.3	97.40	0.78	1.82	82.4
	(97.78,97.88)	(0.49,0.51)	(1.63,1.72)					
E	4.53	94.16	1.31	10.3	4.48	94.15	1.37	12.0
	(4.51,4.56)	(94.01,94.30)	(1.25,1.38)					
U	40.10	3.40	56.51	5.4	40.15	2.81	57.04	5.6
	(38.94, 41.25)	(3.30,3.49)	(55.33,57.70)					

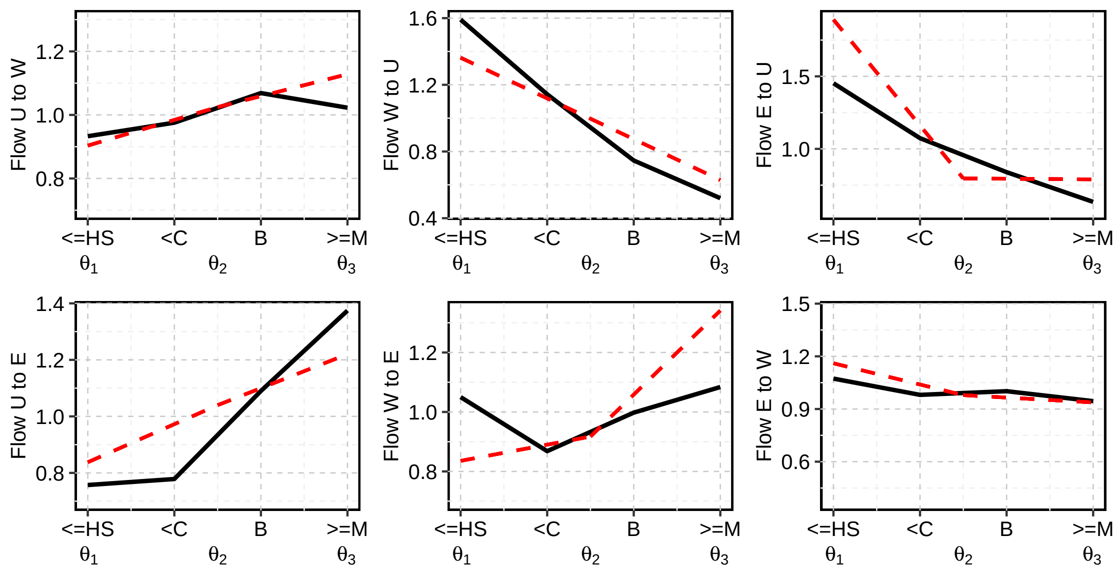
Data sources: authors' computations using CPS data from 1994:IV to 2015:IV. CPS Gross flows are corrected for misreporting. In parenthesis: the 95% confidence interval.

Gross flows by ability Figure 2.4.2 shows the flows by ability level in the model and in the data when we proxy for ability with educational attainment in the CPS. Our use of education is debatable. However, the CPS data do not provide any information concerning business earnings and unemployment compensation that might be relevant to validate the ability dimension. Education is, therefore, the best directly available variable comparable to the model. We divide educational attainment into four groups. $<= HS$: less than or equivalent to a high school degree; $< C$: some college but no degree; B: bachelor's degree; M: master's degree or higher (professional school and doctoral degrees). The above groups respectively account for 31.4%, 29.5%, 23.9%, and 15% of the workers 25 years and older according to the Bureau of Labor Statistics. Overall, the model is able to generate a good fit of the data with perhaps one caveat. Relative to its data counterpart, there is a low flow of workers to entrepreneurship for the lowest ability group. Aside from this, the model accounts exceedingly well for the flow patterns observed in the data. Notably, it fits well the slightly decreasing flow from entrepreneurship to employment, due to the higher risk faced by

target some flows (up to the exogenous entry ζ). Remaining mismatches with respect to the data arise for instance due to transitions out of the labor force, death, or moves to other regions.

low-ability entrepreneurs relative to the fixed cost, which induces them to exit more often. Also, it fits the increasing flow of unemployment to entrepreneurship, most notably because high-ability workers are those with higher levels of wealth. This allows them to wait for the opportunity to run a business and to run a higher scaled more valuable business, conditional on the business shock z.

Figure 2.4.2. Gross labor market flows by ability level ϑ (model in dashed red) and educational attainment (CPS data in straight black).



Note: the data patterns are computed using quarterly flows to obtain a sufficient number of observations by educational attainment. Legend: $<= HS$, less than or equivalent to a high school degree; $< C$, some college but no degree; B, bachelor's degree, M, master's degree or higher (professional school and doctoral degrees).

Gross flows by wealth The decomposition of gross flows by wealth is a natural exercise to consider as wealth drives incentives to quickly exit unemployment due to borrowing constraints or to enter entrepreneurship as personal wealth scales the expected profits of a potential business. Because the CPS does not report individual wealth, we rely on the 1996-2008 (covering the December 1995 to November 2013 period with gaps) panels of the Survey of Income and Program Participation (SIPP). The range of these panels is two to four years. Households are interviewed three times a year about their labor market status over the previous four months in each panel. Every two to four years, they also report their asset holdings. Those assets consist of savings and checking accounts, mutual funds, retirement accounts, real estate, business assets, and other equity. We use total wealth as a measure of individual wealth but our results are robust to choosing only liquid wealth (checking accounts and savings) as a proxy for wealth. Although the definitions of occupations in the SIPP and the CPS slightly differ, we create close enough counterparts based on our explanations in Section 2.2.1.[35]

In the left panel of Table 2.4.2, we report the gross flow rates by wealth quantiles in the SIPP.

[35]In the Online Appendix 1.2.2, we show the aggregate transition matrix between occupations in the SIPP. The latter is quite similar to the CPS analog above.

The right panel reports the gross flow rates by wealth quantiles in the benchmark model and for an alternative case with a Cobb-Douglas entrepreneurial production function ($p \to 0$).

Table 2.4.2. Occupational flow rates by wealth quantiles in the SIPP and the model.

	Data			**Model**					
				Benchmark			Case with $p \to 0$		
	T1	T2	T3	T1	T2	T3	T1	T2	T3
$W \to E$	0.64	0.86	1.50	0.36	0.61	2.03	0.00	0.29	2.71
$W \to U$	1.52	0.85	0.63	1.18	0.98	0.84	1.21	0.98	0.82
$E \to W$	1.17	1.03	0.80	1.56	0.96	0.48	2.49	0.31	0.21
$E \to U$	1.87	0.78	0.34	1.59	0.85	0.56	2.28	0.53	0.19
$U \to E$	0.70	0.96	1.34	0.65	0.85	1.50	0.00	0.59	2.41
$U \to W$	0.96	1.01	1.04	1.25	0.97	0.77	1.27	1.03	0.70

Data sources: authors' computations using the SIPP (1996-2008) panels. The alternative model with a Cobb-Douglas entrepreneurial production function ($p \to 0$) is calibrated to match the K/Y ratio and the masses of occupations.

Overall, the benchmark economy does a qualitatively and quantitatively reasonable job at matching the gross flows by wealth quantiles in the SIPP data. Starting with the flows out of employment, we find that the W to E transition is increasing in wealth. In the model, this is due to the fact that the value of entrepreneurship is higher when a bigger level of business capital is achievable. In that case, conditional on having the opportunity, wealthy individuals find it more valuable to run their own businesses instead of remaining workers. The W to U transition is decreasing. This is due to the selection of high-ability types with lower firing rates at higher levels of wealth. The E to W and E to U transitions are decreasing both in the data and the model. Intuitively, the wealthy can self-insure against the risk of an adverse business shock using their own wealth and are thus less likely to exit. This is also a result of the effect of wealth on the incentive to exit entrepreneurship because businesses are scaled to personal wealth due to the collateral constraint, as illustrated in Figure 2.4.1. The U to E transition is increasing, consistent with the mechanism described for the W to E transition. Finally, the U to W transition is nearly flat in the data but decreasing in the model. This discrepancy can be explained by the fact that, in the model, wealth-poor individuals have a higher incentive to search for a job opportunity.[36] This is related to the fact that the job-finding rate is parsimoniously modeled using the single parameter κ_w. In reality, low-ability individuals self-select at the bottom of the wealth distribution and are less likely to find a job.

As illustrated by the case $p \to 0$ in Table 2.4.2, it is worth noting that a model where entrepreneurs are precluded from producing by using only their own labor significantly changes the flows in and out of this occupation. In such an environment, individuals in the first wealth quantile are unwilling to become entrepreneurs, which is at odds with the SIPP evidence. This negative result is due to the fact that production is scaled to the business capital (and, in turn, to personal wealth). By letting entrepreneurs produce only with their own labor, the benchmark

[36] Krusell et al. (2017) also report a similar discrepancy for the same reasons.

model somewhat disconnects the link between wealth and the propensity to start a business.[37]

2.4.3 Additional validation

Our model captures other moments related to the labor market and entrepreneurship that are not explicitly targeted but that are still reasonably well matched. As argued by Hamilton (2000) and Astebro and Chen (2014), some entrepreneurs create and keep running a business although they would earn more as workers. The share of *out-of-necessity* entrepreneurs, defined as entrepreneurs who started businesses because of a lack of job opportunities, i.e. because $\mathbb{E}_y[W(\mathbf{x})] > E(\mathbf{x}) > U(\mathbf{x})$, is equal to 7.7% in our model and is evaluated by Ali et al. (2008) to be 4.7% of early-stage entrepreneurs for men and 21.4% for women, representing 10% in total. In the pool of previously unemployed new entrepreneurs in the model, in line with Caliendo and Kritikos (2009), this represents a fraction of about 39% *out-of-necessity* entrepreneurs. Relatedly, note that the model captures well the fact that recent entrepreneurs selected out of unemployment have, on average, a lower productivity than those selected out of employment. This is due to a selection effect: individuals selected out of unemployment are more prone to switching to an entrepreneurial situation out of necessity, regardless of the productivity of the business idea. A real-world example would be that, relative to entrepreneurs selected out of employment, those selected out of unemployment are more likely, conditional on their skill-type ϑ, to become low earning entrepreneurs in service activities such as in the hospitality sector or delivering.

The model also has cross-sectional implications regarding the income and wealth distributions. The median ratio of entrepreneurial (resp. worker's) income (including capital gains) to net worth (i.e. total assets minus debt) is 0.20 (resp. 0.88) in the model, while it is 0.19 (resp. 0.73) in the data. In the model, the median ratio of entrepreneurial income over workers' income is 1.7 against 1.4 in the SCF. Entrepreneurs in the model own roughly 29% of total capital, in line with the 33% found by Cagetti and De Nardi (2006b). In terms of wealth distribution, the median ratio of the net worth between entrepreneurs and the whole population is 6.4 in the model against 5.0-6.5 in the SCF. The fraction of zero (or negative) net worth is roughly 8% in the SCF, whereas it is 6% in our model. The model underestimates the wealth Gini: we find 0.67 compared to 0.8 in the SCF.[38] Overall, despite the few limitations that we underlined, to the best of our knowledge, our framework is the first to produce a close fit of the key features of gross labor market flows with entrepreneurship, both at the aggregate and micro levels along the dimensions of ability and wealth.

[37] We explored various mechanisms to generate such a behavior. First, we allowed for the possibility to partially rent capital, by rewriting the borrowing constraint as $k \leq \lambda a + \underline{h}$ with $\underline{h}$ a minimum amount of rentable capital. Second, we assumed two groups of individuals, including one with high non-pecuniary benefits of being an entrepreneur, independently of wealth. In those models, the resulting flows are harder to reconcile with the data.

[38] Regarding the wealth distribution, a version including a *superstar* state of business shock z is able to match the wealth distribution but produces similar results otherwise.

2.5 Impact of UI on Gross Flows: Quantitative Evaluation

We now use our general equilibrium quantitative model to assess the responsiveness of gross labor market flows to UI generosity. We then characterize the effects on occupational masses and aggregate outcomes.

2.5.1 A sample of counterfactual experiments

Our investigations are based on counterfactual stationary economies under alternative UI designs with varying levels of generosity. In the model, total UI generosity is defined by: (i) a maximum duration $\overline{J}$ (this is a bimonthly variable in the model but, for clarity's sake, we express it in weeks equivalent hereafter), (ii) a replacement rate μ, and (iii) a maximum benefit amount $\overline{b}$. To analyze the effects of alternative UI designs, we use the following approach. We run a sufficiently large number of counterfactual deviations from our baseline economy in which the parameters governing the maximum duration $\overline{J}$, the replacement rate μ and the maximum UI benefit amount $\overline{b}$ are uniformly drawn, such that the model variations in those statistics fall in the range of the variations observed across US states and over time, including UI extensions. The replacement rate varies from 30% to 50%, the duration varies from 16 to 99 weeks, and the maximum UI benefit amount varies from 25% to 60% of the mean wage. The generosity of the UI benefits in a given counterfactual i is given by:

$$\text{UI max}_i = C_{adjust} \sum_{j=0}^{\overline{J}} b(\overline{\vartheta}, j) = C_{adjust}(1-\tau) \min\left\{ wh(\overline{\vartheta})\mu_i, \overline{b}_i \right\} \times \overline{J}_i \,, \qquad \overline{\vartheta} = \int_{\mathbf{x}} \vartheta d\Gamma(\mathbf{x}), \tag{2.17}$$

where $\overline{\vartheta}$ captures the average ability level in the economy and $C_{adjust} = \frac{\text{Data wealth median}}{\text{Model wealth median}}$ rescales nominal values in the model relative to the data.

2.5.2 Effects of UI generosity on Occupational Choices

Using observations from the sample of counterfactual experiments, we estimate the elasticity of the flow from a given occupation to another with respect to a variation in UI generosity. To this end, we run the following model-based specification:

$$\log(f_{X\to Y})_i = \varepsilon_{X\to Y} \log(\text{UI max}_i) + err_i, \qquad (X,Y) \in \{E, U_I, U_N, W\}, \tag{2.18}$$

with X any occupation out of which the flow $f_{X\to Y}$ is originating and Y the destination occupation. Additionally, note that similarly to the data section 2.2, we separate the insured unemployment pool U_I from the uninsured pool U_N. $\varepsilon_{X\to Y}$ defines the occupational flow elasticity to UI generosity, i.e. the percentage change in the likelihood to switch to a specific occupation Y out of the occupation X when UI generosity varies by 1%. For instance, $\varepsilon_{U_I \to E}$ is the elasticity of the flow from insured unemployment to entrepreneurship with respect to UI generosity.

As the elasticity $\varepsilon_{X \to Y}$ measures the effects of UI generosity on individual occupational choices using aggregate flows in the model, it is important to rule out any composition effects.[39] To control for this, we estimate equation (2.18) using the stationary distribution of the economy under our baseline UI parameters. To be concrete, agents in this stationary equilibrium face an unanticipated UI shock corresponding to the specific counterfactual $\{\overline{J}, \mu, \overline{b}\}$ set. We then capture the change in the average flow from a given pool of individuals, i.e. $f_{X \to Y}$, arising from variations in search intensities and decisions to switch while keeping the population, prices, and taxes unchanged.

Selection out of unemployment Figure 2.5.1 shows the model-based change in the flow $f_{U \to E}$ from insured (left panel) and uninsured (right panel) unemployment to entrepreneurship. The patterns from unemployment to entrepreneurship are remarkably close to the empirical patterns reported in Figure 2.2.1a and Figure 2.2.1b. The slope of the occupational flow from insured unemployment to entrepreneurship is linear (in log) in UI generosity and is significantly decreasing. Looking at point estimates in Table 2.5.1, it is noticeably steeper than the slope from insured unemployed to employment. The corresponding elasticities, $\varepsilon_{U_I \to E}$ and $\varepsilon_{U_I \to W}$, are again remarkably close to our empirical estimates. In contrast, the elasticities out of the uninsured unemployed pool show no sensitivity to UI generosity: this pool is only slightly less likely to start a business. In section 2.6, we further show that empirical estimates associated to the level of UI benefits are distinct from those on the duration of UI and that our model is able to replicate this feature.

Table 2.5.1. Elasticity of unemployment flows to UI generosity: model and data

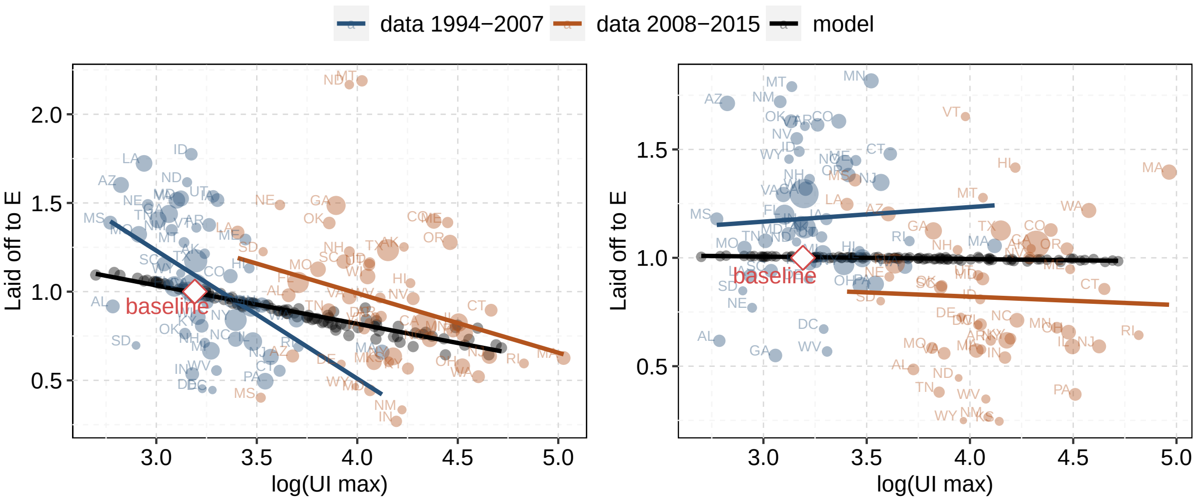
Elasticity $\varepsilon_{X \to Y}$	Data[a]		Model		Model (no monitoring)	
	U to E	U to W	U to E	U to W	U to E	U to W
Insured unemp. workers	-0.200^{**}	-0.056^{***}	-0.247^{***}	-0.044^{***}	-0.346^{***}	-0.214^{***}
	(0.080)	(0.022)	(0.008)	(0.001)	(0.010)	(0.008)
Uninsured unemp. workers	-0.035	0.082	-0.011^{***}	0.001^{***}	-0.027^{***}	0.003^{***}
	(0.193)	(0.053)	(0.000)	(0.000)	(0.001)	(0.000)

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parentheses.
[a] Estimates for the data are taken from the mLogit results in Table 2.2.2.

Additionally, due to the monitoring of program applicants, UI agencies enforce regular job search efforts at the expense of the efforts to start a business. While this channel has been investigated in Hansen and İmrohoroğlu (1992), we quantitatively verify its importance in our setup where it is captured by the probability $\pi^m(s_w)$. To get a sense of the magnitude of this effect,

[39] It is important that the initial distribution of agents remains the same across counterfactuals as otherwise wealth profiles and abilities along the distribution will certainly bias the estimate: depending on wealth and ability, individuals might be more sensitive to UI variations, conditioning the impact we are measuring. In the Online Appendix 2.1 we provide a second metric using the counterfactual long run steady-state masses that yield long run elasticities. The virtue of this measure of elasticities is that it better captures long run GE adjustments. The drawback is that it is based on different population masses in each occupation.

Figure 2.5.1. UI generosity and model average flows from the insured (left panel) and uninsured (right panel) unemployed pools.



Note: each dot corresponds to a UI (duration, benefit level) pair. The red square marks the current average regular UI provision in the US, with $\mu = 0.45$, $\bar{J} = 26$ weeks, and $\bar{b} = 50\%$ of mean wage. The maximum UI generosity here is $\mu = 0.498$, $\bar{J} = 99$ weeks, and $\bar{b} = 60\%$ of mean wage.

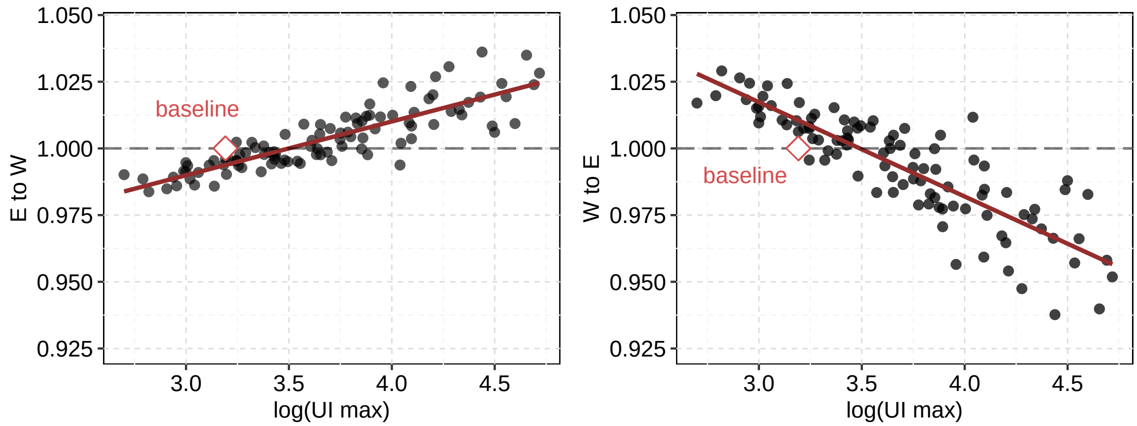
we run our counterfactual experiments in an alternative specification without monitoring. We find that the sensitivities of gross flows from insured unemployment to both employment and entrepreneurship are magnitudes higher and imply resulting elasticities $\varepsilon_{U_I \to E} = -0.346$ and $\varepsilon_{U_I \to W} = -0.214$. Therefore, our monitoring feature, despite its stylized nature, helps in producing the observed lower sensitivity of these flows as estimated in Table 2.2.2.

Selection out of employment and entrepreneurship In the case of the gross flows between entrepreneurship and employment, we find that $\varepsilon_{W \to E} = -0.036$ and $\varepsilon_{E \to W} = 0.020$ and illustrate the resulting slopes in Figure 2.5.2. A consequence of the magnitude of these elasticities is that changes in occupational choices *on-the-business* and *on-the-job* following a change in UI generosity are likely to have long run implications on occupational masses alongside the large and direct effect on the unemployment pool. We further discuss these elements in section 2.5.3.[40]

Selection by ability and wealth Increasing UI benefits has a disproportionate impact on particular groups of individuals in our economy. Table 2.5.2 displays the decomposition by ability and

[40]Using our CPS sample, we verify whether UI generosity is correlated with the likelihood that individuals move from employment to entrepreneurship as well as from entrepreneurship to employment. We restrict our CPS sample to the 25 to 50 years range to focus on individuals most likely to select into employment or self-employment as an alternative life prospect. As in section 2.2.2, we run $\left\{\text{Occ}_1 \text{ to } \text{Occ}_{2ist}\right\} = \alpha + \beta \log(UI\, generosity)_{st} + \boldsymbol{\xi}\mathbf{X}_{it} + \boldsymbol{\eta}\mathbf{Z}_{st} + \boldsymbol{\lambda}_s + \boldsymbol{\mu}_t + \epsilon_{ist}$ with $\text{Occ}_1 \in \{E, W\}$, $\text{Occ}_2 \in \{E, W\}$ and similar controls $\mathbf{X}_{it}$ and $\mathbf{Z}_{st}$ as in the main specification. $\boldsymbol{\lambda}_s + \boldsymbol{\mu}_t$ refers to state and time fixed effects. Despite the fact that we do not have any control groups (akin to the non-eligible unemployed workers in previous regressions), we find that the sign of the effect of increasing UI is significantly positive for the flow from entrepreneurship to employment as $f_{E \to W} = 0.33$ and significantly negative for the reverse flow with $f_{W \to E} = -0.30$. This result is consistent with the model, higher UI generosity reallocates individuals from self-employment to employment.

Figure 2.5.2. UI generosity and model average flows between entrepreneurship and employment.



Note: each dot corresponds to a UI (duration, benefit level) pair. The red square marks the current average regular UI provision in the US, with $\mu = 0.45$, $\bar{J} = 26$ weeks, and $\bar{b} = 50\%$ of mean wage. The maximum UI generosity here is $\mu = 0.498$, $\bar{J} = 99$ weeks, and $\bar{b} = 60\%$ of mean wage.

wealth of elasticities $\varepsilon_{U_I \to E}$ and $\varepsilon_{U_I \to W}$. First, it is noticeable that $\varepsilon_{U_I \to E}$ is less responsive with ability and is almost flat for $\varepsilon_{U_I \to W}$. Second, for both elasticities, wealth poor individuals (relative to the median) have a stronger response than wealthier ones.

Table 2.5.2. Model-based elasticity of insured unemployment to UI generosity by ability and wealth.

Elasticity $\varepsilon_{X \to Y}$	Ability			Net worth	
	$\vartheta = \vartheta_1$	$\vartheta = \vartheta_2$	$\vartheta = \vartheta_3$	$a <$ median	$a \geq$ median
$\varepsilon_{U_I \to E}$	-0.291^{***}	-0.265^{***}	-0.155^{***}	-0.356^{***}	-0.109^{***}
	(0.011)	(0.008)	(0.013)	(0.013)	(0.008)
$\varepsilon_{U_I \to W}$	-0.045^{***}	-0.045^{***}	-0.039^{***}	-0.028^{***}	0.009
	(0.002)	(0.001)	(0.003)	(0.005)	(0.006)

Notes: *p<0.1; **p<0.05; ***p<0.01. Standard errors are reported in parentheses.

The decomposition by wealth is straightforward. On the one hand, and related to the *liquidity effect* in Chetty (2008), wealth-poor individuals are closer to the liquidity constraint and are therefore more sensitive to variations in UI generosity. On the other hand, a higher wealth level lets prospective entrepreneurs run larger and more valuable firms. Again, wealth-poor individuals have a smaller incentive to start a business. As a consequence, the combination of this incentive and the *liquidity effect* makes wealth-poor individuals (below the median) 3 times more responsive to UI generosity when trying to start a business than wealth-rich individuals (above the median). Therefore, the presence of a credit constraint plays an especially important role in understanding the high elasticity of wealth-poor prospective entrepreneurs as investment capability is a key requirement for a valuable business.

Regarding the decomposition by ability ϑ, low ability individuals are on average poorer than higher ability ones. Thus, the liquidity and threshold effects of more UI generosity are stronger for those agents. This explains the decreasing responsiveness of $\varepsilon_{U_I \to E}$ with ability. We do not find, however, that the responsiveness of the gross flows from insured unemployment to employment, $\varepsilon_{U_I \to W}$, differ much by ability. This is due to the effect of monitoring which induces unemployed agents to provide a sufficient amount of job search effort. In an alternative specification without monitoring, the corresponding elasticity decreases with ability and similar to what is obtained here with the gross flows toward entrepreneurship.

Mechanisms Our elasticities results substantiate the idea that higher UI generosity lowers the incentives to exit insured unemployment. Two well-known effects support this interpretation: (i) a *moral hazard effect* which captures the change in the marginal incentive to search following a variation in UI benefits that effectively lowers the expected net income gain of taking a job; (ii) the above-mentioned *liquidity effect*, previously discussed in Browning and Crossley (2001) and Chetty (2008), which captures the variation of the search effort with respect to the loosening of the liquidity constraint following a change in UI generosity. Specifically, for a given level of wealth, the *liquidity effect* is the effect of an extra amount of wealth coming from more UI generosity. This extra amount relaxes the effect of the borrowing constraint and helps with consumption smoothing, thereby lowering the incentive to exit insured unemployment.

On top of those effects, additional considerations appear when analyzing the impact of UI generosity in an entrepreneurial context. As entrepreneurs are not part of the UI system, the value of entrepreneurship is not responsive to an increase in UI generosity, at least not directly, and $\frac{\partial E}{\partial b} \approx 0$ in the current and future periods. As a consequence, insured unemployed individuals significantly reduce their business search effort s_e relative to their job search effort s_w. We view this as an *insurance coverage effect*: relative to a variation in UI generosity, it is the change in the relative riskiness between two asymmetrically covered occupations leading to a change in the incentive to choose one or the other activity. The distance between the entrepreneurial and unemployment values is substantially affected by the change in UI generosity (i.e. $(E - U_I') \ll (E - U_I)$), while the distance between employment and unemployment values is less affected (i.e. $(W' - U_I') < (W - U_I)$), due to the asymmetric UI coverage between employment and entrepreneurship. Therefore, the business search effort of insured unemployed individuals is likely to be more sensitive to a change in the UI relative to the job search effort. The *insurance coverage effect* also concerns the gross flows between employment and entrepreneurship: the risk of a job loss is covered by UI whereas the loss of a self-employed activity is uninsured. As a consequence, the higher the UI generosity, the higher the opportunity cost associated with an entrepreneurial activity relative to employment.[41]

Those results demonstrate that variations in UI generosity have consequences beyond the direct effects on the pool of unemployed individuals and concern gross flows in and out of en-

[41] A similar argument is given in Fuchs-Schündeln and Schündeln (2005): they show that people with lower risk aversion select into civil service occupations. In our paper, the degree of employment coverage distorts the relative riskiness of entrepreneurship relative to employment, which modifies the selection into those occupations.

trepreneurship and employment in general.

2.5.3 Long Run Occupational Masses

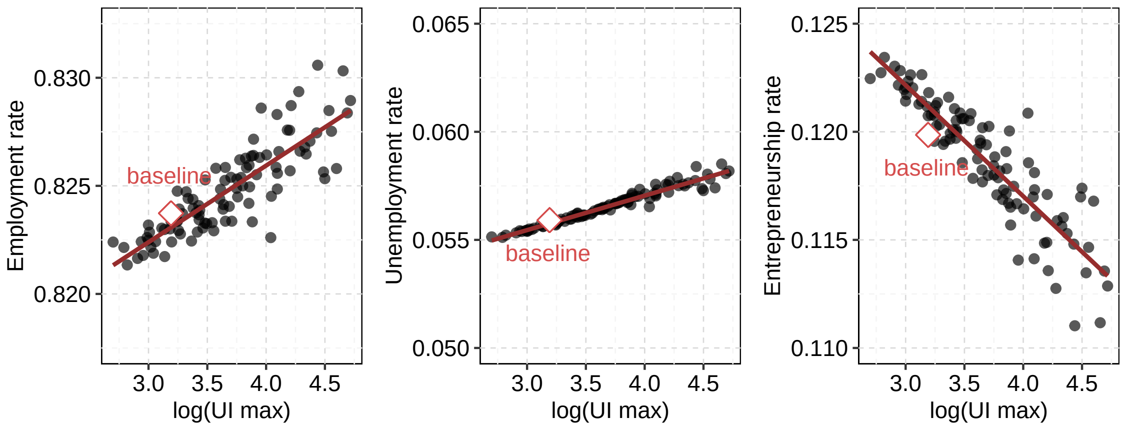
Our model has also additional implications on the long run aggregate masses of individuals in each occupation. To capture this, we compute the long run stationary equilibrium induced by the change in UI generosity in each counterfactual experiment. This analysis provides a characterization of the long run status of the labor market once all transitional adjustments are completed and equilibrium prices are adjusted. This is valuable because the resulting masses are difficult to predict using only information concerning gross flows since the relative mass of individuals in each occupation is different. For instance, even though unemployed agents might have a strong response to a change in the UI, they only represent 5% to 6% of individuals in the model. In contrast, workers account for around 82% to 83% of the population, but, as discussed previously, they react much less to UI variations. The resulting occupational masses are therefore *a priori* ambiguous.

In Figure 2.5.3, we display the long run invariant mass of individuals in each occupation. We find that the employment rate is increasing with UI generosity in the long run, with an elasticity of 0.004. However, it is the masses of entrepreneurs and unemployed individuals that are the most affected (in relative terms) with long run elasticities of the mass to UI generosity of -0.044 and 0.028 respectively. To put this into perspective, a doubling of UI generosity relative to the baseline value would increase the unemployment rate by 0.11 percentage points and decrease the entrepreneurship rate by 0.36 percentage points.

We stress that the patterns discussed above remain robust even when general equilibrium adjustments are neutralized, for instance by fixing prices to the ones in our baseline stationary equilibrium. Our results point out that UI generosity has a particularly large effect on entrepreneurship in the long run. Moreover, if one was to consider total employment as the addition of both self-employment and employment, UI generosity would have a negative and significant effect on this aggregate variable in the model.

Relating our findings to the existing literature is instructive. On the one hand, and as shown by the above results, occupational flows, especially out of insured unemployment, are consistent with those established in the literature and supported by liquidity and moral hazard effects. Notably, UI generosity has a depressing effect on the flow from insured unemployment. On the other hand, another strand of the literature, for instance, Chodorow-Reich et al. (2019) and Boone et al. (Forthcoming), empirically find a small (and non-significant) effect of UI generosity on the aggregate level of employment (in relative terms). This observed disconnect between micro-level transitions from unemployment to employment and the resulting aggregate employment are hard to recon-

Figure 2.5.3. UI generosity and occupational masses.



Note: each dot corresponds to a UI (duration, benefit level) pair. The red square marks the current average regular UI provision in the US, with $\mu = 0.45$, $\bar{J} = 26$ weeks, and $\bar{b} = 50\%$ of mean wage. The maximum UI generosity here is $\mu = 0.498$, $\bar{J} = 99$ weeks, and $\bar{b} = 60\%$ of mean wage.

cile.[42] We show that when taking into account entrepreneurship, adjustments at the micro-level might not reflect adjustments at the macro-level. In the long run, the employment rate increases in our setup with UI generosity because individuals are less likely to enter entrepreneurship and are more likely to exit, generating a reallocation from employment to entrepreneurship.

2.5.4 UI Generosity and Long Run Aggregate Outcomes

We now discuss the effects of UI generosity on long run macro aggregates. As the mass of entrepreneurs decreases, the long run entrepreneurial sector output Y^E and capital K^E are significantly reduced. Conversely, as the mass of workers and aggregate corporate capital remains nearly constant, aggregate corporate output Y is only slightly impacted. Perhaps surprisingly, the average firm size increases with higher UI generosity. As discussed earlier, this is related to the fact that the incentives to create a business and remain entrepreneur are higher for wealth-rich individuals when employment becomes relatively better insured.

Additionally, a higher level of UI generosity reduces precautionary savings overall while selecting wealthier entrepreneurs. Together, these effects lead the ratio of median net worth between entrepreneurs and the rest of the population to rise with UI generosity. Overall, most of the striking effects appear on entrepreneurial margins.

[42]For instance, Boone et al. (Forthcoming) argues that a demand channel following an increase in UI benefits could generate an increase in the aggregate employment rate and dampen the negative effect from micro disincentives. Concerning recent empirical findings with small micro disincentive effects on the job-finding rate, Farber et al. (2015) study the effects of UI extensions during the Great Recession and find little or no effect on job-finding but a reduction in labor force exits due to benefit availability.

2.6 Model Implications and Robustness

In this section, we evaluate key predictions of our framework and perform sensitivity analyses on modeling assumptions and parameter choices.

2.6.1 Policy Experiment: *on-the-business* Unemployment Insurance

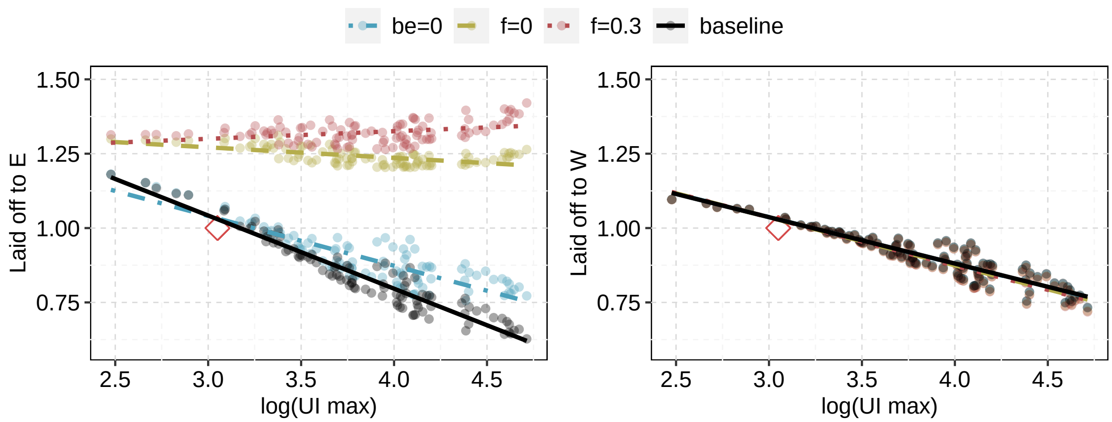
We provide an illustration of the prediction of our model that asymmetric insurance coverage between occupations has large effects on selection and occupational masses by using a counterfactual broader UI policy. We slightly depart from our benchmark specification to let eligible new entrepreneurs be insured *on-the-business*: individuals with UI rights out of the insured unemployed pool are allowed to retain their UI benefits while running a business. The amount of benefit received is contingent on the level of entrepreneurial income π_r. Formally, a benefit $b_e(\theta, j, \pi_r)$ is given to the entrepreneur as a function of her income π_r and UI benefits $b(\theta, j)$. When $\pi_r \leq 0$, full benefits can be claimed. Otherwise, the effective entrepreneurial income is partially or fully deducted from UI benefits, such that

$$b_e(\theta, j, \pi_r) = \begin{cases} b(\theta, j) & , \quad \text{if} \quad \pi_r \leq 0, \\ b(\theta, j) - (1-f)\pi_r & , \quad \text{if} \quad 0 < \pi_r < \frac{b(\theta,j)}{1-f}, \\ 0 & , \quad \text{if} \quad \pi_r \geq \frac{b(\theta,j)}{1-f}, \end{cases} \tag{2.19}$$

where the *deduction parameter* $f \in [0, 1]$ is used to define the specific deduction policy that is implemented. In section 3 of the Online Appendix, a graphical illustration of this policy scheme is given in Figure A2 and **??** illustrates the associated value functions ordering for insured unemployed and entrepreneurial agents. We run our counterfactual UI generosity experiments under three alternative assumptions. The first policy alternative is noted $b_e(\theta, \pi_r) = 0$ because no UI benefits are distributed *on-the-business* and only the right to claim any outstanding benefits once back in the unemployment pool is allowed. The second alternative is noted $f = 0$ because UI benefits are distributed *on-the-business* but if $\pi_r > 0$, the corresponding profit is fully deducted from the benefits. In the third alternative noted $f = 0.3$, UI benefits are distributed *on-the-business* and when $\pi_r > 0$, the corresponding profit is partially deducted following the rule in equation (2.19). The ability to return to the unemployment pool with outstanding UI rights upon business failure is also allowed in the last two alternatives. Each policy alternative is financed with an adjustment in the tax rate τ_w.

