THÈSE de DOCTORAT

de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par

Monsieur Gerard MAIDEU MORERA

Le 19 juin 2025

Essays on Macroeconomics

École doctorale : **TSE Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

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

Thèse dirigée par *Monsieur Christian HELLWIG*

Composition du jury

Rapporteur : M. Árpád ÁBRAHáM
Rapporteur : M. Timo BOPPART
Examinateur : M. Thierry MAGNAC
Directeur de thèse : M. Christian HELLWIG

Essays on Macroeconomics

Gerard Maideu-Morera

May 18, 2025

Aknowledgements

I am indebted to Christian Hellwig for his continuous support, guidance, and generosity. His passion for economics, deep thinking, and taste for good work have been a constant source of inspiration. I am also indebted to Eugenia Gonzalez-Aguado, Nicolas Werquin, and Charles Brendon for their invaluable feedback throughout the PhD and for always being selflessly available whenever I needed help. I also want to thank Fabrice Collard, Patrick Fève, Paul Diegert, Andreas Schaab, Alessandro Pavan, Ulrich Hege, and many other TSE faculty and staff. I am especially thankful to Nour Meddahi for his support and encouragement since the start of the PhD.

I have immensely benefited from discussions and feedback from students of the macro group. For that, I want to thank Alexandre Gaillard, Fernando Stipanicic, Philip Wangner, Jonas Gathen, Oscar Fentanes, Ying Wang, Pablo Mileni, Emil Mortensen, Sergi Barcons, and especially Tanja Linta and Andrei Zaloilo with whom I have shared countless hours of discussions in the office. In particular, I want to thank my coauthor Maria Frech, with whom I have greatly enjoyed working and who has helped me become a better researcher.

I thank Amirreza Ahmadzadeh, Luca Bennati, Michel Biseglia, Enrico Mattia, Alessio Ozzane, Oscar Vilargunter, Wenxuan Xu, and many other fellow students for making these years more enjoyable. I am particularly grateful to Pau Juan and Maxim Sandiumenge for their continuous support and for helping me grow academically and personally. The PhD would have been a very different experience without them.

Finally, I am deeply grateful to my parents, Carmina Morera and Joaquim Maideu, and my brother, Biel. I am very lucky to have their support, and I want to dedicate this thesis to them.

Summary

This thesis is divided into three chapters. Chapter 1, based on my job market paper, aims to understand the importance of non-wage job amenities for measuring economic growth and the distributional effects of technical change. Chapters 2 and 3 use dynamic contracting methods to study the gender wage gap and the motherhood penalty, and entrepreneurship and financing constraints.

Smith (1776) already introduced the idea of compensating differentials—i.e., that wages vary with non-pecuniary job amenities. Since then, the theory and empirical evidence of compensating differentials have become well-established. Amenities are essentially goods in that they enter workers' utility functions and are part of the production process. Yet, traditional growth, distributional, or welfare accounting abstract from non-pecuniary job characteristics. Chapter 1 aims to bridge this gap.

I begin by estimating shadow prices for a range of job amenities that I can measure with occupation-level survey data. With the estimates, I document an amenity-biased shift in labor demand in the US from 1980 to 2015, which reallocated workers from low- to high-amenity occupations. In the remainder of the paper, I make the case that accounting for this amenity-biased reallocation significantly alters our understanding of major macroeconomic changes of the past decades. First, I theoretically show that the shadow value of amenities should be included in output to measure productivity growth. Otherwise, conventional measures—that only account for the costs of amenities—can underestimate it. More formally, Hulten's (1978) theorem—a classical aggregation result for efficient economies—only holds when output includes the shadow value of amenities. Quantitatively, I find that total compensation (wage plus the value of amenities) grew 40% more than wages from 1980 to 2015. Compared to conventional estimates, this implies 25% higher productivity growth but a larger slowdown since 2004. Second, I show that measuring job amenities is also important to understand the distributional effects of technical change. Despite the well-studied polarization along the wage distribution, I find no labor market polarization along the distribution of total compensation.

Chapter 2, together with Maria Frech, is motivated by extensive empirical evidence that highlights women's demand for flexible working hours as a critical cause of the persistent gender disparities in the labor market. We propose a theory of how *hidden* demand for flexibility drives gendered employment dynamics. We develop a dynamic contracting model between an employer and an employee whose time availability is stochastic and unverifiable. In our framework, men and women only to differ in their probability of hav-

ing low time availability, which we measure in the ATUS. We explore contracts designed specifically for each gender (*gender-tailored*) and the polar case where a *male-tailored* contract is given to both men and women. For the latter, we show that contracting frictions endogenously give rise to well-documented gendered labor market outcomes: (i) the divergence and non-convergence of gender earnings differentials over the life-cycle, and (ii) women's shorter job duration and weaker labor force attachment.

Chapter 3 builds on a large literature that uses dynamic contracting methods to understand how agency frictions create financing constraints and drive firm lifecycle dynamics. The literature has typically analyzed these problems assuming a risk-neutral entrepreneur and i.i.d productivity shocks. Using recent advances from the dynamic taxation literature, I study a canonical dynamic cash flow diversion model as Clementi and Hopenhayn (2006), but with a risk-averse entrepreneur who has persistent private information about the firm's productivity. I find that risk aversion fundamentally changes the properties of the optimal contract. The firm's size is always distorted downwards, and its distortions inherit the autoregressive properties of the type process. The entrepreneur's compensation is smoothed and decoupled from the firm size dynamics. Finally, I use numerical simulations to study a quasi-implementation with simpler contracts, which highlights that this class of models is unable to generate realistic firm size and equity share dynamics simultaneously.

Contents

1 Amenity-Biased Technical Change **1**

1.1 Introduction 2

1.2 Data sources 8

1.3 Motivating evidence 10

1.4 Measurement of occupational amenities 16

1.5 Amenity bias and no polarization by total compensation 27

1.6 Growth accounting 32

1.7 Extensions 47

1.8 Conclusion 48

2 The Hidden Demand for Flexibility - A Theory of Gendered Employment Dynamics **51**

2.1 Introduction 52

2.2 Environment 58

2.3 Optimal Contract 62

2.4 Generating Gendered Employment Dynamics 69

2.5 Numerical Simulations 75

2.6 Numerical Results 81

2.7 Recovering Stylized Facts 86

2.8 Concluding Remarks 89

3 **Firm Size and Compensation Dynamics with Risk Aversion and Persistent Private Information** **91**
3.1 Introduction . . . 92
3.2 Model . . . 97
3.3 Optimal allocation . . . 104
3.4 Numerical simulations . . . 115
3.5 Quasi-implementation . . . 119
3.6 Comparison with risk neutral and equity dynamics . . . 125
3.7 Conclusion . . . 127
References . . . 128

A **Appendix of Chapter 1: Amenity-Biased Technical Change** **142**
A.1 Data details . . . 142
A.2 Details amenity pricing method . . . 146
A.3 Gender differences in amenity prices and sorting . . . 161
A.4 Evolution of amenities within occupations . . . 164
A.5 Extra figures . . . 169
A.6 Details growth accounting . . . 174
A.7 Extensions . . . 178

B **Appendix of Chapter 2: The Hidden Demand for Flexibility - A Theory of Gendered Employment Dynamics** **188**
B.1 Extra Tables and Figures . . . 188
B.2 Proofs . . . 190
B.3 Details Numerical Solution . . . 199
B.4 Data . . . 202
B.5 Incentive Compatibility in Male-tailored Contracts . . . 203
B.6 Extensions . . . 206

C Appendix of Chapter 3: Firm Size and Compensation Dynamics with Risk Aversion and Persistent Private Information **212**