In Figure 2.6.1, we extend our counterfactual experiment to capture the effects of increasing UI generosity on this policy. We find a flat flow shape with respect to the UI generosity for the $f = 0$ and $f = 0.3$ policies. In fact the $\varepsilon_{U_I,E}$ elasticity is close to zero for these cases. The reason is quite intuitive. The decreasing shape of the flow from insured unemployment to paid employment is explained by a combination of the *liquidity* and the *moral hazard* effects. Here, these effects are both strongly mitigated. When UI generosity increases not only do UI benefits $b(\theta, j)$ increase but the entrepreneurial income (i.e. $\pi_r + b_e(\theta, j, \pi_r)$) under the $f = 0$ and $f = 0.3$ policies also increases. Therefore, the *moral hazard effect* virtually disappears. It is the same with the *liquidity effect* as the consumption smoothing effect of higher UI will benefit both insured unemployed agents and insured entrepreneurs. The flow shape is still decreasing under the $b_e = 0$ case, since the above *moral hazard* and *liquidity* effects remain effective. Indeed, there is no direct insurance of low entrepreneurial income π_r and as long as the new entrepreneur does not return to the unemployment pool, the cost to the UI system is zero. It is noteworthy that beyond the cost factor, relying only on this limited form of insurance is already making the self-employment activity more attractive while mitigating the distortion in favor of job search.

Figure 2.6.1. On-the-business UI and flow from insured unemployment.



Note: each dot corresponds to a UI (duration, benefit level) pair. The red square marks the current average regular UI provision in the US, with $\mu = 0.45$, $\overline{J} = 26$ weeks, and $\overline{b} = 50\%$ of mean wage. The maximum UI generosity here is $\mu = 0.498$, $\overline{J} = 99$ weeks, and $\overline{b} = 60\%$ of mean wage.

We differ the detailed discussion of the effects on occupational masses to the Online Appendix. However, as expected, *on-the-business* UI policies encourage the start of businesses among the unemployed individuals but, in the long-run, they reallocate individuals from paid-employment to self-employment. To relate these findings to the existing empirical literature, the effects we have discussed can explain the large impact of this type of policies in countries with relatively generous UI schemes. These findings corroborate, for instance, empirical evidence in Hombert et al. (2020) for France detailing a surge in entrepreneurial entries, as well as a surge in the number of small firms (without employees) after the implementation of this type of policy.[43]

[43] In France, the regular UI system through the *Plan d'Aide au Retour à l'Emploi* (PARE) policy lets entrepreneurs partially deduct their entrepreneurial profit, with an extension close to our $f = 0.3$ case.

2.6.2 UI extensions during the Great Recession

As an additional test of the implications of our framework, we now explore, within our model, the response of gross labor market flows to a change in labor market conditions using temporary variations in UI policy during the Great Recession (GR). Starting in late 2008, special UI extensions (the EB and EUC programs) were implemented for about 5 years. We quantitatively evaluate the repercussion on occupational flows and masses and decompose the effects of variations in UI generosity. For the sake of concision, we only comment here a limited set of results while our detailed approach and results are available in section 3 of the Online Appendix.

Using the transitional dynamics of our model, we are able to produce consistent data patterns over the GR as illustrated in Figures **??** and **??** of the Online Appendix. Notably, the adjustments of $f_{E\rightarrow W}$ and $f_{U\rightarrow E}$ flows are consistent with CPS observations over the GR period. We are also able to decompose the contributions of the separation rate, the job-finding rate, the entrepreneurial productivity process as well as the duration of the UI policy to the evolution of gross labor market flows and occupational masses. Focusing on the effects of UI extension adjustments, we find that they have a non-negligible impact on occupational flows and therefore on occupational masses. Echoing Nakajima (2012), UI extensions increases the unemployment rate. Comparing the models with and without the extensions, we find a difference of about 0.3-0.35 percentage points in 2010:2011 that persists until the end of the EB and EUC programs. In line with our previous findings, these UI extensions also decrease the entrepreneurship entry rate and thus the entrepreneurship rate by about 0.4-0.45 percentage points in 2010:2011. However, only a marginal impact on the employment rate is found, consistent with our previous findings. This reallocation effect mechanically translates into a lower GDP as the number of entrepreneurs decreases.

In the data, estimates of the effects of UI duration on the propensity to select into entrepreneurship are lower than those of the effects of UI benefits. This result materializes in Table 2.2.2. In Panel B (2000 to 2007 period), changes in UI generosity are mostly due to benefits adjustments when they are mostly due to duration changes in Panel C (2008 to 2015 period). In Table 2.6.1, we decompose the effects of UI generosity in the overall sample (Panel A) and compare it with the model. We illustrate that both the data and the model corroborates the above finding and that the magnitude of the effects in the model are consistent with the empirics. Overall, the smaller impact of changes in UI duration with respect to UI benefit levels leads us to conclude that the numbers above, notably those about the entrepreneurship rate, might constitute a lower bound of the effects of a temporary change in UI generosity on gross flows and masses during the GR.

2.6.3 Robustness

UI financing We first explore the effects of letting the workers bear the entire cost of UI, i.e. $\tau_p = 0$. In such a case, the elasticities of flows $\varepsilon_{U_I\rightarrow W}$ and $\varepsilon_{U_I\rightarrow E}$ are respectively slightly higher

Table 2.6.1. UI duration versus UI benefits: data and model

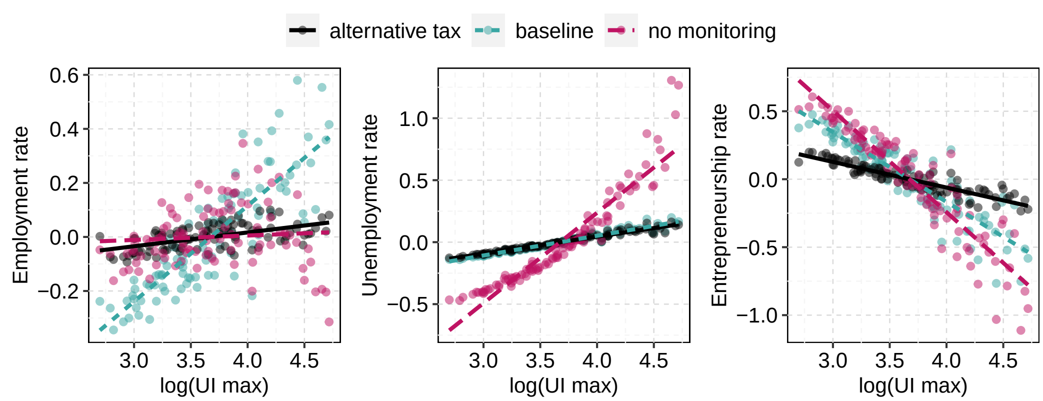
U → E	Data (OLS)	Model
log(UI benefits (WBA))	-0.590^{***}	-0.458^{**}
log(UI duration (EB + EUC))	-0.121^{**}	-0.261^{***}

Notes: *p<0.1; **p<0.05; ***p<0.01. SE clustered by US states. Results are robust using mLogit.

and lower relative to their baseline counterparts. Indeed, a higher labor income tax due to a higher UI generosity changes the incentive to switch between entrepreneurship and employment relative to the benchmark because it reduces the after-tax labor income of employed individuals. Aggregating gross flows, the black dots in Figure 2.6.2 displays the resulting occupational masses under this alternative when we vary UI generosity. With a higher pass-through of the cost of UI toward workers, we find that a higher generosity induces a less positive effect on the aggregate employment rate, which becomes almost irresponsive. This also reduces the negative impact on the self-employment rate, but without overturning our result. This result points out that economies taxing differently employers and employees when UI increases might exhibit differing trade-offs between occupations.

Absence of monitoring The pink dots in Figure 2.6.2 show the resulting masses in the absence of monitoring of program applicants, i.e. $\pi_m(s_w) = 1, \forall s_w$. Under this alternative, the unemployment rate is more sensitive to variations in UI generosity. This leads to a stronger reduction in the self-employment rate and to a more stable aggregate employment rate as UI varies. The main insight that entrepreneurship is highly sensitive to UI generosity remains valid.

Figure 2.6.2. Percent deviation from the mean sample occupational mass for alternative specifications.



Other robustness Non-separable disutility of search, such as $\phi \neq \psi$ with $\phi = 2.5$, produces results close to our benchmark. In particular, there is not much interaction effects between search behaviors. On the entrepreneurial side, we experimented with a high *superstar* business shock z to generate a consistent wealth distribution at the top. While matching the Gini coefficients of wealth, our results were only marginally affected. Finally, under a Cobb-Douglas assumption for the entrepreneurial production function ($p \to 0$), the responsiveness of the flow from unemployment to entrepreneurship to UI generosity is higher with $\varepsilon_{U_i \to E} = -0.321$ due to the fact that entrepreneurship becomes highly sensitive to changes in wealth.

2.7 Conclusion

This paper considers the importance of the entrepreneurial occupation for gross labor market flow dynamics and underlines the relevance of the unemployment insurance design for the selection in and out of this activity. We empirically find a negative and significant relation between UI generosity and the propensity for eligible unemployed individuals to select into entrepreneurship, which is an order of magnitude larger than the one from unemployment to employment. Our parsimonious model of gross labor market flows between employment, unemployment, and entrepreneurship produces an empirically accurate characterization of US aggregate gross labor market flows while accounting for the micro-level decisions that support them. Notably, the model produces an adequate fit of gross flows conditional on individual state variables, in particular by wealth and ability. Our benchmark model is able to generate the empirical responsiveness of gross labor market flows to a typical change in UI generosity. We show that asymmetric insurance coverage of occupations is a significant factor influencing gross labor market flows. A key implication is that reallocations of individuals from entrepreneurship to employment following an increase in UI generosity lead to a faintly increasing aggregate employment rate, providing a channel to explain its observed irresponsiveness to UI variations.

Chapter 3

Buying and Selling Entrepreneurial Assets

Alexandre Gaillard[1,2] Sumudu Kankanamge[3]

Abstract

How are the options to buy and sell a business relevant for entrepreneurs? Prospective entrepreneurs value the purchase of mature firms while incumbents want to recover both the tangible and intangible value of their businesses upon exit. We introduce an empirically-relevant theory of entrepreneurial assets transfer consistent with the life-cycle of entrepreneurs and the dynamics of firms. A businesses for sale market lets entrepreneurs trade their firms. We find that shutting that market down leads to a substantial drop in aggregate output and alters the pool of firms, incentives to enter and exit, and the wealth distribution.

Keywords: Entrepreneurship, Business transfers, Maturity value, Intangible assets.

[1]This paper supersedes a previous version titled "Entrepreneurship, Inter-Generational Business Transmission, and Aging". We are indebted in no specific order to Anmol Bhandari, Henrique S. Basso, Fabrice Collard, Mariacristina De Nardi, Patrick Fève, Christian Hellwig and Ludwig Straub for their useful discussions. We are also very grateful for constructive comments from participants at the 2019 SED conference in St Louis, the 25th CEF conference in Ottawa, and the TSE seminars. We finally thank the University of Science of Toulouse for providing us an access to a supercomputer under the CALMIP project. The authors acknowledge funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program).

[2]Toulouse School of Economics, University of Toulouse Capitole, 31080 Toulouse, France. Email: alexandre.gaillard@tse-fr.eu.

[3]Toulouse School of Economics, University of Toulouse Capitole, 31080 Toulouse, France. Email: sumudu.kankanamge@tse-fr.eu.

3.1 Introduction

Prospective entrepreneurs can acquire a firm by either a new creation or the purchase of an existing business. However, empirical evidence shows a clear performance divide between recently created and purchased firms, giving this choice some significance. For instance, in the pool of entrepreneurs who recently acquired a business, those who purchased an existing business face half the failure rate of those who founded a new one.[4] Moreover, despite accounting for about 20% of the mass of recently acquired firms, purchased firms contribute to about 60% of the employment and the total sales. In fact, this disparity between founded and purchased firms extends over to the main components of heterogeneity the entrepreneurial literature generally considers, namely risk, financial conditions, and productivity. We relate this difference to the maturity of a firm: early-stage firms will face more stringent credit, productivity, and risk conditions as compared to mature firms.[5]

The importance of the maturity of a firm can be explained by the fact that entrepreneurial assets are not limited to tangible physical capital. According to Bhandari and McGrattan (2018), around 60% of business assets are in the form of intangible assets –customer base, client lists, brand value, organization, etc.– most of which, as opposed to tangible assets, cannot be bought directly and take time to accumulate. The option to purchase an existing business is a key factor in shortening that time and preserving the value of intangible assets in the economy. Given this, the question of transferring entrepreneurial assets appears particularly consequential. On the one hand, the exiting entrepreneur has to decide either to sell or liquidate her assets, conditioning whether the accumulated maturity of her firm will persist or not. On the other hand, the entering entrepreneur will find it desirable to purchase an existing mature business but will be subject to borrowing constraints. In this paper, we build a theory of entrepreneurial assets transfer consistent with empirical evidence and introduce a businesses for sale market that values the maturity of entrepreneurial firms.

The agenda of assessing and explaining the purchase and the sale of entrepreneurial assets presents a few challenges. Data on small and medium-sized enterprise (SME) transfers is scarce. Moreover, there is no theoretical framework in the literature to properly consider transfers in a standard entrepreneurial setting. Thus this paper makes two main contributions. First, we provide a theoretical framework with endogenous options to buy or found businesses on the entry side and sell and liquidate them on the exit side. Our model embeds a businesses for sale market allowing firm transfers and is designed to capture the frictions appearing on that market. We

[4]Whenever possible, we define an entrepreneur as an individual actively managing a firm, deriving her primary source of income from it, and holding at least a part of the business assets. We, therefore, exclude passive business owners from the analysis.

[5]Indeed, those margins have been shown to be age-dependent. See, among others, Dunne et al. (1988), Sakai et al. (2010), or Dyrda (2015).

especially consider two distinct margins that interact with the decision to either purchase or found a business on the SME market: (i) the existence of substantial differences between the foundation of a new business and a purchase that we capture with the concept of maturity, (ii) selling and purchasing frictions that limit business transfers in equilibrium. Second, we study the quantitative importance of those margins on aggregate and distributional outcomes.

Our baseline economy is a stylized life-cycle occupational choice model with heterogeneous agents. Individuals choose whether to work in a corporate sector or to be entrepreneurs. We introduce key endogenous choices: each period, an incumbent entrepreneur might need to sell her business and will face an equilibrium selling price as well as a probability to sell. Without an opportunity to sell, the incumbent will be forced to either continue her current activity or liquidate the business assets. Conversely, a prospective entrepreneur might enter the sector by endogenously choosing a firm size and either finding an existing business to buy or founding a new one, these decisions being subject to credit constraints and specific costs. A small and medium-sized enterprises for sale market (SMESM) aggregates selling and buying decisions. Its equilibrium price is designed as an abstract object to account for both the intertemporal and intangible value of a business. Outside this market, the value of intangible assets cannot be recovered. With the concept of maturity, we introduce a very parsimonious measure of intangible assets: for two firms with the same level of tangible assets, the difference in the sale value will reflect specific advantages provided by transferable intangible assets. As most intangibles cannot be directly bought and their accumulation is time-intensive, we argue that the maturity of a firm is a measurement of the intangibles it has built. We assume that all founded businesses are early-stage immature firms with low levels of intangibles whereas purchased ones are well-established and mature with high levels of intangibles. This assumption is supported by empirical evidence: controlling for characteristics of firms and owners, early-stage businesses are, on average, more likely to fail, make lower profits, are charged higher interest rates and have a tighter borrowing constraint.

We support our theoretical contribution with data from the Survey of Small Business Finances (SSBF), the Survey of Business Owners (SBO), and the Survey of Consumer Finances (SCF). With the above data, we first show that business buying and selling constitutes fundamental margins for entrepreneurs. Second, we document notable differences between business acquisition as a result of a purchase as opposed to a new creation and illustrate that, overall, the transfer of business assets over the SMESM results in more efficient acquisition patterns. We use key moments in the SCF, the SBO, and the SSBF data to discipline our model and show that our baseline setting provides a consistent aggregate and cross-sectional representation of the U.S. economy. We carefully validate the properties of our baseline, even outside of specific targets. For instance, the model plausibly accounts for entrepreneurial life-cycle patterns, the increasing survival rate relative to the preceding year as businesses age, and it furthermore generates a consistent concentration at the top of the wealth distribution.

Our results can be organized around four main points, all related to the significance of the SMESM and the maturity of firms. First, we demonstrate the aggregate importance of the SMESM by shutting down that market: under our standard parameterization, the aggregate output drops by a substantial 10.5% with respect to our baseline. This drop is mostly due to an important decrease in the SME sector production. Aggregate savings also decline but at the general equilibrium, the interest rate increases and the wage rate falls, somewhat counteracting potential further output losses. At the same time, the fraction of entrepreneurs decreases despite being mitigated by higher incentives to enter entrepreneurship due to the combined effect of prices. However, the fraction of mature businesses clearly diminishes, changing the composition of the types of firms in the economy: trading on the SMESM generates larger businesses and the ability to transfer maturity preserves the higher survival rates, profitability, and better credit conditions of existing firms.

Second, we decompose the maturity of a firm into its components –namely failure rates, profit rates, and borrowing limit and interest rates– in order to understand the specific impact of each of them on aggregate outcomes. We find that the lower failure rate and higher profit rate of mature businesses are the most important elements embedded in the option value of purchasing a business relative to founding while the other components only have marginal effects. Without the contribution of the first two components, the fraction of business purchasers substantially reduces. Moreover, when the contribution of all components of maturity is removed, we show that there is nothing of value to transfer on the SMESM.

Third, we underline a completely new channel to match wealth concentration and inequality based on the heterogeneity of firms and which is furthermore consistent with empirical evidence. Our baseline model convincingly reproduces the U.S. wealth concentration but the novel aspect is due to the key role of the SMESM and maturity in producing that outcome. Indeed, mature firms accumulate higher returns and, because of lower failure rates, they do so over longer periods. In turn, the SMESM preserves the benefits of maturity between owners, concentrating more wealth into the hands of these individuals.

Finally, we find that matching frictions on the SMESM have a substantial impact on aggregate outcome and the wealth distribution. Increasing the probability to sell a business on the SMESM by one percentage point above our baseline increases the output in the entrepreneurial sector by 6.9% and the wealth Gini by 0.8%.

Related Literature This paper is related to the theoretical study of business transfers that goes back to Holmes and Schmitz Jr (1990). In their paper, the authors shows that business transfers facilitate the division of labor. Recently, David (2017) introduces a model of merger and acquisition linked to the facts documented in Ravenscraft and Scherer (2011) and Lichtenberg et al. (1987). This literature focuses on large businesses and sequential entrepreneurs. In our framework, instead, repeated business owners concern a small sample of entrepreneurs and we focus on the whole population of entrepreneurs, including small businesses. This paper is also related to the extensive literature on SMEs and entrepreneurship with a macroeconomic perspective. This literature generally depicts entrepreneurs as agents adjusting physical capital and hiring employees subject to idiosyncratic business shocks, entrepreneurial abilities, financial frictions, or unexpected capital destruction. Seminal papers in this literature are Quadrini (2000b), Cagetti and De Nardi (2006b), or Buera and Shin (2013): those especially focus on credit constraints and the role of entrepreneurship in shaping the wealth distribution. Along the lines of our paper, Liang et al. (2018) and Engbom (2019) also discuss the relation between age and the decision to enter entrepreneurship. Compared to the above literature, this paper introduces an empirically relevant theoretical framework that accounts for the life-cycle properties of entrepreneurship and the underlying mechanisms of entry and exit while modeling explicitly the market frictions arising upon the transfer of business assets.

Many recent papers highlight the key role of the age of a firm. The argument follows Jovanovic (1982) and Arkolakis et al. (2018): firms acquire knowledge about their environment and learn about the demand addressed to them as they age, which is translated by a higher maturity and a larger stock of intangible assets. For example, among many other studies, Dunne et al. (1988) show that the exit hazard rate decreases with age. In Clementi and Palazzo (2016), this is the case because, on average, entrants are less productive than incumbents. Relatedly, Warusawitharana (2018) shows that profitability evolves with the age of a firm. Moreover, using panel data, Sakai et al. (2010) show that younger small businesses face higher borrowing costs since firms tend to accumulate *reputation* as they age. Dyrda (2015) and Garcia-Macia (2017) show that borrowing constraints faced by entrepreneurs are age-dependent and help to shape the heterogeneous business cycle responses of firms. The older the firm, the less stringent the constraint. The relation between the age of a firm and business performance is modeled, for instance, by Garcia-Macia (2017) and Bhandari and McGrattan (2018) through the accumulation of intangible assets. More generally, compared to the above papers, we explicitly introduce and model the transfer of illiquid business assets. In our case, liquidating a firm lets entrepreneurs recover part of the tangible business assets while selling a (or part of a) business reproduces the transfer of both tangible and intangible assets. Finally, the literature has mainly focused on business transfers through inheritance or gifts, as in Cagetti and De Nardi (2009b). This paper, however, shows that business transfers through a purchase are more common, accounting for over 70% of total business transfers.

The remaining of the paper is organized as follows. Section 3.2 documents empirical elements on business acquisition and transfers, the business for sale market, and the entrepreneurial life-cycle. In Section 3.3, we present our baseline model and Section 3.4 describes how we take the model to the data. We evaluate our model in Section 3.5 and in Section 3.6, we show the importance of the business for sale market. Section 3.7 concludes.

3.2 Business Transfers and Dynamics: Stylized Facts

This section details empirical evidence on business transfers and related entrepreneurs' behavior. Throughout the paper, we gather disparate information from the 2007 Survey of Business Owners (SBO), the 2016 Annual Survey of Entrepreneurs (ASE), and the 2003 Survey of Small Business Finances (SSBF). Evidence are complemented with the Survey of Consumer Finances (SCF), the National Longitudinal Survey of Youth 1979 (NLYS79) and the Panel Study of Income Dynamics (PSID). These datasets provide broad pictures of firm characteristics by acquisition type, and characteristics of purchasers with respect to founders. For consistency, an entrepreneur is defined as an active self-employed business owner whenever possible and as a self-employed business owner otherwise. In the SBO, we additionally focus on individuals declaring that their businesses constitute their primary source of income. As ASE microdata are not publicly available, we report macro estimates for all business owners at least one paid-employee. Finally, unique information on the business for sale markets transaction data are obtained from *bizbuysell* (hereafter BBS) from 2018 to 2019, which is currently the most important online platform to sell a business in the US.[6]

3.2.1 Business Acquisition and Exit

The literature on entrepreneurship has long been interested in the behavior of incumbent entrepreneurs but has been somewhat silent on how businesses come to be in the first place. Throughout this paper, we argue that purchasing and selling a business are important components of entrepreneurship, as evidenced by the behavior of a non-negligible fraction of entrepreneurs in the data. Survey questions define as *acquisition* the way the entrepreneur became the owner of the business: founding a new business or purchasing an existing one. Using the SBO (2007), Table 3.2.1 provides estimates of the types of acquisition: around 20% of all entries into entrepreneurship are the result of the purchase of an existing business, a fact consistent across the SSCF, the SCF and the ASE.[7]. This accounts for 70% of all business transfers, dwarfing gifts and inheritances (see Appendix 3.A.1). Moreover, purchased firms account for a large fraction of total employment and total sales, especially among recently acquired firms.

Concerning the exit out of entrepreneurship, there is little detailed evidence in the literature despite an important body of papers focusing on this subject and its relation to life-cycle aspects. A non-negligible fraction of active business owners sell their firms upon exit: 8% according to the SBO (2007) (varying from 7% for the main owner to 18% for the third and fourth owners and to 16.9% for entrepreneurs with paid employees) and 17% in the 2016 ASE (owners with paid employees). In the NLSY79 (2002-2016), pooling individuals with past ownership, 20% sold their

[6] In appendix 3.A, we compare the representativity of the BBS data to the PSID as well as providing further details regarding all the sample selections.

[7] Interestingly, purchasers are not more likely than founders to have a previous self-employment experience in the SBO (2007) and the SSBF (2003) (see Appendix 3.A.2 for details).

Table 3.2.1. Business acquisition by type [a]

GROUP	ACQUISITION TYPE [b]	METRIC		
		(%) Firms	(%) Employment	(%) Total sales
Of all firms	Purchased	19.6	39.6	42.7
	Founded	80.3	60.4	57.3
Of firms within 3 years of acquisition	Purchased	20.5	60.6	61.4
	Founded	79.5	39.4	38.6

[a] Survey of Business Owners (2007). An entrepreneur is defined as an individual declaring that her business constitutes her primary source of income with an active management role, whenever possible.
[b] These numbers exclude inheritance and gifts as they account for a minor fraction of reported transmissions. These results hold even when only the main/first owner is considered and for the pool of firms with employees.

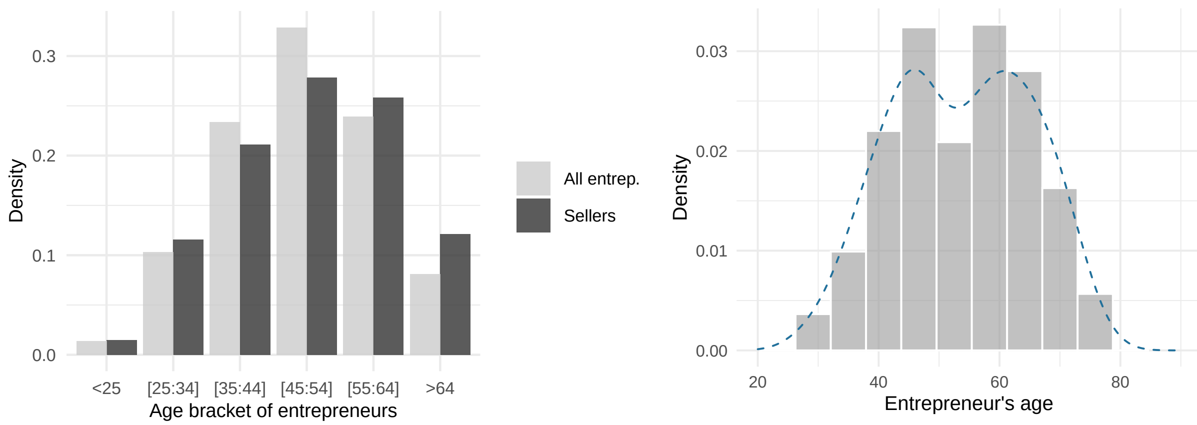
businesses, 70% shut them down and the remaining were in an undefined alternative situation.[8] This selling behavior is largely related to the entrepreneurial age profile and the decision to retire. Using SBO data, we first show in panel (a) of Figure 3.2.1 that the age distribution of sellers is further to the right with respect to the full population of entrepreneurs: 38% of sellers are over 54 years old (in contrast, buyers are relatively young, with a mean age of 44 in the SCF (2007) and 44.5 in the SSBF (2003)). In panel (b) of Figure 3.2.1, we corroborate this evidence using PSID data: the sale of business assets peaks at two age brackets: the 45-50 and the 60-65, close to the typical U.S. average retirement age. This corroborates the fact that retirement is one of the main reasons to cease a business. In the ASE and BBS, 19% to 24% of businesses ceasing and for sale were explained by owners retiring.

Finally, there are substantial difficulties for transferring businesses on the business for sale market (BSM). According to the 2016 ASE, among business owners with paid employees reporting how they planned to exit entrepreneurship, 50% were thinking of selling their businesses to a third-party and 10% to a family member. This is in stark contrast to the much lower number of businesses actually being sold that we report above. Moreover, among business owners with a firm of 16 years of age and more, i.e. businesses much less likely to close due to economic reasons, the main exit strategy is consistently the sale of the entire business (53%). However, even in this population, only 26% declared effectively selling their business *ex post* while 43% declared just retiring and 14% declared failing due to business conditions. To this, our BBS data offer the first direct evidence that selling a process is a long and uncertain process: only 30% of businesses for sale are sold within a year and a non-negligible fraction remain unsold. This shows that there are potentially important transaction delays and frictions (training, screening, the existence of

[8] Note that the NLSY79 included this question only after 2002. One explanation of the gap is that the SBO provides many different options to choose from for the main reason to cease. In contrast, the NLSY79 only offers three options: *selling the business*, *shutting it down* or *other*. For instance, it might be possible that retiring owners in the SBO reported *retirement* as the reason to cease even if the means of exit was selling of the business. Moreover, the SBO treats businesses and business owners differently. Therefore, it is possible that owners exit by selling their shares, while the associated businesses keep on operating. Finally, Appendix 3.A.3 provides further evidence on the exit rate, especially by type of exit.

Figure 3.2.1. Entrepreneurial life cycle, acquisition and business selling

(a) Age distribution of entrepreneurs and sellers, SBO **(b)** Sold business assets by age, PSID



Source: SBO 2007 and PSID averaged over the waves from 1990 to 2015 (adjusted for inflation using the CPI index). The mean age of the distribution is 53.6 and the median is 54.

asymmetric information, etc.) preventing business transfers.

This paper provides two plausible explanations that could generate this *low* observed selling rate. First, selling a business requires the matching of the specific interests and skills of a potential buyer. We refer to this as *selling friction*. Second, purchasing a business requires the payment of physical capital and the value of intangible assets. Borrowing constraints could substantially limit the capacity of potential buyers to purchase existing businesses beyond their traditional effects described in the literature, for instance in Quadrini (1999, 2000b) among others. This is for instance consistent with the fact that the fraction of new purchase among new business owners is increasing with wealth in the SCF.

3.2.2 Sources of Heterogeneity and Maturity of a Business

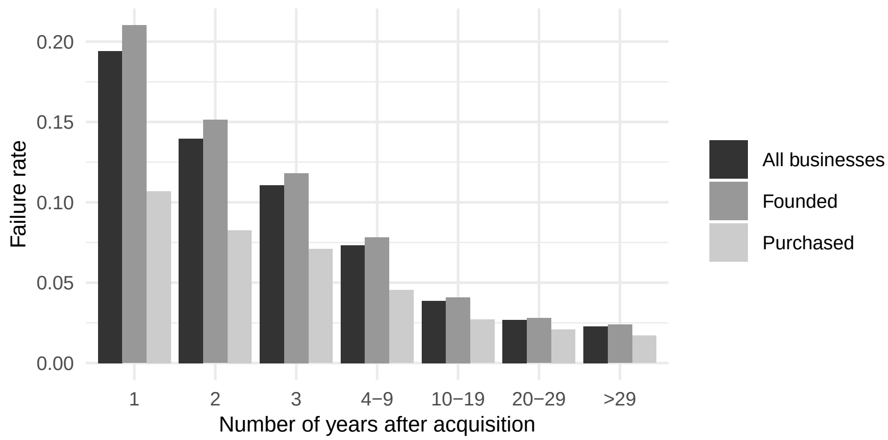
In a seminal theoretical paper, Holmes and Schmitz Jr (1990) show that in a cohort of new businesses developed at a certain date, those that are subsequently involved in a transfer will on average be of higher "quality" and also survive longer than those that are not transferred. In this section, we empirically confirm that transferred businesses (through a purchase) and founded firms differ with respect to the main sources of heterogeneity that the entrepreneurial literature generally considers, namely, heterogeneity in failure risk, credit conditions, and profitability.

In Figure 3.2.2, we show the failure rate of firms in the 2007 SBO by the number of years after acquisition.[9] We observe a stark performance divide between founded and purchased firms benefiting the latter: for instance, one year after the acquisition, the failure rate of newly founded businesses is twice that of purchased ones. As the number of years after acquisition increases, the

[9] The literature sometimes uses the term *exit hazard rate* to cover most of what we call *failure rate*. As in Dunne et al. (1988), we define the failure rate as the ratio of firms exiting due to economic reasons between period t and $t+1$ and the number of operating firms in time t. Results are robust to the use of all exiting reasons.

failure rate decreases and the difference between purchased and founded businesses dissipates. Interestingly, there is still a decreasing shape with respect to age for recently purchased firms, which might be linked to non-transferable intangible assets such as an entrepreneur's talent and knowledge about its environment.[10] In more general terms, without any controls, we find that purchased businesses systematically perform better with respect to all sources of firm heterogeneity that we consider.

Figure 3.2.2. Failure rate by acquisition type and for all businesses.



Source: author's computation using the 2007 SBO. We compute the failure rates using ceasing option linked to either inadequate cash-flows or low sales and lack of business or personal loans/credit.

Using SSBF and SBO data and controlling for characteristics of firms and owners, we estimate the average difference distinguishing recently purchased and founded firms over five components: the failure rate, the credit line interest rate, the credit limit, the credit score, and the profit normalized by the average 2-digits sectoral profit.[11] We report the results in Table 3.2.2. The fourth and fifth columns, under the label Δ *Conditional*, display the conditional difference in the specific components. As a reference, the two first columns display the uncontrolled sample average, respectively for purchased and founded firms, while the third column is simply the difference between those two values. We compare our results between a pool of recent firms within 3 years of their acquisition and one of older firms over 15 years of theirs.

Starting with the failure rate, our estimates using the SBO show that purchased firms are a significant 6.3% less likely to fail. Importantly, contrarily to many surveys considering the acquisition and establishment date as equivalent, the SBO contains information about the true establishment year of a business.[12] Although imperfect, the establishment year lets us control for the contribu-

[10] Guiso et al. (Forthcoming) show that that type entrepreneurial knowledge is important for the decision to enter the sector.

[11] Specifically, we use the following OLS regression: $C_i = \alpha + \beta D_i(\text{purchased}) + \gamma \mathbf{X}_i + \epsilon_i$ with $D_i(\text{purchased})$ a dummy indicating whether the business was initially purchased, C_i the specific component and $\mathbf{X}$ the vector of controls. β that captures the difference associated with purchasing relative to founding a businesses.

[12] This information is available yearly between 2003 to 2007 and is bracketed prior to that as [2000:2002], [1990:1999],

tion of a firm established many years prior to its sale with respect to a recently established and sold firm. Controlling for the establishment age, the associated failure rate wedge falls to 2.4% (last column), implying that this age captures (part of) the difference between a founded and a purchased firm. A key consequence of this is that the true age of a firm appears to be a critical factor. Quite contrastingly, in the pool of older firms, the wedge between purchased and founded businesses is virtually negligible, and the effect of the establishment date also disappears. Overall, our result points out that recently purchased firms are less likely to fail than recently founded firms. Consistently with Dunne et al. (1988), this difference is partly captured by the fact that purchased firms are in general older (conditional on size, and other characteristics). In a model of firm dynamics with an entry margin, older firms are those that survived and were selected over time. Following Jovanovic (1982), we argue that an entrepreneur learns the characteristics and the potential of a new firm mostly *on the business* but directly observes the characteristics of older (and purchasable) firms.

Table 3.2.2. Key Heterogeneity Components by Acquisition Type

	Acquisition Type		Δ Unconditional	Δ Conditional	
	Purchased	Founded		Controls[a]	+Estab. date
Firms within 3 years of acquisition					
Risk: Failure rate	0.087	0.164	−0.077	−0.063*** (0.004)	-0.024*** (0.005)
Financial: Int. rate	10.89	12.12	−1.234	−1.98** (0.831)	
Financial: $\frac{\text{Granted loan}}{\text{Applied for}}$	0.994	0.950	0.043	0.061** (0.029)	
Financial: Credit score	3.316	2.964	0.352	0.526***(0.173)	
Norm. Profitability	0.710	0.364	0.346	0.431* (0.232)	
Firms over 15 years after acquisition					
Risk: Failure rate	0.023	0.033	−0.010	−0.005*** (0.001)	−0.002*** (0.001)
Financial: Int. rate	12.29	12.38	-0.081	0.229 (0.548)	
Financial: $\frac{\text{Granted loan}}{\text{Applied for}}$	0.975	0.975	-0.002	0.006 (0.019)	
Financial: Credit score	4.024	3.913	0.111	0.155 (0.143)	
Norm. Profitability	1.616	1.176	0.420	0.030 (0.326)	

Note: *p<0.1; **p<0.05; ***p<0.01. Source: SBO, SSBF (2003). The sample comprises owners managing their firms and declaring they constitutes their primary source of income. The first owner is used to select the firm acquisition year.
[a] Firm and owner controls: depending on the survey, characteristics of owners include age, experience as entrepreneur, education, sex, home-based dummy, home equity and other net worth, and number of owners. Characteristics of firms: sector and FIPS dummies, employment, legal structure, equity, past bankruptcy indicator, urban dummy, payroll and franchise indicator.

Our findings for the failure rate extends to financial and productivity components and corroborates the findings of Holmes and Schmitz Jr (1990) that purchased business are, on average, of better quality.[13] In the SSBF data, recently founded businesses tend to pay, on average, a higher interest rate on their credit line with respect to recently purchased businesses. The premium is non-negligible at around 2.0%. Concerning the credit limit, we find that recently purchased businesses obtain a higher fraction of the loan they applied for. This is also confirmed when using

[1980:1989], and [before 1980].

[13]However, this paper mainly focus on large businesses that eventually lead to M&A.

the credit score as a proxy for credit constraints: recently purchased businesses get a significantly higher score, of 0.5 on average, on a scale from 1 (lowest score) to 6 (highest score). These findings suggest that founded businesses face tighter financial constraints, consistently with Sakai et al. (2010) and Dyrda (2015).[14] Finally, regarding productivity, the ratio of profit relative to the average profit in the corresponding industry is higher by about 0.4 for purchased firms. Consistently, in the older pool of firms, the gap between purchased and founded firms either vanishes or is not significant for financial and productivity components. All those effects can be driven by two key mechanisms. First, it is consistent with a *mean-reverting productivity* in which new firms enter with an, on average, lower productivity level and then converge to their long-run productivity level as they age.[15] Second, it is consistent with the selection of the best firms through time: if purchased firms are in general older, they have already undergone the selection process that occurs at the early-stage. In general, it is difficult to distinguish between the two mechanisms without a full panel characterization of a sample of recently acquired firms. Recent evidence by Alon et al. (2018), however, tend to support that most of the cross-sectional patterns between age and performance is driven by selection.

By distinguishing recently purchased from founded businesses, our results show that maturity components are crucial and allow potential purchasers to overcome the selection process. We argue that our findings concerning the performance divide between purchased and founded firms for a prospective entrepreneur is a key element to incorporate in a model examining business assets transfers. We relate these findings to a growing literature documenting the importance of intangible assets (customer bases, client lists, organizational methods, brand value, etc.) in explaining the market value of a firm. However, as evidenced by Bhandari and McGrattan (2018), the direct measurement of most intangible assets is a difficult task. But basically, as they cannot be directly bought and their accumulation is time-intensive, early-stage firms are immature with low levels of intangibles whereas well-established ones are mature with high levels of intangibles.[16], We, here, adopt a parsimonious approach supported by our empirical evidence: an immature firm will face more stringent credit, productivity and risk conditions whereas a mature firm will have better perspectives on these components. From the point of view of a prospective entrepreneur, a business purchase will convey the specific value of maturity and the business for sale market will play a crucial role in transferring it between owners.[17]

[14]As additional evidence, in both the SCF and the SSBF, the main reason given to explain why early-stage businesses face difficulties in obtaining the credit they apply for is *"the firm was not in business long enough"* (see Appendix 3.A.2 for details).

[15]Unfortunately, the SSBF data does not include the establishment date of firms, making it impossible to control by this element for financial and profit components.

[16]Bhandari and McGrattan (2018) also discuss the dynamics of intangible assets accumulation and show how it is related to the age of a firm. The relation with age is also consistent with the literature on firm dynamics (see Dunne et al. (1988)).

[17]This paper focuses solely on transferable assets, whether tangible or intangible. We abstract from non-transferable intangible asset, such as the managerial value of a specific retiring business owner.

3.3 Model

The economy consists of a corporate sector and a unit measure of *ex post* heterogeneous agents who are either entrepreneurs or workers. The former hold small and medium-sized businesses while the latter occupy wage-paying jobs in a corporate sector. Individuals who enter the SME sector have to either found a new business or purchase an existing one. Upon exit, entrepreneurs can either sell their business or liquidate the physical business assets. Purchasing and selling are subject to financial and selling frictions or the *SME for sale* market (hereafter *SMESM*). Finally, a government levies a menu of taxes to cover for old-age pensions and other public expenditures.