C.1 Appendix: Derivations and proofs . . . 212

C.2 Additional tables and figures . . . 217

C.3 Extensions . . . 222

C.4 Details numerical simulations . . . 229

Chapter 1

Amenity-Biased Technical Change

Gerard Maideu-Morera[1]

Abstract

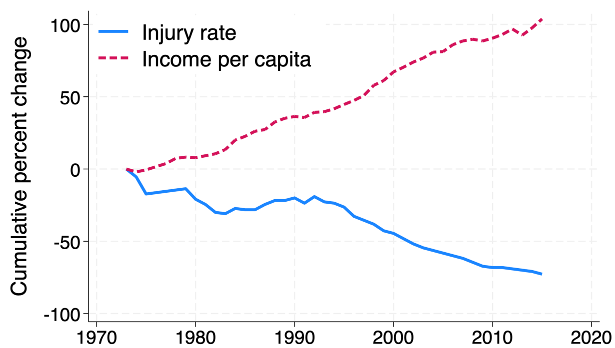
I argue that technical change has raised living standards not only by increasing wages but also by making work more pleasant and safer. Yet, traditional growth, distributional, or welfare accounting abstract from non-pecuniary job characteristics. By estimating shadow prices for job amenities, I first document an amenity-biased shift in labor demand in the US from 1980 to 2015, which reallocated workers from low- to high-amenity occupations. This reallocation significantly alters our understanding of several major macroeconomic changes. First, I theoretically show that the shadow value of amenities should be included in output to measure productivity growth. Otherwise, conventional measures—that only account for the costs of amenities—can underestimate it. Quantitatively, I find that total compensation (wage plus the value of amenities) grew 40% more than wages from 1980 to 2015. Compared to conventional estimates, this implies 25% higher productivity growth but a larger slowdown since 2004. Second, I find no labor market polarization along the distribution of total compensation; employment and relative wages declined the most at the bottom of the distribution instead of in the middle.

[1]I am grateful to Christian Hellwig, Nicolas Werquin, Eugenia Gonzalez-Aguado, and Charles Brendon for their advice and guidance through the project. I have greatly benefited from advice from Paul Diegert and Víctor Ríos-Rull. I would also like to thank George-Marios Angeletos, Matteo Bobba, Fabrice Collard, Olivier De Groote, Jason Faberman, Patrick Fève, Maria Frech, Alexandre Gaillard, Jonas Gathen, Robert Gordon, Pau Juan-Bartroli, Narayana Kocherlakota, Martí Mestieri, Alessandro Pavan, Mathias Reynaert, François Salanié, Wenxuan Xu, Andrei Zaloilo, and seminar participants at the Vigo macro workshop and TSE workshops.

1.1 Introduction

Technological progress is understood to be the long-run driver of living standards, which are typically measured with wages. However, workers also value non-pecuniary job characteristics (amenities) such as safety, schedule flexibility, or how meaningful a job is. The theory of compensating differentials states that workers trade off lower wages for better amenities, and recent research finds that these differentials are large (Rosen, 1986; Lavetti, 2023). Hence, improved amenities, such as safer jobs, can also raise living standards. Figure 1.1 shows that in the US, from 1973 to 2015, while real income per capita doubled, average occupational injury rates decreased nearly fourfold. Traditional output or income measures do not account for the improved safety or potentially many other improvements in job amenities. Yet, any welfare assessment based only on pecuniary compensation is incomplete if workers value amenities. Moreover, increases in amenities can decrease output and productivity estimates because the costs of providing amenities are typically measured but not their value.

Figure 1.1: Changes in occupational injury rates and income per capita (US, 1973-2015)



Note: The injury rate is the aggregate occupational injury and illnesses rate constructed from historical news releases of the Survey of Occupational Injuries and Illnesses (SOII).

In this paper, I introduce the notion of amenity-biased technical change, defined as a change in production technology that leads to increases in non-pecuniary compensation or differentially impacts jobs based on their amenities. By estimating shadow prices for job amenities, I first document an amenity-biased reallocation of labor demand from 1980 to 2015. Employment and relative wages decreased in low-amenity occupations and increased in high-amenity ones.

Accounting for this amenity-biased reallocation significantly alters our understanding of major macroeconomic changes of the past decades. First, I show that augmenting output to include the shadow value of amenities yields a more welfare-relevant growth estimate and properly measures productivity improvements. Quantitatively, I estimate that total compensation (wage plus the value of amenities) grew 40% more than wages between 1980 and 2015. Compared to the conventional total factor productivity (TFP) measure, this implies 25% larger productivity growth but a larger slowdown between 2004 and 2015. Second, the US labor market has experienced a polarization since the 1980s, where employment and relative wages have decreased in middle-wage occupations (Acemoglu, 1999; Autor *et al.*, 2006). By contrast, I find that there is no polarization if occupations are ranked by total compensation. I argue that this suggests an aggregate shift in demand from low- to high-skill workers (as in Bound and Johnson, 1992 or Katz and Murphy, 1992).

Motivating evidence. I begin by providing indirect evidence of compensating differentials by comparing wages and proxies for workers' skills (average AFQT score and average years of schooling) across occupations.[2] Wage differences unexplained by skill differences suggest compensating differentials. I find a significant mismatch between wages and skill proxies across occupation groups that is consistent with differences in several measured amenities. Specifically, blue-collar occupations have similar skill proxy values as services despite earning substantially higher wages and score low in multiple job characteristics, suggesting that part of their wages reflect compensating differentials. Blue-collar occupations experienced the largest employment declines since the 1980s. I argue that their declines led to an amenity-biased reallocation and aggregate improvements in non-pecuniary compensation. Additionally, I show that—despite the polarization by wages—there is no polarization if we rank occupations by the skill proxies.

Amenity prices. To formally assess the hypothesis of the amenity-biased reallocation and quantify aggregate improvements in non-pecuniary compensation, I then turn to the estimation of amenity prices. First, I lay down the occupational choice setup used throughout the paper. Workers have heterogeneous productivities and preferences, and occupations have heterogeneous amenities and wage functions mapping productivities to wages. I define the worker's frontier as the set of undominated combinations of wages and amenities available to the worker, given her productivity type. The slope of these frontiers at each point defines the local amenity prices.

The identification challenge results from not observing workers' frontiers, especially since higher-skilled workers tend to have frontiers with higher wages and higher ameni-

[2]The AFQT score is a general aptitude test that was, in particular, administered to the subjects of the NLSY79 and is commonly used as an ability proxy in labor economics. I also use non-cognitive proxies and find that they deliver similar comparisons across occupation groups.

ties. To tackle this, I use the methodology developed by Bell (2022) and Bell *et al.* (2024), which extends the hedonic pricing (Rosen, 1974) to account for heterogeneous frontiers.[3] By placing assumptions on the frontiers, I can map the occupational model to their reduced form setup. Their approach uses an imprecise proxy for workers' skills as a shifter of the frontier to recover its tangent vector—i.e., the amenity prices. Intuitively, a valid proxy must correlate with the worker's skills but be unrelated to preferences. I use occupation-level amenity data from the O*NET and the ATUS (from Kaplan and Schulhofer-Wohl, 2018) and worker-level data from the NLSY79. The data allows me to measure a broad range of amenities, from tangible ones like exposure to contaminants to subjective ones such as feelings of meaning. Moreover, for the latter periods, I can use different releases of the O*NET to measure changes in amenities over time within occupations.

Results. Using the amenity price estimates, I can aggregate to occupation-level amenity and total compensation measures, which I use to document:

1. *Amenity-biased reallocation of labor demand between 1980 and 2015.* I document that employment and relative wages decreased in low-amenity occupations and increased in high-amenity ones. This result is largely explained by shifts across broad occupation groups—especially by the declines in blue-collar occupations, which tend to have the lowest amenities. The simultaneous decrease in wages and employment suggests a decrease in the labor demand for low-amenity occupations.[4]

2. *No polarization along the total compensation distribution.* Employment and relative wages decreased at the bottom of the total compensation distribution rather than in the middle. This finding is a byproduct of the amenity-biased reallocation. In particular, the polarization disappears on aggregate because blue-collar and service occupations have roughly the same total compensation level.[5] Moreover, total compensation can be interpreted as a skill measure that corrects for compensating differentials. Thus, we find no polarization along the skill distribution, suggesting that, on aggregate, there has been a decrease in the demand for low-skill workers—as in the older skill-biased technical change literature (Bound and Johnson, 1992; Katz and Murphy, 1992).

[3] Bell (2022) first proposed the method for the estimation of compensating differentials in the labor market, and then Bell *et al.* (2024) extended it in the housing context.

[4] If employment decreased in low-amenity occupations because workers now value more amenities, we should observe an increase in wages in these occupations. Conversely, if relative wages in low-amenity occupations decrease due to improved working conditions, employment should have increased.

[5] Similar total compensation levels between blue-collar and service occupations do not imply that workers are roughly indifferent due to preference heterogeneity. However, the changes in utility for blue-collar workers who move to services would be less than the observed reduction in wages.

Growth accounting. Next, I show that TFP underestimated productivity improvements because conventional output measures do not account for increases in amenities. To this end, I extend the empirical setup and introduce the production side of the economy. I model biased changes in labor demand by assuming that each occupation's output can be produced with labor and/or an occupation-specific new technology, which could capture, for example, industrial robots or foreign workers (offshoring).

First, I illustrate the key intuitions using a simple two-occupation example. Since the low-amenity occupation must pay a higher wage to attract workers, its marginal product of labor (MPL) is higher. Consequently, output, income, and TFP decrease after workers reallocate to the high-amenity occupation. The difference in MPLs between the low- and high-amenity occupations is the shadow cost of providing higher amenities, which is measured as a misallocation because the shadow value of amenities is not accounted for. When workers move to the high-amenity occupation, they enjoy better amenities and reduce their consumption of normal goods, but only the value of the latter is measured. Therefore, by measuring total compensation growth, we can (i) obtain a more welfare-relevant growth estimate and (ii) properly measure productivity improvements by correcting for the misallocation in measured TFP.

More generally, worker reallocations across occupations have first-order effects on TFP akin to changes in allocative efficiency in inefficient economies where Hulten's theorem (Hulten, 1978) does not hold (Baqaee and Farhi, 2020). Hulten's theorem—i.e., only changes in technology have first-order effects on TFP—is often understood as a macro envelope condition from the first welfare theorem. Although the equilibrium is Pareto efficient, Hulten's theorem does not hold because output and TFP are not maximized when workers value amenities. The main result is a Hulten's theorem for augmented TFP growth, which is defined as usual but with output augmented to include the shadow value of amenities. This yields growth accounting formulas that allow us to measure productivity improvements with amenity prices, including gains from improved amenities within occupations (e.g., safer machines). Additionally, I define the corrected TFP growth as the growth in conventional output from the changes in technology, which can be computed by adding the change in amenity value to conventional TFP growth.[6]

The growth accounting formulas derived are nonparametric and, under assumptions on workers' frontiers, can be measured with the amenity price estimates. I find that since

[6] I study the growth accounting in two extensions. First, a model where firms provide amenities at a cost. Increased spending on amenities reduces measured TFP by raising costs without increasing workers' measured compensation or sales. Second, an inefficient economy with occupation-level wage markdowns where I characterize TFP and corrected TFP growth.

1980, due to the reallocation of workers, the per-worker amenity value increased by 6424 dollars in 2012 prices, implying that total compensation grew 40% more than wages. As a result, the corrected TFP growth has been 25% larger than the conventional TFP growth, with an average growth rate of 1.9% compared to 1.5% for TFP. That is, the estimates suggest that productivity improvements were 25% larger than indicated by the conventional measure. For the latter years, I can use different releases in the O*NET to estimate augmented TFP growth. Between 2006 and 2015, the augmented TFP growth was equal to 8.5%, compared to a TFP growth of 7.9% and a corrected TFP growth of 9.7%.[7] Of the 8.5% growth, 0.6 percentage points are explained by improvements in amenities within occupations.[8]

Related literature. In this paper, I introduce job amenities and the theory of compensating differentials, which are well-established in labor economics, to the broad literatures on economic growth and technical change. However, more specifically, I contribute to the literatures on growth accounting and labor market polarization.

The growth accounting literature aims to quantify the sources of productivity growth (Hulten, 1978; Hall (1990); Basu and Fernald, 2002; Petrin and Levinsohn, 2012; Baqaee and Farhi, 2020; Dávila and Schaab, 2023).[9,10] The main contribution is to show how the conventional TFP measure does not properly measure productivity improvements when jobs have heterogeneous amenities and to quantify that, as a result, technological progress has been understated. Formally, I show that, in an efficient economy with job amenities, Hulten's theorem does not hold for the conventional output measure but holds if output is augmented to include the shadow value of amenities.

An extensive literature has documented and studied the causes and consequences of the polarization of the labor market. Earlier contributions include Acemoglu (1999), Autor *et al.* (2003), Autor *et al.* (2006), or Goos and Manning (2007).[11] Accounting for

[7] The augmented TFP growth is smaller than the corrected TFP growth because it is relative to the augmented output, which is larger than the measured output.

[8] This estimate should be interpreted as a lower bound as I cannot measure changes in all amenities.

[9] Closer to the technical change literature, Acemoglu (2024) uses Hulten's theorem to evaluate the productivity effects of AI in a task-based model.

[10] A related literature quantifies the aggregate effects of misallocation in inefficient economies (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). On growth accounting, see Baqaee and Farhi (2020), and on the labor market, see Hsieh *et al.* (2019) and Berger *et al.* (2022). I show how, with amenities, we can measure misallocation in an efficient economy and correct it with amenity prices. Additionally, I explore how the interaction between inefficiencies from markdowns and compensating differentials affects TFP growth.

[11] Böhm (2020) estimates changes in task prices in a Roy model with manual, routine, and abstract tasks and also allows for differences in non-pecuniary utility in each of the tasks. Consistent with my estimates, he finds a lower non-pecuniary utility for routine tasks.

amenities provides a more nuanced understanding of the impact of polarization. First, I find that there is no polarization by total compensation.

Second, the polarization literature often refers to wages and workers' skills interchangeably, so that polarization has impacted the middle-skilled occupations (e.g., Autor *et al.*, 2006; Acemoglu and Autor, 2011; Autor and Dorn (2013)). By contrast, I find no evidence of polarization along the skill distribution—both with skill proxies and using total compensation as a skill measure—suggesting an aggregate decrease in relative demand for low-skill workers.

Some papers also examine changes in non-pecuniary job characteristics over time. Hamermesh (1999) analyzed changes in injury rates and nighttime work and documented that injury rates fell more in industries with larger income growth.[12] Using survey data from the ATUS on workers' feelings about work, Kaplan and Schulhofer-Wohl (2018) document changes in the disutility of work due to the reallocation of workers across occupations. Boar and Lashkari (2021) construct an index of non-pecuniary quality of occupations and show that children of richer parents are more likely to sort into occupations with higher index. Additionally, they document a reallocation of employment and wages toward occupations with higher non-pecuniary quality and quantify its impact on intergenerational mobility and welfare. By explicitly measuring amenity prices, I can aggregate multiple amenities into a unidimensional measure.[13] Moreover, this enables me to quantify the implications for labor market polarization and growth accounting.