3.3.1 Agents

Households live through J stages of life and the total population, of unit mass, is divided among J generations indexed with $j \in [1; J]$. Groups 1 through $J-1$ are called Juniors and have access to the labor market. The J^{th} group, called Seniors, is comprised of individuals beyond the retirement age. Each period, a fraction $p_d(j_t)$ of individuals pass away and exit the model. Over the life-cycle, households belong in an occupation $o \in \{o_e, o_w, o_r\}$. Junior households can be entrepreneurs (o_e) or occupied in the workforce (o_w) whereas Senior households are either retired (o_r) or are old age entrepreneurs.

Households have preference described by the life-time utility

$$E_0 \sum_{t=0}^{\infty} [1 - p_d(j_t)]\, \beta^t \mathcal{U}(c_t, j_t, o_t), \tag{3.1}$$

where we drop the dependence on time t in the following. The age and occupation argument in the utility translates the fact that being active beyond the retirement age or being entrepreneur might generate respectively some disutility costs or non-pecuniary benefits (Hurst and Pugsley, 2015).

Depending on its occupation, a household can possess liquid savings, a, and/or illiquid (business) assets, k, that are used to produce with the entrepreneurial technology to produce the homogeneous consumption good. The liquid asset can be freely used to purchase it but not the illiquid asset. To obtain liquid assets from illiquid assets, individuals have to either sell their firm contingent on finding a buyer or liquidate partially or totally subject to an adjustment cost. Conversely, acquiring illiquid capital using liquid capital is subject to an adjustment cost but can be also achieved by buying a firm with a specific illiquid capital amount.

The state-space for an entrepreneur are savings a, illiquid business capital k, and $\mathbf{x}_e = \{j, m\}$, where $m = \{0, 1\}$ indicates whether the business is mature. A newly founded firm is assumed to start immature ($m = 0$) and has a probability P_m to mature. Only mature firms can be sold

Figure 3.3.1. Timing.
[width=]media/timingtikz

on the SMESM. Thus, all purchased firms are preexisting mature businesses ($m = 1$).[18] Maturity translates the accumulation of intangible assets and provides specific benefits. Entrepreneurs are precluded from possessing multiple firms. The state-space for a worker is a, and $\mathbf{x}_w = \{j, y, \iota\}$, with y her working productivity and ι her potential ability to manage a business. Both ι and y follow first-order Markov processes. We note $\mathcal{Y}(j, y)$ the income of a worker. The entrepreneurial income derives from entrepreneurial production using technology $f(k, m)$.[19]

For convenience, the full individual states vector is noted $\mathbf{X} := (a, k, j, m, \iota, y, o) \in \mathbb{X}$. The states of an entrepreneur are $\mathbf{X}^e := (a, k, j, m) \in \mathbb{X}^e$ and those of a worker are $\mathbf{X}^w := (a, y, j, \iota) \in \mathbb{X}^w$. Let $\{\Phi(\mathbf{X}), \Phi(\mathbf{X}^e), \Phi(\mathbf{X}^w)\}$ be measures over all agents and each occupation respectively.

3.3.2 Dynamic Problem

We decompose and solve backward the agent's intra-period decision process into a sequence of three subperiods. In the last subperiod, the consumption-saving and entrepreneurial investment problems are tackled. In the middle subperiod, the buying and selling problems are addressed contingent on occupational changes and the maturity of a business. Finally, in the first subperiod occupational choices are made. Given that $\mathcal{W}(a, \boldsymbol{x}_w)$ and $E(a, k, \boldsymbol{x}_e)$ are respectively the value function of a worker and an entrepreneur, Figure 3.3.1 summarizes this decomposition.

The Last Subperiod: Consumption-Saving Problem

Consumption and saving decisions in the last subperiod can be distinguished into those of workers either continuing or exiting their activity and those of entrepreneurs continuing or exiting theirs. For the sake of simplicity, continuing workers can not borrow. Similarly to an incumbent entrepreneur, an exiting worker entering an entrepreneurial activity can borrow to invest in a level of business assets k, as long as a minimum amount θ is pledged. Thus those individuals are subject to the following borrowing constraint:

$$a' \geq -(1-\theta)\mathcal{P}(k, m) \tag{3.2}$$

[18]In detailed business transaction data, we find that only 15% of businesses for sale have been established in the 5 preceding years. In the SBO 2007, only about 23% of sellers declare selling a firm established in the 4 years preceding the sale. In pratice, it would be straightforward to allow a fraction of firms to mature directly upon entry. We believe our results are robust to such an assumption.

[19]Note that for the sake of parsimony, we make a number of simplifying assumptions. First, we abstract from entrepreneurs hiring workers and leave it for a straightforward extension. Second, the fact that only mature firms can be sold on the *SMESM* reflects that in the data the average age of businesses for sale is much higher than the average of all firms. Using a dataset of business selling transactions detailed in Appendix 3.A.4, we find that 85% of businesses for sale are older than 5 years. Moreover, as it takes time to create a valuable business, maturity captures the process of accumulating intangibles for early-stage firms.

where $\mathcal{C}(k(1-\delta),0)$ refers to the liquidation cost (i.e. the adjustment cost from reducing the capital level from $(1-\delta)k$ to 0). This borrowing constraint tilted to the value of the intangible value of a firm (future cash-flow, for instance) is in line with recent evidence in Lian and Ma (2021). We interpret $\theta(m)$ as a maturity-specific downpayment requirement translating the minimum fraction of business assets an entrepreneur has to provide in order to get a loan. This formulation implicitly assumes that, under a liquidation procedure, the creditor can only resell the business assets net of depreciation δ. An indebted entrepreneur faces an interest $r_b(m)$ that depends on her net worth, a, and the maturity of her business, m, such that $\tilde{r}(m) = \mathbb{1}_{a'\geq 0} r - \mathbb{1}_{a'<0} r_b(m)$.

Continuing entrepreneurs An incumbent entrepreneur continuing her activity chooses next period's illiquid capital k' and saving a' given her current income $f(k,m)$. The consumption-saving problem of this entrepreneur is thus:

$$E^{cont}(a,k,\mathbf{x}_e) = \max_{c>0,\ a'\geq -\psi(k'),\ k'\geq 0} \left\{ \mathcal{U}(c,j,o_e) + \beta \mathbb{E}_{j',m'|j,m} E(a',k',\mathbf{x}_e') \right\} \tag{3.3}$$

$$s.t. \quad c + a' + k' = (1+\tilde{r}(m))a + f(k,m) + k(1-\delta) - \mathcal{C}(k(1-\delta),k') \tag{3.4}$$

with E^{cont} the subperiod specific value function of this continuing entrepreneur and τ_w the tax rate on entrepreneurial income.

Exiting entrepreneurs When exiting, an entrepreneur has to choose savings a' subject to the no-borrowing constraint. The value function of an exiting entrepreneur depends on the exit option z: voluntarily or business failure liquidation ($z=0$) or sale of the business ($z=1$).

$$E_z^{exit}(a,k,\mathbf{x}_e) = \max_{c>0,\ a'\geq 0} \left\{ \mathcal{U}(c,j,o_e) + \beta \mathbb{E}_{j',\iota'|j} \mathcal{W}(a',\widetilde{\mathbf{x}}_w') \right\} \tag{3.5}$$

$$\begin{aligned} s.t. \quad c + a' &= (1+\tilde{r}(m))a + f(k,m) \\ &+ (1-z)\underbrace{\left[k(1-\delta)(1-\mathcal{C}(k(1-\delta),0))\right]}_{\text{Business liquidation}} + z\underbrace{\left(\mathcal{P}(k(1-\delta))\right)}_{\text{Business sale}} \end{aligned} \tag{3.6}$$

with E^{exit} the subperiod specific value function of this entrepreneur and $\widetilde{\mathbf{x}}_w'$ the specific exogenous worker state of an exiting entrepreneur.[20] Liquidating is identical to adjusting the business capital to zero by fully paying the corresponding adjustment cost $\mathcal{C}(k(1-\delta),0)$. Alternatively, by successfully selling the business the entrepreneur recovers the total amount $\mathcal{P}(k(1-\delta))$.

[20]The main specificity is the assumption that any new worker coming from the entrepreneurial sector starts with the lowest level of worker productivity. The argument is that the productivity state y is strongly related to the experience of a worker in a specific corporate job. This seniority on a job cannot be randomly obtained but has to be *earned*. Recall, however, that there is an age-component in the determination of the wage process. Finally, the entrepreneurial ability of a new worker is drawn from the invariant distribution of the associated process.

Continuing workers Such a worker has to choose savings a' subject to the no-borrowing constraint and solves:

$$\mathcal{W}^{cont}(a,\mathbf{x}_w) = \max_{c>0,\ a'\geq 0} \left\{ \mathcal{U}(c,j,o_w) + \beta \mathbb{E}_{j',y',\iota'|j,y,\iota} \mathcal{W}(a',\mathbf{x}'_w) \right\} \tag{3.7}$$

$$s.t. \quad c + a' = (1+r)a + \mathcal{Y}(j,y)(1-\tau_w) \tag{3.8}$$

with $\mathcal{W}^{cont}$ the subperiod specific value function of this worker and τ_w the tax on labor income.

Exiting workers An exiting worker enters entrepreneurship by either purchasing an existing mature business ($d = 1$ and $m' = 1$) and paying the total amount $\mathcal{P}(k')$ plus fixed cost F_b or by founding a new business ($d = 0$ and $m' = 0$) and paying the adjustment cost $\mathcal{C}(0,k')$.[21] Depending on whether the agent is currently buying ($d = 1$) or founding ($d = 0$) a firm, her problem is to choose the optimal next period capital size k', savings a' and consumption c. Such a worker solves:

$$\mathcal{W}_d^{exit}(a,\mathbf{x}_w) = \max_{c>0,\ a'\geq -\psi(k'),k'\geq 0} \left\{ \mathcal{U}(c,j,o_w) + \beta \mathbb{E}_{j'|j} E(a',k',\mathbf{x}'_e) \right\} \tag{3.9}$$

$$s.t. \quad c + a' = (1+r)a + \mathcal{Y}(j,y)(1-\tau_w) - d\underbrace{\left(\mathcal{P}(k') + F_b\right)}_{\text{Purchasing}} - (1-d)\underbrace{k'(1+\mathcal{C}(0,k'))}_{\text{Founding}} \tag{3.10}$$

with $\mathcal{W}^{exit}$ the subperiod specific value function.

The Middle Subperiod: Acquisition and Selling Problems

In the middle subperiod, the buying/founding and selling/liquidating problems are solved. When the sale of a business is unsuccessful, entrepreneurs can continue operating the business or may liquidate. Similarly, in the case of an unsuccessful business purchase, a buyer can remain a worker or found a new business.

The seller's problem An entrepreneur with a mature business ($m = 1$) can try to sell ($z = 1$) it on the *SMESM*. A buyer is found with probability h_s. Otherwise, the entrepreneur chooses whether to liquidate ($z = 0$) or to continue operating the business. Depending on whether the entrepreneur exits endogenously or is forced to exit, the following problem is solved:[22]

$$\mathcal{S}^{ee}(a,k,\mathbf{x}_e) = \left(h_s E^{exit}_{z=1} + (1-h_s) \max\left\{ E^{exit}_{z=0}, E^{cont} \right\} \right)(a,k,\mathbf{x}_e) \qquad \text{(endogenous exit)} \tag{3.11}$$

$$\mathcal{S}^{fe}(a,k,\mathbf{x}_e) = \left(h_s E^{exit}_{z=1} + (1-h_s) E^{exit}_{z=0} \right)(a,k,\mathbf{x}_e) \qquad \text{(forced exit)} \tag{3.12}$$

[21] We use the purchase specific fixed cost to bring the model closer to the data by capturing the fact that purchased firms are twice as large as founded ones in terms of start-up capital. A side effect is that buyers are prevented from purchasing very small businesses unless their credit constraint can afford this cost. The fixed cost could capture costs associated to brokerage, screening, negociation or training.

[22] We introduce exogenous shocks to capture entrepreneurial exits unrelated to business failure: migration, death, divorce, etc.

where $\mathcal{S}^{ee}$ and $\mathcal{S}^{fe}$ are the subperiod specific value functions for the endogenous exit and the forced exit problems.

The buyer's problem A buyer has a probability h_b of finding a seller and purchasing a business ($d = 1$). Otherwise, she chooses whether to found a new business ($d = 0$) or to keep being a worker. Thus, the following problem is solved:

$$\mathcal{B}(a, \mathbf{x}_w) = \Big(h_b \mathcal{W}^{exit}_{d=1} + (1 - h_b) \max\left\{\mathcal{W}^{exit}_{d=0}, \mathcal{W}^{cont}\right\}\Big)(a, \mathbf{x}_w) \tag{3.13}$$

with $\mathcal{B}(a, \mathbf{x}_w)$ the subperiod specific value function for this problem.

The First Subperiod: Occupational Choice and Exit Strategy

Worker A worker starts the period with states $\{a, \mathbf{x}_w\}$ and, provided she has an entrepreneurial ability (i.e. $\iota = 1$), chooses whether to try to purchase an existing business (with value $\mathcal{B}(a, \mathbf{x}_w)$), to found a new business ($d = 0$), or to remain a worker, such that:

$$\mathcal{W}(a, \mathbf{x}_w) = \Big((1 - \iota)\mathcal{W}^{cont} + \iota \max\left\{\mathcal{B}, \mathcal{W}^{exit}_{d=0}, \mathcal{W}^{cont}\right\}\Big)(a, \mathbf{x}_w) \tag{3.14}$$

Entrepreneur An entrepreneur starts the period with states $\{a, k, \mathbf{x}_e\}$ and decides whether to sell, liquidate or continue her business endogenously unless she is forced to exit. $\chi(m)$ is the probability of entrepreneurial exit due to business failure, which is a function of the maturity of the business, and ζ is the exogenous exit probability, conditional on not failing. Only businesses that do not fail can be sold. In the end, the following problem is solved:

$$\begin{aligned} E(a, k, \mathbf{x}_e) = \Big(& \underbrace{\chi(m) E^{exit}_{z=0}}_{\text{Failure}} + (1 - \chi(m)) \Big[\zeta \underbrace{\left(m \max\{\mathcal{S}^{fe}, E^{exit}_{z=0}\} + (1 - m) E^{exit}_{z=0}\right)}_{\text{Exogenous exit}} \\ & + (1 - \zeta) \underbrace{\left(m \max\{\mathcal{S}^{ee}, E^{exit}_{z=0}, E^{cont}\} + (1 - m) \max\{E^{exit}_{z=0}, E^{cont}\}\right)}_{\text{Endogenous exit/continue decision}}\Big]\Big)(a, k, \mathbf{x}_e) \end{aligned} \tag{3.15}$$

Contingent on the entrepreneur not failing (with probability $(1 - \chi(m))$), she has a probability ζ to be forced to exit, and a probability $(1 - \zeta)$ to choose whether to stay entrepreneur, liquidate the business or sell the business if the business is mature ($m = 1$).

3.3.3 Corporate Sector

The corporate sector output Y_t is produced by a single competitive representative firm using a Cobb-Douglas technology with capital share $\alpha \in (0, 1)$ and total factor productivity A, capital level $K_{c,t}$ and labor $L_{c,t}$, such that: $Y_t = F(K_{c,t}, L_{c,t}) = AK^{\alpha}_{c,t} L^{1-\alpha}_{c,t}$. Capital depreciates at rate δ

in both the corporate and the SME sectors. The interest rate and the wage rate equalize their respective marginal products: $r_t = F'_{K_{c,t}}(K_{c,t}, L_{c,t})$ and $w_t = F'_{L_{c,t}}(K_{c,t}, L_{c,t})$.

3.3.4 The Small and Medium Sized Enterprises for Sale Market (SMESM)

On the *SMESM*, businesses sellers and buyers meet in a frictional decentralized market where transaction failures may result in business liquidation on the side of sellers and, on the other side, may compel prospective entrepreneurs into founding their businesses. For tractability, we make a number of assumptions:

Only intangible assets are acquired on the SMESM market. Physical capital are priced at their face value (discussed below). The market clearing condition on the *SMESM* imposes the following condition:

$$\int_{\mathbf{X}^e} z(\mathbf{X}^e) h_s m(\mathbf{X}^e) d\Phi(\mathbf{X}^e) = \int_{\mathbf{X}^w} d(\mathbf{X}^w) h_b m(\mathbf{X}^w) d\Phi(\mathbf{X}^w) \tag{3.16}$$

where the price p_m adjust to satisfy the condition.

There is a fraction ωY of physical capital produced by the corporate sector, and which is sold on the capital market together with sellers selling their physical capital, such that:

$$Y_K + \int_{\mathbf{X}^e} z(\mathbf{X}^e) h_s m(\mathbf{X}^e) d\Phi(\mathbf{X}^e) = \int_{\mathbf{X}^w} d(\mathbf{X}^w) h_b m(\mathbf{X}^w) d\Phi(\mathbf{X}^w) \tag{3.17}$$

With these assumption, we avoid the challenging multidimensional dynamic sorting problem of the direct matching between heterogeneous buyers and sellers, which may require each individual to forecast the dynamics of the entire distribution of sellers and purchasers. Instead, with the above assumptions, the equilibrium condition requires the price p to clear the exchange of maturity value on the SMESM:

$$\begin{aligned} &\int_{\mathbf{X}^e} z(\mathbf{X}^e) h_s \big[\mathcal{V}\big(k(\mathbf{X}^e), 1\big) - \mathcal{V}\big(k(\mathbf{X}^e), 0\big)\big] d\Phi(\mathbf{X}^e) \\ &\quad = \int_{\mathbf{X}^w} d(\mathbf{X}^w) h_b \big[\mathcal{V}\big(k'(\mathbf{X}^w), 1\big) - \mathcal{V}\big(k'(\mathbf{X}^w), 0\big)\big] d\Phi(\mathbf{X}^w) \end{aligned} \tag{3.18}$$

with h_s and h_b the respective frictions (or mismatch probability) on the sellers' and buyers' side of the market.

In this specification, cash flow units are indistinguishable. Therefore, selling a firm is here consistent with providing to the market all cash flow units and tangible business assets owned by the entrepreneur at the same time. Conversely, buying a firm is equivalent to collecting available cash flow units until the endogenously decided capital size k is attained and then paying the total price $\mathcal{P}(k)$. We argue that a number of elements support this specification. First, it lets us recover in a stylized manner that entire businesses are exchanged without changing global value. Second, it stresses that businesses can be bought not only by a single individual but by several individuals associated together.

These assumptions let us capture the fact that selling a business cannot be reduced to selling only its tangible assets. Instead, the value recovered after a transaction should cover the discounted value of future profits.[23] We convey this idea here through the fact that the price p is an abstract object. It is determined at the global equilibrium between cash flow units sold and bought translating at the same time the intertemporal (since holding businesses provide an expected stream of future profits) and intangible (with maturity affecting the relative value of buying versus founding) values of a business.

Finally, this pricing specification ensures that the value associated with selling a business is always higher than the liquidation value. A price $p = 0$ would mean that businesses are sold at their liquidation value and that the market does not price maturity. Consequently, in equilibrium, it should be that $p > 0$ whenever the mass of sellers and buyers is positive.

3.3.5 Demography and Bequest

The model features multiple generations of individuals. An individual in the last age bracket has a probability p_{die} to die. In such a case, the individual is assumed to be reborn as a worker with age $j = 0$, with the ownership of the net of estate taxation bequest. Estate taxation is defined by the tax rate τ_a on every unit of a bequest left to the descendant. When an entrepreneur dies, we assume that the business is liquidated and that the debt is reimbursed. The remaining becomes initial wealth for the newly born worker.[24]

3.3.6 Government

The government collects revenues from labor income taxes and pensions (defined as the amount $\mathcal{Y}(J, y)$), as well as from estate taxation and taxes on the sale of businesses. Government expenditures comprise an exogenous government spending proportional to aggregate output, $G = \overline{G}Y$ and pensions. The government budget constraint is:

$$\int_{\mathbf{X}_w} \Big(\mathcal{Y}(y,j)\tau_w + \int_{\mathbf{X}^e} \Big(z(\mathbf{X}^e)p\big[\mathcal{V}((1-\delta)k(\mathbf{X}^e)) - \mathcal{V}((1-\delta)k(\mathbf{X}^e),0)\big]\tau_s\Big)\, d\Phi(\mathbf{X}^e)\Big) + \int_{\mathbf{X}} \mathbb{1}_{j=J}\, p_{die}\tau_a a(\mathbf{X})\, d\Phi(\mathbf{X}) = \overline{G}Y + \int_{\mathbf{X}^w} \mathcal{Y}(J,y)\, d\Phi(\mathbf{X}^w) \tag{3.19}$$

[23] Using the ValuSource business for sale transaction data, we estimate a ratio of intangible assets over the business price of about 38% for the median and 54% for the mean. Moreover, Bhandari and McGrattan (2018) find that there is little cross-sectional dispersion in intangible assets valuation, supporting our choice of a single price p for all cash flow units.

[24] We also studied a version with a voluntary bequest motive in which older individuals value the utility of their descendants with a *warm-glow* utility function of the form $\mathcal{V}(a)$. The results are qualitatively similar.

3.3.7 Equilibrium

A recursive competitive equilibrium consists of value functions for entrepreneurs and workers $\left\{E^{cont}(\mathbf{X}^e), E_z^{exit}(\mathbf{X}^e), \mathcal{S}^{ee}(\mathbf{X}^e), \mathcal{S}^{fe}(\mathbf{X}^e), E(\mathbf{X}^e)\right\}$, $\left\{\mathcal{W}^{cont}(\mathbf{X}^w), \mathcal{W}_d^{exit}(\mathbf{X}^w), \mathcal{B}(\mathbf{X}^w), \mathcal{W}(\mathbf{X}^w)\right\}$, decisions rules $\left\{a'(\mathbf{X}), k'(\mathbf{X}), d(\mathbf{X}^w), z(\mathbf{X}^e), c(\mathbf{X}^e)\right\}$ and occupational choices, factor prices $\{w, r\}$, a price p for a unit of business profit, and government spending $\overline{G}$ such that:

1. Household optimize value functions and decision rules by solving problems (3.3)-(3.15).

2. The labor and capital markets clear. Total labor demand by the corporate sector equals household labor supply. The wage is determined by the marginal productivity of labor in the corporate sector, such that $L_c = \int_{\mathbf{X}^w} h(j) y d\Phi(\mathbf{X}^w)$. Corporate capital and the total entrepreneurial capital equate total agent's net worth in the economy: $K_c + \int_{\mathbf{X}^e} k(\mathbf{X}^e) d\Phi(\mathbf{X}^e) = \int_{\mathbf{X}} a(\mathbf{X}) d\Phi(\mathbf{X})$. The interest rate is determined by the marginal productivity of capital in the corporate sector.[25]

3. The government budget constraint in (3.19) is balanced with $\overline{G}$.[26]

4. The *SMESM* clears such that the price p in the pricing function (**??**) equates the value of relative expected cash flow units sold to those bought in equilibrium.

5. The distribution of agents $\Phi(\mathbf{x})$ is induced by decision rules and exogenous shocks and is summarized by the transition matrix of the system $M(\mathbf{X}', \Phi' | \mathbf{X}, \Phi)$. A steady state implies a stationary measure $\Phi(\mathbf{X})$.

This problem has no analytical solution and has to be solved numerically. Two main computational challenges arise. First, the dimensionality of the problem with two assets is large and fast optimization methods are required. Second, due to the presence of both discrete (occupational choices) and continuous choices, first-order conditions are no longer sufficient while still necessary. Our computation strategy follows a version of the Discrete Continuous Endogenous Grid Method (DC-EGM) developed in Iskhakov et al. (2017) with taste shocks to smooth kinks. Our specific algorithm is discussed in Appendix 3.B.1.

3.4 Parameterization

We parameterize the model to match microdata on the purchasing and selling margins, occupational choices, and life-cycle patterns. We compute the moments using the Current Population Survey (CPS) averaged from 2000 to 2008, the SCF averaged over the 2001, 2004 and 2007 waves,

[25] By a no arbitrage condition factor prices are identical in the entrepreneurial and the corporate sectors.

[26] In the benchmark economy, we set τ_w and let $\overline{G}$ adjust. In counterfactual experiments, we keep $\overline{G}$ to its benchmark value and adjust the tax rate τ_w.

and finally the 2007 SBO. We pin down a number of parameters by normalizing them or by relying on values widely used in the literature. We then jointly set the rest of the parameters to match key moments in the data with their model counterpart.

3.4.1 Fixed Parameters

Demography and preferences We set $J = 9$, with 8 stages to represent adult working life, of 5 years each, and a last bracket to capture all ages beyond the retirement threshold. We use the following utility function:

$$\mathcal{U}(c, j, o) = \frac{(c^{1-\sigma} - 1)}{1 - \sigma} - \mathbb{1}_{j=J} u_R + \mathbb{1}_{o=o_e} u_E \tag{3.20}$$

with relative risk aversion $\sigma = 1.5$ and u_R and u_E are jointly endogenously calibrated.[27] Senior households face an additional utility cost u_R when operating a business, in order to translate the difficulty of still being active in old age.

Earnings and retirement The labor income process is particularly important for the decision to become an entrepreneur as it lets workers accumulate sufficient wealth to run valuable businesses when they are endowed with the entrepreneurial ability.[28]

We define labor earnings as a function of the wage level w, an age-dependent component $h(j)$ and a persistent stochastic process for labor productivity y such that:

$$\log(\mathcal{Y}_{i,t}(j, y)) = \log(w_t) + \log(y_{i,t}) + \log(h_{i,t}(j)) \qquad \forall j \in \{0, ..., J\} \tag{3.21}$$

$$\log(y_{i,t}) = \rho_y \log(y_{i,t-1}) + \epsilon^y_{i,t}; \qquad \epsilon^y_{i,t} \sim \mathcal{N}(0, \sigma_y) \tag{3.22}$$

We discretize the process for y by setting $\rho_y = 0.96$ and adjusting the variance to $\sigma_y = 0.2$ to generate an earnings Gini of 0.36. When $j = J$, $h(j)$ defines the retirement pension that we set to 40% of the average income. Once retired, an individual keeps the same component y forever and her offsprings' productivity is drawn from the invariant distribution.

Otherwise, the components $h(j)$ for $j \in \{1, ..., J-1\}$ are chosen in order to replicate the average lifetime earning profile within each earning percentile as in Guvenen et al. (2015).[29] Additionally, the probability of dying, p_{die}, is set to 0.091 (corresponding to an expected retirement period of 11 years). The benchmark labor tax rate τ_w is set to 0.15.

[27] Since a complete characterization of preference heterogeneity is outside the scope of this paper, we assume a unique IES-risk aversion parameter σ. However, risk aversion has been shown to have a key role on entrepreneurial decisions (see for instance Herranz et al. (2015b)). In our setup, due to maturity-specific risk, high risk aversion individuals would rather purchase than found. We leave this very relevant issue for future research.

[28] Three saving motives arise in the model. A precautionary one due to the inherent productivity risk, a life-cycle one, and an entrepreneurial motive in order to acquire and run a larger, more profitable firm.

[29] We provide the values in Appendix 3.B.2.

Adjustment costs and liquidation value Incumbent and entering business owners pay a cost $\mathcal{C}(k, k')$ to adjust entrepreneurial capital from k to k'. For tractability, we assume those adjustment costs are linear, with ϕ_u the per-unit cost.[30] At the other end, there are transaction costs when selling business assets. In particular, we assume that for each unit of business capital liquidated outside of the SMESM, entrepreneurs recover only a fraction $(1 - \phi_d)$. To summarize, we have:

$$\mathcal{C}(k, k') = \begin{cases} \phi_u(k' - k) & k' > k \\ \phi_d(k - k') & k < k' \end{cases} \tag{3.23}$$

We set ϕ_d to 30%, corresponding to a business capital recovery rate of the of 70%, which is in the range of the average liquidation costs reported in Alderson and Betker (1995). For the sake of parsimony, we normalize ϕ_u to 0.[31]

Business maturity and intangible value This paper quantifies the importance of the SMESM when owners can sell both the physical capital assets and the intangible value of their firm. To quantify the importance of the SMESM, we must consistently match that intangible value, translated in the model by maturity components. An immature business switches from early-stage to mature with a yearly probability of 20% (about 5 years in operation). Maturity allows businesses to be sold and implies some additional benefits: (i) a lower interest rate charged on the debt, (ii) a higher borrowing limit, (iii) higher profitability, and (iv) a lower probability to fail. All these elements have been highlighted in Section 3.2 and we discuss in Section 3.6.2 the importance of each component for the intangible value of a business.

A mature business pays a lower interest rate on its financing, translating the higher amount of information that a creditor has access to (i.e. history of past transactions, client lists, etc.), which is intrinsically part of the intangible business value. We therefore define the debtor interest rate as $r_b(m) = r + v_s + v_m \mathbb{1}_{m=0}$, where v_s is a wedge common to all businesses while v_m is the additional interest rate premium charged on early-stage businesses. We set $v_s = 2\%$, the usual value used in the literature. The wedge charged on immature firms is set to 1.5% in line with our estimates.

The borrowing limit tightness $\theta(m = 1)$ is set to 0.3: entrepreneurs have to provide a down payment of 30%. In total, entrepreneurs can borrow up to $(1 - \theta(0))(1 - \phi_d) = 49\%$ of the business assets k, and therefore provide the remaining 51%, which is close to the 50% assumed in Herranz et al. (2015b).[32] We estimate that for firms within 4 years of their purchase or foundation, the former ones are granted about 3-4 percentage points higher loans than the latter. Accordingly, we choose $\theta(m = 0) = 0.35$, corresponding to a borrowing limit of 0.455%.

[30] Assuming linear cost greatly simplify the complexity of the problem by avoiding to keep track of the past capital level when investing or disinvesting.

[31] We perform sensitivity analysis on the effect of this parameter in Appendix 3.C.

[32] In Appendix 3.C we provide sensitivity analyses on this parameter and show that the model properties are broadly unaffected by reasonable changes to this parameter.

We adopt a conservative 10% profit rate wedge which is lower than our estimate from the data but is closer to the wedge in Clementi and Palazzo (2016) for new entrants relative to old incumbents.[33] We therefore normalize $f(k,0) = \gamma(0)k^{\nu}$ and $f(k,1) = \gamma(1)k^{\nu}$, with $\nu < 1$ and $\gamma(1) = 1.1\gamma(0)$. Parameters ν and $\gamma(0)$ are part of the joint calibration.

Finally, we discussed in Section 3.2 that a key advantage of purchasing an existing business is a substantial reduction in the probability of failure in the first years after the acquisition. We convey this idea in the model by pinning down $\chi(m)$, the probability of failure. In the 2016 ASE, the fraction of early-stage business owners (within 5 years of acquisition) exiting for reasons related only to business conditions account for 50% of total exits.[34] We, therefore, set $\chi(0)$ to 50% of the average exit rate of newly created businesses, the latter being around 20 to 25%. Consequently, $\chi(0) = 0.12$. Then, using the 2007 SBO, we estimate a difference of 7 percentage points in the likelihood to fail of mature firms with respect to early-stage ones for recently acquired businesses. We thus set $\chi(1) = 0.05$. We then endogenously adjust ζ, the probability of exogenous entrepreneurial exit (independently of business maturity), to match a realized entrepreneurial exit rate of 15%.

Matching probabilities Buyers and sellers are subject to selling frictions captured by the respective probabilities of finding a seller (h_b) and a buyer (h_s). Given the scarcity of business transactions data, measuring those probabilities pose a challenge, in particular for small and middle-sized businesses. On the seller side, we circumvent this issue by relying on a new dataset of business selling transactions from a leading U.S. online marketplace. This dataset includes more than 90,000 observations and provides various business-specific characteristics such as age, size, cash-flow, EBITDA, the fraction of fixed assets, the number of employees, etc. In contrast to other business transactions data, we continuously observe businesses for sale and closed transactions over time, allowing us to construct a panel dataset of businesses for sale. We provide a detailed overview of this dataset in Appendix 3.A.4.

We use the above data to infer the probability that a business is sold within a year by constructing a daily panel of businesses for sale between 2018 and 2019. We then construct cohorts of those businesses and compute the total number of sold businesses over time. We exclude from the cohorts all the businesses that were removed from the listings without resulting in a sale.[35] Then, a year after the first listing dates, we compute the fraction of sold businesses relative to the total initial number of businesses for sale within the cohort. The resulting indicator provides the fraction of businesses for sale that is actually sold after a year. Using this indicator, we find that

[33] In Appendix 3.C, we double the wedge associated to this component as a robustness exercise. All the quantitative results are magnified following the larger gap in maturity value between early-stage and mature firms.

[34] The ASE survey question regarding the reasons to cease has multiple choices. This number is the total of *business failure* responses over the total number of responses excluding *other reasons*.

[35] Results are qualitatively similar if we include those businesses, while the magnitude of the probabilities is lowered by around 15-20%. We posit that excluding vanishing businesses from the stock of businesses for sale lets us exclude businesses that are not performing well from the sample.

the average probability of selling a business is about 30% after a year. This probability displays a slight size-dependence. Taking the price as a proxy for size, we find that firms with a listing price below 500K dollars (resp. above 1000K dollars) have a probability of being sold of 35% (resp. 27%) after a year. Therefore, we pick a conservative estimate for the probability of finding a buyer with $h_s = 0.3$.

Other parameters The corporate sector features a constant returns to scale Cobb-Douglas production function with capital share $\alpha = 0.34$. Total factor productivity is normalized to $A = 1$ and capital depreciates at rate $\delta = 0.07$ in both sectors. The estate taxation is set to 30%, consistently with the statutory tax rate in the U.S. and the value used in Cagetti and De Nardi (2009b).

In the U.S., capital gains following the sale of a business are taxed at a statutory tax rate between 0% and 20%. We choose a benchmark tax of 10%. Finally, we calibrate the transition probability of the entrepreneurial ability process ι: we endogenously determine $p_\iota = P(\iota' = 1|\iota = 0)$ and we restrict $P(\iota' = 0|\iota = 1) = \chi(0)$.

Table 3.4.1. Fixed parameters

Parameter	Value	Description
σ	1.5	Risk-aversion coefficient
$\{\delta, \alpha\}$	$\{0.07, 0.34\}$	Depreciation rate, Corporate returns to scale
$\{\rho_y, \sigma_y\}$	$\{0.96, 0.2\}$	Earnings process
$h(j)$	*See Appendix 3.B.2*	Life-cycle earnings
p_{die}	0.091	Probability of dying during retirement
h_s	0.3	Probability of selling the business within a year
ϕ_d	30%	Liquidation recovery rate
$\{\tau_s, \tau_a\}$	$\{10\%, 30\%\}$	Selling and estate tax rates
P_m	20%	Probability of maturing
$\{v_s, v_m\}$	$\{2\%, 1.5\%\}$	Interest rate wedge for immature/mature businesses
$\chi(m)$	$\{12\%, 5\%\}$	Exogenous probability to fail
$\gamma(m = 1)$	$1.1\gamma(m = 0)$	Profitability wedge (10%)

3.4.2 Joint Parameterization

The remaining nine parameters are chosen jointly so that the model matches nine moments of the U.S. economy related to the small business market, entrepreneurship, and the wealth distribution. The discount factor β helps to match a capital-output ratio of 3.2, computed using the Penn World Table 9.1. The probability of being endowed with an entrepreneurial ability p_ι captures the fraction of entrepreneurs in the working-age population, which ranges between 7% to 12% in the data, depending on the survey, the period considered and the definition. We choose a target of 11%. The probability to fail for exogenous reasons ζ helps to match the exit rate of entrepreneurs, which is equal to 15% in the PSID, according to Mankart and Rodano (2015). Entrepreneurial ability scale $\gamma(0)$ helps to match a share of small business GDP of 46%, as reported in Kobe (2012) for 2008, while the return to scale ν helps to match the wealth Gini coefficient of 0.81. The purchasing fixed cost captures the ratio of the mean capital of purchased business relative to founded ones which is 2.2 in the SSBF, and h_b helps to recover a fraction of purchased businesses of 22% (SCF) upon entry.[36] Finally, preference parameters u_E and u_R help to capture the ratio of the median net worth between entrepreneurs and workers of 7.0 which is closely what is observed in the SCF, and the about 5% of entrepreneurs in the last age bracket.

Our model is exactly identified, with nine parameters used to pin down nine moments. The resulting parameter values are reported in Table 3.4.2.

Table 3.4.2. Model parameters calibrated within the model [a]

Description	Symbol	Value	Data	Model	Source/Main moment [c]
Discount factor	β	0.910	3.20	3.24	Capital-output ratio (Penn World Table 9.1)
Returns to scale priv. bus.	ν	0.845	0.81	0.81	Wealth Gini coefficient
Buyer's matching friction	h_b	0.331	22.0	24.0	% purchasing bus. (SCF)
Prob. to fail for exo reasons	ζ	0.081	15.0	14.9	% exiting self-employed (PSID)
Disutility of working (retired)	u_R	1.601	4.87	4.84	% retired entrepreneurs (SCF)
Non-pecuniary benefits	u_E	1.490	7.00	6.80	Ratio median net worth E/W
Buying fixed cost	F_b	1.271	2.20	2.22	Ratio mean K buying/founding (SSBF)
Probability entrep. ability [b]	p_ι	0.020	11.0	11.3	% share of entrepreneurs to workers (SCF)
Entrepreneurial ability scale	$\gamma(0)$	0.570	46.0	45.1	% share of small business GDP, SBA

[a] The main moments are indicative. Changing one endogenous parameter affects the whole equilibrium. All targets are matched within an interval lower than 10%.
[b] Computations using the CPS are averaged from 2001 to 2008.
[c] The share of GDP attributable to small businesses (less than 250 employees) in the U.S. is taken from the OECD estimates.

3.5 Properties of the Baseline Model

To add:

- in a footnote: our model have prediction regarding the sorting between ability and purchase (and thus quality of business – quality of entrepreneur). Because wealthier individuals are on average more talented, self-selection arises naturally in our setting.
- wealthier households tend to be richer, on average. Use the SCF to demonstrate this.

3.5.1 Model Validation

In this section, we validate our framework by reporting key model generated statistics that were not targeted in the joint parameterization. In our baseline model, the entrepreneurial sector holds around 47% of total capital which is slightly higher than the 40% reported in Quadrini (2000b). We find a fraction of mature businesses of 66%. This is comparable to the statistics reported by the Bureau of Labor Statistics: 64% of businesses were 5 years or more in 2003 and the corresponding

[36] The mean capital of purchased business relative to founded ones is computed by comparing firms within 5 years of acquisition.

number is 62% in 2010. This fraction is generated by the probability to mature (P_m) and the fraction of mature businesses that are transferred between individuals through the SMESM. The fraction of agents with zero net worth in the population is 14%, against 12% in Quadrini (2000b).