The quantitative results also speak to the recent literature studying the US productivity slowdown from 2004 to 2015 (Fernald, 2015; Syverson, 2017; Byrne *et al.*, 2016; Gordon, 2018).[14] I find that the corrected TFP estimates indicate an even larger productivity slowdown because most of the amenity-biased reallocation of workers had already occurred by 2006. Note that this is only suggestive as increases in expenditures on amenities within occupations or firms, which I do not measure, could have contributed to the slowdown.

Finally, long since Nordhaus and Tobin (1972), economists have aimed at constructing

[12]Relatedly, Hamermesh (2001) used survey responses on job satisfaction from the NLSY79 and documented a widening of the distribution of job satisfaction.

[13]Boar and Lashkari (2021) use survey data from the Quality of Work-life Module of the General Social Survey and principal. They use principal component analysis to construct a unidimensional intrinsic quality index and estimate the compensating differential for the index with a structural model. The O*NET data also allows me to use less aggregated occupation codes.

[14]Relatedly, Rachel (2024) develops a theory of growth with leisure-enhancing technologies (e.g., social media). He shows that faster growth in the leisure economy is associated with slower growth of traditional TFP. As I do, he argues that GDP becomes a less accurate measure of welfare. However, he finds that GDP is not significantly mismeasured because these goods have zero or very low marginal costs. By contrast, my amenity price estimates indicate substantial shadow costs of amenities, so I find a large mismeasurement.

better measures of economic welfare than GDP.[15] More recently, there has been a renewed interest in this line of work. For instance, Jones and Klenow (2016) create a measure incorporating consumption, leisure, mortality, and inequality and use it to compare welfare across countries and over time.[16] Similarly, I use theory to construct a measure that better captures welfare than GDP. However, ignoring the amenities causes mismeasurement in the conventional TFP. Moreover, by defining augmented output, I can follow the same approach of the traditional growth accounting literature; I define productivity growth in the usual way and derive a Hulten's theorem.

Outline. The paper is organized as follows. Section 1.2 describes the data sources used throughout the paper. Section 1.3 provides motivating evidence for the amenity-biased reallocation and its implications. Section 1.4 lays down the occupational choice setup and explains the estimation of amenity prices and the construction of the amenity and total compensation measures. Section 1.5 presents the main results on the amenity-biased reallocation and the absence of polarization along the total compensation distribution. Section 1.6 studies growth accounting; I first go through the theory and then do the quantitative analysis. Finally, Section 1.7 discusses the extensions, and Section 1.8 concludes.

1.2 Data sources

I now describe the data sources used throughout the paper. I use occupation-level data on job characteristics (amenities) from the Occupational Information Network (O*NET) and the American Time Use Survey (ATUS). I obtain individual-level survey data on workers' income, occupations, and characteristics (mainly skill proxies) from the National Longitudinal Survey of Youth 1979 (NLSY79). In addition, I use the census and American Community Survey (ACS) to construct occupation-level employment and wage data. All the occupation-level data is cross-walked into the occupation codes developed by Autor and Dorn (2013), which provide consistent coding for 330 occupations across all census and ACS samples.

Amenities data. I use three sources of data for amenities: (i) the O*NET context file (as Bell, 2022), (ii) the O*NET interests file, and (iii) data from Kaplan and Schulhofer-Wohl

[15] Nordhaus and Tobin (1972) developed a *Measure of Economic Welfare* that, in particular, deducted the disamenities from urbanization by using income differentials across localities with different population densities.

[16] Other contributions include Becker *et al.* (2005), Cordoba and Verdier (2008), Boarini *et al.* (2006), Fleurbaey and Gaulier (2009), or Fleurbaey (2009) (a review). Also related are Basu *et al.* (2022) and Durán and Licandro (2024), who show conditions under which conventional output or TFP measure welfare.

(2018), which is constructed from the ATUS.

The O*NET is a comprehensive database of worker attributes and job characteristics constructed and maintained by the U.S. Department of Labor. Its data is constructed through a combination of occupational surveys, expert reviews, and input from a broad range of workers and occupations. The O*NET context file provides detailed occupation-level characteristics such as the frequency of exposure to contaminants, the importance of teamwork, or the time spent standing. However, not all characteristics may be considered amenities, as it is reasonable to assume that they do not affect workers' utility. Therefore, I manually exclude these types of characteristics. For example, I exclude variables such as the frequency with which a job requires written letters or memos. In total, I select 17 characteristics out of 57 available. In Table A.1 in Appendix A.1, I describe all the variables from the context file used and how I define each amenity. With each new release of the O*NET, data on some of the occupations are updated. I use data from 2006, which is the earliest year with complete data available. Additionally, I use data from 2022 to measure changes in amenities in the later years and assess the sensitivity of the estimates to changes in amenities over time.

The O*NET interests file is based on the RIASEC model, and it constructs a score on how realistic, investigative, artistic, social, enterprising, or conventional every occupation is (see Table A.2 in Appendix A.1). Finally, I use the ATUS data from Kaplan and Schulhofer-Wohl (2018), who use survey responses on workers' feelings during working hours to construct measures of the disutility of work across occupations. I use the feeling responses about stress, tiredness, pain, and meaning.[17]

The three sources combined give a total of 27 amenities across 330 occupations. The measured amenities do not capture all the potential occupational characteristics that workers may value. However, they cover a broad range of characteristics, from more tangible ones, such as the exposure to contaminants, to more subjective ones, like the feelings of meaning.[18] Further details on the amenities data can be found in Appendix A.1.1.

Income, occupation, and skill proxy data (NLSY79). I use data on workers' income, occupation, and skills from the NLSY79. The NLSY79 is a longitudinal study tracking the labor market activities and life events of Americans born between 1957 and 1964 since 1979. The NLSY79 is useful because it provides several variables that can be used as prox-

[17] I do not include the responses on whether the worker is happy or sad due to the concern that these may be directly influenced by the worker's wage.

[18] There are a few missing occupations in each amenities dataset. Whenever this is the case, I impute the corresponding amenities using the average of the occupation group of the missing occupation.

ies for workers' skills. The main one that I use is the AFQT (Armed Forces Qualification Test) score. The AFQT is a general aptitude test score derived from the math and verbal subsets of the Armed Services Vocational Aptitude Battery (ASVAB), and it was administered to subjects of the NLSY79 panel in 1980. The AFQT test is a commonly used skill proxy in labor economics (e.g., Farber and Gibbons, 1996; Altonji and Pierret, 2001; Neal and Johnson, 1996). I also use the variables self-esteem and mastery as well as height (residualized on sex) as proxies (Bell, 2022). I use data from 1986 (ages 23-29), 1996 (ages 33-39), 2006 (ages 43-49), and 2016 (ages 53-59). More details on the data and sample selection can be found in Appendix A.1.2.

Census and ACS data on wages and employment by occupation. Finally, I obtain employment and wage data across occupations and years from the census and the ACS. I use data for the years 1980-2015 and closely follow Deming (2017) for the data cleaning and aggregation. More details can be found in Appendix A.1.3.

1.3 Motivating evidence

In this section, I provide motivating evidence for the main findings of the paper. First, I discuss some recent labor literature estimating compensating differentials and the magnitude of their estimates. Second, I present suggestive evidence for the presence of compensating differentials across broad occupational categories. Third, I explain how changes in the occupational composition in the U.S. since the 1980s may have led to an amenity-biased reallocation. Fourth, I show that this amenity-biased reallocation may imply an absence of polarization across the skill distribution. Finally, I briefly discuss how wage declines due to shifts from blue-collar to service occupations may lead to biased productivity estimates.

Estimates of compensating differentials. The theory of compensating differentials is an old idea (Smith, 1776; Lucas, 1977; Rosen, 1986). Yet, despite its intuitive appeal and indirect evidence, estimating compensating differentials has proved challenging. The identification challenge is that workers' skills are not directly observable, and higher-skilled workers tend to be employed in jobs with higher wages and higher amenities. As a result, workers' wages and job amenities are typically positively correlated in the data. The flip side is that we also cannot directly infer skills from wages. For instance, a low-skill worker may earn more than a higher-skilled one if compensated for lower amenities.

In recent years, improved data and methods have renewed interest and shown that compensating differentials can be substantial (see Lavetti (2023) for a historical overview of the literature). Possibly the best evidence of sizable valuations for job amenities comes from discrete choice experiments as in Mas and Pallais (2017).[19] They find that, on average, workers are willing to give up 20% of their wages to avoid schedules set by an employer on short notice. Using a similar approach, Maestas *et al.* (2023) estimate that going from a job requiring heavy physical activity to moderate physical activity is equivalent to a 14.5% wage increase. They also find that a job where the person is sitting is equivalent to an 11.6% increase, and a relaxed environment adds another 4.3%. Overall, transitioning from the worst to the best job, based on the possible survey responses, is equivalent to a 55% wage increase.

The magnitude of these estimates suggests that job amenities must also be important at the macro level. For example, suppose that improvements in production technology reduce the physical activity of the jobs of 25% of workers from heavy to moderate. According to the estimates of Maestas *et al.* (2023), in aggregate terms, this would be equivalent to an increase in average wages of 3.6%, which is well above the annual average wage growth in the U.S. Moreover, since these estimates are based on a limited set of survey questions, accounting for all differences in job amenities across occupations could lead to substantially larger compensating differentials.

Compensating differentials across major occupation groups. I now argue that differences in job amenities can lead to significant compensating differentials across broad occupation groups. In particular, I provide evidence that blue-collar occupations generally have the lowest amenities and that part of their wages may reflect compensating differentials.

I use proxies for workers' skills as a first piece of evidence. If workers in certain occupations systematically earn more than predicted by their skill proxies, they may be receiving compensation for lower amenities. I use two occupation-level skill proxies: the average AFQT score and the average years of schooling. In Table 1.1, I compute the average percentile of the proxies by major occupation groups. Blue-collar occupations (Production/craft, Transportation/construction... and Machine operators/assemblers) that populated the middle of the wage distribution, rank relatively lower in both skill distributions. By contrast, at the bottom of the wage distribution, service occupations become

[19] Another recent strand of the literature uses structural models to estimate compensating differentials across firms. For instance, Sorkin (2018) estimates that compensating differentials explain at least 15% of the variance of earnings.

relatively higher skilled—at around the same level as the blue-collar.[20] Therefore, the higher wages of blue-collar workers could partly represent a compensating differential for lower amenities.[21]

Table 1.1: Wage, skill proxy and amenity percentiles by major occupation groups

		Skill proxies		Amenities			
	Wage (1980)	AFQT	Years schooling (1980)	No hazardous conditions	No contaminants	Artistic	No pain
Managers/professionals/ technicians/finance/ public safety	73.9	79	80.2	65.6	73.3	66.6	59.3
Production/craft	70.0	41.2	38.6	27.8	28.2	35.5	36
Transportation/ construction/mechanics/ mining/farm	52.4	29.7	25.5	31,8	25.5	27.2	32.5
Machine operators/ assemblers	36.7	16.9	17.3	24.4	18	39.5	30.2
Clerical/retail sales	30.6	55.5	57	74.7	65.5	32.1	53.9
Service occupations	11.0	23	24.7	35.2	44.7	48.7	31

Note: For every occupation, I first compute its percentile in the distribution of each variable (wage, AFQT, years of schooling and each of the amenities). Then, I compute average percentiles by major occupation groups weighting by the employment shares of each occupation in 1980.

Table 1.1 also shows the average percentiles in some amenities by major occupation groups. We indeed find evidence of blue-collar occupations having lower amenities. Blue-collar workers are among the most likely to be exposed to hazardous conditions and contaminants. Moreover, these occupations also score low in the artistic characteristic, and the workers are among the most likely to experience pain. Conversely, clerical occupations—also middle-wage—score higher in the skill proxies and generally have higher amenities.

[20] That workers in blue-collar (routine-manual) occupations had low education levels is well known (see Cortes *et al.*, 2017 for example). However, to my knowledge, the connection to compensating differentials has not been made.

[21] We may be concerned that the AFQT and years of schooling are mainly measures of cognitive skills and that these may be relatively less important in blue-collar occupations than, for example, in services. To assess this concern, in Table A.11 in Appendix A.5, I compute the average of other (non-cognitive) skill proxies by major occupation groups. I generally find similar results where the blue-collar occupations tend to have the lowest scores along with the service ones.

Changes in occupational composition and amenity bias. The US labor market has undergone a large occupational reshuffling since the 1980s. A sizable share of the labor force has moved away from middle-wage blue-collar and clerical occupations into low-wage service and high-wage abstract occupations (see Table 1.2). In particular, the share of the labor force employed in blue-collar occupations declined from approximately 34% in 1980 to 22% in 2015. The evidence provided above suggested that blue-collar occupations generally had the lowest amenities. Therefore, their declines in employment can lead to an amenity-biased reallocation of workers towards higher amenity occupations, thereby increasing aggregate non-pecuniary compensation.

Table 1.2: Changes in wages and employment by major occupation groups

	Employment share (1980)	Wage percentile (1980)	(100x) Change employment share (1980-2015)	Change log. hourly wage (1980-2015)
Managers/professionals/ technicians/finance/public safety	33.1	73.9	12.1	0.25
Production/craft	5.1	70.0	-2.35	-0.1
Transportation/construction/ mechanics/mining/ farm	19.4	52.4	-3.9	-0.02
Machine operators/assemblers	10.2	36.7	-6.2	0.0
Clerical/retail sales	23	30.6	-4.0	0.1
Service occupations	9.2	11.0	4.3	0.08

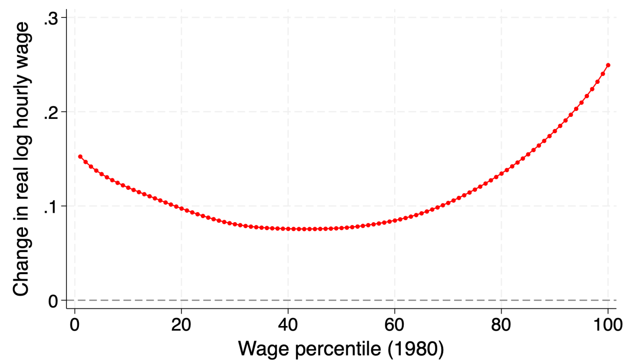
Note: To compute the wage percentile, for every occupation, I first compute its percentile in the wage distribution. Then, I compute average wage percentiles by major occupation groups weighting by the employment shares of each occupation in 1980. The change in real wages in the last column is with 2012 prices, and it is also computed weighting by the employment shares.

Amenity bias and polarization. Shifts in labor demand away from blue-collar and clerical occupations led to a polarization of the labor market in the US and most developed economies (Autor *et al.*, 2006; Goos *et al.*, 2009). I depict this polarization in Panels (a) and (b) of Figure 1.2. I rank occupations by their wage percentile in 1980 and compute the (smoothed) change in employment and wages at every percentile. Employment and relative wages decreased in occupations in the middle of the wage distribution and increased at the top and bottom tails. The literature has shown that this polarization is mainly driven by the automation of routine jobs (Autor and Dorn, 2013; Goos *et al.*, 2014), but also by offshoring (Goos *et al.*, 2014) and structural change (Bárány and Siegel, 2018; Comin *et al.*, 2020).