While we pin down the probability of selling a business within a year, the baseline generates a ratio of sellers to exiting entrepreneurs of about 9.9%, against 7% to 20% in the 2007 SBO, the 2014-2016 ASE, and the NLSY79. Regarding life-cycle characteristics, Figure Figure 3.5.1a displays the baseline density of entrepreneurs by age bracket compared to the distribution in the 2007 SCF. Similarly, Figure Figure 3.5.1b compares the baseline density by age bracket of entrepreneurs exiting by selling their business to both the 2007 SBO and to business assets in the PSID averaged over the 1989-2015 waves. The model replicates reasonably well the life-cycle patterns of an average entrepreneur in the economy. It is especially relevant here as we are interested in characterizing who sells and buys businesses. We also find that the ratio of business assets sold in the last age bracket (65 and over) relative to the total business assets being sold is about 44% in the PSID averaged over the 1989-2016 waves. The corresponding number in the baseline is 33%. However, note that in the PSID, we can not distinguish between business assets sold as part of the sale of entire businesses and liquidations of fractions of business assets. Our baseline number is only about entire business sales. Regarding the age of entry into entrepreneurship, we find a mean age of 44 for both founders and purchasers in the 2007 SCF while it is respectively 45 and 46.8 in the model.[37] On the exit side, the mean age of business asset sellers is 53.7 in the PSID, while it is 52.4 in the model. Thus, consistently with the data, the entrepreneurial life-cycle appears to be a key component of both the model and the SMESM. In an alternative economy where we double the disutility cost u_R associated with working while in retirement, we observe a decrease in the fraction of individuals in the last age bracket from 4.8% in the baseline to 1.4%. At the same time, the fraction of business sellers substantially increases to reach 12.3% and the business price p reduces from 0.17 to 0.16 (see Appendix 3.C for details).

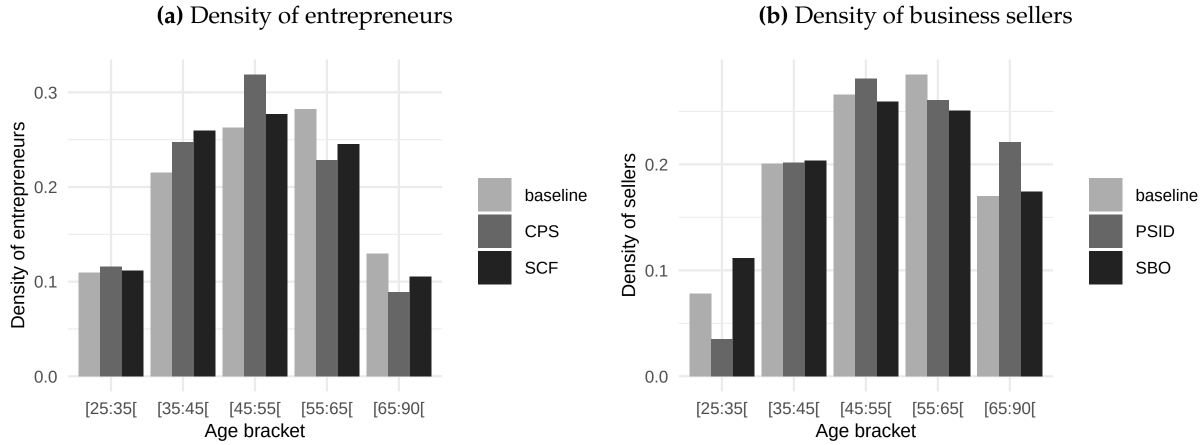
Finally, the baseline model is also able to closely reproduce wealth concentration and inequality statistics. We delay this discussion to Section 3.6.4.

3.5.2 The Decision to Enter Entrepreneurship

The selection in and out of entrepreneurship is a key element of our model. The main drivers leading individuals to select into entrepreneurship is wealth and entrepreneurial ability. Following the literature, non-pecuniary benefits appear as an additional driver. Concerning the type of acquisition, fixed costs F_b lead wealth-poor individuals to enter entrepreneurship by founding instead of

[37] Our comparison point is new entrepreneurs within 4 years of firm ownership. Notice that in the CPS, the average entry age into entrepreneurship is 43 in 1996 and 48 in 2016. Using the SSBF, we find no clear difference between the age of buyers and founders, as shown in table 3.A.2 of Appendix 3.A.2. In the model, we use the midpoint within an age bracket to compute the mean age.

Figure 3.5.1. Life-cycle pattern of entrepreneurship



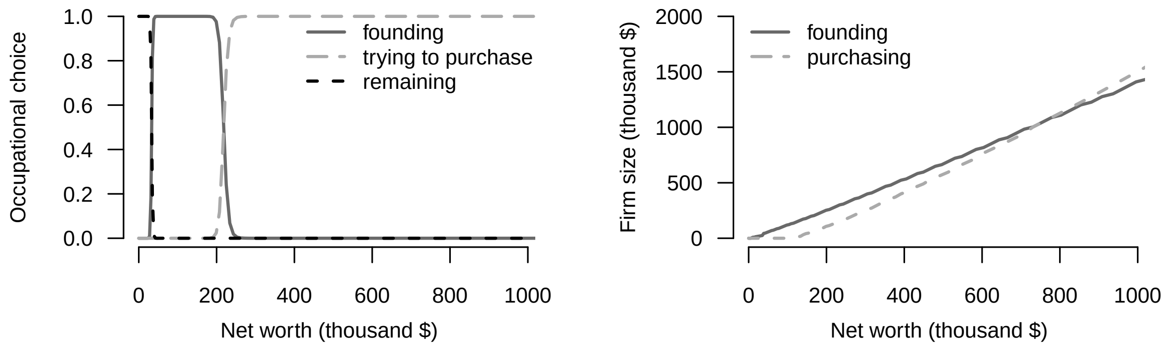
Note: we report the survey weighted density for the 2007 Survey of Consumer Finance (SCF) and the 2007 Survey of Business Owners (SBO) and the PSID (averaged from 1989 to 2015, deflated). Baseline reports the exact same densities in the model.

purchasing. Figure Figure 3.5.2a displays the model-based decision to enter entrepreneurship and the type of acquisition as a function of wealth for an entrant with average working productivity. Conditional on always finding an existing business, wealthy individuals would rather purchase an existing mature business instead of founding a new one. However, the existence of matching frictions ($h_b > 0$) suggests that only 33% of potential purchasers will match a seller. Consistently with the fact that purchasers are in general wealthier, the ratio of the mean net worth between recent purchasers and founders is about 2.8 in the SCF (2007). In the model, this ratio is about 2.1.

Alongside the decision to buy or found, a prospective entrepreneur also chooses the amount of start-up capital. In Figure Figure 3.5.2b, we display a typical model-based start-up capital policy function for a new entrepreneur deciding either to buy an existing business or found a new one. There is a threshold below which agents found and above which they purchase. The exact position of the threshold is state-dependent. We illustrate the case of an individual in a middle-age bracket for whom the threshold is at a net worth around 750K dollars. Purchasing a size k business is more expensive than founding one of the same size. This is due to the intangible value embedded in a purchased firm which incentivizes entrepreneurs to buy an existing mature business even when the initial size is smaller. Finally, the slope difference between the purchasing and the founding curves is generated by two components. First, founders face a tighter borrowing constraint because $\theta(0) > \theta(1)$. Second, the nature of the business price leads to decreasing returns to scale when buying larger businesses. This is due to the concavity of the production function reflected in the maturity value that enters the pricing formula $\mathcal{P}(k)$ together with the fact that the liquidation value is $(1 - \phi_d) < (1 + \phi_u)$. As a consequence, as shown in Figure Figure 3.5.2c, the average price $\mathcal{P}(k)/k$ declines as the purchased business capital increases, leading to higher start-up capital when entrepreneurs are able to purchase larger businesses.

Figure 3.5.2. Occupational choice and start-up capital as a function of wealth for new entrants.

(a) Occupational choice **(b)** Start-up capital



(c) Cost of founding and purchasing and start-up capital
k

[width=8.5cm]media/costfoundingbuyingtikz

Note: panel (a) displays the behavior of an hypothetical worker ($j = 4$, $y = 2$ and $\iota = 1$) when faced with the alternatives of entering entrepreneurship either by founding or purchasing and remaining a worker. Panel (b) displays the behavior of an hypothetical entrant with $j = 4$. Notice that the probability to switch occupations is not binary due to the perturbation method used to smooth the kinks generated by the occupational choices.

We finally show in Figures Figure 3.5.3a and Figure 3.5.3b the distribution of start-up capital as a function of business acquisition type: foundation or purchase. The model generates a consistent right-skewed distribution of start-up capital. Moreover, the distribution of purchased businesses is shifted to the right relative to that of founded ones.[38] To further relate the above points to empirical observations, in the SSBF, among firms within 3 years of their acquisition, the ratio of total firms assets between purchased and founded firms is about 3.7. In terms of the number of employees, profit and total sales, these ratios are respectively 2.1, 2.1 and 3.8 (see Appendix 3.A.2 for details). We observe similar evidence using the median ratio. From this, we infer that purchased businesses are indeed substantially larger upon acquisition, a feature that the model captures well.

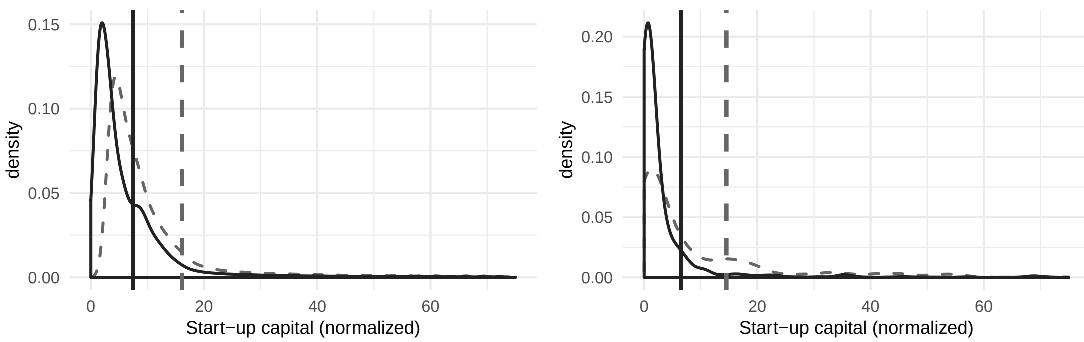
3.5.3 The Decision to Exit Entrepreneurship

Our baseline model is consistent with several features of entrepreneurial exit. Most notably, it is able to capture the exit behavior of firms with respect to their maturity. In Table 3.5.1, we compare the survival rates of firms in our model to both the Bureau of Labor Statistics and Knaup and Piazza (2007) data. We start with the survival rate relative to the establishment year: our baseline replicates it well in the first years but underestimates it after 6 years. This comes from the fact

[38] Notice however that our resulting distribution of start-up capital for purchased businesses is less dispersed relative to the data. An extension with an heterogeneous fixed cost F_b in the range of Clementi and Palazzo (2016) would generate such a distribution. For instance, increasing F_b to 2.0 raises the average firm size of purchased businesses from 20.2 in the baseline to 22.4.

Figure 3.5.3. Density of start-up capital by type of acquisition.

(a) Baseline model **(b)** 2003 SSBF data



Note: we normalize start-up capital by the median net worth in the baseline model and the data. The straight line corresponds to founded businesses while the dashed line to purchased businesses. The corresponding vertical lines indicate the mean normalized start-up capital for each density.

that for mature businesses, the failure rate, given by $\zeta + \chi(1)$ is somewhat constant.[39] However, a model without maturity and constant exit rate, such as the one in Cagetti and De Nardi (2006b), would underestimate the survival rate even more: after 5 years, such a model has a survival rate lower by 5 percentage points relative to our baseline. Moreover, the model captures quite well the increasing survival rate as firms mature.[40] In Section 3.6, we discuss the central role of maturity in our model, relating it to the intangible value of a firm embedded in the purchasing option. Interestingly, in the data, the survival rate relative to the preceding year substantially increases between 4 and 5 years, which may indicate that firms start to be well-established after that. This supports our calibration of the probability to switch from immature to mature with an average of 5 years in the early-stage.

Overall, the baseline model's ability to appropriately replicate a number of key features of the data that were not targeted during the joint parameterization seems sound to us.

3.6 Businesses for Sale Market: Quantitative Analysis

This model is the first to feature a businesses for sale market (SMESM) letting entrepreneurs transfer their business assets to a different owner. Thus, this section presents our quantitative assessment of the importance of that market and the associated maturity components. We show that

[39] In a recent paper, Fairlie et al. (2018) document survival rate differences between start-ups with and without employees, as well as the dependence on the legal form. Our estimates are in the range of theirs.

[40] The fact that the model survival rate is lower as compared to the BLS data can be explained by their specific definition: their survival rate is constructed using establishment openings (new businesses consisting of both establishments that are created and establishments that are reopening, including establishments that open on a seasonal basis). Moreover, the difference with the data is quite large after 10 years. Nevertheless, the model does match the difference in survival rates one year after acquisition between founded businesses and purchased ones, which is more likely to be relevant for agents deciding between those two options. Moreover, the data reports the survival rate of firms with at least one employee. In practice, the model also accounts for self-employed businesses that may have lower survival rates. See Knaup and Piazza (2007) for further details.

Table 3.5.1. Survival rate: model versus data

	Number of years						
	1y	2y	3y	4y	5y	6y	10y
Survival rate relative to first year							
U.S. data (BLS)	80.1	68.7	60.2	52.6	46.8	43.2	33.8
U.S. data (Knaup and Piazza (2007))	81.2	65.8	54.3	44.4	38.3	34.4	–
Baseline model	80.8	66.4	55.3	46.5	39.5	33.7	19.6
Zero probability to mature [a]	80.8	65.4	52.8	42.7	34.5	27.9	11.9
Survival rate relative to the preceding year							
U.S. data (BLS)	80.1	85.8	87.6	87.4	89.0	92.3	94.1
U.S. data (Knaup and Piazza (2007))	81.2	81.0	82.6	81.7	86.3	89.9	–
Baseline model	80.8	82.1	83.2	84.1	84.9	85.4	86.7
Zero probability to mature [a]	80.8	80.8	80.8	80.8	80.8	80.8	80.8

[a] This is the survival rate when entrepreneurs are not allowed to mature using the same panel of entrepreneurs.

the ability to build a firm's maturity is a critical component of that market and its outcomes. We emphasize two dimensions: the aggregate outcomes and the cross-sectional implications on occupational choices, the distribution of firms, and wealth inequality.

3.6.1 Assessing the Importance of Business Transfers

We first investigate the counterfactual in which the SMESM is missing. In this alternative economy, all new entrants must found a new immature business and all exiting entrepreneurs must liquidate their assets. We consider two situations: a general equilibrium (GE) case where prices and labor tax τ_w adjust and a partial equilibrium (PE) case where prices and taxes are kept at the baseline level. Table 3.6.1 reports the results.

Without the SMESM and, thus, in the absence of any business transfers, the steady-state decline in aggregate output is substantial at 10.5%. This drop is mostly due to the 18.7% decrease in the SME sector production, accounting only for 41% of total production against 45% in the baseline. As a side effect, aggregate savings also decline by almost 15%, which leads to a corporate sector production drop of 3.8%. Overall, 80% of the loss is attributable to the decrease in the SME sector production, while 20% is coming from a lower corporate output.

Interestingly, the decline in aggregate savings has important effects on prices: the interest rate is higher and the wage rate is lower in equilibrium.[41] These GE effects somewhat counteract the potential corporate output loss: in the PE case, aggregate savings losses are much larger. In the absence of price adjustments, the steady-state aggregate output would decrease by 13% with a 21% drop of aggregate savings.

In Table 3.6.2, we compare our baseline to the alternatives described above but along the lines of occupational decisions and firm size distribution. Without the SME for sale market, the fraction of entrepreneurs diminishes by about 0.3 percentage points. A larger drop is partly offset by

[41] Moreover, notice that the labor income tax increases since a higher fraction of older individuals are not working and there are no government revenues from business transfers.

Table 3.6.1. Aggregate outcomes with and without the SMESM

	Y	$\frac{\Delta Y}{Y}$	Y_c	Y_{SME}	K	$\frac{Y_{SME}}{Y}$	$\frac{K}{Y}$	r_s	w	τ_w
U.S. data [a]	–	–	–	–		46.0	3.20	–	–	–
Baseline with SMESM	2.38	–	1.31	1.07	7.70	45.1	3.24	4.9 %	1.18	15.0 %
No SMESM (GE)	2.13	-10.5%	1.26	0.87	6.56	40.8	3.07	5.6 %	1.15	16.0 %
No SMESM (PE)	2.07	-13.0%	1.21	0.86	6.07	41.4	2.94	4.9 %	1.18	15.0 %

[a] The U.S. share of output produced in the SME sector is taken from Kobe (2012). Y_c and Y_{SME} refer respectively to the corporate and the SME sector output.

higher incentives to enter entrepreneurship coming from lower wages and a higher interest rate, the latter letting workers accumulate more capital. Surprisingly, even without price changes, the entry rate increases. This is directly linked to the absence of a SMESM. Indeed, we find that entrepreneurial entry decisions are not comparable with and without the SMESM. One key reason is the natural incentive to wait for a future opportunity after an unsuccessful attempt to buy an existing business. Such a mechanism is absent in current entrepreneurial models in the literature. As we illustrate in Figure 3.6.1, this mechanism conditions the nature of entry in the sector: the dashed (blue) line reports the probability to enter the sector by founding a business in a setting without the SMESM whereas the dotted (green) line reports the probability to enter by founding conditional on not finding a business to purchase in a setting with a SMESM. With respect to net worth, there is a significant gap between these two lines relating to the incentive to wait. Upon not finding a business to purchase, for a large range of net worths, prospective entrepreneurs will choose to rather wait for a future purchase opportunity. For the same range of net worth, without a SMESM, prospective entrepreneurs can only found and enter resulting in an appreciable difference in the type of entrepreneurial firms that can be generated by the two settings. In this range of net worth, the setting with a SMESM would produce mature firms entry corresponding to business transfers instead of new firms of similar sizes resulting in a diverging pool of firms in the economy. As a comparison point, the solid (red) line indicates the baseline probability to purchase instead of founding. For similar reasons, Table 3.6.2 shows that the exit rate increases: in the baseline sellers who are unable to find a buyer might postpone their exit. In the alternative, exiting entrepreneurs can only liquidate their assets and all exits are immediate. This feature also leads old owners (above 65 years of age) to exit entrepreneurship earlier. The fraction of entrepreneurs in the last bracket falls from 4.83% in the baseline to 4.58% without the SMESM.

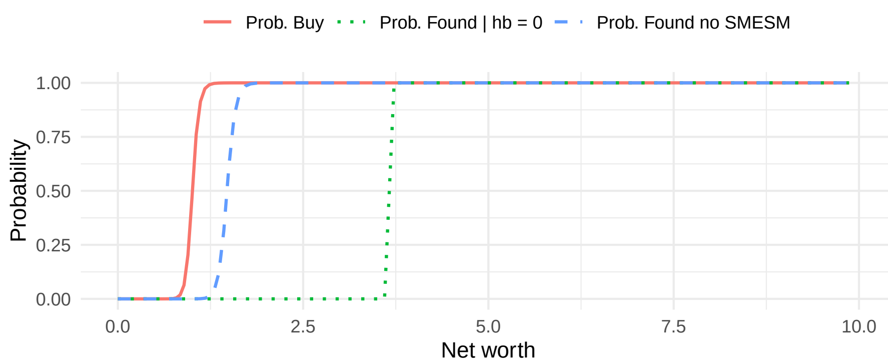
Without the SMESM, valuable existing businesses are liquidated instead of being transferred, resulting in two especially remarkable findings concerning the distribution of firms. First, the share of mature businesses falls by a considerable 13 percentage points, implying that the SMESM alone is a key market in transferring overall business maturity. As a result, the economy is faced with a substantial loss in intangible assets due to the impossibility to transfer them. Second, there is a significant decrease in the average firm size due to both the overall increase in the failure rate

Table 3.6.2. Occupational choice and distribution of firms with and without a SMESM

	$\% \frac{E}{W+E}$	Mature [a] %	Entry [b] %	Exit [b] %	Average size k	k entry	k exit
Baseline with SMESM	11.27	66.0	1.797	14.94	45.0	11.8	41.3
No SMESM (GE)	10.99	53.2	1.838	15.74	36.1	9.1	32.0
No SMESM (PE)	10.90	53.2	1.824	15.74	35.9	8.5	31.7

[a] This column measures the share of mature ($m = 1$) businesses in the economy.
[b] The entry rate is computed as the ratio of workers entering entrepreneurship relative to the total population of workers. The exit rate is computed as the ratio of entrepreneurs exiting entrepreneurship relative to the total population of entrepreneurs.

Figure 3.6.1. Entrepreneurial entry decision



and the lower average start-up capital upon entry. Indeed, without the SMESM, immature businesses are founded resulting in a higher failure rate as compared to mature ones. Consequently, a smaller number of businesses are accumulating productive capital over time, reducing aggregate output. As smaller firms enter, the overall distribution of firm sizes is shifted to the left and the average firm size upon exit is also significantly lower (by about 22%). In other words, the possibility to transfer business assets through the SMESM creates an environment where firms expand without loosing intangible value.

3.6.2 Decomposing Maturity Effects on the Businesses for Sale Market

In this section, we further detail the importance of maturity on aggregate outcomes and entry decisions. To that end, based on our finding documented in Section 3.2, we decompose the wedges generating components of maturity: (i) the profit rate, (ii) the interest rate charged on the debt, (iii) the failure rate and (iv) the borrowing constraint tightness. Our approach is to set each maturity component to its average value in the benchmark economy.[42] To help our decomposition, we distinguish the following alternative economies with respect to our baseline case: case (1) removes the SMESM entirely, case (2) removes all maturity effects making the SMESM inoperative and

[42] For instance, concerning the failure rate, as 53% of businesses are mature and 47% are immature without differential failure rate, we set $\chi(1) = \chi(0) = 0.915$ such that we broadly recover the same failure rate as in the benchmark economy.

cases (3) through (6) remove each of the specific maturity components one at a time. Table 3.6.3 displays the results of this decomposition.

We find the failure rate and the profit rate to be the most significant components of maturity as emphasized by case (6) and then (5). When removing those two components, the proportion of business buyers in the population of new entrepreneurs is substantially reduced. While the profit rate component substantially increases the returns associated with running a business, the lower failure rate of a mature business largely decreases the risks associated with entrepreneurial activities. The latter generates *higher returns* while the former increase the *persistence* of entrepreneurial returns. Consistently, our results confirm that the option to buy an existing business offers an appealing and empirically relevant mechanism through which prospective entrepreneurs can reduce the risks associated with early-stage entrepreneurship. As a consequence, the fraction of mature businesses in the economy falls, which, in turn, lowers aggregate output. In contrast, the interest rate and borrowing limit components have marginal aggregate effects despite slightly reducing the fraction of buyers and the average business size. In fact, the lower average firm size coming from more stringent borrowing constraints and financial conditions does not significantly impact aggregate outcomes. Consistently, in case (2), where we remove all four of the components above, we find substantial deviations from the baseline. This supports the fact that not taking into account the importance of firm maturity and the accumulation of intangibles would significantly lower aggregate outcomes as well as the fraction of entrepreneurs in the economy.[43]

Overall, maturity components and their potential transfer on a market is a fundamental interaction supporting the underlying mechanisms of entry and exit in the entrepreneurial sector, the firm distribution and aggregate outcomes.

Table 3.6.3. SME for sale market, maturity effects and intangible value decomposition

	$\frac{\text{E}}{\text{W+E}}$%	Buy %	Sell %	Mature %	Avg. size	Y	$\frac{Y_{SME}}{Y}$	$\frac{K}{Y}$	r_s
Baseline with SMESM	11.3%	24.0	9.96	66	45	2.4	0.45	3.24	4.9%
(1) No SMESM	11.0%	–	–	53	36	2.1	0.41	3.07	5.6%
(2) No maturity components	11.1%	–	–	49	27	1.9	0.35	2.93	6.0%
(3) No int. rate component r_b	11.3%	23.0	9.89	66	43	2.3	0.44	3.21	5.0%
(4) No borr. cst component θ	11.3%	22.6	9.74	65	43	2.3	0.44	3.21	5.0%
(5) No profit rate component γ	11.4%	19.6	9.59	64	38	2.2	0.42	3.15	5.4%
(6) No failure rate component χ	11.0%	10.7	7.6	55	31	2.0	0.38	3.02	5.6%

3.6.3 Businesses for Sale Market and Matching Efficiency

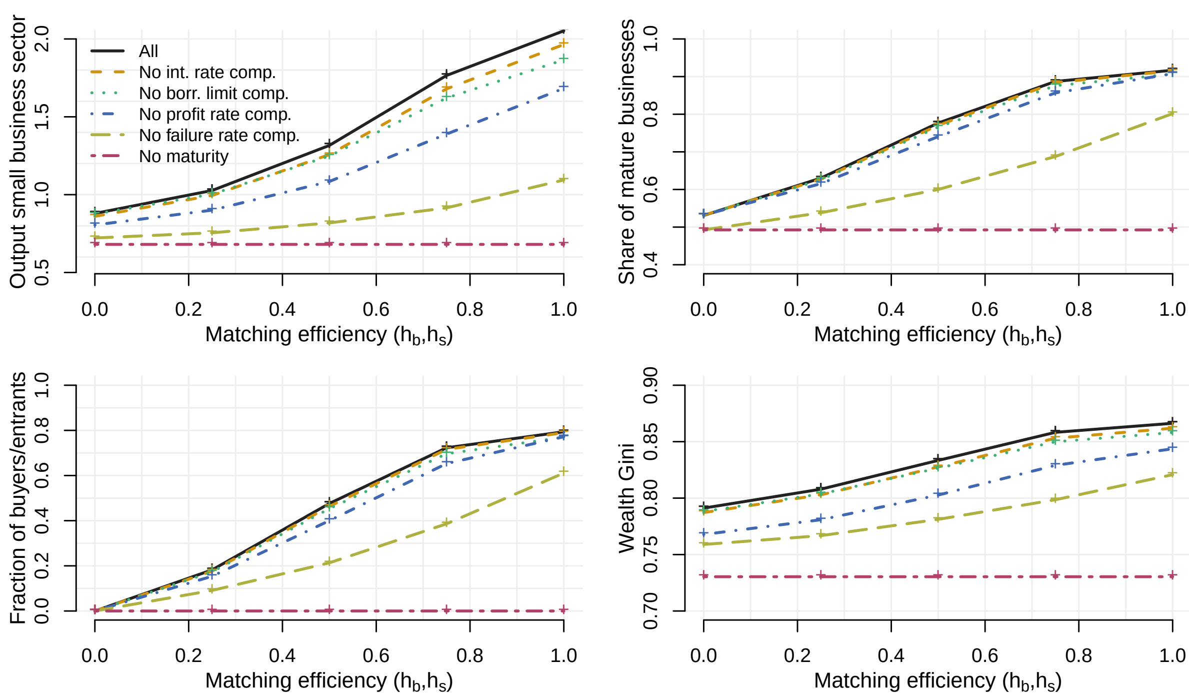
In our parameterization section, we provided evidence of selling frictions on the SMESM. In the model, both the actions of buying and selling a business are dependent on the occurrence of a

[43]This finding is related to Bhandari and McGrattan (2018) and a number of other papers showing the importance of intangibles when measuring contributions to aggregate output. Moreover, we solely focus on transferable intangible assets. In practice, non-transferable intangible assets might have large additional effects on output.

match on the SMESM. Thus, we investigate the role of the matching efficiency on the propensity to transfer businesses between owners. Figure 3.6.2 displays the equilibrium outcomes implied by the counterfactual experiment of increasing the matching probabilities (h_b, h_s) simultaneously and in the same proportion. Furthermore, as we stressed above that the effects of maturity were significant, we provide a decomposition by either removing all maturity components or each one, one by one.

We find that lowering the frictions on the SMESM by increasing the matching efficiency yields large aggregate outcomes. In our baseline with all maturity effects, increasing the probability to sell businesses by one percentage point on the SMESM increases output in the entrepreneurial sector by 6.9% and the wealth Gini by 0.8%. This is due to the large increase in the number of mature businesses in the economy. Moreover, consistently with our previous findings, most of the effect on the output of a higher matching efficiency comes from lower failure rates and higher profit rate.[44] Notice that even in the absence of mismatches between buyers and sellers (i.e. $h_b = h_s = 1$) the fraction of buyers to entrants is capped at 80%. This emphasizes the importance of both borrowing constraints and the fixed cost F_b in reducing the ability to purchase businesses. Interestingly, the last panel of Figure 3.6.2 shows that wealth concentration is sensitive to the component of the maturity and matching efficiency. We discuss the role of maturity in shaping wealth inequality in the next section.

Figure 3.6.2. The importance of the matching efficiency on the SME for sale market



[44]In the details, there are important general equilibrium effects. We underline them by comparing with a partial equilibrium model in a Supplementary Appendix available upon request.

3.6.4 Wealth Concentration and Inequality

In this section, following the important literature relating entrepreneurship and wealth concentration, we investigate the impact of maturity and the SMESM on wealth inequality. In most survey data but also in general business valuation approaches, the entrepreneur is required to value her business assets based on their market value.[45] In the model, depending on the maturity m of a firm, we value business assets using the following approach:

$$\begin{aligned}\text{Business assets}(k,m) &= \text{Tangible assets}(k) + \text{Transferable intangible value}(k,m) \\ &= (1-m)[k - C(k,0)] + m\mathcal{P}(k) \end{aligned} \tag{3.24}$$

where only mature businesses possess a transferable intangible value.

Given the above, we find that the SMESM and maturity effects have remarkable consequences on wealth concentration and inequality. First, in Table 3.6.4 we compare wealth (defined as net worth) distribution statistics in the model and in the SCF data. The baseline model with entrepreneurs matches the U.S. wealth concentration extremely well, while a comparable model without entrepreneurs is unable to do so. Previous entrepreneurial models, for instance Cagetti and De Nardi (2006b), were also able to match the wealth distribution. However, the novel aspect here is that the SMESM and the maturity value of businesses enhance wealth concentration in our model. We illustrate this by comparing our baseline to an alternative without the SMESM in case (1). Two results emerge. First, wealth concentration is more pronounced in the baseline case because business owners transfer the value of maturity on the SMESM. As such, the average firm size is higher and entrepreneurs are richer. This also helps in better reproducing the wealth Gini with respect to the data. Second, the absence of a SMESM substantially impacts the ratio of median wealth between workers and entrepreneurs as then businesses are valued only based on their tangible assets.[46]

We further explore the maturity components, by applying the decompositions above to wealth concentration and inequality. As illustrated by case (2), in the absence of any maturity effects (and thus of an operative SMESM), the ratio of median wealth between entrepreneurs and workers decreases appreciably, translating that the share of wealth held by entrepreneurs is significantly reduced. Wealth concentration and inequality drop by a sizable margin as evidenced by both the share of wealth held by the top percentiles and the wealth Gini. For instance, the wealth held by the top 1% falls from 35.2% in the baseline case to 24.8%. Comparing case (1) and (2), we note that removing maturity elements are enough to generate most of the effects on wealth concentration.

[45] For instance, the SCF evaluates business assets based on their market value using the following question: *"What could you sell it for?"*.

[46] Notice that inequality would still diminish if we were to assume that businesses keep the same valuation of maturity even in the absence of a SMESM.

Table 3.6.4. Wealth concentration and inequality

	Ratio E/W medians	Gini	Wealth share of top 1%	5%	10%	20%	40%
U.S. data [a]	7.0	0.810	33.6	60.3	71.4	83.4	94.6
Baseline with SMESM	6.8	0.814	35.2	62.8	73.8	83.7	93.9
Baseline without entrepreneurs	-	0.522	5.1	18.2	32.1	53.2	80.1
(1) No SMESM	4.1	0.782	32.4	58.5	69.1	80.2	92.0
(2) No maturity	3.6	0.730	24.8	49.9	61.9	75.2	90.0
(3) No interest rate component	6.7	0.808	34.3	62.0	72.9	83.3	93.7
(4) No borr. cst component	6.5	0.809	34.3	62.1	73.0	83.3	93.7
(5) No profit rate component	6.0	0.785	31.1	58.3	69.5	80.4	92.4
(6) No failure rate component	5.1	0.770	30.0	55.8	67.4	79.0	91.8

[a] We report values for the U.S. wealth shares from Benhabib et al. (2019).

Interestingly, the literature has provided a number of mechanisms to match wealth concentration, from the introduction of entrepreneurs into a worker-based economy (Quadrini (2000b)), to heterogeneous patience between individuals (Krusell and Smith (1998)) or the existence of voluntary bequest motives (Cagetti and De Nardi (2006b)).[47] We underline a completely new channel based on the heterogeneity of firms absent in the literature and that is furthermore consistent with the behavior of individuals in the economy and empirical evidence: maturity effects can help shape a significant portion of wealth concentration beside their importance for the SMESM. Maturity generates two key features: (i) more dispersed returns and income inequality between entrepreneurs with early-stage small businesses and those running large mature ones, (ii) more income persistence for entrepreneurs with mature businesses as the exit rate (the failure rate) is reduced. Those elements, in turn, translate into more wealth inequality and wealth concentration in the hands of very few individuals. Consequently, the effects of removing the SMESM (and hence the possibility to transfer maturity) on inequality is larger when the accumulation of intangible assets over time translates into more maturity, as shown by comparing our baseline and case (1).

To close our analysis of wealth concentration and inequality, we decompose the effects of maturity along the components mentioned above in case (3) through (6). Again the components on profits and failure rates reported in cases (5) and (6) are the most striking, with a significant reduction in the wealth Gini and wealth concentration at the top percentiles. This result mirrors our previous one on aggregate outcomes: most of the maturity effects come from those two margins.

3.6.5 Sensitivity Analyses

To conclude our quantitative analysis of the businesses for sale market, we perform a number of sensitivity and robustness exercises regarding model parameters. For reasonable parameter value changes, we find that the properties of the model remain relatively stable and that the main results

[47] For a more general discussion on relevant margins to generate wealth concentration consistent with social mobility in the U.S., we refer to Benhabib et al. (2019).

of the paper are valid. The details of these exercises are discussed in Appendix 3.C.

3.7 Conclusion

In this paper, we build a life-cycle heterogeneous agents model with occupational choices and introduce two key margins: prospective entrepreneurs must either buy or found their businesses upon entering the sector while incumbents must either sell or liquidate theirs upon exit. At the equilibrium, an endogenous business price clears a small and medium-sized enterprises for sale market (SMESM). The option to purchase lets entrepreneurs acquire well-established mature businesses, with a lower probability to fail, higher profits, and better financial conditions. We argue that maturity relates to the intangible value of a firm and show why it is a key component of entrepreneurship and entrepreneurial entry and exit.

Our baseline model provides a consistent cross-sectional and aggregate representation of the U.S. economy. We first demonstrate that the SMESM has important implications for aggregate outcomes. By allowing the transfer of the value of the maturity of a firm between owners, overall survival rates, firm sizes, and aggregate production are increased. Without this market, aggregate production drops by about 10% and the consequences on aggregate savings and prices are also severe. Second, we find that the decreasing failure rate over time is the most important component embedded in the option value of a purchase relative to that of a new firm creation. Third, we show that entrepreneurial life-cycle and matching frictions are important determinants of entry and exit. While the literature has evaluated the former, we establish that the latter can significantly shape aggregate outcomes and the composition of the entrepreneurial pool. Finally, we uncover a novel channel to match wealth concentration and inequality that is consistent with individual behavior and empirical evidence and is furthermore directly linked to the SMESM and business maturity.

Our contributions might be particularly relevant for future research on the aging of entrepreneurs and the decline of the start-up rate where we expect business transfers to play a predominant role.

Appendix

Appendix

3.A Empirical appendix

In this section of the Appendix, we provide additional empirical evidence supporting the fact that recently purchased businesses are substantially different from newly founded ones.

3.A.1 Acquisition type

Table 3.A.1 provides estimates of the proportion of firms by acquisition type for various U.S. survey data. Broadly speaking, one out of five entrepreneurs enter the sector through the purchase of an existing business. This number remains consistent across survey data and time.

Table 3.A.1. Business acquisition by type in U.S. surveys

			Acquisition (%)		Transmission (%)	
Survey [a]	Year	Sample selection [b]	Founded	Purchased	Inherited [c]	Other/Gift [c]
SCF	2016	All entrepreneurs	74.4	18.2	3.5	3.9
ASE	2016	Only employers	68.1	20.8	4.0	7.1
SSBF	2003	All entrepreneurs	79.8	16.7	– 3.5 –	
SSBF	2003	Entrepreneurs ($<$ 5y)	77.4	20.8	– 1.8 –	
SBO	2007	All entrepreneurs	74.6	18.2	2.3	4.9
SBO	2007	Entrepreneurs ($\leq$ 3y)	74.4	19.1	1.2	5.3

[a] An entrepreneur is defined as an individual declaring that her business constitutes her primary source of income (with an active management role, whenever possible). The ASE reports macro data for all business owners with at least one paid employee.
[b] The estimates are based on self-employed entrepreneurs defining themselves as business owners. Early-stage entrepreneurs are those who acquired their businesses within the last 5 years.
[c] When possible, we distinguish the acquisition type between gift/other and inheritance.

3.A.2 Business performances and owner characteristics

Table 3.A.2 shows characteristics of firms within 3 years of their acquisition and the characteristics of their owners. Purchased firms systematically perform better than their founded counterparts: the former display 4 times the average total assets and total sales with respect to the latter and twice the average number of employees and profit. Concerning the owners of recently purchased or founded firms, they appear very similar in terms of age, years of experience as entrepreneurs and education. To complement, in the SBO 2007, entrepreneurs were asked *Whether the owner*

previously owned a business or had been self-employed. 60% of founders reported having no prior experience compared to 63% for the purchasers.

The above statistics slightly differ from those in the 2007 SCF. In that survey, for a sample of entrepreneurs within 4 years of the acquisition of their firms, we find a mean age of about 44 for both founders and purchasers. The fraction of purchasers (resp. founders) with a degree above high school is 84% (resp. 66%). Finally, the ratio of the mean net worth held by purchasers relative to founders is about 2.8.

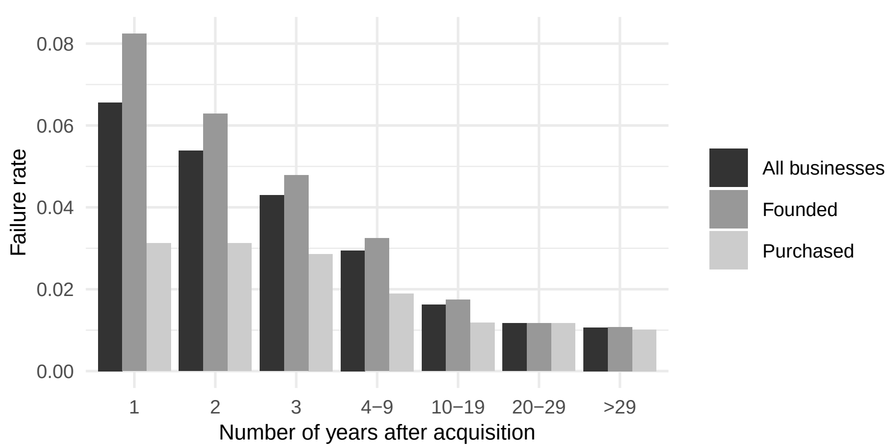
Table 3.A.2. Characteristics of firms within 3 years of acquisition by type

	Mean			Median		
	Purchased	Founded	Ratio	Purchased	Founded	Ratio
Firms						
Total assets (USD)	766K	191K	4.0	74K	26K	2.8
Avg. number employees	9.2	4.1	2.2	5	2	2.5
Profit (USD)	118K	52K	2.3	20K	2K	10
Total sales (USD)	1093K	264K	4.1	300K	61K	4.9
Owners						
Age	44.4	44.9	1.0	44	45	1.0
Years experience	10.30	10.13	1.0	6	6	1.0
≥High school deg. (%)	64	63	1.0			

Source: Survey of Small Business Finances (2003)

Figure 3.A.1 complements Figure 3.2.2 in Section 3.2 on the failure rate of purchased versus founded firms. The evidence reported here only concern firms with paid employees. Broadly speaking, our results appear consistent when focusing on this group.

Figure 3.A.1. Failure rate by acquisition type and for all businesses, employer group



Source: author's computation using the 2007 SBO. We compute the failure rates using ceasing option linked to either inadequate cash-flows or low sales and lack of business or personal loans/credit.

Finally, we show in Table 3.A.3 and Table 3.A.4 additional evidence on the financial constraints faced by the group of recent purchasers and founders (within 3 years of acquisition). First, founded firms are more likely to be denied a loan they required. Second, the main reason why recent founded firms are denied access to credit is related their status as *not in business long enough*, implying that, for creditors, the sales history is an important piece of information influencing credit conditions.

Table 3.A.3. Loan acceptance rate for firms within 4 years of acquisition by type (%)

Acquisition type	Access to loan is		
	Always accepted	Always denied	Sometimes accepted or denied
Purchased	85.7	12.1	2.2
Founded	71.5	21.4	7.1

Source: Survey of Small Business Finances (2003). We use 4 years in order to increase the number of observations.

Table 3.A.4. Main reasons why credit was denied (%)

Acquisition type	Main reason for denial was			
	Not in business long enough	Credit history	Insufficient collateral	Other
Purchased	0.0	26.2	24.2	49.6
Founded	34.2	19.7	7.7	38.4

Source: Survey of Small Business Finances (2003). We take 4 years in order to increase the number of observations.

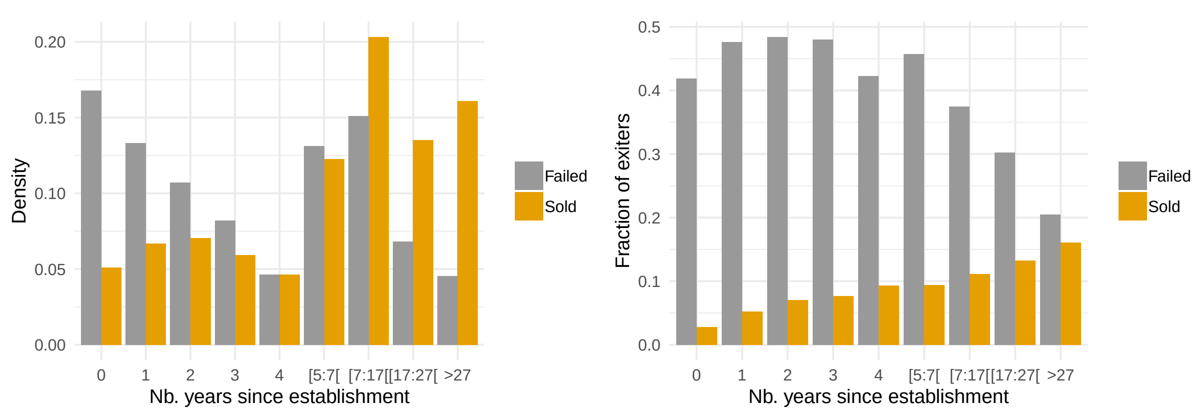
3.A.3 Exit rate and type of exit

We display in this subsection the exit rate by exit option: business failure versus successful sale. First, we note in Figure 3.A.2 panel (a) that sold businesses are generally older. This point is also confirmed in Appendix 3.A.4 using business for sale transaction data. Second, in Figure 3.A.2 panel (b), as opposed to early-stage business owners, old business owners are substantially more likely to exit after a successful sale of their firm rather than business failure. This additional evidence suggest that sold businesses are generally older well-established ones.

Figure 3.A.2. Type of exit and establishment data

(a) Density of exiters by firm's age and type

(b) Exit rate by type of exit



3.A.4 Small and medium-sized enterprises for sale market

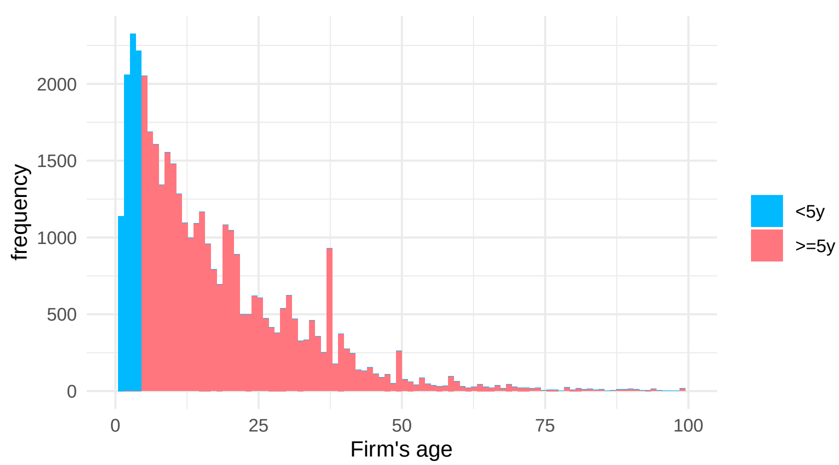
In order to characterize the selling frictions, we collected transactions data from the online platform *Bizbuysell.com* (hereafter BBS), one of the oldest and largest online marketplace dedicated to business selling transactions in the U.S.. The available data correspond to over 90,000 observations of businesses for sale or sold. In the details, we have two sets of data: (i) a panel of businesses for sale from 2018 to 2020 including any changes in business information; (ii) a collection of closed transactions from 2010 to 2020. Table 3.A.5 provide summary statistics regarding this BBS database.

Table 3.A.5. Descriptive statistics: BBS data on businesses for sale (2018-2020)

Statistic	N	Min	Pctl(25)	Median	Pctl(75)	Max	Mean	St. Dev.
Listing price	93,270	1,100	120,000	250,000	595,000	620,000,000	713,556	4,124,977
Cash flow	57,925	0	70,000	120,812	221,000	50,000,000	200,146	405,272
Gross revenue	69,267	0	260,000	520,500	1,066,730	9,969,000	910,338	1,161,825
EBITDA	9,842	12	62,549	120,000	265,878	435,000,000	314,871	4,423,594
Nb. employees	51,855	0	2	4	9	913	8	15
Inventory	35,065	4	5,000	15,000	54,728	35,000,000	88,503	486,542
Ceasing for retirement	24,643	-	-	-	-	-	0.20	-

Figure 3.A.3 shows that only around 15% of businesses for sale are younger than 5 years. This confirms that many businesses for sale are actually well-established mature businesses.

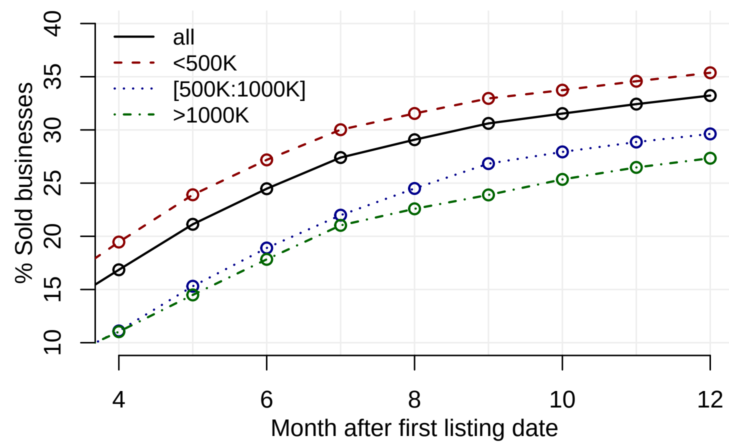
Figure 3.A.3. Age profile of businesses for sale



Source: Authors' own computation using BBS data.

Given the remarkably low number of entrepreneurs reporting having sold their business, especially with respect to initial intentions, important failures when trying to sell are potentially occurring. The inability of SME owners to reach any business selling deal results in liquidations: the sale at very low prices and usually restricted to tangible assets. Thus, to investigate whether

Figure 3.A.4. Probability to sell with listing price as a proxy for size.



To be read as follows: *after a year, 32% of all businesses have been sold*. The price corresponds to that at the initial listing date. *Source*: Authors' own computation using collected BBS data.

there are mismatches on the SMESM, we use the BBS data to infer the probability of selling a business as well as the time needed to sell.[48]

Using the same strategy as in the core of the paper, Figure 3.A.4 displays the probability to sell a business after a number of months from the listing date and by listing price brackets. After a full year, only around 25%-35% of businesses for sale are actually sold. While this number is fairly similar for any business size (as proxied by the listing price), it seems to be slightly easier to sell a smaller business.

3.B Model: further details

3.B.1 Numerical solution method: discrete-continuous model with the endogenous grid method (DC-EGM) under taste shocks

To tackle the issue of the high dimensionality of our problem, we adapt the recently developed DC-EGM solution method introduced in Iskhakov et al. (2017): it solves the occupational choice problem while still accommodating the fast endogenous grid method developed in Carroll (2006). On top of a substantially increased computation speed, this method is also well-adapted to the context of occupational choices. As shown in Hurst and Pugsley (2011), the decision to enter

[48]Concerning the comparability of the BBS dataset with respect to existing surveys, we note that 23% to 25% of sold businesses were sold because the owner(s) retired, against 19% in the ASE (2016). About the distribution of listing prices, the mean price is 495K USD and the median is 165K USD in BBS against respectively 682K USD and 95K USD in the PSID; meaning that the BBS price distribution is comparatively shifted to the right. However, this comparison is indicative: there are only 357 observations concerning sold businesses in the PSID (1990 to 2015) against 80,000 in BBS (and 93,000 businesses for sale). Moreover, many BBS listings are broker mediated and announcers have to pay a monthly premium membership to list their entry. This might be constraining enough for very small businesses. Overall, we believe that the BBS dataset provides a reasonable representation of U.S. business transactions.

entrepreneurship is also driven by non-pecuniary benefits that are, to some extent, not observable by the econometrician. In the model, we, therefore, introduce taste shocks to both smooth the value functions when applying the DC-EGM algorithm, as well as to get closed-form expressions for the probability to switch from one occupation to another through the available options of continuing, selling, purchasing, founding or liquidating a business.

Our algorithm has been implemented in C/C++. The details of the computation are provided in the Online Appendix.

3.B.2 Calibration details

Earning process We take the life-cycle average earning profile in Guvenen et al. (2015). We fit a life-cycle earning profile with a third order polynomial. Table 3.B.1 provides the corresponding values.

Table 3.B.1. Life cycle earning profile

[25 : 30[	[30 : 35[	[35 : 40[	[40 : 45[	[45 : 50[	[50 : 55[	[55 : 60[	[60 : 65[	65 and over
$h(1)$	$h(2)$	$h(3)$	$h(4)$	$h(5)$	$h(6)$	$h(7)$	$h(8)$	$h(9)$
0.58	0.72	0.84	0.93	0.98	0.97	0.90	0.75	0.40

3.C Sensitivity Analyses

The baseline model embeds a number of features that let us match empirical elements. In this section, we investigate the importance of key features interacting with the option values of selling and purchasing a business that might have an impact on aggregate outcomes. In the following, prices clear the markets and taxes are adjusted to balance the government budget constraint. Results are displayed in Table 3.C.1.

Table 3.C.1. Sensitivity analyses on the SME market and aggregate outcomes.

	$\frac{E}{W}$	Buy %	Sell %	Mature %	Avg. size	Y	$\frac{Y_{SME}}{Y}$	$\frac{K}{Y}$	p	r_s	τ_w
U.S. data [a]	11.0%	21.0	7-14	67	–	–	0.40	3.2	–	4-6%	–
Baseline model	11.3%	24.0	10.0	66	45	2.4	0.45	3.2	0.17	4.9%	15.0%
Adj. cost, $\phi_{up} = 0.1\Delta k$	11.0%	18.9	9.8	63	22	1.8	0.33	2.8	0.41	6.4%	16.9%
Doubling u_R cost	10.0%	26.3	12.3	65	38	2.1	0.39	3.2	0.16	5.1%	16.3%
No fixed cost $F_b = 0$	11.7%	51.0	13.1	79	48	2.5	0.47	3.3	0.27	4.8%	14.7%
Doubling profit rate wedge	11.0%	27.0	10.4	68	43	2.3	0.43	3.2	0.18	5.1%	15.2%
Borr. limit, $\begin{bmatrix}\theta(0)\\ \theta(1)\end{bmatrix} = \begin{bmatrix}0.50\\ 0.45\end{bmatrix}$	11.1%	22.0	9.8	65	32	2.1	0.39	3.1	0.20	5.4%	15.9%
No sales tax $\tau_s = 0$	11.2%	24.1	10.0	66	45	2.4	0.45	3.2	0.17	4.9%	15.2%

[a] We report value for the U.S. economy using Penn World Table 9.1, treasury bond interest rate, Current Population Survey (1998:2008), Survey of Consumer Finance (2001:2007) and the 2007 Survey of Business Owners.

We first investigate the role of adjustment costs. In the baseline model, we normalized the upsizing cost ϕ_{up} to zero. Thus the main source of capital illiquidity came from the downsizing cost of liquidating business assets. We test the sensitivity of increasing ϕ_{up} to 10%. That is, for each unit of capital k bought, the cost is $(1 + 0.1)k$. The resulting equilibrium implies a lower average business size. With the increase in the relative cost of founding a new business relative to purchasing an existing one, business price level increases and the fraction of buyers able to purchase business drops. Due to fewer business transfers and a lower level of accumulated business capital, aggregate output drops. In the end, it comes as no surprise to us that adjustment costs have important equilibrium effects, but the main mechanisms of the model remain similar.

As we argued in Section 3.2, the entrepreneurial life-cycle matters since an important fraction of entrepreneurs sell their business assets upon retirement. In the model, we capture this behavior by having a disutility cost u_R of working in the retirement age bracket. Doubling this utility cost reduces the fraction of entrepreneurs in the last age bracket from 4.8% to 1.4%, and the fraction of entrepreneurs from 11.5% to 10.0%. The fraction of sellers substantially increases, since many old entrepreneurs are now trying to sell, lowering the business price and therefore increasing the fraction of buyers in the economy. Because older entrepreneurs want to exit earlier, they accumulate fewer capital assets, and the average firm size and production fall. We argue that the aging of entrepreneurs, accelerating since the 2010s, might be a first-order concern on the SMESM.

Buying a business incurs the payment of a fixed cost F_b. Our third sensitivity test sets F_b to zero. We find that the share of entrepreneurs increases by 0.3 percentage points and that the share of buyers rises from 21% to 51%. As purchasing a business is now less costly with respect to founding, the share of mature businesses increases significantly. Business price level increases since now a larger fraction of entrepreneurs are able to buy an existing business. Overall, this leads to a substantial increase in aggregate output.

Next, we benchmark the effect of the profit margin by doubling the profit component between immature and mature firms (we let $\gamma(0) = 0.9\gamma^{\text{benchmark}}(0)$ and we keep $\gamma(1) = 1.1\gamma^{\text{benchmark}}(0)$ such that the difference in terms of profit rate is about 20%). This leads to an increase in the propensity of new entrant to buy an existing business. As mature firms are now more valuable, more of them are bought and their proportion in the economy increases, leading to an increase in business price. However, due to the lower profit rate of early-stage businesses, aggregate output decreases..

Finally, additional sensitive tests include lowering the tightness of the borrowing constraint and setting the sales tax τ_s to zero. Those two margins are shown to have only marginal effects.

Chapter 4

Value-Added and Productivity Linkages Across Countries

Alexandre Gaillard[1,2] François de Soyres[3]

Abstract

Traditional international real business cycle models produce a weak relationship between trade and cross-country real GDP correlations, contradicting widespread empirical findings. This puzzle can be resolved by defining real GDP in the model using double deflation, exactly mirroring how the data is constructed. Whenever imported input's base period price does not reflect their marginal revenue product, real GDP movements become mechanically linked to fluctuations in imported inputs. Focusing on the cases of markups and love of variety, we quantitatively show that input trade is associated with the synchronization of real GDPs, measured productivities and profits, consistent with data.

Keywords: International Trade, Comovement Puzzle, Real GDP Measurement, Networks, Input-Output Linkages, Solow Residual.

[1]The views in this paper are solely the responsibility of the authors and should not necessarily be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. We are indebted to Thomas Chaney and Christian Hellwig for their guidance. We are also grateful to Manuel Amador, Ariel Burstein, Patrick Fève, Simon Fuchs, Rob Johnson, Tim Kehoe, Martí Mestieri, Carter Mix, Alban Moura, Franck Portier, Ana-Maria Santacreu, Constance de Soyres, Shekhar Tomar, Guillaume Sublet, Robert Ulbricht, Kei-Mu Yi, and seminar/workshop participants from various venues for their comments. We thank Henry Young for his research assistance. Finally, we thank the Federal Reserve Bank of Minneapolis, where part of this research has been conducted, for their hospitality and the ERC grant N°337272–FiNet for financial support. All errors are our own.

[2]Toulouse School of Economics, University of Toulouse Capitole, 31080 Toulouse, France. Email: alexandre.gaillard@tse-fr.eu.

[3]Email: francois.m.desoyres@frb.gov.

4.1 Introduction

Since the seminal contribution of Frankel and Rose (1998), the role of trade in propagating shocks across countries has been the subject of substantial empirical attention. Two countries with stronger trade linkages tend to experience more synchronized business cycles.[4] Yet, Kose and Yi (2001, 2006) show that standard international real business cycle (IRBC) models are unable to quantitatively account for the empirical relationship by an order of magnitude. This failure of standard models is referred to as the *Trade Comovement Puzzle* (TCP) and remains an important open question in international macroeconomics.[5]

In this paper, we take a step back and argue that the mismatch between the observed empirical association and standard models' predictions can be resolved by improving the mapping between macroeconomic models and the data. Once real GDP is defined in the model with the same procedure used by statistical agencies, we show that it becomes more sensitive to foreign shocks and the puzzle vanishes. In contrast, stronger trade linkages are not associated with higher synchronization of a theory-consistent index of real value added. Our results highlight that, when comparing a macroeconomic model to the data, it is key to define aggregate variables in a way that is consistent with statistical agencies' procedures.

What is real value added in an economy that imports intermediate inputs? In a seminal contribution, Fabricant (1940) described real value added as an ideal index of the *net physical output* of an industry. This concept has then been refined by Sims (1969) and Arrow (1974), arguing that if the production function for gross output is separable between primary factors and intermediate inputs, real value added can be defined implicitly from the production function itself. Taking a simple example, if a country produces gross output by combining a domestic input A and an imported input B, the Sims (1969)-inspired definition of real value added corresponds to the quantity of input A only. We call this object the "physical value added". By construction, changes in the quantity of imported input B impact physical value added only insofar as it changes the quantity of input A produced in the economy.

In practice, statistical agencies do not observe physical value added and construct a measure of "statistical value added" using double deflation, a method consisting of taking the difference between gross output and intermediate inputs, both valued using base period prices. This statistical measure is what is called real GDP in most national accounts database. Our core argument is that real GDP measured by statistical agencies is, in general, equal to physical value added only if

[4] For empirical studies supporting the association between international trade and business cycle synchronization, see Clark and van Wincoop (2001), Imbs (2004), Baxter and Kouparitsas (2005), Kose and Yi (2006), Calderon et al. (2007), Inklaar et al. (2008), Di Giovanni and Levchenko (2010), Ng (2010), Liao and Santacreu (2015), Duval et al. (2015), Di Giovanni et al. (2018) and Avila-Montealegre and Mix (2020).

[5] For quantitative studies of the TCP, see for instance Kose and Yi (2001, 2006), Burstein et al. (2008), Arkolakis and Ramanarayanan (2009), Johnson (2014), Liao and Santacreu (2015) and Avila-Montealegre and Mix (2020).

prices used in the construction of real GDP are un-distorted – that is, if the base period price used for intermediate input reflects both its marginal cost and the marginal revenue that can be derived from its usage. If instead the base period price used in the valuation of imported inputs is lower (higher) than the marginal revenue product of these inputs, then an increase in foreign input usage is associated with higher (lower) statistical value added even for fixed domestic factors and fixed technology. In such case, real GDP and physical value added fluctuations differ.

To be clear from the outset, we do not wish to compare all possible mechanisms through which shocks can propagate across countries. Instead, our paper highlights and clarifies how the method used to *measure real GDP* impacts the propagation property in an international macroeconomic model. In particular, when base period prices are used in real GDP construction, any distortion between the price of imported input and their marginal revenue product generates a link between input usage and movements in real GDP, which in turns increases the strength of propagation of shocks across countries. Specifically, we focus on two sources of such a distortion, markups and love of variety, which are commonly used in the macro and trade literature. While the inclusion of input-output linkages together with markup and/or love of variety has been previously examined in the literature, we show that it is the conjunction of these elements with a statistically-consistent real GDP measurement (using double deflation) that allows for a solution to the Trade Comovement Puzzle.

With markups, which imply non-zero profits in the domestic economy, intermediate inputs generate more revenues than their cost. Hence, using more imported inputs results in higher statistical value added, even when domestic factors and technology are unchanged. Importantly, our argument does not rely on variable markups: the sheer presence of constant markups in the base-period prices used in the construction of real GDP creates a mechanical link between domestic real GDP and fluctuations in imported input usage.

In the presence of love of variety, accessing a larger range of foreign inputs is associated with efficiency gains that are not reflected in imported input prices. Again, using more imported inputs leads to higher statistical value added, over and beyond any change in domestic factors or technology and, as a result, measured productivity is directly affected by foreign shocks. Within a simple accounting framework, we show that accounting for real GDP fluctuations that arise from changes in imports is key to generating a strong link between trade linkages and business cycle synchronization.

We then turn to quantification of our theory and demonstrate that measuring real GDP in a way that is consistent with the data helps reconcile theory and empirical findings, thereby solving the Trade-Comovement Puzzle. To do so, we build an IRBC model with 15 countries, monopolistic competition and firms' entry and exit. Keeping technology shocks unchanged, we calibrate the model to different levels of trade flows and assess its ability to produce a strong trade comovement slope. Our empirical counterpart is constructed using a panel dataset which allows for dyadic and time windows fixed effects. We document that the positive relationship between trade and GDP comovement is mostly driven by *trade in intermediate inputs*, whereas trade in final goods is found to be insignificant or negative. This finding is consistent with our theory and further confirmed in our quantitative model. When real GDP is constructed using double deflation, our model reproduces almost exactly the *trade-comovement (TC) slope* observed in the data, hence offering a resolution to the Trade-Comovement Puzzle. To put things into perspective: both in the data and in our simulations, our point estimates imply that the observed increase in input trade between the 1970s and the 2000s is associated with an 11 percentage points increase in international real GDP correlation. As expected, the association between input trade and the synchronization of physical value added (as opposed to real GDP) is significantly lower in our simulations, reaching less than a tenth of the slope obtained with real GDP.

Finally, in our model as well as in the data, higher input trade is also associated with an increase in the synchronization of aggregate profits and measured productivity. This supports our view that real GDP comovement across countries is not solely driven by correlated factor supply. Additionally, we find that higher business cycle synchronization is associated with movements in the number of traded varieties, with both the range and the variance of extensive margin fluctuations being associated with a surge in GDP correlation.

Relationship to the literature Our work builds on a number of previous papers that helped refine our understanding of the relationship between bilateral trade and GDP comovement. Starting with Frankel and Rose (1998), many studies confirmed the positive association between trade and real GDP synchronization in the cross-section. If the empirical link between trade and real GDP comovement has long been known, the underlying economic mechanisms of this relationship are still unclear. Using the workhorse IRBC model with three countries, Kose and Yi (2006) have shown that their model can explain at most 10% of the *slope* between trade and business cycle synchronization, leading to what they call the *Trade Comovement Puzzle* (TCP). Since then, many papers have refined the puzzle, highlighting ingredients that could bridge the gap between theory and data.

Burstein et al. (2008) show that allowing for international production sharing can deliver tighter business cycle correlation if the elasticity of substitution between home and foreign intermediate inputs is extremely low. Arkolakis and Ramanarayanan (2009) analyze the impact of vertical specialization on the relationship between trade and business cycle synchronization. Their model with perfect competition does not generate significant dependence of business cycle synchronization on trade intensity, but they show that the introduction of time-varying price distortions improves the model's fit with the data. Incorporating trade in inputs in an otherwise standard multi-country IRBC model, Johnson (2014) shows that input-output linkages *alone* is not sufficient to solve the trade-comovement puzzle, but points that such production linkages do synchronize input usage. The three papers above feature perfectly competitive models in which real GDP is measured as "physical value added". Compared to them, we add firms' entry/exit and monopolistic competition and argue that, once real GDP is measured using double deflation, those ingredients reconcile models' predictions and the data.

The role of markups in generating a link between intermediate input and measured productivity has been discussed in several papers such as Hall (1988) and Basu and Fernald (2002), and more recently in Gopinath and Neiman (2014). With markups, intermediate inputs generate more revenues than their cost. Hence, statistical real value added can be created by simply using more inputs, even with fixed domestic factors and technology. The importance of love of variety and fluctuations in the number of imported varieties has been pioneered by Feenstra (1994). Most related to our international comovement focus, Liao and Santacreu (2015) uses a two-country model to show that when measured productivity is scaled by the number of varieties, a country-specific shock generates cross-country TFP comovement from its effects on firms' entry and exit. Compared to this paper, we highlight the importance of real GDP measurement and show that while the inclusion of price distortions and/or extensive margin adjustments significantly alters real GDP fluctuations as measured in the data, it does not materially change the model's propagation property if one looks at physical value added. To the best of our knowledge, we are the first to highlight and quantify the key importance of this distinction when evaluating the trade comovement puzzle. Finally, Drozd et al. (2020) develops a complementary approach by introducing dynamic trade elasticity in the presence of convex capital adjustment costs. This approach sheds lights on the mechanisms that enable a model to generate a link between international trade and factor supply synchronization.

4.2 On the definition and measurement of real GDP

This section illustrates how real GDP, as measured by statistical agencies, is disconnected from the ideal theoretical definition of real value added. In particular: when a country imports intermediate inputs and pays a price that does not reflect the marginal revenue product of those inputs, the usual statistical definition of real GDP is not equal to the net physical production in the economy. When marginal revenue derived from input is above their marginal cost, real GDP captures not only the physical value added but also accounts for the "accounting value added" derived from imported input usage. In turn, this measurement issue creates a wedge between measured productivity and actual technology.

4.2.1 A Simple Accounting Framework

Consider an economy that produces a gross output (GO) using domestic factors (K, L) and imported inputs (X). According to Sims (1969) and Arrow (1974), if the production function for gross output is separable between primary factors and imported inputs, real value added can be defined implicitly from the production function itself. If $GO = Q(K, L, X)$ can be re-written as $GO = Q(V(K, L), X)$, then we can "*imagine capital and labor cooperating to produce an intermediate good called real value added (V), which in turn cooperates with materials to produce the final product*".[6] Using this definition, real value added can be thought of as "physical value added" and its fluctuations are only attributed to changes of the value added bundle $V(.)$.

In practice, real value added is measured using double-deflation as the difference between gross output and intermediate inputs, both valued using base period prices. As a result, we show in this section that real GDP equals physical value added if and only if the base period price used for intermediate inputs reflects both their cost and the marginal revenue that can be derived from their usage.

We focus on two widely-used ingredients that create a wedge between imported inputs' base period price and their marginal revenue product: markups or/and love of variety. With these features, we show that real GDP fluctuations, as measured in the data, are not limited to changes of the theory-consistent physical value added, but are also the direct result of changes in the quantity and variety of imported input. By creating a mechanical link between real GDP and imports, those features allow for a quantitative resolution of the TCP.

Setup Consider an economy with N countries. Gross output in country n is produced using domestic factors (L_{nt} and K_{nt}) and imported inputs (X_{nt}) according to:

$$GO_{nt} = \Big[\underbrace{Z_{nt} L_{nt}^{\alpha} K_{nt}^{1-\alpha}}_{\text{Physical Value Added}} \Big]^{\gamma} \cdot \Big[\underbrace{X_{nt}}_{\text{Imported Inputs}} \Big]^{1-\gamma}, \tag{4.1}$$

where γ is the value added share of gross output and Z_{nt} is value-added TFP.

Markup Let μ_{nt} be the ratio between sales (i.e. Gross Output valued at current price) and total cost in country n, defined as:

$$\mu_{nt} = \frac{P_{nt} GO_{nt}}{TC_{nt}}. \tag{4.2}$$

There are many reasons why μ_{nt} could be above one. For example, monopoly power could allow gross output price to be above its marginal cost. Alternatively, any tax collected on value added and passed on prices would also imply $\mu_{nt} > 1$.

Extensive margin We introduce love of variety in gross output production in the form of a Dixit-Stiglitz aggregation of many varieties of imported inputs. Let the imported input bundle X_{nt} be a

[6] This quote is taken from Arrow (1974), pp 4-5.

CES aggregate of $\mathcal{M}_{nt}$ varieties, such that:

$$X_{nt} = \left(\int_0^{\mathcal{M}_{nt}} x_i^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} . \tag{4.3}$$

Assuming foreign producers are symmetric and denoting by x_{nt} their (common) production level, X_{nt} reduces to $X_{nt} = \mathcal{M}_{nt}^{\sigma/(\sigma-1)} x_{nt} = \mathcal{M}_{nt}^{1/(\sigma-1)} \cdot \mathcal{M}_{nt} x_{nt}$. Moreover, denoting by p_{nt}^x the (common) price of a given variety, the ideal price index dual to the CES aggregation is given by $\mathcal{P}_{nt} = \mathcal{M}_{nt}^{1/(\sigma-1)} \cdot p_{nt}^x$. Denoting $\widehat{Y}_t = \frac{\Delta Y_t}{Y_{t-1}} \approx d\ln(Y_t)$ the proportional change of any variable Y, changes in the imported input bundle can be expressed as:

$$\widehat{X}_{nt} = \underbrace{\widehat{\mathcal{M}_{nt} x_{nt}}}_{\text{Change in total imports}} + \underbrace{\frac{1}{\sigma-1} \widehat{\mathcal{M}}_{nt}}_{\text{Entry/Exit Effect}} , \tag{4.4}$$

In equation (4.4), the first term is simply the change in total imported inputs. It is completed by a second term that measures the additional variation in X_{nt} associated with changes in the number of available varieties. As discussed in Feenstra (1994), when the production function exhibits love of variety, any increase in the mass of input suppliers leads to a surge in efficiency. As we will see, this channel amplifies the quantitative impact of imported input movements on (measured) real GDP fluctuations.

Value Added We define two concepts of real value added in this economy. First, in line with Sims (1969) and others, we define *Physical Value Added* (PVA) implicitly from the production function as $PVA_{nt} = Z_{nt} L_{nt}^{\alpha} K_{nt}^{1-\alpha}$. Second, we follow the procedure used by statistical agencies and construct real GDP (RGDP) using double deflation. More precisely, RGDP growth between $t-1$ and t is constructed by valuing quantity changes with $t-1$ prices. Proportional changes of the two real value added indices can be expressed as:[7,8]

$$\text{Physical Value Added}: \quad \widehat{PVA}_{nt} \quad = \widehat{Z}_{nt} + \alpha \widehat{L}_{nt} + (1-\alpha)\widehat{K}_{nt}\,, \tag{4.5}$$

$$\text{Real GDP}: \quad \widehat{RGDP}_{nt} \quad = \frac{P_{nt-1}\Delta GO_{nt} - p^x_{nt-1}\Delta(\mathcal{M}_{nt}x_{nt})}{P_{nt-1}GO_{nt-1} - p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}\,. \tag{4.6}$$

Proposition 3. *Consider a production economy described by (4.1) to (4.3) and two definitions of real value added described by (4.5) and (4.6). Real GDP and Physical Value Added are related by:*

$$\widehat{RGDP}_{nt} = \omega_{nt-1}\left[\gamma\Big(\underbrace{\widehat{PVA}_{nt}}_{\text{Physical Value Added}}\Big) + \frac{1-\gamma}{\mu_{nt-1}}\Big(\underbrace{(\mu_{nt-1}-1)\cdot\widehat{X}_{nt}}_{\text{Markup Effect}} + \underbrace{\frac{1}{\sigma-1}\widehat{\mathcal{M}}_{nt}}_{\text{Variety Effect}}\Big)\right]. \tag{4.7}$$

Proof: Equation (4.6) can be written as:

$$\begin{aligned}
\widehat{RGDP}_{nt} &= \underbrace{\frac{P_{nt-1}GO_{nt-1}}{P_{nt-1}GO_{nt-1} - p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}}_{=\,\omega_{nt-1}}\left[\frac{\Delta GO_{nt}}{GO_{nt-1}} - \frac{p^x_{nt-1}\Delta(\mathcal{M}_{nt}x_{nt})}{P_{nt-1}GO_{nt-1}}\right] \\
&= \omega_{nt-1}\left[\gamma\left(\widehat{Z}_{nt} + \alpha\widehat{L}_{nt} + (1-\alpha)\widehat{K}_{nt}\right) + (1-\gamma)\widehat{X}_{nt} - \frac{p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}{P_{nt-1}GO_{nt-1}}\widehat{\mathcal{M}_{nt}x_{nt}}\right].
\end{aligned}$$

Using the definition of PVA as well as equation (4.4) then leads to:

$$\begin{aligned}
\widehat{RGDP}_{nt} \quad = \quad \omega_{nt-1}\Bigg[\gamma\left(\widehat{PVA}_{nt}\right) + \left((1-\gamma) - \frac{p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}{P_{nt-1}GO_{nt-1}}\right)\widehat{X}_{nt} \\
+ \frac{p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}{P_{nt-1}GO_{nt-1}}\frac{1}{\sigma-1}\widehat{\mathcal{M}}_{nt}\Bigg].
\end{aligned} \tag{4.8}$$

The definition of μ_{nt} in (4.2), provides a relationship between the intermediate input share of total cost, $(1-\gamma)$, and the share of total sales, $\frac{p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}{P_{nt-1}GO_{nt-1}}$, such that:

$$\frac{p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}{P_{nt-1}GO_{nt-1}} = \frac{1-\gamma}{\mu_{nt-1}}\,. \tag{4.9}$$

[7]As discussed in Burstein and Cravino (2015), the BEA does not use $t-1$ prices to construct real GDP, but rather a Fisher chain-weighted price index, according to:

$$\widehat{RGDP}_{nt} = \left(\frac{P_{t-1}GO_t - P^X_{t-1}X_t}{P_{t-1}GO_{t-1} - P^X_{t-1}X_{t-1}}\right)^{0.5}\left(\frac{P_tGO_t - P^X_tX_t}{P_tGO_{t-1} - P^X_tX_{t-1}}\right)^{0.5}.$$

Intuitively, the Fisher index is a geometric average between two base period pricing methods, alternatively using $t-1$ and t prices. We simplify the discussion and use $t-1$ prices, also known as the Laspeyres index.

[8]Equations (4.5) and (4.6) are expressed in terms of growth rate (and not in *levels*), which is consistent with all our quantitative results in section 4.4 where real GDP is HP-filtered. In practice, the *level* of RGDP at time t is constructed iteratively using the level at $t-1$ and the the growth rate as defined in (4.6).

Finally, using (4.9) in (4.8) delivers equation (4.7). ■

Discussion Two features are worth noting in proposition 3. First, the scaling term ω_{nt-1} is the ratio of sales over nominal value added, which is a standard Domar weight. From an accounting perspective, one can decompose total sales into total cost and profits (Π_{nt}), such that: $P_{nt-1}GO_{nt-1} = w_{nt-1}L_{nt-1} + r_{nt-1}K_{nt-1} + p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1}) + \Pi_{nt-1}$. When total sales equal total cost, $\Pi_{nt-1} = 0$ and the Domar weight is simply equal to the inverse of the value added share in gross output, $\omega_{nt-1} = \frac{P_{nt-1}GO_{nt-1}}{w_{nt-1}L_{nt-1}+r_{nt-1}K_{nt-1}} = 1/\gamma$. In the presence of a wedge between sales and cost, we can use equations (4.2) and (4.9) to rewrite the Domar weight as:

$$\omega_{nt-1} = \frac{P_{nt-1}GO_{nt-1}}{P_{nt-1}GO_{nt-1} - p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})} = \frac{1}{1 - \frac{p^x_{nt-1}(\mathcal{M}_{nt-1}x_{nt-1})}{P_{nt-1}GO_{nt-1}}} = \frac{\mu_{nt-1}}{\gamma + \mu_{nt-1} - 1}.$$

As long as μ_{nt-1} is close to one, the Domar weight ω_{nt-1} is close to $1/\gamma$. However, when profits are large and $\mu_{nt-1} > 1$, we have $\gamma\omega_{nt-1} < 1$ and equation (4.7) implies that $RGDP$ reacts less than one-for-one with PVA. Second, Real GDP as measured by statistical agencies and Physical Value Added are identical if there is no price distortions ($\mu_{nt} = 1$) and there is no love of variety in the production function ($\sigma = +\infty$).

Practical Implications The difference between physical value added and measured real GDP as expressed in (4.7) has important implications for the interpretation of real GDP fluctuations and for our understanding of international business cycle synchronization. With $\mu_{nt} > 1$ and $\sigma < +\infty$, real GDP as measured in the data is not only tied to movements in technology or factor supply, but also reflects changes in the quantity and variety of imported inputs. If both $\widehat{X}_{nt}$ and $\widehat{\mathcal{M}}_{nt}$ fluctuate with foreign technology shocks, equation (4.7) implies that an increase in the share of imported input in domestic production (i.e. a decrease in γ) raises the association between foreign shocks and domestic real GDP. International macroeconomic models that identify real GDP to physical value added cannot account for this relationship.

To better see this point, consider first a model with perfect competition and no love of variety, meaning that $\mu_{nt} = 1$ and $\sigma = +\infty$. In this case, equation (4.7) shows that real GDP and physical value added are identical and real GDP fluctuations can only arise from changes in factor supplies and changes in technology. When simulating such a model and assuming exogenous technology shocks, a researcher would find that foreign shocks could impact domestic real GDP only to the extent that it affects factor supply. This simple observation is at the heart of the negative result presented in Kehoe and Ruhl (2008): in a model where firms take prices as given, profit maximization insures that the marginal benefit of using an additional unit of imported input is equal to its marginal cost. Hence, foreign shocks can only affect real GDP to the extent that it triggers a change in domestic factor supply. This result lies at the heart of the trade co-movement puzzle and explains why trade is not a powerful channel of propagation in standard IRBC models. In frameworks where real GDP is equal to physical value added, real GDP changes in response to a foreign shock can only arise from variations in factors supply which, in turn, are disciplined by (i) the elasticity of domestic factor supply and (ii) the complementarity between domestic factors and foreign inputs.[9] As shown in Johnson (2014), complementarity in production factors *alone* is not sufficient to solve quantitatively the TCP.

Consider now a situation where $\mu_n > 1$ and $\sigma < +\infty$. Equation (4.7) reveals that changes in intermediate input usage have a first order impact on real GDP fluctuations beyond the movements of domestic factors or technology. In such a case, imported inputs yield more gains than what is reflected in their price and using more foreign inputs is associated with profits (when $\mu_n > 1$) or with efficiency gains (when $\sigma < +\infty$).[10] All told, constructing real GDP using "base period prices" that do not reflect imported inputs' marginal revenue product creates a wedge between physical value added and real GDP fluctuations.