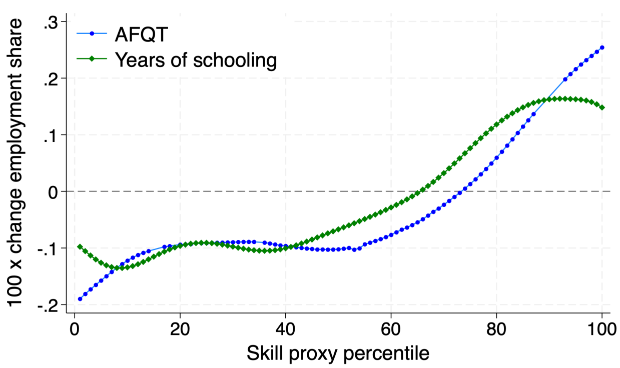
Figure 1.2: Smoothed changes in employment and wages by wage and skill proxy percentiles, 1980-2015

(a) Employment changes by wage percentile

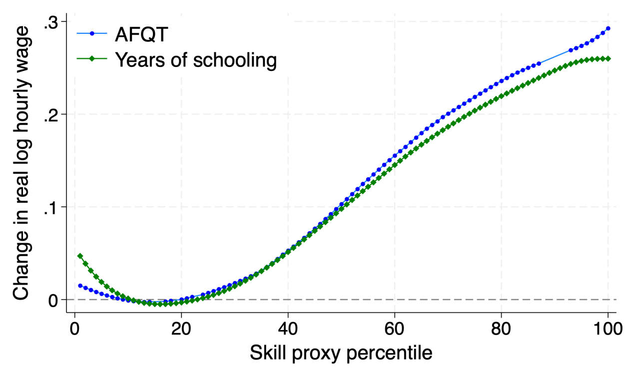
(b) Wage changes by wage percentile



(c) Employment changes by skill proxy percentiles



(d) Wage changes by skill proxy percentiles



Note: First, I compute the percentile in the distribution of each variable (AFQT, years of schooling, and wage) for every occupation. Then, I compute the changes at every percentile, weighting by the employment shares of each occupation. Finally, I smooth the changes with a LOWESS regression using the same bandwidth in all cases.

The literature on labor market polarization usually interprets the wages of workers in an occupation as a measure of their skills. Through this lens, polarization impacted middle-skilled occupations (Autor *et al.*, 2006; Acemoglu and Autor, 2011; Autor and Dorn, 2013).[22] As discussed, compensating differentials break the link between wages and skills. The right definition of skills is not straightforward, but, for example, we may not want to label a worker as higher-skilled after she takes a less safe job at a higher wage. Since Table 1.1 shows that blue-collar and service occupations have similar values for the proxies, this raises the question of whether there is actually polarization along the distribution of the skills proxies.

[22]For instance, in the same type of plot as Figure 1.2, Autor and Dorn (2013) write in the x-axis: "skill percentile (ranked by 1980 occupational mean wage)".

To this end, I redo the exercise but instead ranking occupations by skill proxies. Panels (c) and (d) of Figure 1.2 plot the smoothed changes in employment and wages by the AFQT score and years of schooling percentiles.[23] There is no polarization when ranking occupations by the skill proxies; occupations at the bottom of the distribution—instead of at the middle—have experienced the largest declines. This finding challenges the notion that polarization has affected middle-skilled occupations, suggesting instead a pure skill-biased change—as in Bound and Johnson (1992) or Katz and Murphy (1992)—where labor demand shifts towards higher-skill occupations.

Amenity bias and productivity estimates. The following thought experiment illustrates how compensating differentials can lead to misleading productivity assessments when workers switch occupations. Suppose that the differences in average wages between service and blue-collar occupations are purely a compensating differential for lower amenities. Imagine that one percent of workers switch from blue-collar occupations to services and suppose these workers were indifferent between the two occupations. If they earned the average wages in the two occupations, their income in 1980 would decrease by approximately 17,000 dollars in 2012 prices (from 40,000 to 23,000). Abstracting from any general equilibrium adjustment, GDP per worker would fall by $\frac{0.01 \times 17.000}{\text{GDP per worker 1980}} \approx 2.7\%$.

Although we would measure a 2.7% decline in GDP per worker, and so a decrease in labor productivity, neither workers' welfare nor the production technology has changed. This apparent misallocation of workers results from not accounting for changes in non-pecuniary compensation. Imagine that instead of only measuring workers' wages, we also measured the value of their job amenities, i.e., we measure their total compensation. Since we assumed that the wage differences between blue-collar and service occupations are entirely explained by compensating differentials, total compensation in the two occupations is the same. As a result, total compensation remains constant after workers reallocate—the drop in wages in exactly offset by the increase in non-pecuniary compensation. Therefore, if we measure total compensation instead of wages, we would not observe any decline in labor productivity.

Tacking stock and roadmap. To sum up, since the 1980s, decreased demand for middle-wage occupations has led to a polarization of the labor market across the wage distribution. Among these occupations are blue-collar jobs, which I have provided evidence to generally have the lowest amenities. Thus, the US labor market may not only have experienced a polarization along the wage distribution but also an amenity-biased reallocation.

[23] In Figure A.20 in Appendix A.5, I redo this figure separately by the periods 1980-1990, 1990-2000, and 2015. Qualitatively, the figures are similar in most periods.

In turn, the amenity bias can imply an absence of polarization along the skill distribution. Moreover, only measuring wage changes of workers who move from low-amenity occupations can lead to misleading inferences about productivity growth. The objective of the following section is to estimate amenity prices, which will allow us to formally evaluate and quantify these hypotheses in the rest of the paper.

1.4 Measurement of occupational amenities

In this section, I describe how I estimate amenity prices and how I use them to construct amenity and total compensation measures. Section 1.4.1 lays down the occupational choice setup that I will use throughout the paper. Then, Section 1.4.2 describes the main identification challenge and how I tackle it with the skill proxy method. Finally, Section 1.4.3 explains the estimation, and Section 1.4.4 discusses the robustness exercises.

1.4.1 Set up

Consider an economy with a set of workers $\mathcal{I}$, a set of occupations $\mathcal{J}$, and a set of amenities $\mathcal{N}$. Workers are heterogeneous in their productivity in a set of tasks $\mathcal{L}$. The vector of task-specific productivities of worker $i \in \mathcal{I}$ is $\theta_i = \{\theta_{i1}, ..., \theta_{il}, ..., \theta_{iL}\}$, where $\theta_{il} \in \mathbb{R}_+$ denotes the productivity in task $l \in \mathcal{L}$. Every occupation $j \in \mathcal{J}$ has a wage function mapping task-specific productivities to wages $w_j : \mathbb{R}^L_+ \to \mathbb{R}_+$. Later (Section 1.6), I will specify the production side of this economy so the wage function will be determined in equilibrium, but we can take it as given for the estimation of amenity prices. Occupations are also heterogeneous in the value of each amenity $n \in \mathcal{N}$. I denote by A_{jn} the value of amenity $n \in \mathcal{N}$ in occupation $j \in \mathcal{J}$. Workers value wages and amenities. I denote the utility of worker i in occupation j by $U_i(w_j(\theta_i), \{A_{jn}\}_n)$ and assume that it is strictly increasing in all arguments.[24] Workers choose to work in the occupation that maximizes their utility.