Importantly, the disconnect between real GDP and physical value added does not rely on the cyclicality of markups: even with constant markups, a wedge between the marginal cost and marginal revenue product of imported inputs leads to a first order impact of intermediate input usage on measured real value added. In a sense, the wedge is a purely measurement issue: when constructing real GDP, statistical agencies do not simply measure the quantity of goods produced, but use observed prices (fixed at a base period) to assign a value to measured quantities. If base period prices used in valuing quantities contain a markup, it creates a wedge between the marginal revenue generated by an additional unit of imported input x_n and its marginal cost p_n^X.

Finally, our results bears important implications for the calibration of macroeconomic models. As revealed by proposition 3, a researcher who builds a model with distorted prices cannot equate a model-based measure of physical value added to real GDP data in the calibration process: doing so would attribute the markup and variety effects in (4.7) to changes in PVA, either through technology shocks ($\widehat{Z}_{nt}$) or through factor supply ($\widehat{L}_{nt}$ or $\widehat{K}_{nt}$).[11]

4.2.2 Productivity and technology

The real GDP decomposition presented in equation (4.7) also bears important implications for the measure of productivity based on the Solow Residual (SR). As standard, we define SR so that it

[9]The role of input complementarity is discussed at length in Burstein et al. (2008) or in Boehm et al. (2019).

[10]In practice, many models feature a single parameter governing both the size of the markup and the degree of Love of Variety – this is obviously the case with CES aggregation and monopolistic competition. For clarity, equation (4.7) shows a specification in which the two channels are perfectly distinguishable.

[11]In the present paper, we focus on the implication of precise real GDP definition for the resolution of the Trade Comovement Puzzle and hence emphasize the role of *imported* inputs. Our results can also be extended to a closed economy with multiple sectors and input-output linkages. When real GDP is computed at the sector level, proposition 3 holds and implies a disconnect between real GDP and PVA with markups and/or love of variety.

captures fluctuations in real GDP that are not explained by changes in domestic factors:

$$\widehat{SR}_{nt} = \widehat{RGDP}_{nt} - \alpha \widehat{L}_{nt} - (1-\alpha)\widehat{K}_{nt} \,. \tag{4.10}$$

Proposition 4. *When real GDP is constructed as in (4.6), the relationship between SR and technology is given by:*

$$\begin{aligned} \widehat{SR}_{nt} &= \gamma\omega_{nt-1} \cdot \widehat{Z}_{nt} + \underbrace{(\gamma\omega_{nt-1} - 1) \cdot \left(\alpha \widehat{L}_{nt} + (1-\alpha)\widehat{K}_{nt}\right)}_{\text{Scale Effect}} \\ &\quad + \frac{\omega_{nt-1} \cdot (1-\gamma)}{\mu_{nt-1}} \Bigg(\underbrace{(\mu_{nt-1} - 1) \cdot \widehat{X}_{nt}}_{\text{Imported Input Effect}} + \underbrace{\frac{1}{\sigma - 1} \widehat{\mathcal{M}}_{nt}}_{\text{Variety Effect}} \Bigg) . \end{aligned} \tag{4.11}$$

Proof: The result follows from replacing (4.7) in (4.10). ■

Discussion When $\mu_n = 1$, we have $\omega_{nt-1} = 1/\gamma$ and both the scale effect and markup effect terms in equation (4.11) vanish. Additionally, in absence of love of variety ($\sigma = +\infty$), the variety effect also disappears, implying that productivity is an accurate measure of technology.

In general, when $\mu_n > 1$ and $\sigma < +\infty$, equation (4.11) makes it clear that productivity, as measured by the Solow Residual, is disconnected from the true technology shock Z_{nt}. First, with positive markups, fluctuations in real GDP can result from movement in profits which are captured by Solow Residual fluctuations. Such profits movements can arise either from changes in domestic factors (the scale effect) or from changes in foreign input usage (the imported input effect). Second, an additional term captures the gains associated with accessing more variety from abroad whenever $\sigma < +\infty$. With love of variety, this change in efficiency is also reflected in measured productivity.

All told, the above decomposition highlights the disconnect between standard measures of productivity, such as the *Solow Residual*, and actual technology. The introduction of markups and love of variety creates new channels through which foreign shocks impact measured domestic productivity. As a result, two countries that trade intermediate inputs should have correlated Solow Residuals, a prediction we later test in the data and which our quantitative model is able to reproduce. Finally, our results can be seen as continuation of insights from Basu and Fernald (2002) or Feenstra et al. (2009) who highlight the risk of identifying the Solow Residual to technology shocks in the calibration of a macroeconomic model.

4.3 A Model of International Trade with Cross-Border Input Linkages

We now put more structure on our insights and quantitatively assess the role of markups and love of variety, in conjunction with a statistically-consistent real GDP measurement, in generating a plausible trade comovement slope. We depart from the standard IRBC model and develop a many-country international business cycle model that features trade in both final and intermediate goods, imperfect competition and extensive margin adjustments. The model is related to Ghironi and Melitz (2005) and Alessandria and Choi (2007) extended to multiple countries with homogeneous firms that are able to export and import, which implies that intermediate goods cross borders multiple times.[12]

4.3.1 Consumption and Labor Supply

We consider a multi-period world economy with many countries ($i, j \in \{1, ..., N\}$). Each country is populated by a representative consumer who consumes final goods and supplies labor $L_{i,t}$ for production. Consumers' preferences are described by the following utility function:

$$U_0 = \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log\left(C^F_{i,t}\right) - \frac{\chi_i}{1+\nu} L_{i,t}^{1+\nu} \right], \tag{4.12}$$

$$\text{with} \quad C^F_{i,t} = \left(\sum_{j=1}^{N} \omega^F_i(j)^{\frac{1}{\rho^F}} \cdot C_{j,i,t}^{\frac{\rho^F-1}{\rho^F}} \right)^{\frac{\rho^F}{\rho^F-1}}, \quad \text{and} \quad C_{j,i,t} = \left(\int\limits_{s\in\Omega^F_{j,i,t}} c_{j,i,t}(s)^{\frac{\sigma_j-1}{\sigma_j}} ds \right)^{\frac{\sigma_i}{\sigma_i-1}}, \tag{4.13}$$

where χ_i is a scaling parameter, β is the rate of time preference, ν the inverse of the Frisch elasticity of labor supply and σ_i the elasticity of substitution between different varieties of final goods originating from country i. $\omega^F_i(j)$ measures the share of country j in the *consumption* bundle of country i, and $\Omega^F_{j,i,t}$ is the endogenous set of firms from country j that serve the *final good* market in country i. Finally, ρ^F is the *final goods* Armington elasticity of substitution. Final good price indexes are defined as:

$$\mathcal{P}^F_{i,t} = \left(\sum_{j=1}^{N} \omega^F_i(j) \cdot \left(\widetilde{\mathcal{P}}^F_{j,i,t}\right)^{1-\rho^F} \right)^{\frac{1}{1-\rho^F}}, \quad \text{and} \quad \widetilde{\mathcal{P}}^F_{j,i,t} = \left(\int\limits_{s\in\Omega^F_{j,i,t}} p^F_{j,i,t}(s)^{1-\sigma_i} ds \right)^{\frac{1}{1-\sigma_i}}, \tag{4.14}$$

where $p^F_{j,i,t}(s)$ is the price charged by firm s in the set $\Omega^F_{j,i,t}$ when selling in the *final good* market in country i. As we will see below, given our assumptions, firms charge the same price in both final

[12] Alternatively, the model presented here can be thought of as an extension of the IRBC model presented in Johnson (2014) with two new elements: markups and extensive margin adjustments. It is also related to the static small open economy model in Gopinath and Neiman (2014)

and intermediate good markets in a given country.

Asset markets Our benchmark economy assumes financial autarky between countries, so that agents choose consumption, investment and labor, subject to:[13,14]

$$\mathcal{P}^F_{i,t}\left(C_{i,t}+\mathcal{I}_{i,t}\right)=w_{i,t}L_{i,t}+r_{i,t}K_{i,t}-\mathcal{T}_i\,, \tag{4.15}$$

$$K_{i,t+1}=(1-\delta)K_{i,t}+\mathcal{I}_{i,t}\left[1-\Psi\left(\frac{\mathcal{I}_t}{\mathcal{I}_{t-1}}\right)\right]\,, \tag{4.16}$$

where the term $\mathcal{T}_i$ captures potential trade imbalance in country i, i.e. $\mathcal{T}_i<0$, corresponds to a trade deficit meaning that country i consumes more than the value of its production. Following Christiano et al. (2005), we introduce investment adjustment costs that satisfy the following properties: $\Psi(1)=\Psi'(1)=0$ and $\Psi''(1)>0$. We use the following function form: $\Psi\left(\frac{I_t}{I_{t-1}}\right)=\frac{\psi}{2}\left(\frac{I_t}{I_{t-1}}-1\right)^2$, where ψ governs the degree of the adjustment costs. Given prices, consumers choose $\{C_{i,t},L_{i,t},K_{i,t+1}\}$ to maximize (4.12) subject to (4.15) and (4.16).

4.3.2 Production side

In country i, production is performed by a continuum of homogeneous firms with productivity $Z_{i,t}$. Firms produce with a Cobb-Douglas technology using labor $\ell_{i,t}$, capital $k_{i,t}$ and intermediate inputs $I_{i,t}$ bought from both home and foreign firms. The intermediate input index, $I_{i,t}$, is a CES aggregation of country specific bundles $M_{j,i,t}$, with an *intermediate goods* Armington elasticity ρ^I. Each country specific bundle is itself a CES aggregation of many varieties, with an elasticity of substitution σ_j. Production technology for a firm in i writes:

$$q_{i,t}=\left(Z_{i,t}\ell^{\alpha}_{i,t}k^{1-\alpha}_{i,t}\right)^{\gamma_i}I^{1-\gamma_i}_{i,t}\,, \tag{4.17}$$

$$\text{with}\quad I_{i,t}=\left(\sum_{j=1}^{N}\omega_i(j)^{\frac{1}{\rho^I}}M_{j,i,t}^{\frac{\rho^I-1}{\rho^I}}\right)^{\frac{\rho^I}{\rho^I-1}},\quad\text{and}\quad M_{j,i,t}=\left(\int_{s\in\Omega_{j,i,t}}m_{j,i,t}(s)^{\frac{\sigma_i-1}{\sigma_i}}ds\right)^{\frac{\sigma_i}{\sigma_i-1}}, \tag{4.18}$$

where γ_i is the share of value added in gross output, $\omega^I_i(j)$ measures the share of country j in the *production process* of country i, and $\Omega^I_{j,i,t}$ is the endogenous set of firms based in j and serving

[13] Heathcote and Perri (2002) have shown that IRBC models with financial autarky yield a closer fit to some key business cycle moments than a version under complete markets. In Appendix 4.B, we present a version with complete financial markets. In such case, the magnitude of the Trade Comovement slope is only slightly lower and the main message is qualitatively similar.

[14] Note that the right hand side of (4.15) includes firms' profits since, as explained below, firms pay entry costs using domestic labor. It should then be understood that $L_{i,t}$ includes both production and "entry cost" workers.

the *intermediate input* market in country i.[15] Similarly to the *final good* market, we have

$$\mathcal{P}^I_{i,t} = \left(\sum_{j=1}^{N} \omega^I_i(j) \left(\widetilde{\mathcal{P}}^I_{j,i,t} \right)^{1-\rho^I} \right)^{\frac{1}{1-\rho^I}}, \quad \text{and} \qquad \widetilde{\mathcal{P}}^I_{j,i,t} = \left(\int_{s\in\Omega^I_{j,i,t}} p^I_{j,i,t}(s)^{1-\sigma_i} ds \right)^{\frac{1}{1-\sigma_i}}, \tag{4.19}$$

$$\text{and } \mathcal{P}^{IB}_{i,t} = \left(\frac{w_{i,t}}{\alpha\gamma_i} \right)^{\alpha\gamma_i} \left(\frac{r_{i,t}}{(1-\alpha)\gamma_i} \right)^{(1-\alpha)\gamma_i} \left(\frac{\mathcal{P}^I_{i,t}}{1-\gamma_i} \right)^{1-\gamma_i}, \tag{4.20}$$

where $\mathcal{P}_{j,i,t}$ denotes the price of the country-pair specific bundle $M_{j,i,t}$ and $\mathcal{P}^{IB}_{i,t}$ is the unit cost of the Cobb Douglas bundle aggregating $I_{i,t}$, $k_{i,t}$ and $\ell_{i,t}$ (called the *input bundle*) and represents the price of the basic production factor in country i. $p^I_{j,i,t}(s)$ is the price charged by any firm s in the set $\Omega^I_{j,i,t}$ when selling in the *intermediate input* market in country i.

Finally, there is an overhead entry cost f^E_i, sunk at the production stage. Based on their expected profit in all markets, firms enter the economy until the expected value of doing so equals the overhead entry cost. This process determines the mass of firms $M_{i,t}$.

4.3.3 Equilibrium

Let us define $X_{i,t}$, the aggregate consumers' revenue, and $S_{i,t}$, the total firms' spendings in country i. Given prices, total demand faced by a firm in country i is the sum of demand stemming from *final good* and *intermediate input* markets in all countries:

$$q_{i,t} = \underbrace{\sum_j \left(\frac{p^F_{i,j,t}}{\widetilde{\mathcal{P}}^F_{i,j,t}} \right)^{-\sigma_i} \left(\frac{\widetilde{\mathcal{P}}^F_{i,j,t}}{\mathcal{P}^F_{j,t}} \right)^{-\rho^F} \frac{\omega^F_j(i) X_{j,t}}{\mathcal{P}^F_{j,t}}}_{\text{Final goods demand}} + \underbrace{\sum_j \left(\frac{p^I_{i,j,t}}{\widetilde{\mathcal{P}}^I_{i,j,t}} \right)^{-\sigma_i} \left(\frac{\widetilde{\mathcal{P}}^I_{i,j,t}}{\mathcal{P}^I_{j,t}} \right)^{-\rho^I} \frac{\omega^I_j(i)(1-\gamma_j) S_{j,t}}{\mathcal{P}^I_{j,t}}}_{\text{Intermediate goods demand}}. \tag{4.21}$$

Firms choose their price to maximize profits. With constant price elasticity of demand, they charge a constant markup over marginal cost. For a firm from country i, the only elasticity that is relevant for pricing is σ_i, capturing the fact that their individual pricing decision has no impact on country-specific price indexes. As a result, firms charge the same markup in the final and intermediate good markets, and we have: $p^F_{i,j,t} = p^I_{i,j,t} = p_{i,j,t}$ and $\widetilde{\mathcal{P}}^F_{i,j,t} = \widetilde{\mathcal{P}}^I_{i,j,t} = \widetilde{\mathcal{P}}_{i,j,t}$. The

[15] In an earlier version of this paper, we also introduced heterogeneity with firm's idiosyncratic productivity as in Ghironi and Melitz (2005) or Fattal Jaef and Lopez (2014). The results are quantitatively similar to those obtained in this version and we therefore dropped this layer of complexity.

marginal cost of a firm in country i is $\mathcal{P}_{i,t}^{IB}/(Z_{i,t}^{\gamma_i})$ and its optimal price in country j is:

$$p_{i,j,t} = \tau_{ij} \frac{\sigma_i}{\sigma_i - 1} \frac{\mathcal{P}_{i,t}^{IB}}{Z_{i,t}^{\gamma_i}} . \tag{4.22}$$

Unlike Krugman (1980) or Ghironi and Melitz (2005), one needs to jointly solve for all prices in the economy. Through $\mathcal{P}_{i,t}^{IB}$, the price charged by a firm in country i depends on the prices charged by all firms supplying country i which in turn depend on the prices charged by their suppliers and so on and so forth. Determining prices requires solving jointly for all country-pair specific price indexes. Using the fact that $\widetilde{\mathcal{P}}_{i,j,t} = \tau_{ij}\widetilde{\mathcal{P}}_{i,i,t}$, the definition of price indexes in every country yields a system of N equations which jointly defines all *inner* price indexes:

$$(\widetilde{\mathcal{P}}_{i,i,t})^{1-\rho^I} = \mu_i \left(\sum_{j=1}^{N} \omega_i^I(j) \left(\tau_{ji} \widetilde{\mathcal{P}}_{j,j,t} \right)^{1-\rho^I} \right)^{1-\gamma_i} , \tag{4.23}$$

with μ_i depending on the mass of firms and parameters.[16] For given mass of firms, this system admits a unique non-negative solution.[17]

Finally, the mass of firms is determined by the free entry condition defined as:

$$\Pi_{i,t} = M_{i,t} \frac{w_{i,t}}{Z_{i,t}^{\gamma_i}} . f_i^E \quad \text{for all } i , \tag{4.24}$$

where the sunk cost f_i^E is labeled in labor units and $\Pi_{i,t}$ dis aggregate profits in country i.

Closing the model involves standard market clearing conditions for capital, labor and goods. Total revenues of all firms from country i can be written as:

$$R_{i,t} = \sum_{j=1}^{N} \left[\left(\frac{\widetilde{\mathcal{P}}_{i,j,t}}{\mathcal{P}_{j,t}^F} \right)^{1-\rho^F} \omega_j^F(i) X_{j,t} + \left(\frac{\widetilde{\mathcal{P}}_{i,j,t}}{\mathcal{P}_{j,t}^I} \right)^{1-\rho^I} \omega_j^I(i)(1-\gamma_j) S_{j,t} \right] . \tag{4.25}$$

Total exports are the sum of final goods and intermediate inputs exports, defined as:

$$T_{i \to j} = \left(\frac{\widetilde{\mathcal{P}}_{i,j,t}}{\mathcal{P}_{j,t}^F} \right)^{1-\rho^F} \omega_j^F(i) X_{j,t} + \left(\frac{\widetilde{\mathcal{P}}_{i,j,t}}{\mathcal{P}_{j,t}^I} \right)^{1-\rho^I} \omega_j^I(i)(1-\gamma_j) S_j . \tag{4.26}$$

Consumer's revenues $X_{i,t}$ are equal to the sum of the payment to production workers $\alpha\gamma_i S_{i,t}$, rent from capital $(1-\alpha)\gamma_i S_{i,t}$, total firms' profits $\Pi_{i,t}$ (which, at the free entry equilibrium, is

[16]From equation (4.19) and (4.22), we have: $\mu_i^{\frac{1-\sigma_i}{1-\rho^I}} = M_{i,t} \left(\frac{\sigma_i}{\sigma_i - 1} \left(\frac{w_{i,t}}{\alpha\gamma_i} \right)^{\alpha\gamma_i} \left(\frac{r_{i,t}}{(1-\alpha)\gamma_i} \right)^{(1-\alpha)\gamma_i} \left(\frac{1}{1-\gamma_i} \right)^{1-\gamma_i} \frac{1}{Z_{i,t}^{\gamma_i}} \right)^{1-\sigma_i}$.

[17]Following Kennan (2001) and denoting $G_k = (\widetilde{\mathcal{P}}_{i,i,t})^{1-\rho^I}$ and G the associated $N \times 1$ vector, it suffices to show that the system is of the form $G = f(G)$ with $f : \mathbb{R}^N \to \mathbb{R}^N$ a vector function which is strictly concave with respect to each argument, which is obvious as long as $0 < \gamma_i < 1$ for all countries.

completely used to pay the entry cost f_i^E), and potential trade imbalances $-\mathcal{T}_{i,t}$. Moreover, in absence of fixed cost of exporting, note that both profits and spending can be expressed as a function of revenues with $\Pi_{i,t} = \frac{1}{\sigma_i} R_{i,t}$ and $S_{i,t} = \frac{\sigma_i - 1}{\sigma_i} R_{i,t}$. Using $X_{i,t} = w_{i,t} L_{i,t} + r_{i,t} K_{i,t} - \mathcal{T}_{i,t} = \gamma_i S_{i,t} + \Pi_{i,t} - \mathcal{T}_{i,t}$, equation (4.25) can be written in compact form as:

$$\mathbf{G} \cdot \mathbf{R}' = - \left[\mathbf{W}_F' \circ \mathbf{P}_F \right] \boldsymbol{\mathcal{T}}', \tag{4.27}$$

where $\boldsymbol{\mathcal{T}}$ and $\mathbf{R}$ are vectors that stack respectively trade imbalances and total revenues of all firms. $\mathbf{W}_F$ s the weighting matrix associated with final good aggregation and whose elements are defined as $W_{ij}^F = \omega_i^F(j)$, $\mathbf{P}_F$ is a matrix defined by elements $P_{i,j,t}^F = \left(\frac{\widetilde{\mathcal{P}}_{i,j,t}}{\mathcal{P}_{i,t}^F} \right)^{1-\rho^F}$, and $\circ$ is the element-wise (Hadamard) product. The matrix $\mathbf{G}$ is defined at any time t as:

$$G_{i,j,t} = \mathcal{I}_{i,j} - \left(\frac{\widetilde{\mathcal{P}}_{i,j,t}}{\mathcal{P}_{j,t}^F} \right)^{1-\rho^F} \omega_j^F(i) \frac{\gamma_j \cdot (\sigma_j - 1) + 1}{\sigma_j} - \left(\frac{\widetilde{\mathcal{P}}_{i,j,t}}{\mathcal{P}_{j,t}^I} \right)^{1-\rho^I} \omega_j^I(i)(1-\gamma_j) \left(\frac{\sigma_j - 1}{\sigma_j} \right) \tag{4.28}$$

Finally, labor can be used either for production ($L_{i,t}^p$) or for the entry cost ($L_{i,t}^e$) so that $L_{i,t} = L_{i,t}^p + L_{i,t}^e$. Setting $w_1 = 1$, implying $S_1 = L_1^p / (\alpha_1 \gamma_1)$, provides a unique solution for all variables by solving together the consumer problem (4.12), the price system (4.23), the Free Entry system (4.24) and the Revenue system (4.27).

4.3.4 Real Value Added definitions

As in section 4.2, we introduce two measures of value added: a *model-based* measure of physical value added (PVA) and a *statistical* measure of real GDP ($RGDP$). Only the latter index is comparable to the data produced by statistical agencies.

Physical Value Added Thanks to separability between domestic factor and inputs in the firm-level production function (4.17), aggregating physical value added across all firms yields $PVA_{i,t} = Z_{i,t} L_{i,t}^{\alpha} K_{i,t}^{1-\alpha}$. This measure of real value added is unit-less and, following Arrow (1974), one can interpret it as a measure of a purely theoretical bundle that is used, in combination with intermediate input, to produce the gross output.

Real GDP Real GDP as constructed by statistical agencies is not unit-less but uses prices at their base period level to express each component in a commonly accepted unit of account.[18]

In most databases, real GDP is defined using the Fisher ideal quantity index which is a geometric mean of the Laspeyres and Paasche indices. Hence, for any period t, the base period price

[18] In most cases, real GDP is constructed using chain weighted prices, as discussed here. Some database report real GDP in "constant prices" where prices used in the construction of real value added are fixed at a reference year. Obviously, no database reports real GDP by "counting the number of goods" produced in a country.

used in the construction of real GDP growth from $t-1$ to t is a geometric mean between period t and period $t-1$ prices. To be as close as possible to the method used in the construction of the data while simplifying the analysis, we define real GDP (RGDP) using steady state prices as base-period prices. Real GDP is obtained by deflating nominal spending using price indexes that are corrected from product variety effects to measure "quantity indices", and then by valuing these "quantity indices" using steady-state prices. More precisely:

$$RGDP_{i,t} = \underbrace{\underbrace{\widehat{\mathcal{P}_i^{F,ss}} \frac{X_{i,t}}{\widehat{\mathcal{P}_{i,t}^{F}}}}_{\text{Consumption + Investment}} + \underbrace{\sum_j \widehat{\mathcal{P}_{i,j}^{ss}} \frac{T_{i\to j,t}}{\widehat{\mathcal{P}_{i,j,t}}}}_{\text{Total exports (final+ inputs)}}}_{\text{= Gross Output + Imported Final Goods}} - \underbrace{\sum_j \widehat{\mathcal{P}_{j,i}^{ss}} \frac{T_{j\to i,t}}{\widehat{\mathcal{P}_{j,i,t}}}}_{\text{Total imports (final + inputs)}} \tag{4.29}$$

where, in order to be consistent with the way actual data are collected, we defined variety-corrected price indexes as $\widehat{\mathcal{P}_{i,j,t}} = (M_{i,t})^{1/(\sigma_i-1)} \widetilde{\mathcal{P}}_{i,j,t}$ and $\widehat{\mathcal{P}_{i,t}^{F}} = \left(\sum_j \omega_i^F(j) \cdot \left(\widehat{\mathcal{P}_{j,i,t}}\right)^{1-\rho^F}\right)^{\frac{1}{1-\rho^F}}$. Since both consumers' utility and production functions have a CES component, it is well known that the associated price indexes can be decomposed into components reflecting average prices (captured by statistical agencies) and product variety (which is not taken into account in national statistics).[19]

In equation (4.29), we defined RGDP from the expenditure side. One could also define RGDP from the production side, by summing gross domestic output sold in all markets and subtracting imported inputs. All our results are unchanged using such alternative measure.

Taking the Model to Data We conclude the model section by discussing two broad points about the quantitative exercise that deserve more attention. First, it is important to note that IRBC models where the production function is expressed in value-added terms use physical value-added as a measure of real GDP. In the context of the TCP literature, papers such as Kose and Yi (2006), Burstein et al. (2008) and Johnson (2014) evaluate the Trade-Comovement slope using PVA. In these perfectly competitive models, our discussion in section 4.2 shows that $RGDP$ and PVA are equal. A notable exception is Liao and Santacreu (2015) who find that including an extensive margin component can account for part of the Trade-Comovement Slope. In their simulations, the TC slope obtained when using a statistical measure of GDP is only 37% of our empirical estimates, and is mostly generated by real effects while we highlight the importance of a measurement channel.[20] In this paper, we formally compare both real value added measures and investigate the difference in their properties. This distinction clarifies how, in the presence of imperfect compe-

[19]See for example the illuminating discussion in Feenstra (1994) or Ghironi and Melitz (2005).

[20]The 37% is obtained using Liao and Santacreu (2015)'s highest model-based slope with a consistent statistical measure of real GDP, in Table 8, pp 276. When using an elasticity of substitution of 3.1 and above (i.e. a markup lower or equal than 48%), their trade comovement slope using statistical GDP turns negative, as shown in tables 9 and 10 of the paper.

tition and extensive margin adjustments, $RGDP$ measured by statistical agencies is disconnected from a theory-consistent measure of physical value-added.

Second, our proposition 4 in section 4.2.2 shows that one cannot use the standard Solow Residual to calibrate the technology process. Indeed, the Solow Residual also captures changes in profits and changes in efficiency in response to movements of the number of foreign varieties. We circumvent this issue by calibrating the technology shocks in our model using real GDP targets: we set the shocks so that real GDP in our simulation matches moments observed in the real GDP data. In particular, our simulations deliver an average correlation of real GDP over all country-pairs equal to the co-movement observed in our sample.

4.3.5 Calibration of the Model Parameters

We solve and parameterize the model with 14 countries and a composite *Rest-Of-the-World* for the time period 1980-1990 using standard linearization techniques.[21] Table 4.3.1 reports fixed parameters and Table 4.3.2 reports parameters that are calibrated to match empirical moments. We now describe our parameterization in detail.

We first choose values for parameters that are common across countries. Starting with preferences, we set $\beta = 0.99$ and $\nu = 0.5$, leading to a Frisch elasticity of 2. Regarding the production function, we choose $\alpha = 0.67$ and $\delta = 0.025$. The macro (Armington) elasticities ρ^I and ρ^F are set to unity, which is in the range of the literature. For comparison, Saito (2004) provides estimations from 0.24 to 3.5 for the Armington elasticity.[22] There is also a theoretical convenience to use $\rho^I = \rho^F = 1$, as it allows the model to take the same form as other network models such as Acemoglu et al. (2012). The degree of investment adjustment costs, ψ, is chosen to obtain a volatility of investment with respect to real GDP consistent with the data.

We then move to country-specific parameters. χ_i is chosen to replicate the *relative difference* of working age population with a normalization ensuring an average capital-output ratio of 13 in the model. We set a value of $\sigma_i = \sigma = 4.3$, $\forall i$ for the micro elasticity of substitution in the baseline simulation. Anderson and van Wincoop (2004) reports estimates in the range of 3 to 10. Following Bernard et al. (2003), Ghironi and Melitz (2005) choose a micro elasticity of 3.8 and recently, papers such as Barrot and Sauvagnat (2016) or Boehm et al. (2019) argue that firms' ability to substitute between their suppliers can be very low. This choice leads to markups of 30%. As a robustness, we also consider alternative elasticities for σ_i in section 4.5.3, and defer discussion of those cases till then.

The value added shares, γ_i, are calibrated using data on cost of intermediates and total sales in

[21] The set of countries is: Australia, Austria, Canada, Denmark, France, Germany, Ireland, Italy, Japan, Mexico, Netherlands, RoW, Spain, United Kingdom and United States. They represent around 78% of total trade flows, 79% of total trade in final goods and 77% of total trade flows in intermediate goods.

[22] Studying macro and micro elasticities for final goods, Feenstra et al. (2014) finds that, for the majority of goods, the macro elasticity is lower than the micro elasticity.

Table 4.3.1. Fixed parameters of the model.

	Parameter	Value	Moment / Source
Discount factor	β	.99	Annual discount rate of 4%
Labor curvature	ν	0.5	Frisch elasticity of 2.0
Investment adjustment cost	ψ	1.25	Volatility of σ_I/σ_{RGDP} between 3 and 4.5
Labor Supply Scaling	χ_i	$[5.4-5, 0.16]$	Relative working age population
Labor share	α	0.67	Standard value
Argminton elasticities	ρ^I, ρ^F	1.0	Saito (2004), Feenstra et al. (2014)
Micro elasticity of substitution	$\sigma_i, \forall i$	4.35	Markup of 30%, De Loecker and Eeckhout (2018)
Sunk entry cost	f_i^E / f_{US}^E	[0.4 - 3.9]	*Doing Business Database* - World Bank
Fixed trade cost	f_{ij}^c	[3.3 - 18]	*Doing Business Database* - World Bank
Iceberg trade cost	τ_{ij}	[1 - 2.8]	ESCAP - World Bank

the WIOD database at the 2-digits sector level. Specifically, $(1-\gamma_{i,s}) = \frac{cost\_intermediates_s}{total\_sales_s}$, represents the share of intermediate inputs in total costs in a given sector. We use the fact that $total\ sales_s = \mu_i \times total\ cost_s$ with μ_i the markups in country i. Therefore, we fix $\gamma_{i,s} = 1 - \frac{cost\_intermediates_s}{total\_sales_s}\frac{\sigma_i}{\sigma_i-1}$. The implied mean values of γ_i, weighted by the sector importance in total sales, range from 0.31 to 0.45 for the considered countries (we set the value for *RoW* to the mean value), which is consistent with values reported in Halpern et al. (2015).

The sunk entry costs f_i^E are computed from the *Doing Business Indicators*.[23] We measure the relative entry fixed costs by using the information on the amount of time required to set up a business in the country relative to the US, where we normalize f_{US}^E in order to generate a ratio of total number of firms divided by the working population, $\frac{M}{L}$, of about 12%.[24]

Finally, we move to country-pair specific parameters. Variable (iceberg) trade costs $\tau_{ij} > \tau_{ii}$, are taken from the ESCAP World Bank: *International Trade Costs Database*, where we normalize $\tau_{ii} = 1$. This database features symmetric bilateral trade costs in its wider sense, including not only international transport costs and tariffs but also other trade cost components discussed in Anderson and van Wincoop (2004).

Data on bilateral trade flows, $\{T_{j\to i}^I/RGDP_i, T_{j\to i}^F/RGDP_i\}$, are sufficient to identify the shares $\omega_i^I(j)$ and $\omega_i^F(j)$. We use trade data from Johnson and Noguera (2017) dis-aggregated into final and intermediate goods. Moreover, since complete financial autarky is inconsistent with the trade balances observed in the data, we calibrate the model trade imbalances $\{\mathcal{T}_1, ..., \mathcal{T}_N\}$ to match trade imbalances relative to real GDP, and then hold those nominal imbalances constant during the

[23] As shown later, f_i^E plays a little role in the correlation between trade and real GDP comovement.

[24] There is about 22-24 millions of non-employer businesses and 5.5 millions of employer businesses in the US, while the working age population represents around 180 millions of individuals during the considered period. Consistently, the self-employment rate is around 12% for the US between 1990 and 2000 (BLS). Results are not sensitive to this assumption. We provide a comparison of this rate and the self-employment rate in each economy in the online appendix C.

simulation. By construction, the model's steady state matches relative bilateral trade flows and trade imbalances exactly.

Aggregate Technology Process The *level* of real GDP comovement in our simulations is driven both by correlated technology shocks and by the transmission of those shocks across countries. As discussed earlier, productivity measures available in usual macro database can not be used as a proxy for the country-specific technology process $Z_{i,t}$. To generate an international correlation of real GDP consistent with the data, we pin down the off diagonal elements of the technology covariance matrix so that the average correlation of real GDP in the model matches exactly the one observed in the data, which is 0.27. We also calibrate the variance $\sigma_Z(i)$ and persistence ρ_Z of technology shocks to match the (de-trended) real GDP volatility and an average auto-correlation of 0.84. This allows us to generate real GDP fluctuations in the simulated economy that are similar to those observed in the data.[25] The targeted moments reported in Table 4.3.2 are all perfectly matched.[26]

Table 4.3.2. Calibrated parameters of the model.

	Parameter	Value	Main target
Inputs spending weights	$\omega_i^I(j)$	in sup. app.	Import shares in inputs
Final goods spending weights	$\omega_i^F(j)$	in sup. app.	Import shares in final goods
Trade imbalance	$\{\mathcal{T}_i, ..., \mathcal{T}_N\}$	in sup. app.	Trade imbalance over GDP
Persistency of Techno. shocks	ρ_Z	.71	Avg. RGDP auto-correlation
Std. of Techno. shocks	$\sigma_Z(i)$	[.0012, .0050]	RGDP volatility (de-trended)
Covariance of Techno. shocks	$\sigma_Z(i,j), \forall i \neq j$	.22	Avg. RGDP correlation of 0.27

Interestingly, in order to match an observed international real GDP correlation ($RGDP$) of 0.27, the correlation of technology shocks is only 0.22. This implies that, according to our framework, trade linkages explain 20% of the international real GDP correlations ($= 0.27 - 0.22 = 0.05$), with shock correlation explaining the remaining 80%. In contrast, the international correlation of physical value added (PVA) is only 0.24, implying that the correlation of shocks explains more than 92% (=0.22/0.24) of PVA comovement. The rest of the paper clarifies the channels through which trade linkages propagates international shocks and synchronizes business cycle, by studying and decomposing the trade comovement slope.

4.3.6 Quantification of the Empirical Trade-Comovement Slope

To assess the quantitative relevance of markups and love of variety in conjunction with measurement procedure for real GDP, we need an empirical counterpart against which our model can be

[25] Again, recall that the goal of this exercise is not to explain the ***level*** of comovement across countries, but its ***slope*** following a change in trade intensities.

[26] We report standard business cycle statistics of our model in the online appendix, section C.

compared. To do so, we now update the seminal Frankel and Rose (1998) (henceforth, FR) analysis on the relationship between bilateral trade and GDP comovement.

We depart from FR in two ways. First, we decompose total trade into trade in intermediate inputs and trade in final goods. Second, we make use of a panel estimation and exploit within country-pair variations to estimate the relationship between changes in trade linkages and changes in GDP co-movement. As predicted by our results in section 4.2, trade in intermediate inputs is associated with an increase of both real GDP and measured productivity synchronization.

Our panel is composed of 40 countries from 1970 to 2009 and accounts for 90% of world GDP. We use real GDP measured at chained PPPs from the 9th Penn World Table, which is transformed in two ways: (i) HP filter with smoothing parameter 6.25 to capture the business cycle frequencies and (ii) log first difference. Similar to our model calibration, bilateral trade flows are taken from Johnson and Noguera (2017) and we separate between trade in final goods and trade in intermediate inputs.[27] As standard in the literature, we construct symmetric measures of bilateral trade intensity using the sum of total exports ($T^d_{i\to j}$) from country i to j in category $d \in \{\text{input}, \text{final}\}$ and total imports ($T^d_{j\to i}$) relative to GDP, such that: $\text{Trade}^d_{ij} = \frac{T^d_{i\to j} + T^d_{j\to i}}{GDP_i + GDP_j}$.

The extent to which countries have correlated GDP can be influenced by many factors beyond international trade, including correlated shocks, financial linkages, common monetary policies, etc. Because those other factors can themselves be correlated with the index of trade proximity in the cross section, using cross-section identification could yield biased results.[28] To separate the effect of trade linkages from other unobservable elements, we construct a panel dataset by creating four periods of ten years each. Within each time window, we compute GDP correlation (Corr GDP) as well as the average trade intensities defined above.

Our empirical strategy relies on the estimation of the following specifications:

$$\text{Corr GDP}_{ijt} = \beta_1 \ln(\text{Trade}^{\text{input}}_{ijt}) + \beta_2 \ln(\text{Trade}^{\text{final}}_{ijt}) + \mathbf{X}_{ijt} + \text{CP}_{ij} + \text{TW}_t + \epsilon_{ijt} \tag{4.30}$$

where i and j denote the two countries and t the time window. CP_{ij} and TW_t stand for country-pair and time windows fixed effects. The set of controls $\mathbf{X}_{ijt}$ include dummy variables for countries among trade union: the different waves of the European Unions, the Euro Area, and the USSR. We later include a set of additional controls that we discuss below.

Table 4.3.3 shows that when using within country-pair variations, trade in intermediate inputs is significantly and positively associated with higher GDP co-movement (columns (1) and (4)). This result confirms the cross-sectional estimates in Di Giovanni and Levchenko (2010), who investigate the role of vertical linkages in output synchronization using I/O matrices from the BEA.

[27] We provide additional details on data sources and the list of countries in the Online Appendix A.

[28] This limitation of cross sectional analysis has also been discussed by Imbs (2004), who notes that bilateral trade intensity can be a proxy of country-pair similarity, and thus of correlated shocks.

Columns (2) and (5) reveal that the relationship between trade in intermediate inputs and GDP co-movement is robust to the inclusion of time-windows fixed effect. Our estimates are also economically significant. Based on estimates in column (2) and noting that the median increase of the log trade intensity in intermediate goods between 1970-1979 and 2000-2009 is about 1.84, the slope coefficient implies a surge of GDP correlation of 11%, a non negligible increase. In contrast, trade in final goods is insignificant, or weakly negatively correlated.[29]

Table 4.3.3. Panel estimation: Trade proximity and GDP correlation [a]

	Corr GDP$^{\text{HP filter}}$			Corr ΔGDP		
	(1)	(2)	(3)	(4)	(5)	(6)
ln(Trade$^{\text{input}}$)	0.068***	0.060**	0.063***	0.053**	0.055**	0.053**
	(0.025)	(0.024)	(0.024)	(0.023)	(0.022)	(0.022)
ln(Trade$^{\text{final}}$)	−0.020	−0.038	−0.048**	−0.024	−0.038	−0.036
	(0.023)	(0.024)	(0.025)	(0.020)	(0.023)	(0.023)
Country-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Time windows fixed effects	No	Yes	Yes	No	Yes	Yes
Additional controls	No	No	Yes	No	No	Yes
N	2,900	2,900	2,900	2,900	2,900	2,900
Within R^2	0.076	0.169	0.172	0.082	0.162	0.169

Notes: *p<0.1; **p<0.05; ***p<0.01. SE clustered on country-pairs.

Robustness of the empirical slope Our results are robust to additional controls that capture the similarity of *trade networks*, which measures common exposure to third countries, and *sectoral composition* of trade. Our "network proximity" index is motivated by the fact that two countries with similar trade partners could co-move because of their common exposure to other countries. We define $network_{prox}(i,j) = 1 - \frac{1}{2}\sum_{k\neq i,j}\left|\frac{T^{\text{total}}_{i\to k}+T^{\text{total}}_{k\to i}}{\sum_s T^{\text{total}}_{i\to s}+T^{\text{total}}_{s\to i}} - \frac{T^{\text{total}}_{j\to k}+T^{\text{total}}_{k\to j}}{\sum_s T^{\text{total}}_{j\to s}+T^{\text{total}}_{s\to j}}\right|$. It measures the degree of similarity in the geographical distribution of trade shares between country i and country j, and is equal to 0 if countries i and j have completely separated trade partners while it is equal to 1 if all trade shares are equal.[30] The "sectoral proximity" index is defined as $sector_{prox}(i,j) = 1 - \frac{1}{2}\sum_{s\in\mathcal{S}}\left|\frac{T_i(s)}{\sum_{s\in\mathcal{S}} T_i(s)} - \frac{T_j(s)}{\sum_{s\in\mathcal{S}} T_j(s)}\right|$ with $T_i(s)$ the total export of country i in the specific sector s in the set of all sectors $\mathcal{S}$, and controls for changes in specialization. If two countries export exactly the same share of each product, then the index is equal to 1. If shocks have a sectoral component, then two countries that tend to specialize over time in the same sectors could have an increase in business cycle comovements over and beyond any trade effects. For those two indexes, we use bilateral trade data (SITC4 REV. 2) from the Observatory of Economic Complexity.

[29]Interestingly, the estimate obtained using a cross-section regression, that is, without country-pair fixed effects, differs only slightly from the one implied by the within country-pair variations.

[30]A complementary approach to this common exposure term has recently been proposed in Avila-Montealegre and Mix (2020). Their analysis measures the exposure to correlated trade partners, which captures a possible high common exposure even for country-pairs that do not share the same partners.

As shown in Table 4.3.3, the results are robust to the inclusion of sector proximity and network proximity (columns (3) and (6)). Finally, we test a wide range of alternative specifications and filtering methods, different sample selection varying country and time coverage, variable definitions as well as additional controls in section A of the online Appendix. In all our specifications, trade in intermediate inputs is significantly and positively associated with higher cross-country GDP correlation, a finding consistent with our measurement theory.

4.4 Results

This section presents our main quantitative results. We frame this section by focusing on three questions: (i). Is the model able to reproduce the magnitude of the trade comovement slope observed in the data? (ii) What role do markups and extensive margin adjustments play in generating the trade comovement slope? (iii) What role does real GDP measurement play in rationalizing these findings?

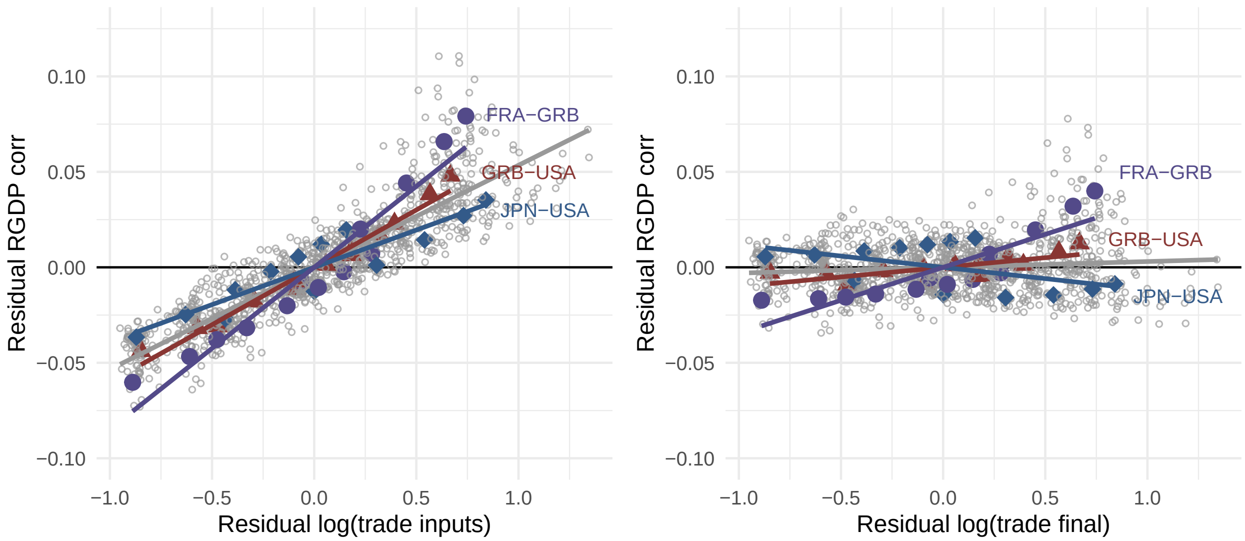
4.4.1 A Quantitative Resolution of the Trade Comovement Puzzle

To assess the model's ability to replicate the link between trade and real GDP-comovement, we simulate the exact same sequence of 5,000 shocks in different configurations in which we vary bilateral trade intensities. Starting with our baseline calibration of trade flows as described above, we simulate our model under 12 alternative calibration targets, using varying levels of final goods and intermediate goods bilateral trade intensities, each time resulting in new values for the shares $\omega_i^F(j)$ and $\omega_i^I(j)$. We use the data to discipline the variations of bilateral trade intensities across all configurations. Relative to the first time window (1970:1979) of our sample, the median bilateral trade intensity in intermediate goods and final goods increased by 5-6 and 8-9 in the last time-windows (2000:2009), respectively. For each configuration, we feed in the *exact same* technology shocks from our baseline calibration and record the pairwise correlation of logged and HP-filtered real GDP.

All told, this procedure gives rise to a panel dataset of $14 \times 13/2 = 91$ country-pairs (excluding *RoW*) for each of the 13 configurations, hence a total of 1,183 observations. Consistent with the procedure used in our empirical estimates, we run specification (4.30) and exploit within country-pair variation to gauge how changes in trade intensity impact bilateral real GDP comovement. Since only country-pairs trade intensities are modified, each configuration can be thought of as a different time-window.

We first focus on the relationship between measured real GDP correlations (corr $RGDP$) and bilateral trade intensities in intermediate inputs ($\text{Trade}_{ij}^{\text{input}}$) and final goods ($\text{Trade}_{ij}^{\text{final}}$). In Figure 4.4.1, we plot the residual relationship between bilateral trade in final and intermediate inputs and $RGDP$ synchronization, after controlling for country-pair fixed effects and change in the other good bilateral trade intensity (either final or intermediate goods). The model is found to capture well the high and significant correlation between real GDP comovement and bilateral trade in intermediate goods, while trade in final goods has a very small effect.[31]

Figure 4.4.1. Model-based association between RGDP correlation and Trade intensities. Left chart shows Intermediate Inputs trade, right chart shows Final goods trade.



Note: residual relationship in the model after controlling for other covariates, i.e. country-pair fixed effects and the other trade intensities (either final or intermediate goods). The grey solid line reports the Trade Comovement slope including all country-pairs.

The magnitude of the implied slope in our simulated data is reported in Table 4.4.1 (1^{st} row, columns (1)-(2)). The model generates a significant and positive trade comovement slope of 0.056, which is quantitatively in line with our empirical estimates in Table 4.3.3.[32] To get a sense of this number, our point estimates suggest that an increase in the log input trade index by 1.84, which is the average increase between 1970-1979 and 2000-2009, generates an increase of real GDP correlation in our simulations of about 10.3%. To put this into perspective, in our sample, the observed increase in cross-country GDP correlation between 1980-1990 and 2000-2009 in the data was about 25 percentage points. In our model, increasing trade intensities in the same proportion as what was observed during this period would lead to a 6 percentage points increase in cross-country GDP correlation. Therefore, trade linkages would explain roughly 24% of the observed increase in cross-country GDP correlations.

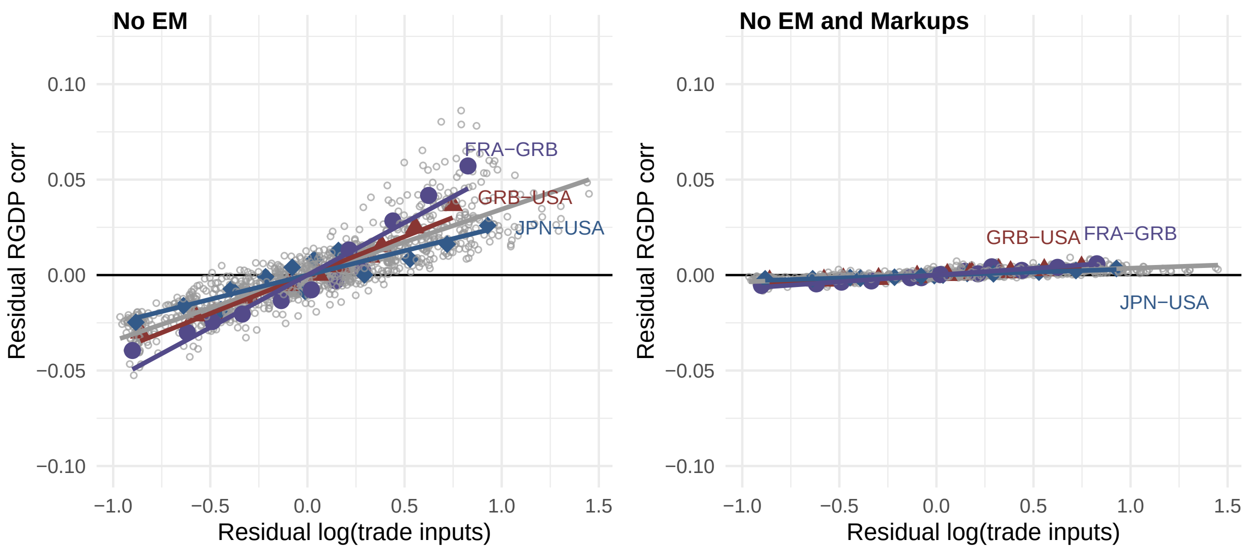
To understand the quantitative success of the model in replicating the large trade comovement slope, we turn off one by one the key elements of the model: (i) movements along the extensive margin, and (ii) monopolistic competition. Figure 4.4.2 depicts the implied relationship

[31] In appendix 4.B, we show that a higher Armington elasticity for final goods aggregation can account for the insignificant and negative slope between cross-country GDP correlations and trade in final goods.

[32] Using total bilateral trade intensity which captures the sum of input and final good trade, we find a slope of 0.07, comparable to the range of values $[4.8\% - 11\%]$ reported in the literature.

between $RGDP$ and bilateral trade intensity in intermediate inputs under these two alternatives, with points estimates in Table 4.4.1. With markups but no extensive margin (EM) adjustments, the trade comovement slope in intermediate inputs is 0.035. In a model with neither markups nor extensive margin adjustments (right panel of Figure 4.4.2 and the 3^{rd} row of Table 4.4.1), the model delivers a virtually flat trade comovement slope of 0.004. By comparing the implied slopes under those alternatives, Table 4.4.1 provides a decomposition of our result: Input-Output links alone explain only 7.0% of the trade comovement slope (=0.004/0.056) while the markup and extensive margin channels contribute 55.5% and 37.5% respectively.

Figure 4.4.2. Decomposition of the Input Trade-comovement slope using alternative model specifications. Left chart shows a model without Extensive Margin adjustments, right chart shows a model with neither extensive margin nor markups.



Note: residual relationship in the model after controlling for other covariates, i.e. country-pair fixed effects and the other trade intensities (either final or intermediate goods).

Table 4.4.1. Quantitative assessment of the Trade Comovement Slope: Data versus Model.

Trade – comovement slope [a]	Based on $RGDP$		Based on PVA	
	Input	*Final*	*Input*	*Final*
	(1)	(2)	(3)	(4)
Data: CP & TW fixed effects	.060**	−.038		
Model:				
1. IO link. + Markups + EM	.056***	.006***	.001***	.004***
2. IO link. + Markups	.035***	.001*	.001***	.002***
3. IO link.	.004***	.004***	.001***	.005***

*p<0.1; **p<0.05; ***p<0.01. SE clustered on country-pairs. Results are robust to the inclusion of our $network_{prox}$ index.
[a]The trade indexes used in those experiments are $(T_{i\to j} + T_{j\to i})/(GDP_i + GDP_j)$.

4.4.2 The Importance of Measurement

As discussed in section 4.2, the introduction of price distortions and extensive margin adjustments generate a disconnect between measured real GDP ($RGDP$) and physical value added (PVA). In our simulations, the association between trade and the synchronization of physical value added is low, consistent with earlier findings in the literature. Columns (3)-(4) in Table 4.4.1 show that, for all model specifications, using the PVA measure consistently results in a negligible estimated trade comovement slope. These results highlight the importance of defining real GDP in the model in a way that is consistent with data construction procedure: a researcher using PVA as a proxy for real GDP would mistakenly conclude that the model is not consistent with the data.

To sum up, our results show that adding intermediate inputs crossing multiple borders to an otherwise standard IRBC model is not sufficient to solve the TCP, as highlighted in Johnson (2014). This is because in these models, RGDP is equal to PVA, which reacts only modestly to foreign shocks. However, a model combining I/O linkages with markups and extensive adjustments in conjunction with statistically-consistent measured real GDP provides a quantitative solution to the Trade Comovement Puzzle.

4.5 Further Investigations

4.5.1 Measured Productivity, Profits and Trade in the Data

As our theory demonstrates, movements in productivity reflect movements in the profits derived from imported inputs. It follows that an increase in intermediate input trade should be associated with an increase in the co-movement of both measured productivity and aggregate profits. We now test those predictions in the data and discuss their counterpart in the model.[33]

We first study the relation between international trade and measured productivity. We calculate productivity as the Solow Residual (SR) in a standard Cobb-Douglas production function, that we then transform using, alternatively, HP-filter and log difference. As before, we separate bilateral trade intensity into intermediate inputs and final goods and test:

$$\text{Corr SR}_{ijt} = \beta_1 \ln(\text{Trade}^{\text{input}}_{ijt}) + \beta_2 \ln(\text{Trade}^{\text{final}}_{ijt}) + \mathbf{X}_{ijt} + \text{CP}_{ij} + \text{TW}_t + \epsilon_{ijt} \tag{4.31}$$

Second, we study the relation between international trade and profits, as measured by the Net Operating Surplus (NOS). We use the NOS measured by the OECD at quarterly frequency, that we then transform using HP-filter and log-difference. To avoid any bias due to inflation synchronization, we also deflate these nominal value by the consumer price index. We then test

[33]Note that Kose and Yi (2006) and Liao and Santacreu (2015) study the relationship between measured TFP and bilateral trade in the cross-section. We add to their analysis by separating bilateral trade into intermediate inputs and final goods and use a panel estimation.

the following specification:

$$\text{Corr NOS}_{ijt} \quad = \beta_1 \ln(\text{Trade}^{\text{input}}_{ijt}) + \beta_2 \ln(\text{Trade}^{\text{final}}_{ijt}) + \mathbf{X}_{ijt} + \text{CP}_{ij} + \text{TW}_t + \epsilon_{ijt} \tag{4.32}$$

Results are gathered in Table 4.5.1. Consistent with our predictions, higher trade intensity in intermediate inputs tend to have a positive and statistically significant effect on the co-movement of measured aggregate productivity and net operating surplus. Trade in final goods is found to have a negative and statistically insignificant effect. Moreover, according to our theoretical results in section 4.2, controlling for SR correlation in specification (4.30) should capture both the markup and love of variety effects highlighted in proposition 3. In the online appendix A.4, we show that once we control for SR correlation, we obtain a lower and insignificant point estimate for trade in intermediate inputs, which in line with our theoretical predictions.[34,35]

Table 4.5.1. Trade, Measured Productivity and Net Operating Surplus

	Corr Productivity$^{\text{HP}a}$		Corr ΔProductivity^{a}		Corr NOS$^{\text{HP}b}$		Corr ΔNOS^{b}	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\ln(\text{Trade}^{\text{input}})$	.055**	.065**	.045*	.051**	.371***	.175*	.288***	.250***
	(.027)	(.026)	(.025)	(.024)	(.104)	(.095)	(.083)	(.086)
$\ln(\text{Trade}^{\text{final}})$	.002	−.032	.001	−.022	−.085	−.105	−.100	−.157**
	(.025)	(.025)	(.022)	(.022)	(.079)	(.073)	(.062)	(.079)
Country-pair fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time windows fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
N	2,340	2,340	2,340	2,340	364	364	364	364
R^2	.113	.217	.130	.200	.104	.263	.121	.273

Notes: *p<0.1; **p<0.05; ***p<0.01. In parenthesis: std. deviation. SE clustered on country-pairs. Controls include third indexes and trade unions. Results are robust without third indexes.
^{a} Data for measured productivity are constructed using the PWT 9.1 using total employment and measured capital. We use 2/3 for the labor share.
^{b} Data for Corr NOS are measured at quarterly frequency, for 16 consecutive quarters in time-windows of 4 years. Due to data limitation, our sample covers 1999Q1-2014Q4. Data for trade flows are taken from OEC data. Result are robust without the Great Recession time-window (2007-2010) and additional controls: dummies for trade unions and similarity indexes.

4.5.2 Measured productivity in the Model

Using our theoretical framework to further investigate the above empirical finding, we analogously measure productivity in the model using the Solow Residual such that: $SR_{it} = \log(RGDP_{it}) -$

[34] As discussed in Huo et al. (2020), unobserved factor utilization can create a measurement error for SR. In the online appendix A.4, we introduce an unobserved component in labor input leading to measurement errors in SR. Using model-based simulations, we show that such unobserved component leads to a downward bias in the slope between trade and $corr(SR)$. This finding implies that the positive and significant results obtained in the data using specification (4.31) may be a lower bound of the slope's true value.

[35] In the online Appendix A, we also show that the comovement of labor input (as measured in the Penn World Tables 9.1) is not statistically associated with trade integration. While there might be issues with measuring labor input in the data, the absence of statistical association between trade links and measured labor input synchronization suggests that looking beyond factor supply comovement is important when looking at cross country real GDP correlation

$\alpha \log(L_{it}) - (1-\alpha)\log(K_{it})$. Table 4.5.2 presents the results of regressing changes in productivity on changes in trade proximity in the baseline model as well as a version of the model without extensive margins and/or without markups. The first row reveals that, in our baseline model, measured productivity correlation increases when country-pairs trade more intermediate inputs, and the magnitude of this association is in line with the data. In line with the insights emerging from equation (4.10), this association dissipates in a version of the model without extensive margin adjustments or without markup: without these ingredients, productivity is simply equal to technology shocks, which are identical in all simulations. Moreover, if one constructs productivity as the Solow Residual of PVA (defined as: $SR_{it}^{PVA-based} = \log(PVA_{it}) - \alpha\log(L_{it}) - (1-\alpha)\log(K_{it})$, in columns (3)-(4)), none of the additional ingredients generate a disconnect between technology and measured productivity. This is in sharp contrast with earlier studies. For instance, Liao and Santacreu (2015) do find that extensive margin adjustments help a model get a stronger link between cross-country TFP correlation and trade linkages, but in their simulations all effects go through $SR^{PVA-based}$.[36] In this paper, we instead highlight a measurement channel once imported input's base period price does not reflect their marginal revenue product, which is reflected in SR but not in $SR^{PVA-based}$.

Table 4.5.2. Association between trade and SR comovement: Data versus Model.

Trade – Productivity comovement slope:[a]	SR based on $RGDP$		SR based on PVA	
	Input	*Final*	*Input*	*Final*
	(1)	(2)	(3)	(4)
Data: CP & TW fixed effects	.065**	−.032		
Model:				
1. IO link. + Markups + EM	.051***	.003***	−.000***	.000***
2. IO link. + Markups	.033***	−.000	.000*	.000***
3. IO link.	.000**	.000***	.000	.000***

*p<0.1; **p<0.05; ***p<0.01. SE clustered on country-pairs. Results are robust to the inclusion of our $network_{prox}$ index.
[a]The trade indexes used in those experiments are $(T_{i\to j} + T_{j\to i})/(GDP_i + GDP_j)$. Productivity is measured as the Solow Residual using either $RGDP$ or PVA as a measure for real value added.

This discussion underscores the importance of measuring real GDP in a way that is consistent with actual data construction procedures. Indeed, the international correlation of measured productivity in our simulations is about 0.264, while the correlation of technology shocks, captured by Z_{it}, is only 0.22. This illustrates that using measured productivity in the data as a calibration target for technology would overstate the actual correlation of shocks between countries.

[36]In their simulations, the trade-productivity comovement slope obtained using statistically measured productivity row is less than 1% (see in their table 8, 2nd, pp 276).

4.5.3 A Focus on the Role of Markups and Profits in the Model

Our baseline calibration assumes a common lower-tier elasticity of substitution $\sigma = 4.35$. In this section, we confirm the quantitatively important role of markups in the cross-country correlation of profits and productivity observed in Table 4.5.1.[37]

Table 4.5.3 presents our simulation-based trade comovement slope using alternative values for the markups. We first test the implication of higher ($\sigma = 3.5$) and lower ($\sigma = 6.0$) markups in the 2^{th} and 3^{rd} row, respectively. As expected, an increase in markups leads to a higher association between trade and real GDP comovement. Increasing markups from 20% to 33% almost doubles the association between input trade and real GDP synchronization. Unsurprisingly, we find that physical value added PVA does not respond more to foreign shocks when we change the value of the (constant) markup, as shown in the last two columns of the table.

We then relax the homogeneous assumption and introduce heterogenous market power across countries. We evaluate whether plausible heterogeneity in observed markups – for the set of countries we consider – have important consequences for the implied trade comovement slope. Specifically, we simulate the model with heterogenous σ_i estimated from the data using two different sources. We first use OECD STAN's database and construct the Price Cost Margin (PCM) (4^{th} row) as an estimate of markups within each industry, which measures the difference between revenue and variable cost. Second, we use direct markup estimates from De Loecker and Eeckhout (2018) (DLE) (5^{th} row). In each experiment, we center the heterogenous markups $\{\sigma_1, ..., \sigma_N\}$ around the baseline value. The implied standard deviation of σ are .63 and .73 for the PCM and the DLE experiments respectively. Interestingly, adding heterogeneous markups centered around our baseline value does not substantially affect the trade comovement slope, which suggests that accounting for cross-country heterogenous markups does not change the aggregate strength of international propagation in our model.

Those results are interesting because they help rationalize the large heterogeneity in the estimated trade comovement slopes found in the literature, with estimates ranging from 4.8% to 11%, depending on country and time coverage. Consequently, our model can generate such an heterogeneity through different levels of market power over time and between countries.

4.5.4 Extensive Margin fluctuations and Real GDP Comovement in the Data

We finally investigate the role of extensive margin adjustments in generating the observed association between trade and real GDP comovement. We use the Hummels and Klenow (2005) (HK) decomposition and investigate the relation between the average and the volatility *within* each time window of the Extensive Margin (EM) and Intensive Margin (IM) of bilateral trade intensities in

[37] In section A of the online appendix, we further validate the role of markups in generating a link between terms of trade and real GDP fluctuations.

Table 4.5.3. The role of price distorsions and heterogenous markups [a]

	Elasticity	Trade - GDP comovement slope coefficients: based on *RGDP*		based on *PVA*	
		Input	*Final*	*Input*	*Final*
Data: CP & TW fixed effects	-	.060**	-.038		
Model:					
1. Baseline	$\sigma = 4.35$	.056***	.006***	.001***	.004***
2. Low markups	$\sigma = 6.0$	.037***	.005***	.001***	.004***
3. High markups	$\sigma = 3.5$	.075***	.007***	.001***	.003***
4. Heterogenous markups, PCM	$\sigma_i \in [3.20, 5.65]$	.054***	.006***	.001***	.004***
5. Heterogenous markups, DLE	$\sigma_i \in [3.68, 6.07]$	.056***	.006***	.001***	.004***

Notes: *p<0.1; **p<0.05; ***p<0.01. Results are robust to the inclusion of our $network_{prox}$ index.
[a]The simulations are based on the exact same sequence of shocks, under the five variations of trade indexes used in the benchmark.

intermediate inputs and real GDP comovement. We use a panel estimation with 38 countries from 1971 to 2010 and focus on trade in inputs.[38] Trade data comes from the NBER United Nations Trade Data from 1971 to 2010 and the UN COMTRADE data from 2001 to 2010. We use the bilateral trade flows as categorized under the SITC (rev. 2, 4-digits) classification and we classify intermediate inputs as Johnson and Noguera (2017).

Using the HK decomposition, we construct the EM and IM of trade in intermediate inputs for each directed pair of country $(i \rightarrow j)$. The *Rest-of-the-World*, indexed k, is taken as a reference country. The EM is defined as a weighted count of varieties of intermediate inputs exported from j to m relative to those exported from k to m, i.e. $EM_{jm} = \sum_{i \in I_{jm}} T^{\text{input}}_{k \rightarrow m}(i) / \sum_{i \in I} T^{\text{input}}_{k \rightarrow m}(i)$, where I_{jm} is the set of observable categories in which j has a positive shipment to m and I is the set of all categories of intermediate inputs exported by the reference country. If all categories are of equal importance and the reference country k exports all categories to m, then the extensive margin is simply the fraction of categories in which j exports to m. The corresponding IM is the ratio of nominal shipments from j to m and from k to m in a common set of intermediate goods, i.e. $IM_{jm} = \sum_{i \in I_{jm}} T^{\text{input}}_{j \rightarrow m}(i) / \sum_{i \in I_{jm}} T^{\text{input}}_{k \rightarrow m}(i)$. Note that the product of the two measures provide a measure of the overall trade from j to m relative to the overall trade from k to m. Finally, since those measures are not symmetric within a country-pair we sum, for each country pair (i, j), the IM and EM from i to j and from j to i, and normalize this by the sum of GDP.

We compute the *within* time-window average and std. deviation of EM and IM and test:

$$\text{Corr GDP}_{ijt} = \beta_1 \ln(\text{EM}_{ijt}) + \beta_2 \ln(\text{IM}_{ijt}) + \text{CP}_{ij} + \text{TW}_t + \epsilon_{ijt} \tag{4.33}$$

$$\text{Corr GDP}_{ijt} = \beta_1 \ln(std(\text{EM})_{ijt}) + \beta_2 \ln(std(\text{IM})_{ijt}) + \text{CP}_{ij} + \text{TW}_t + \epsilon_{ijt} \tag{4.34}$$

[38]We drop Czechoslovakia, Estonia, Russia, Slovenia and Slovakia due to missing data.

Results are reported in Table 4.5.4. First, using specification (4.33) in columns (1) and (3), we recover a result in line with Liao and Santacreu (2015) who use an IV estimator instead of a panel estimation: the correlation between the level of the extensive margin of trade in intermediate inputs and real GDP comovement is positive and significant. In contrast, the intensive margin of trade is found not significantly related with GDP comovement.

Second, we use specification (4.34) and test more directly our insights from section 4.2. Indeed, our theory does not say that the *level* of the extensive margin is important for GDP comovement, but rather that movement in real GDP are related to *variations* in the number of input varieties that are traded. In columns (2) and (4), we relate GDP comovement to the standard deviation of extensive and intensive margin movements. Consistent with the theory, the results point to the fact that larger fluctuations along the extensive margin are positively and significantly correlated with higher GDP comovement, while the estimates for the intensive margin implies no significant relationship.[39]

Table 4.5.4. Real GDP correlations and the margins of intermediate inputs trade

	Corr $GDP_{ijt}^{\text{HP filter}}$		Corr ΔGDP_{ijt}	
	(avg measure) (1)	(std measure) (2)	(avg measure) (3)	(std measure) (4)
EM	0.050**	0.049**	0.045***	0.068***
	(0.022)	(0.021)	(0.020)	(0.021)
IM	−0.010	−0.013	−0.003	−0.015
	(0.017)	(0.011)	(0.016)	(0.043)
CP + TW fixed effects	Yes	Yes	Yes	Yes
N	2,347	2,347	2,347	2,347
Within R^2	0.083	0.084	0.102	0.107

Notes: *p<0.1; **p<0.05; ***p<0.01. In parenthesis: std. dev. SE clustered on country-pairs.

4.5.5 Robustness of the Quantitative Results.

As our result is quantitative in nature, it is important to check the validity of our results under alternative model specifications and calibrations. In Table 4.B.1 of appendix 4.B, we investigate the model ability to generate the trade comovement puzzle under alternative parameter values with respect to the elasticity of labor supply, international trade costs, capital adjustment costs and the Argminton elasticity in the CES good aggregation. In those experiments, our results are virtually unchanged. Our results also hold under complete financial markets, which constitutes the other extreme relative to financial autarky. The findings are also robust to alternative calibration periods and when we (wrongly) infer the correlation of TFP shocks from the data. Finally, a model

[39]This result is particularly striking given that most of the variations in trade at business cycle frequency is explained by variations along the intensive margin. In the online appendix A.7, we further investigate the role of the extensive and intensive margins of trade using an alternative dataset using a direct measure of the number of firms and find that the extensive margin of trade is positive and significant.

without correlated TFP shocks produces a weaker trade comovement slope although it remains significantly high.

4.6 Conclusion

This paper analyzes the relationship between international trade and business cycle synchronization across countries, with a focus on improving the mapping between real GDP in the data and its counterpart in standard macroeconomic models. We show that real GDP, as constructed by statistical agencies, is not equal to the theory-consistent "physical value added". This disconnect appears when the base period price used to value imported input does not reflect their marginal product, for example in presence of markup and love of variety. With those ingredients, real GDP fluctuations are not only tied to movements of technology and factor supply, but can also fluctuate as a result of changes in imported input usage.

We show that a quantitative model with markups and love of variety delivers a strong link between input trade and business cycle comovement, with a magnitude in line with empirical estimates, offering the first quantitative solution for the *Trade Comovement Puzzle*. Conversely, model-based simulations show that trade is much less associated with the synchronization of physical value added. We finally confirm the predictions based on the data. First, higher trade in intermediate inputs is associated with an increase in the bilateral correlation of both real GDP and productivity. Second, higher trade in intermediate input is also associated with synchronized profits. Third, real GDP is sensitive to fluctuations in the number of varieties imported, implying that the extensive margin of trade plays an important role in business cycle synchronization.

To conclude, this paper seeks to draw attention to real GDP measurement in macroeconomic models and how it should be compared with its data counterpart. In the context of the *Trade-Comovement Puzzle*, IRBC models that generate weak cross-country propagation properties in terms of physical value-added can actually feature strong propagation in terms of real GDP. More generally, recognizing that real GDP fluctuations are not only tied to physical value added movements could be a promising research avenue for business cycle analysis.

Appendix

4.A Empirical Appendix: Markup Measures

We used two different markup index estimates. We first used aggregated micro markups from De Loecker and Eeckhout (2018), who estimate aggregate markups using a cost-based approach in 134 countries from 1980 to 2016. This method defines markups as the ratio of the output price to the marginal costs, and therefore relies solely on information from the financial statements of firms (sales value and cost of goods sold). Aggregating all firms specific markups for each country, De Loecker and Eeckhout (2018) provide a detailed and comparable measure of market power between countries. The sample that we use from their estimates includes 29 countries from 1980 to 2016.[40]

Second, we use Price Cost Margin (PCM) as an estimate of markups within each industry using data from 22 countries from 1971 to 2010.[41] Widely used in the literature, PCM is the difference between revenue and variable cost (the sum of labor and material expenditures, over revenue): $PCM = \frac{\text{Sales}-\text{Labor exp.}-\text{Material exp.}}{\text{Sales}}$. Data at the industry level come from the OECD STAN database, an unbalanced panel covering 107 sectors for 34 countries between 1970 and 2010. Due to missing data for many countries in the earliest years, we restrict the analysis for 22 countries. We compute PCM for each industry-country-year and then construct an average of PCM within each country-year by taking the sales-weighted average of PCM over each industry. Finally, the average PCM for a given time window is simply the mean of country-year PCM over all time periods.

4.B Theoretical and Quantitative appendix

Our results are robust to a number of alternative specifications, as presented in table 4.B.1.

Parameters We restate our baseline result in the 1^{st} row for reference. As shown in rows 2 and 3, the Frisch elasticity has a significant impact on the magnitude of the overall trade comovement slope, while preserving the relative importance of final goods versus intermediate inputs. The level of trade frictions in the calibrated steady state $\{\tau_{ij}, f^c_{ij}, f^E_{ij}\}$ does not affect the implied TC

[40] The list of countries is: Austria, Belgium, Canada, Colombia, Denmark, Finland, France, Germany, Greece, Ireland, Iceland, Indonesia India, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, Turkey, the United-Kingdom and the United-States.

[41] The list of countries is: Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Iceland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, Portugal, Spain, Sweden, the United-Kingdom and the United-States.

slope, as seen in rows 4 and 5. Regardless of the initial level of those trade frictions, increasing trade proximity is associated with the same reaction for real GDP comovement. In row 6 we artificially remove all trade imbalances observed in the data. Under all those alternative specifications, our main results hold. In rows 7 and 8 we vary ψ and find that lower adjustment costs results in more volatile real value added and investments $\mathcal{I}$, which magnifies the Trade Comovement slope. In row 9, we set a higher Armington elasticity in the CES final good aggregation, $\rho_F = 1.5$, and show that it can rationalize the negative and insignificant slope with respect to trade in final goods.

Productivity process We conduct robustness on the properties of technology processes $\{Z_i\}$. We first use the observed estimated covariance matrix of standard TFP data computed as the Solow residual in Penn World Tables ($\widetilde{\sum}$). While this approach is sometimes used in the literature, it leads to overshooting the *level* of cross-country real GDP correlation. Results regarding the trade comovement (TC) slope remain similar to the benchmark calibration. Furthermore, we simulate the model under the counterfactual assumption that technology shocks are uncorrelated across countries and set the off-diagonal elements of the covariance-variance matrix to zero (i.e. $cov(Z_{i,t}, Z_{j,t}) = 0, \forall i \neq j$), in row 11. In this case, the TC slope decreases to 0.035. This finding echoes the discussion about the consequences of using correlated vs. uncorrelated shocks in Johnson (2014), namely that the TC slope decreases with uncorrelated shocks.[42] However, it is important to recall that our exercise are very different: while Johnson (2014) uses a cross-section regression, our panel estimation includes country-pair FE implying that differences in average shock correlation across different country-pair are controlled for.

Reference period In rows 12 to 13 we use a different reference period for the calibration of the CES weights ω^I and ω^F, a variation that does not materially alter our key messages.

Alternative Financial Markets Our benchmark specification assume financial autarky. We verify if the results of our quantitative model hold under complete financial markets, which can be thought as the other extreme modelling assumption. We assume that there are complete contingent claims dominated in units of one of the countries' tradable final good. Let s_t denote the state of an economy in period t, with transition probability density $f(s_{t+1}, s_t)$. We denote $B_i(s_{t+1})$ denote the country i's holdings of a one-period state-contingent bonds, paying off one unit of the numeraire good in state s_{t+1}, and let $b(s_{t+1}, s_t)$ be the price of that security in state s_t at date t. Furthermore, these state-contingent bonds are in zero net supply in all states: $\sum_i B_i(s_{t+1}) = 0$. In

[42] Looking at panel B in table 1 of Johnson (2014), the TC slope generated by a model with correlated shocks is about 15% of its empirical counterpart. Results with uncorrelated shocks (in panel C) are only about 3% of the empirical one. In contrast, while we also find a small decrease when using uncorrelated shocks, the magnitude of the drop is significantly smaller.

this case, the consumer budget constraint is given by:

$$P^F_{i,t}(C_{i,t} + \mathcal{I}_{i,t}) + \int b(s_{t+1}, s_t)B_i(s_{t+1})ds_{t+1} = r_{i,t}K_{i,t} + w_{i,t}L_{i,t} + B_i(s_t) \tag{4.35}$$

Consumers choose $\{C_{i,t}, L_{i,t}, K_{i,t+1}\}$ and asset holdings $\{B_i(s_{t+1})\}$ given prices and initial asset endowments $\{B_i(s_0)\}$ to maximize equation (4.12). Results are provided in Table 4.B.1 (in rows 14 and 15). We find that the trade comovement slope persist even under complete markets and the main message of the paper remains unchanged.

Table 4.B.1. Model-based simulations: sensitivity analysis [a]

Robustness/Alternative specification	Parameter change	Trade - RGDP slope	
		Input	*Final*
1. Baseline	-	.056***	.006***
A. Model parameter & trade frictions			
2. High Frisch elasticity	$\nu = .25$	.064***	.010***
3. Low Frisch elasticity	$\nu = .75$	.052***	.004***
4. Iceberg costs	+10%	.056***	.006***
5. Fixed costs	+10%	.056***	.006***
6. No trade imbalance	$\mathcal{T}_i = 0,\ \forall i$	.056***	.006***
7. Low adjustment cost	$\psi = 1.0$	.061***	.009***
8. High adjustment cost	$\psi = 1.5$	.052***	.004***
9. Alternative CES elasticity	$\rho^F = 1.5$	.055***	$-.001$
B. Productivity process			
10. PWT-estimated TFP shocks	$\widetilde{\Sigma}$	.046***	.006***
11. Uncorrelated techno. shocks	$cov(Z_{i,t}, Z_{j,t}) = 0, \forall i \neq j$	.030***	.005***
C. Reference period for $\{\omega^I, \omega^F\}$			
12. Baseline	1990-2000	.084***	.015***
13. Baseline, no EM/markups	1990-2000	.007***	.009***
D. Asset Market			
14. Complete Markets	–	.055***	.010***
15. Complete Markets, no EM/markups	–	.005***	.005***

[a] The simulations are based on the exact same sequence of shocks, under the five variations of trade indexes used in the benchmark.

Bibliography

Acemoglu, Daron, Vasco Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, Vol. 80, pp. 1977–2016.

Agrawal, Ashwini K and David A Matsa (2013): "Labor unemployment risk and corporate financing decisions," *Journal of Financial Economics*, Vol. 108, pp. 449–470.

Aiyagari, S. Rao (1994): "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, Vol. 109, pp. 659–684.

Aiyagari, S Rao (1995): "Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting," *Journal of Political Economy*, Vol. 103, pp. 1158–1175.

Alderson, Michael J and Brian L Betker (1995): "Liquidation Costs and Capital Structure," *Journal of Financial Economics*, Vol. 39, pp. 45–69.

Alessandria, George and Horag Choi (2007): "Do Sunk Costs of Exporting Matter for Net Export Dynamics?," *The Quarterly Journal of Economics*, Vol. 122, pp. 289–336.

Ali, Abdul, I. Elaine Allen, Candida Brush, William D. Bygrave, Julio De Castro, Julian Lange, Heidi Neck, Joseph Onochie, Ivory Phinisee, Edward Rogoff, and Albert Suhu (2008): "What Entrepreneurs Are Up To,"Technical report, Global Entrepreneurship Monitor.

Alon, Titan, David Berger, Robert Dent, and Benjamin Pugsley (2018): "Older and slower: The startup deficit's lasting effects on aggregate productivity growth," *Journal of Monetary Economics*, Vol. 93, pp. 68–85.