Instead of analyzing workers' choices over occupations, we can focus on choices over wage-amenity arrangements $(w, \{A_n\})$. To this end, I define the worker's offer set as all the combinations of wages and amenities available to a worker given her productivity θ_i:

$$\Phi(\theta_i) = \{(w, \{A_n\}_n) : \exists j \in \mathcal{J} \text{ such that } (w_j(\theta_i), \{A_{jn}\}_n) = (w, \{A_n\}_n)\}. \quad (1.1)$$

Because workers do not have specific preferences for occupations, they choose from the

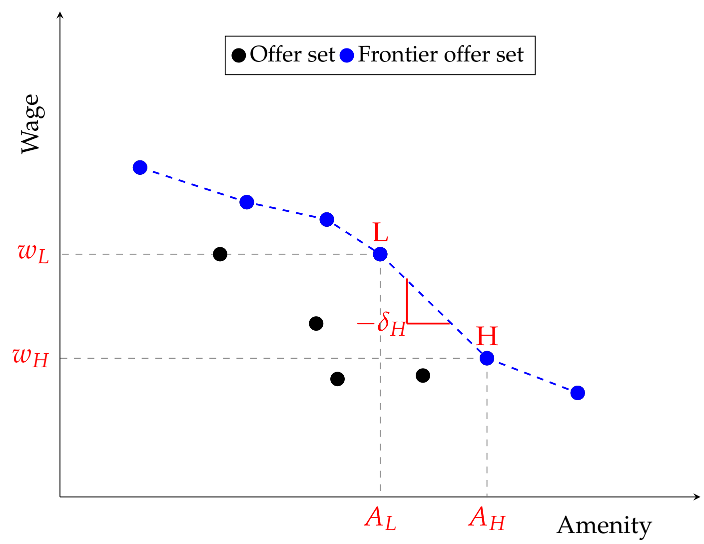
[24]Note that any job characteristic D_n disliked by workers can always be redefined as $A_n = -D_n$.

arrangements $(w, \{A_n\}_n)$ in their offer set that are undominated, i.e., that no alternative job offers weakly higher wage and amenities, with at least one being strictly higher. Formally, I define the frontier of the offer set as

$$\overline{\Phi}(\theta_i) = \quad (1.2)$$
$$\{(w, \{A_n\}_n) \in \Phi(\theta_i) : \nexists (w', \{A'_n\}_n) \in \Phi(\theta_i) \text{ s.t } w' \geq w \ \& \ A'_n \geq A_n \forall n, \text{ with } > \text{for some } n \text{ or } w'\}.$$

Figure 1.3 depicts the offer set and the corresponding frontier in an economy with one amenity. Each dot represents a combination of wage and amenity (w, A) in the worker's offer set. The blue dots are the undominated combinations, i.e., in the frontier. In this example with one amenity, we can define the (local) price of the amenity from the slope of the dashed line. Consider workers employed in occupation H. To receive $A_H - A_L > 0$ extra units of the amenity, they have given up $w^L - w^H > 0$ in wages. Hence, as illustrated in the graph, the price paid for the extra amenity by these workers is $\delta_H = -\frac{w_H - w_L}{A_H - A_L}$. Note that these prices are local because they vary with the amenity level.

Figure 1.3: Offer set, frontier and local amenity prices



Model discussion. I model amenities as exogenous characteristics of an occupation. This is natural for amenities that may be intrinsic to the occupation, such as whether an occupation involves artistic tasks or not. However, we may think of other amenities as being endogenously provided by firms, e.g., office perks. I focus on the first type of amenities because I am interested on the reallocation of employment and relative wages across occupations.

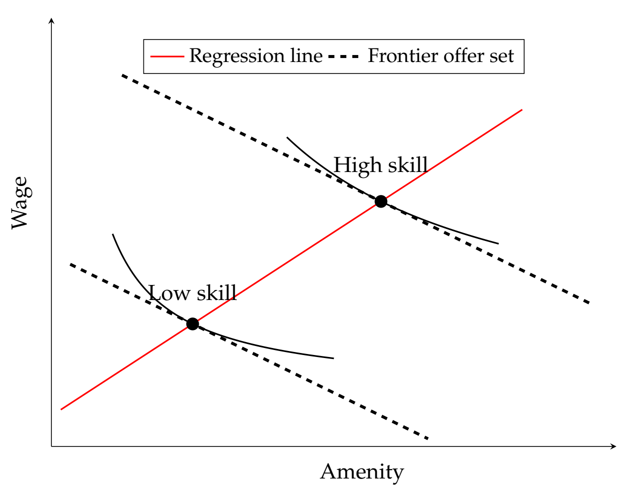
Throughout the paper, I do not need to impose any restrictions on preference heterogeneity, except that all workers agree on which amenities are desirable. In particular, preferences can be arbitrarily correlated with the workers' productivity types θ_i. As I explain in the following section, the identification of amenity prices requires assumptions that restrict the way in which θ_i can affect the worker's frontier but not on preferences.

1.4.2 Amenity pricing with skill proxy method

Identification challenge and approach. The identification challenge is that we do not observe the workers' frontiers. In particular, more skilled or able workers tend to have frontiers with both higher wages and amenities. As a result, the data usually shows a positive correlation between wages and amenities (Bell, 2022). Therefore, Rosen (1974)'s approach of directly regressing prices on characteristics to estimate the hedonic price function is not valid.

Figure 1.4 illustrates this in an economy with two workers, one amenity, and where I assume that each worker's frontier of the offer set is a line. The high-skill worker is employed in a job with a higher wage and amenity, and her frontier contains jobs with higher wages for every value of the amenity. Hence, regressing wages on the amenity would give a positive relationship and so a wrong-sided compensating differential. However, each worker's frontier of the offer set is downward sloping, implying a positive price for the amenity.

Figure 1.4: Heterogeneous frontiers and bias in regression of wages on amenities



I follow the methodology proposed by Bell (2022) and Bell *et al.* (2024). It consists of estimating amenity prices using a skill proxy as a shifter of the workers' frontiers.[25] This

[25] Bell, 2022 first proposed this methodology in the labor market context. Then, Bell *et al.* (2024) extended

method is akin to hedonic pricing (Rosen, 1974) but takes into account that workers have heterogeneous offer sets.[26]

Mapping to Bell (2022) and Bell *et al.* (2024) representation and main assumption. I now describe how the general setup introduced in Section 1.4.1 can be adapted to estimate amenity prices with the proxy method of Bell (2022) and Bell *et al.* (2024). First, because workers have preferences for wages and amenities but no specific preference for each occupation, they choose work arrangements $(w, \{A_n\})$ at the frontiers of their offer sets.[27] Hence, we can restrict attention to the workers' frontiers.

Then, I make a continuity assumption in order to represent the frontiers as surfaces (or lines if $n = 1$). This assumption can be rationalized either in the limit where there are infinitely many work arrangements or by assuming that the workers can choose combinations of the available arrangements. That is, I assume the frontier of worker type θ_i satisfies

$$\overline{\Phi}(\theta_i) = \{(w, \{A_n\}) : \phi(w, \{A_n\}; \theta_i) = \overline{\phi}(\theta_i)\}, \tag{1.3}$$

for some continuous, differentiable and strictly increasing function $\phi(.; \theta_i)$ and parameter $\overline{\phi}(\theta_i) \in \mathbb{R}$. The following is the main assumption on the structure of the frontiers that will allow identification of the amenity prices.

Assumption 1. *There exists a function $\phi(w, \{A_n\})$ such that $\phi(w, \{A_n\}) = \phi(w, \{A_n\}; \theta_i)$ for all $i \in \mathcal{I}$.*[28]

Assumption 1 allows us to summarize the work arrangements available to workers by the scalar $\overline{\phi}(\theta_i)$, as two workers i and i' with $\overline{\phi}(\theta_i) = \overline{\phi}(\theta_{i'})$ face the same options.[29] Thus, this assumption reduces the dimensionality of the worker type from $\mathbb{R}^L$ to $\mathbb{R}$. Practically, the restriction is that the frontiers of the offer sets do not intersect. That is, if worker i earns a higher wage than worker i' with a high amenity so that $\overline{\phi}(\theta_i) > \overline{\phi}(\theta_{i'})$, she must also earn a higher wage with a low amenity. From now on, we can thus work directly with $\overline{\phi}$. Throughout, I call the index $\overline{\phi}$ the total compensation level.

it and applied it to the estimation of amenities in housing. Although it is in a slightly different setting, my exposition of the method will be closer to that in the second paper. For the labor market, this method has also been used by Folke and Rickne (2022), Burbano *et al.* (2023) and De Schouwer and Kesternich (2023).