Alpanda, Sami and Geoffrey Woglom (2007): "The Case Against Power Utility and a Suggested Alternative: Resurrecting Exponential Utility," MPRA Working Paper No. 5897.

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman (2018): "Who owns the wealth in tax havens? Macro evidence and implications for global inequality," *Journal of Public Economics*, Vol. 162, pp. 89–100.

Anderson, James E. and Eric van Wincoop (2004): "Trade Costs," *Journal of Economic Literature*, Vol. 42, pp. 691–751.

Anderson, Patricia M and Bruce D Meyer (2000): "The effects of the unemployment insurance payroll tax on wages, employment, claims and denials," *Journal of Public Economics*, Vol. 78, pp. 81–106.

Angeletos, George-Marios (2007): "Uninsured Idiosyncratic Investment Risk and Aggregate Saving," *Review of Economic Dynamics*, Vol. 10, pp. 1–30.

Angeletos, George-Marios and Laurent-Emmanuel Calvet (2006): "Idiosyncratic Production Risk, Growth and the Business Cycle," *Journal of Monetary Economics*, Vol. 53, pp. 1095–

1115.

Arkolakis, Costas, Theodore Papageorgiou, and Olga A Timoshenko (2018): "Firm Learning and Growth," *Review of Economic Dynamics*, Vol. 27, pp. 146–168.

Arkolakis, Costas and Ananth Ramanarayanan (2009): "Vertical Specialization and International Business Cycle Synchronization*," *The Scandinavian Journal of Economics*, Vol. 111, pp. 655–680.

Arrow, Kenneth J. (1974): "The Measurement of Real Value Added," in Paul A. David and Melvin W. Reder eds. *Nations and Households in Economic Growth: Essays in honor of Moses Abramovitz*: Academic Press, pp. 3–19.

Asenjo, Antonia, Clemente Pignatti et al. (2019): *Unemployment insurance schemes around the world: Evidence and policy options*: International Labour Office.

Astebro, Thomas and Jing Chen (2014): "The entrepreneurial earnings puzzle: Mismeasurement or real?" *Journal of Business Venturing*, Vol. 29, pp. 88–105.

Avila-Montealegre, Oscar and Carter Mix (2020): "Common Trade Exposure and Business Cycle Comovement," International Finance Discussion Papers 1306, Board of Governors of the Federal Reserve System (U.S.).

Bach, Laurent, Laurent E Calvet, and Paolo Sodini (2020): "Rich pickings? Risk, return, and skill in household wealth," *American Economic Review*, Vol. 110, pp. 2703–47.

Banerjee, Abhijit V and Andrew F Newman (1993): "Occupational choice and the process of development," *Journal of Political Economy*, Vol. 101, pp. 274–298.

Barro, Robert J (2000): "Inequality and Growth in a Panel of Countries," *Journal of Economic Growth*, Vol. 5, pp. 5–32.

——— (2008): "Inequality and growth revisited,"Technical report, ADB Working paper series on regional economic integration.

Barrot, Jean-Noël and Julien Sauvagnat (2016): "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks," *The Quarterly Journal of Economics*, Vol. 131, pp. 1543–1592.

Bassi, Vittorio, Raffaela Muoio, Tommaso Porzio, Ritwika Sen, and Esau Tugume (2021): "Achieving Scale Collectively,"Technical report, National Bureau of Economic Research.

Basu, Susanto and John G. Fernald (2002): "Aggregate productivity and aggregate technology," *European Economic Review*, Vol. 46, pp. 963–991.

Baxter, Marianne and Michael A Kouparitsas (2005): "Determinants of business cycle comovement: a robust analysis," *Journal of Monetary Economics*, Vol. 52, pp. 113–157.

Been, Jim, Susann Rohwedder, and Michael Hurd (2020): "Does home production replace consumption spending? Evidence from shocks in housing wealth in the Great Recession," *Review of Economics and Statistics*, Vol. 102, pp. 113–128.

Benhabib, Jess, Alberto Bisin, and Mi Luo (2019): "Wealth Distribution and Social Mobility in the US: A Quantitative Approach," *American Economic Review*, Vol. 109, pp. 1623–47.

BENHABIB, JESS, ALBERTO BISIN, AND SHENGHAO ZHU (2011): "The distribution of wealth and fiscal policy in economies with finitely lived agents," *Econometrica*, Vol. 79, pp. 123–157.

BERNARD, ANDREW B., JONATHAN EATON, J. BRADFORD JENSEN, AND SAMUEL KORTUM (2003): "Plants and Productivity in International Trade," *American Economic Review*, Vol. 93, pp. 1268–1290.

BEWLEY, TRUMAN (1986): "Stationary monetary equilibrium with a continuum of independently fluctuating consumers," *Contributions to mathematical economics in honor of Gérard Debreu*, Vol. 79.

BHANDARI, ANMOL AND ELLEN R MCGRATTAN (2018): "Sweat Equity in U.S. Private Business," Working Paper 24520, National Bureau of Economic Research.

BOAR, CORINA AND VIRGILIU MIDRIGAN (2019): "Markups and inequality,"Technical report, National Bureau of Economic Research.

——— (2020): "Efficient Redistribution,"Technical report, National Bureau of Economic Research.

BOEHM, CHRISTOPH E., AARON FLAAEN, AND NITYA PANDALAI-NAYAR (2019): "Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake," *The Review of Economics and Statistics*, Vol. 101, pp. 60–75.

BOONE, CHRISTOPHER, ARINDRAJIT DUBE, LUCAS GOODMAN, AND ETHAN KAPLAN (Forthcoming): "Unemployment insurance generosity and aggregate employment," *American Economic Journal: Economic Policy*.

BOURGUIGNON, FRANÃ§OIS (1981): "Pareto Superiority of Unegalitarian Equilibria in Stiglitz' Model of Wealth Distribution with Convex Saving Function," *Econometrica*, Vol. 49, pp. 1469–1475.

BROWNING, MARTIN AND THOMAS F CROSSLEY (2001): "Unemployment insurance benefit levels and consumption changes," *Journal of Public Economics*, Vol. 80, pp. 1–23.

BRÜGGEMANN, BETTINA (2020): "Higher taxes at the Top: The role of entrepreneurs," *American Economic Journal: Macroeconomics, Forthcoming*.

——— (2021): "Higher taxes at the top: The role of entrepreneurs," *American Economic Journal: Macroeconomics*, Vol. 13, pp. 1–36.

BRUNNERMEIER, MARKUS K. AND STEFAN NAGEL (2008): "Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-Evidence on Individuals' Asset Allocation," *The American Economic Review*, Vol. 98, pp. 713–736.

BUERA, FRANCISCO J. AND YONGSEOK SHIN (2013): "Financial Frictions and the Persistence of History: A Quantitative Exploration," *Journal of Political Economy*, Vol. 121, pp. 221–272.

BUERA, FRANCISCO, JOSEPH KABOSKI, AND YONGSEOK SHIN (2011): "Finance and Development: A Tale of Two Sectors," *American Economic Review*, Vol. 101, pp. 1964–2002.

BURSTEIN, ARIEL AND JAVIER CRAVINO (2015): "Measured Aggregate Gains from International

Trade," *American Economic Journal: Macroeconomics*, Vol. 7, pp. 181–218.

BURSTEIN, ARIEL, CHRISTOPHER KURZ, AND LINDA TESAR (2008): "Trade, production sharing, and the international transmission of business cycles," *Journal of Monetary Economics*, Vol. 55, pp. 775–795.

CAGETTI, MARCO AND MARIACRISTINA DE NARDI (2006a): "Entrepreneurship, frictions, and wealth," *Journal of Political Economy*, Vol. 114, pp. 835–870.

——— (2006b): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, Vol. 114, pp. 835–870.

——— (2009a): "Estate taxation, entrepreneurship, and wealth," *American Economic Review*, Vol. 99, pp. 85–111.

——— (2009b): "Estate Taxation, Entrepreneurship, and Wealth," *American Economic Review*, Vol. 99, pp. 85–111.

CALDERON, CESAR, ALBERTO CHONG, AND ERNESTO STEIN (2007): "Trade intensity and business cycle synchronization: Are developing countries any different?" *Journal of international Economics*, Vol. 71, pp. 2–21.

CALIENDO, MARCO AND ALEXANDER KRITIKOS (2009): ""I Want to, But I Also Need to": Start-Ups Resulting from Opportunity and Necessity," IZA Discussion Papers 4661, Institute for the Study of Labor (IZA).

CALVET, LAURENT, JOHN CAMPBELL, AND PAOLO SODINI (2009): "Fight or Flight? Portfolio Rebalancing by Individual Investors," *The Quarterly Journal of Economics*, Vol. 124, pp. 301–348.

CALVET, LAURENT E, JOHN Y CAMPBELL, FRANCISCO J GOMES, AND PAOLO SODINI (2019): "The cross-section of household preferences,"Technical report, Working Paper, Harvard University.

CARROLL, CHRISTOPHER D (2006): "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems," *Economics letters*, Vol. 91, pp. 312–320.

CASTRO, RUI, GIAN LUCA CLEMENTI, AND YOONSOO LEE (2015): "Cross sectoral variation in the volatility of plant level idiosyncratic shocks," *The Journal of Industrial Economics*, Vol. 63, pp. 1–29.

CHETTY, RAJ (2008): "Moral hazard versus liquidity and optimal unemployment insurance," *Journal of Political Economy*, Vol. 116, pp. 173–234.

CHETTY, RAJ, NATHANIEL HENDREN, PATRICK KLINE, AND EMMANUEL SAEZ (2014): "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, Vol. 129, pp. 1553–1623.

CHODOROW-REICH, GABRIEL, JOHN COGLIANESE, AND LOUKAS KARABARBOUNIS (2019): "The macro effects of unemployment benefit extensions: a measurement error approach," *The Quarterly Journal of Economics*, Vol. 134, pp. 227–279.

CHRISTIANO, LAWRENCE J, MARTIN EICHENBAUM, AND CHARLES L EVANS (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of political Economy*, Vol. 113, pp. 1–45.

CLARK, TODD AND ERIC VAN WINCOOP (2001): "Borders and business cycles," *Journal of International Economics*, Vol. 55, pp. 59–85.

CLEMENTI, GIAN LUCA AND BERARDINO PALAZZO (2016): "Entry, Exit, Firm Dynamics, and Aggregate Fluctuations," *American Economic Journal: Macroeconomics*, Vol. 8, pp. 1–41.

CONESA, JUAN CARLOS, SAGIRI KITAO, AND DIRK KRUEGER (2009): "Taxing capital? Not a bad idea after all!," *American Economic Review*, Vol. 99, pp. 25–48.

DAVID, JOEL (2017): "The aggregate implications of mergers and acquisitions," *Available at SSRN 2033555*.

DE LOECKER, JAN AND JAN EECKHOUT (2018): "Global Market Power," Working Paper 24768, National Bureau of Economic Research.

DE NARDI, MARIACRISTINA AND FANG YANG (2016): "Wealth inequality, family background, and estate taxation," *Journal of Monetary Economics*, Vol. 77, pp. 130–145.

DI GIOVANNI, JULIAN AND ANDREI A LEVCHENKO (2010): "Putting the parts together: Trade, vertical linkages, and business cycle comovement," *American Economic Journal: Macroeconomics*, Vol. 2, pp. 95–124.

DI GIOVANNI, JULIAN, ANDREI A. LEVCHENKO, AND ISABELLE MEJEAN (2018): "The Micro Origins of International Business-Cycle Comovement," *American Economic Review*, Vol. 108, pp. 82–108.

DIAMOND, ROBERT M (1998): *Designing and Assessing Courses and Curricula: A Practical Guide. The Jossey-Bass Higher and Adult Education Series.*: ERIC.

DILLON, ELEANOR W AND CHRISTOPHER T STANTON (2017): "Self-employment dynamics and the returns to entrepreneurship,"Technical report, National Bureau of Economic Research.

DROZD, LUKASZ A., SERGEY KOLBIN, AND JAROMIR B. NOSAL (2020): "The Trade-Comovement Puzzle," *American Economic Journal: Macroeconomics, Forthcoming*.

DUNNE, TIMOTHY, MARK J ROBERTS, AND LARRY SAMUELSON (1988): "Patterns of Firm Entry and Exit in US Manufacturing Industries," *The RAND journal of Economics*, pp. 495–515.

DUVAL, ROMAIN, NAN LI, RICHA SARAF, AND DULANI SENEVIRATNE (2015): "Value-added trade and business cycle synchronization," *Journal of International Economics*, pp. –.

DYRDA, SEBASTIAN (2015): "Fluctuations in Uncertainty, Efficient Borrowing Constraints and Firm Dynamics," *Society for Economic Dynamics, 2015 Meeting Papers*.

ENGBOM, NIKLAS (2019): "Firm and Worker Dynamics in an Aging Labor Market," Working Paper 756, Federal Reserve Bank of Minneapolis.

EVANS, DAVID AND LINDA S LEIGHTON (1989): "Some Empirical Aspects of Entrepreneurship," *American Economic Review*, Vol. 79, pp. 519–35.

EVANS, DAVID S. AND BOYAN JOVANOVIC (1989): "An Estimated Model of Entrepreneurial Choice under Liquidity Constraints," *Journal of Political Economy*, Vol. 97, pp. 808–827.

FABRICANT, SOLOMON (1940): "Changes in Total Manufacturing Output," in *The Output of Manufacturing Industries, 1899-1937*: National Bureau of Economic Research, Inc, pp. 663–685.

FAGERENG, ANDREAS, CHARLES GOTTLIEB, AND LUIGI GUISO (2017): "Asset market participation and portfolio choice over the life-cycle," *The Journal of Finance*, Vol. 72, pp. 705–750.

FAGERENG, ANDREAS, LUIGI GUISO, DAVIDE MALACRINO, AND LUIGI PISTAFERRI (2020): "Heterogeneity and persistence in returns to wealth," *Econometrica*, Vol. 88, pp. 115–170.

FAGERENG, ANDREAS, MARTIN BLOMHOFF HOLM, BENJAMIN MOLL, AND GISLE NATVIK (2019): "Saving behavior across the wealth distribution: The importance of capital gains,"Technical report, National Bureau of Economic Research.

FAIRLIE, ROBERT W, JAVIER MIRANDA, AND NIKOLAS ZOLAS (2018): "Job Creation and Survival Among Entrepreneurs: Evidence from the Universe of US Startups," *mimeo*.

FARBER, HENRY S, JESSE ROTHSTEIN, AND ROBERT G VALLETTA (2015): "The effect of extended unemployment insurance benefits: Evidence from the 2012-2013 phase-out," *American Economic Review*, Vol. 105, pp. 171–76.

FATTAL JAEF, ROBERTO N AND JOSE IGNACIO LOPEZ (2014): "Entry, trade costs, and international business cycles," *Journal of International Economics*, Vol. 94, pp. 224–238.

FEENSTRA, ROBERT C (1994): "New Product Varieties and the Measurement of International Prices," *American Economic Review*, Vol. 84, pp. 157–177.

FEENSTRA, ROBERT C, PHILIP A LUCK, MAURICE OBSTFELD, AND KATHERYN N RUSS (2014): "In search of the Armington elasticity,"Technical report, National Bureau of Economic Research.

FEENSTRA, ROBERT C, BENJAMIN R MANDEL, MARSHALL B REINSDORF, AND MATTHEW J SLAUGHTER (2009): "Effects of Terms of Trade Gains and Tariff Changes on the Measurement of U.S. Productivity Growth," Working Paper 15592, National Bureau of Economic Research.

FELDSTEIN, MARTIN S (1969): "The effects of taxation on risk taking," *Journal of Political Economy*, Vol. 77, pp. 755–764.

FLAVIN, MARJORIE AND TAKASHI YAMASHITA (2002): "Owner-occupied housing and the composition of the household portfolio," *American Economic Review*, Vol. 92, pp. 345–362.

DA FONSECA, JOAO GALINDO (2021): "Unemployment, entrepreneurship and firm outcomes," *Review of Economic Dynamics*.

FORBES, KRISTIN (2000): "A Reassessment of the Relationship between Inequality and Growth," *The American Economic Review*, Vol. 90, pp. 869–887.

FRANK, MARK W (2009): "Inequality and growth in the United States: Evidence from a new state-level panel of income inequality measures," *Economic Inquiry*, Vol. 47, pp. 55–68.

FRANKEL, JEFFREY A AND ANDREW K ROSE (1998): "The Endogeneity of the Optimum Currency Area Criteria," *Economic Journal*, Vol. 108, pp. 1009–25.

FUCHS-SCHÜNDELN, NICOLA AND MATTHIAS SCHÜNDELN (2005): "Precautionary savings and Self-Selection: Evidence from the German reunification "experiment"," *The Quarterly Journal of*

Economics, Vol. 120, pp. 1085–1120.

GABAIX, XAVIER (2011): "The granular origins of aggregate fluctuations," *Econometrica*, Vol. 79, pp. 733–772.

GABAIX, XAVIER, JEAN-MICHEL LASRY, PIERRE-LOUIS LIONS, AND BENJAMIN MOLL (2016): "The dynamics of inequality," *Econometrica*, Vol. 84, pp. 2071–2111.

GALOR, ODED AND JOSEPH ZEIRA (1993): "Income Distribution and Macroeconomics," *Review of Economic Studies*, Vol. 60, pp. 35–52.

GARCIA-MACIA, DANIEL (2017): "The Financing of Ideas and the Great Deviation," Working Paper 17/176, International Monetary Fund.

GERRITSEN, AART, BAS JACOBS, ALEXANDRA VICTORIA RUSU, AND KEVIN SPIRITUS (2020): "Optimal taxation of capital income with heterogeneous rates of return."

GHIRONI, FABIO AND MARC J MELITZ (2005): "International trade and macroeconomic dynamics with heterogeneous firms," *The Quarterly Journal of Economics*, Vol. 120, pp. 865–915.

GOLLIER, CHRISTIAN (2001): "Wealth Inequality and Asset Pricing," *The Review of Economic Studies*, Vol. 68, pp. 181–203.

GOMEZ, MATTHIEU ET AL. (2016): "Asset prices and wealth inequality," *Unpublished paper: Princeton*.

——— (2018): "Decomposing the Rise in Top Wealth Inequality," in *2018 Meeting Papers*, No. 1116, Society for Economic Dynamics.

GOPINATH, GITA AND BRENT NEIMAN (2014): "Trade Adjustment and Productivity in Large Crises," *American Economic Review*, Vol. 104, pp. 793–831.

GRUBB, DAVID (2001): "Eligibility criteria for unemployment benefits," *Labour Market Policies and Public Employment Service*.

GUISO, LUIGI AND MONICA PAIELLA (2000): "Risk Aversion, Wealth and Financial Market Imperfections," Mimeo.

GUISO, LUIGI, LUIGI PISTAFERRI, AND FABIANO SCHIVARDI (Forthcoming): "Learning Entrepreneurship From Other Entrepreneurs?," *Journal of Labor Economics*.

GUVENEN, FATIH, GUEORGUI KAMBOUROV, BURHANETTIN KURUSCU, SERGIO OCAMPO-DIAZ, AND DAPHNE CHEN (2019): "Use it or lose it: Efficiency gains from wealth taxation,"Technical report, National Bureau of Economic Research.

GUVENEN, FATIH, FATIH KARAHAN, SERDAR OZKAN, AND JAE SONG (2015): "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?," Working Paper 20913, National Bureau of Economic Research.

——— (2021): "What Do Data on Millions of US Workers Reveal About Lifecycle Earnings Dynamics?" *Econometrica*, Vol. 89, pp. 2303–2339.

HALL, ROBERT E (1988): "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, Vol. 96, pp. 921–47.

HALPERN, LÁSZLÓ, MIKLÓS KOREN, AND ADAM SZEIDL (2015): "Imported Inputs and Productivity," *American Economic Review*, Vol. 105, pp. 3660–3703.

HALTER, DANIEL, MANUEL OECHSLIN, AND JOSEF ZWEIMÜLLER (2014): "Inequality and growth: the neglected time dimension," *Journal of Economic Growth*, Vol. 19, pp. 81–104.

HALVORSEN, ELIN, JOACHIM HUBMER, SERDAR OZKAN, AND SERGIO SALGADO (2021): "Why are the Wealthiest So Wealthy?".

HAMILTON, BARTON H (2000): "Does entrepreneurship pay? An empirical analysis of the returns to self-employment," *Journal of Political economy*, Vol. 108, pp. 604–631.

HAMILTON, BARTON H, NICHOLAS W PAPAGEORGE, AND NIDHI PANDE (2019): "The right stuff? Personality and entrepreneurship," *Quantitative Economics*, Vol. 10, pp. 643–691.

HANSEN, G.D. AND A. İMROHOROĞLU (1992): "The role of unemployment insurance in an economy with liquidity constraints and moral hazard," *Journal of political economy*, Vol. 100, pp. 118–142.

HEATHCOTE, JONATHAN AND FABRIZIO PERRI (2002): "Financial autarky and international business cycles," *Journal of Monetary Economics*, Vol. 49, pp. 601–627.

HEATHCOTE, JONATHAN, KJETIL STORESLETTEN, AND GIOVANNI L VIOLANTE (2010): "The macroeconomic implications of rising wage inequality in the United States," *Journal of Political Economy*, Vol. 118, pp. 681–722.

HEATHCOTE, JONATHAN, KJETIL STORESLETTEN, AND GIOVANNI L. VIOLANTE (2017): "Optimal Tax Progressivity: An Analytical Framework," *The Quarterly Journal of Economics*, Vol. 132, pp. 1693–1754.

HERRANZ, NEUS, STEFAN KRASA, AND ANNE P VILLAMIL (2015a): "Entrepreneurs, risk aversion, and dynamic firms," *Journal of Political Economy*, Vol. 123, pp. 1133–1176.

HERRANZ, NEUS, STEFAN KRASA, AND ANNE P. VILLAMIL (2015b): "Entrepreneurs, Risk Aversion, and Dynamic Firms," *Journal of Political Economy*, Vol. 123, pp. 1133 – 1176.

HOLMES, THOMAS J AND JAMES A SCHMITZ JR (1990): "A theory of entrepreneurship and its application to the study of business transfers," *Journal of political Economy*, Vol. 98, pp. 265–294.

HOMBERT, JOHAN, ANTOINETTE SCHOAR, DAVID SRAER, AND DAVID THESMAR (2020): "Can unemployment insurance spur entrepreneurial activity? Evidence from France," *The Journal of Finance*, Vol. 75, pp. 1247–1285.

HSU, JOANNE W, DAVID A MATSA, AND BRIAN T MELZER (2018): "Unemployment insurance as a housing market stabilizer," *American Economic Review*, Vol. 108, pp. 49–81.

HUBMER, JOACHIM, PER KRUSELL, AND ANTHONY A SMITH JR (2020): "Sources of US wealth inequality: Past, present, and future," in *NBER Macroeconomics Annual 2020, volume 35*: University of Chicago Press.

HUGGETT, MARK (1993): "The risk-free rate in heterogeneous-agent incomplete-insurance economies," *Journal of Economic Dynamics and Control*, Vol. 17, pp. 953–969.

HUMMELS, DAVID AND PETER J. KLENOW (2005): "The Variety and Quality of a Nation's Exports," *American Economic Review*, Vol. 95, pp. 704–723.

HUMPHRIES, JOHN ERIC (2017): *The causes and consequences of self-employment over the life cycle*: The University of Chicago.

HUO, ZHEN, ANDREI A LEVCHENKO, AND NITYA PANDALAI-NAYAR (2020): "Utilization-Adjusted TFP Across Countries: Measurement and Implications for International Comovement," Working Paper 26803, National Bureau of Economic Research.

HURST, ERIK G AND BENJAMIN W PUGSLEY (2015): "Wealth, Tastes, and Entrepreneurial Choice," Working Paper 21644, National Bureau of Economic Research.

HURST, ERIK AND ANNAMARIA LUSARDI (2004): "Liquidity constraints, household wealth, and entrepreneurship," *Journal of Political Economy*, Vol. 112, pp. 319–347.

HURST, ERIK AND BENJAMIN WILD PUGSLEY (2011): "What Do Small Businesses Do?," Working Paper 17041, National Bureau of Economic Research.

IMBS, JEAN (2004): "Trade, finance, specialization, and synchronization," *Review of Economics and Statistics*, Vol. 86, pp. 723–734.

İMROHOROĞLU, SELAHATTIN (1998): "A quantitative analysis of capital income taxation," *International Economic Review*, pp. 307–328.

INKLAAR, ROBERT, RICHARD JONG-A-PIN, AND JAKOB DE HAAN (2008): "Trade and business cycle synchronization in OECD countries: A re-examination," *European Economic Review*, Vol. 52, pp. 646–666.

ISKHAKOV, FEDOR, THOMAS H JØRGENSEN, JOHN RUST, AND BERTEL SCHJERNING (2017): "The Endogenous Grid Method for Discrete-continuous Dynamic Choice Models With (or Without) Taste Shocks," *Quantitative Economics*, Vol. 8, pp. 317–365.

JAKOBSEN, KATRINE, KRISTIAN JAKOBSEN, HENRIK KLEVEN, AND GABRIEL ZUCMAN (2020): "Wealth taxation and wealth accumulation: Theory and evidence from Denmark," *The Quarterly Journal of Economics*, Vol. 135, pp. 329–388.

JOHNSON, ROBERT C. (2014): "Trade in Intermediate Inputs and Business Cycle Comovement," *American Economic Journal: Macroeconomics*, Vol. 6, pp. 39–83.

JOHNSON, ROBERT C. AND GUILLERMO NOGUERA (2017): "A Portrait of Trade in Value Added over Four Decades," *The Review of Economics and Statistics*, Vol. 99(5), pp. 896–911.

JOVANOVIC, BOYAN (1982): "Selection and the Evolution of Industry," *Econometrica*, pp. 649–670.

KACPERCZYK, MARCIN, JAROMIR NOSAL, AND LUMINITA STEVENS (2019): "Investor sophistication and capital income inequality," *Journal of Monetary Economics*, Vol. 107, pp. 18–31.

KALDOR, NICHOLAS (1956): "Alternative Theories of Distribution," *The Review of Economic Studies*, Vol. 23, pp. 83–100.

KAPLAN, GREG, BENJAMIN MOLL, AND GIOVANNI L VIOLANTE (2018): "Monetary policy according to HANK," *American Economic Review*, Vol. 108, pp. 697–743.

KEHOE, TIMOTHY J. AND KIM J. RUHL (2008): "Are shocks to the terms of trade shocks to productivity?" *Review of Economic Dynamics*, Vol. 11, pp. 804 – 819.

KEKRE, ROHAN AND MORITZ LENEL (2020): "Monetary Policy, Redistribution and Risk Premia," Working Paper.

KENNAN, JOHN (2001): "Uniqueness of Positive Fixed Points for Increasing Concave Functions on Rn: An Elementary Result," *Review of Economic Dynamics*, Vol. 4, pp. 893 – 899.

KIHLSTROM, RICHARD E AND JEAN-JACQUES LAFFONT (1979): "A general equilibrium entrepreneurial theory of firm formation based on risk aversion," *Journal of Political Economy*, Vol. 87, pp. 719–748.

KINDERMANN, FABIAN AND DIRK KRUEGER (2014): "High marginal tax rates on the top 1%? Lessons from a life cycle model with idiosyncratic income risk,"Technical report, National Bureau of Economic Research.

KITAO, SAGIRI (2008a): "Entrepreneurship, taxation and capital investment," *Review of Economic Dynamics*, Vol. 11, pp. 44–69.

——— (2008b): "Entrepreneurship, taxation and capital investment," *Review of Economic Dynamics*, Vol. 11, pp. 44–69.

KLASS, OREN S, OFER BIHAM, MOSHE LEVY, OFER MALCAI, AND SORIN SOLOMON (2006): "The Forbes 400 and the Pareto wealth distribution," *Economics Letters*, Vol. 90, pp. 290–295.

KLEVMARKEN, N ANDERS, JOSEPH P LUPTON, AND FRANK P STAFFORD (2003): "Wealth Dynamics in the 1980s and 1990s Sweden and the United States," *Journal of Human Resources*, Vol. 38, pp. 322–353.

KNAUP, AMY E. AND MERISSA C. PIAZZA (2007): "Business Employment Dynamics Data: Survival and Longevity, II," *Monthly Labor Review*.

KOBE, KATHRYN (2012): "Small Business GDP: Update 2002-2010,"Technical report, Small Business Administration, Office of Advocacy.

KOSE, M. AYHAN AND KEI-MU YI (2001): "International Trade and Business Cycles: Is Vertical Specialization the Missing Link?" *The American Economic Review*, Vol. 91, pp. 371–375.

KOSE, M AYHAN AND KEI-MU YI (2006): "Can the standard international business cycle model explain the relation between trade and comovement?" *Journal of international Economics*, Vol. 68, pp. 267–295.

KRUGMAN, PAUL (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, Vol. 70, pp. 950–59.

KRUSELL, PER, TOSHIHIKO MUKOYAMA, RICHARD ROGERSON, AND AYŞEGÜL ŞAHIN (2017): "Gross worker flows over the business cycle," *American Economic Review*, Vol. 107, pp. 3447–76.

KRUSELL, PER, TOSHIHIKO MUKOYAMA, AND AYŞEGÜL ŞAHIN (2010): "Labour-market matching with precautionary savings and aggregate fluctuations," *The Review of Economic Studies*, Vol. 77, pp. 1477–1507.

Krusell, Per and Anthony A Smith, Jr (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of political Economy*, Vol. 106, pp. 867–896.

Lentz, Rasmus and Torben Tranaes (2005): "Job search and savings: Wealth effects and duration dependence," *Journal of labor Economics*, Vol. 23, pp. 467–489.

Lian, Chen and Yueran Ma (2021): "Anatomy of corporate borrowing constraints," *The Quarterly Journal of Economics*, Vol. 136, pp. 229–291.

Liang, James, Hui Wang, and Edward P Lazear (2018): "Demographics and Entrepreneurship," *Journal of Political Economy*, Vol. 126, pp. S140–S196.

Liao, Wei and Ana Maria Santacreu (2015): "The trade comovement puzzle and the margins of international trade," *Journal of International Economics*, Vol. 96, pp. 266 – 288.

Lichtenberg, Frank R, Donald Siegel, Dale Jorgenson, and Edwin Mansfield (1987): "Productivity and changes in ownership of manufacturing plants," *Brookings Papers on Economic Activity*, Vol. 1987, pp. 643–683.

Lockwood, Benjamin B, Charles G Nathanson, and E Glen Weyl (2017): "Taxation and the Allocation of Talent," *Journal of Political Economy*, Vol. 125, pp. 1635–1682.

Lucas Jr, Robert E (1978): "On the size distribution of business firms," *The Bell Journal of Economics*, pp. 508–523.

Lusardi, Annamaria, Pierre-Carl Michaud, and Olivia S Mitchell (2017): "Optimal financial knowledge and wealth inequality," *Journal of Political Economy*, Vol. 125, pp. 431–477.

Makarov, Dmitry and Astrid V. Schornick (2010): "A Note on Wealth Effect under CARA Utility," *Finance Research Letters*, Vol. 7, pp. 170–177.

Mankart, Jochen and Giacomo Rodano (2015): "Personal Bankruptcy Law, Debt Portfolios, and Entrepreneurship," *Journal of Monetary Economics*, Vol. 76, pp. 157–172.

Meeuwis, Maarten (2019): "Wealth Fluctuations and Risk Preferences: Evidence from U.S. Investor Portfolios," Job Market Paper Version.

Merton, Robert C (1969): "Lifetime portfolio selection under uncertainty: The continuous-time case," *The review of Economics and Statistics*, pp. 247–257.

Meyer, Bruce D (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, Vol. 58, pp. 757–782.

Moffitt, Robert (1985): "Unemployment insurance and the distribution of unemployment spells," *Journal of Econometrics*, Vol. 28, pp. 85–101.

Moll, Benjamin (2014): "Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?" *The American Economic Review*, Vol. 104, pp. 3186–3221.

Moll, Benjamin and Oleg Itskhoki (2019): "Optimal Development Policies With Financial Frictions," *Econometrica*, Vol. 87, pp. 139–173.

Mortensen, Dale T and Christopher A Pissarides (1994): "Job creation and job destruction in the theory of unemployment," *The Review of Economic Studies*, Vol. 61, pp. 397–415.

MOSKOWITZ, TOBIAS J AND ANNETTE VISSING-JØRGENSEN (2002): "The returns to entrepreneurial investment: A private equity premium puzzle?" *American Economic Review*, Vol. 92, pp. 745–778.

NAKAJIMA, MAKOTO (2012): "A quantitative analysis of unemployment benefit extensions," *Journal of Monetary Economics*, Vol. 59, pp. 686–702.

NG, ERIC CY (2010): "Production fragmentation and business-cycle comovement," *Journal of International Economics*, Vol. 82, pp. 1–14.

PENCIAKOVA, VERONIKA (2018): "Diversification & Risky Innovation among US Firms."

PERESS, JOEL (2004): "Wealth, information acquisition, and portfolio choice," *The Review of Financial Studies*, Vol. 17, pp. 879–914.

PEROTTI, ROBERTO (1996): "Growth, income distribution, and democracy: What the data say," *Journal of Economic Growth*, Vol. 1, pp. 149–187.

PFEFFER, FABIAN T, ROBERT F SCHOENI, ARTHUR KENNICKELL, AND PATRICIA ANDRESKI (2016): "Measuring wealth and wealth inequality: Comparing two US surveys," *Journal of economic and social measurement*, Vol. 41, pp. 103–120.

PIKETTY, THOMAS (2018): *Capital in the twenty-first century*: Harvard University Press.

PIKETTY, THOMAS AND EMMANUEL SAEZ (2003): "Income inequality in the United States, 1913–1998," *The Quarterly journal of economics*, Vol. 118, pp. 1–41.

PIKETTY, THOMAS, EMMANUEL SAEZ, AND STEFANIE STANTCHEVA (2014): "Optimal taxation of top labor incomes: A tale of three elasticities," *American economic journal: economic policy*, Vol. 6, pp. 230–71.

POSCHKE, MARKUS (2013): "Who becomes an entrepreneur? Labor market prospects and occupational choice," *Journal of Economic Dynamics and Control*, Vol. 37, pp. 693–710.

——— (2019): "Wage employment, unemployment and self-employment across countries."

QUADRINI, VINCENZO (1999): "The Importance of Entrepreneurship for Wealth Concentration and Mobility," *Review of Income and Wealth*, Vol. 45, pp. 1–19.

——— (2000a): "Entrepreneurship, saving, and social mobility," *Review of economic dynamics*, Vol. 3, pp. 1–40.

——— (2000b): "Entrepreneurship, Saving and Social Mobility," *Review of Economic Dynamics*, Vol. 3, pp. 1–40.

RAVENSCRAFT, DAVID J AND FREDERIC M SCHERER (2011): *Mergers, sell-offs, and economic efficiency*: Brookings Institution Press.

ROBINSON, BAXTER (2012): "Risky business: The choice of entrepreneurial risk under incomplete markets," *Job Market Paper*.

RØED, KNUT AND JENS FREDRIK SKOGSTRØM (2014): "Unemployment insurance and entrepreneurship," *Labour*, Vol. 28, pp. 430–448.

ROTHSCHILD, CASEY AND FLORIAN SCHEUER (2016): "Optimal taxation with rent-seeking," *The Review of Economic Studies*, Vol. 83, pp. 1225–1262.

ROTHSTEIN, JESSE (2011): "Unemployment insurance and job search in the Great Recession,"Technical report, National Bureau of Economic Research.

ROUSSANOV, NIKOLAI (2010): "Diversification and its discontents: Idiosyncratic and entrepreneurial risk in the quest for social status," *The Journal of Finance*, Vol. 65, pp. 1755–1788.

RUBINSTEIN, YONA AND ROSS LEVINE (2020): "Selection into entrepreneurship and self-employment."

SAEZ, EMMANUEL (2001): "Using elasticities to derive optimal income tax rates," *The review of economic studies*, Vol. 68, pp. 205–229.

SAEZ, EMMANUEL AND GABRIEL ZUCMAN (2016): "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data," *The Quarterly Journal of Economics*, Vol. 131, pp. 519–578.

——— (2019): "Progressive wealth taxation," *Brookings Papers on Economic Activity*, Vol. 2019, pp. 437–533.

SAITO, MIKA (2004): "Armington elasticities in intermediate inputs trade: a problem in using multilateral trade data," *Canadian Journal of Economics*, Vol. 37, pp. 1097–1117.

SAKAI, KOJI, IICHIRO UESUGI, AND TSUTOMU WATANABE (2010): "Firm Age and the Evolution of Borrowing Costs: Evidence from Japanese Small Firms," *Journal of Banking & Finance*, Vol. 34, pp. 1970–1981.

SAMUELSON, PAUL A (1969): "Lifetime portfolio selection by dynamic stochastic programming," *The review of economics and statistics*, pp. 239–246.

SCHEUER, FLORIAN AND JOEL SLEMROD (2020): "Taxation and the Superrich," *Annual Review of Economics*, Vol. 12, pp. 189–211.

——— (2021): "Taxing our wealth," *Journal of Economic Perspectives*, Vol. 35, pp. 207–30.

SCHULZ, KARL (2021): "Redistribution of Return Inequality."

SHOURIDEH, ALI ET AL. (2012): "Optimal taxation of wealthy individuals," *Work. pap. U. of Pennsylvania*.

SIMS, CHRISTOPHER A (1969): "Theoretical Basis for a Double Deflated Index of Real Value Added," *The Review of Economics and Statistics*, Vol. 51, pp. 470–471.

SMITH, MATTHEW, DANNY YAGAN, OWEN ZIDAR, AND ERIC ZWICK (2019a): "Capitalists in the Twenty-first Century," *The Quarterly Journal of Economics*, Vol. 134, pp. 1675–1745.

——— (2019b): "Capitalists in the Twenty-First Century*," *The Quarterly Journal of Economics*, Vol. 134, pp. 1675–1745.

STIGLITZ, J. E. (1969): "Distribution of Income and Wealth Among Individuals," *Econometrica*, Vol. 37, pp. 382–397.

STORESLETTEN, KJETIL, CHRIS I TELMER, AND AMIR YARON (2004): "Cyclical dynamics in idiosyncratic labor market risk," *Journal of Political Economy*, Vol. 112, pp. 695–717.

STRAUB, LUDWIG (2018): "Consumption, savings, and the distribution of permanent income," *Unpublished manuscript, Harvard University*.

VERESHCHAGINA, GALINA AND HUGO A HOPENHAYN (2009): "Risk taking by entrepreneurs," *American Economic Review*, Vol. 99, pp. 1808–30.

VERMEULEN, PHILIP (2016): "Estimating the Top Tail of the Wealth Distribution," *The American Economic Review*, Vol. 106, pp. 646–650.

VOITCHOVSKY, SARAH (2005): "Does the profile of income inequality matter for economic growth?" *Journal of Economic Growth*, Vol. 10, pp. 273–296.

WARUSAWITHARANA, MISSAKA (2018): "Profitability and the Lifecycle of Firms," *The BE Journal of Macroeconomics*, Vol. 18.

WASMER, ETIENNE AND PHILIPPE WEIL (2004): "The macroeconomics of labor and credit market imperfections," *American Economic Review*, Vol. 94, pp. 944–963.

XAVIER, INES (2020): "Wealth Inequality in the US: the Role of Heterogeneous Returns."

XU, WENJIAN (2019): "Entrepreneurship Among the Unemployed: The Effects of Unemployment Benefits," *Available at SSRN 3403783*.