[26]After the hedonic price function has been estimated, the preferences can be recovered (under some parametric assumptions) in a second step using choice data (Rosen, 1974; Bajari and Benkard, 2005).

[27]This would not be the case if I introduced occupation-level taste shocks, as is common.

[28]Although stated slightly differently, Assumption 1 is equivalent to the single index assumption of Bell *et al.* (2024).

[29]However, this assumption does not necessarily imply that all workers with the same $\overline{\phi}$ must earn the same in all occupations.

In Appendix A.2.1, I provide more intuition on which cases Assumption 1 may not hold. Intuitively, the frontiers do not cross if either the θ_ls are sufficiently correlated for every worker or there are a few θ_ls that are very important determinants of wages. Moreover, if we split the data based on observables, Assumption 1 only has to hold within each group. For instance, we can allow workers of different ages or genders to face different frontiers. As I explain in Section 1.4.3, I do this for the workers' age by estimating the amenity prices separately for each NLSY79 sample.[30]

Assumption 1 allows us to invert and write a wage function $w(\{A_n\},\overline{\phi})$, which assigns a unique wage for every total compensation level and amenities combination. Our object of interest is the local price of each amenity A_n (i.e., the compensating differential), which is defined as the decrease in the wages from an infinitesimal increase in the amenity:

$$\delta_n(\overline{\phi},\{A_n\}) \equiv -\frac{\partial w}{\partial A_n}|_{(\{A_n\},\overline{\phi})}= \frac{\frac{\partial \phi(w,\{A_n\})}{\partial A_n}}{\frac{\partial \phi(w,\{A_n\})}{\partial w}}. \tag{1.4}$$

Notice that the amenity prices are local to the total compensation $\overline{\phi}$ and amenity $\{A_n\}$ levels. If amenity prices are linear, i.e., $-\frac{\partial w}{\partial A_n}|_{(\{A_n\},\overline{\phi})}= \delta_n$, the total compensation level is equal to the wage plus the monetary value of amenities:

$$\overline{\phi} = w + \sum_{n\in\mathcal{N}} \delta_n A_n. \tag{1.5}$$

This formulation gives an ordinal ranking of work arrangements by total compensation that does not impose restrictions on preference heterogeneity. Although worker i may not prefer the allocation of i', if $\overline{\phi}_i < \overline{\phi}_{i'}$, worker i would be strictly better off if she faced the same options as i'. It is important to remark that workers are not indifferent between all the work arrangements with the same total compensation level. In particular, if workers i and i' have the same total compensation $\overline{\phi}_i = \overline{\phi}_{i'}$, but choose different wage-amenity combinations, i.e., $(w_i,\{A_{n,i}\}) \neq (w_{i'},\{A_{n,i'}\})$, worker i may strictly prefer her worker arrangement over that of i', and conversely.

As discussed in Section 1.3, compensating differentials imply that we cannot properly measure workers' skills using only wages. The total compensation provides a way to measure workers' skills accounting for the fact that workers can split their total compensation differently between wages and amenities. Generally, we can think of all the work arrangements $(w,\{A_n\})$ with the same total compensation level $\overline{\phi}$ as requiring the same

[30] For the main estimates, I do not estimate the amenity prices separately by gender to not reduce the sample size. However, in Appendix A.3, I show that doing so delivers similar results.

skill level. For example, consider an economy where jobs are heterogeneous in their wage and safety. If we observe two workers earning the same wage but in jobs with different safety, then it is natural to think of the worker in the safer job as higher-skilled, and vice versa.

Skill proxy methodology. With a wage function represented by $w(\{A_n\},\overline{\phi})$, Bell (2022) and Bell *et al.* (2024) propose a proxy method to estimate the amenity prices. The idea is to use an imprecise proxy for the worker's skill x as a shifter of the frontier, which allows us to recover its tangent vector, i.e., the amenity prices. The following steps follow Bell (2022) and Bell *et al.* (2024), so I keep the description purposely brief.

Assumption 2 provides conditions on the validity of the skill proxy.

Assumption 2. *The skill proxy x satisfies the following assumptions.*

1. *Conditional independence: The skill proxy x is independent of $(w, \{A_n\})$ conditional on $\overline{\phi}$, i.e.,*
$$x \perp (w, \{A_n\})|\overline{\phi}. \tag{1.6}$$
2. *Strict monotonicity: $\mathbb{E}[x|\overline{\phi}]$ is strictly monotone in $\overline{\phi}$.*

The conditional independence assumption states that, conditional on a worker's total compensation, the skill proxy should not affect the worker's choices. In particular, the skill proxy should be independent of preferences for wages and amenities. The strict monotonicity requires that the proxy is, on average, informative about the total compensation level.[31]

Under these conditions, we can recover the ordinal total compensation ranking and identify the amenity prices.

Theorem 1. *(Bell, 2022 and Bell et al., 2024) Let*

$$\hat{x}(\overline{w}, \{\overline{A}_n\}) = \mathbb{E}[x|w = \overline{w}, \{A_n = \overline{A}_n\}] \tag{1.7}$$

denote the predicted proxy function. Under assumptions 1 and 2,

1. *We have $\hat{x}(\overline{w}, \{\overline{A}_n\}) = \mathbb{E}[x|\overline{\phi}_0]$, where $\overline{\phi}_0 = \phi(\overline{w}, \{\overline{A}_n\})$.*

[31] Bell *et al.* (2024) label this an anti-IV approach because a valid proxy must be correlated with the unobserved error (strict monotonicity) but uncorrelated with the endogenous variables $\{A_n\}$ (conditional independence).

2. *The amenity price δ_n can be identified and consistently estimated by:*

$$\hat{\delta}_n(\overline{w}, \{\overline{A}_n\}) = \frac{\frac{\partial \hat{x}(\overline{w}, \{\overline{A}_n\})}{\partial A_n}}{\frac{\partial \hat{x}(\overline{w}, \{\overline{A}_n\})}{\partial w}}.$$

Part 1 (Lemma 2.4 in Bell *et al.*, 2024) follows directly from assumptions 1 and 2.[32] It implies that the set of combinations $(w, \{A_n\})$ with the same value of the predicted proxy (i.e., such that $\hat{x}(w, \{A_n\}) = x_0$ for some x_0) must have the same total compensation level (i.e., $\phi(w, \{A_n\}) = \overline{\phi}_0$ for some $\overline{\phi}_0$). Hence, with the predicted proxy function, we can recover the (ordinal) total compensation ranking.

Because the predicted proxy gives the sets of $(w, \{A_n\})$ in the same frontier, the tangent vector—and so the amenity prices—must also be the same (Part 2). The price of some amenity A_n will be larger the more the predicted proxy grows with the amenity relative to with the wage.

Therefore, to estimate amenity prices and recover the total compensation ranking we only need to project the skill proxy on wages and amenities. In general, this can be estimated non-parametrically, where the amenity prices depend on the total compensation level. By contrast, a linear regression imposes the same prices for all workers.

1.4.3 Estimation

Following Bell (2022), I use the AFQT score from the NLSY79 as the skill proxy x for the main specification. Because the test measures basic skills, it should not affect how the workers trade off wages and amenities. For example, years of schooling would not be a good proxy because a higher preference for amenities could induce workers to obtain more education. As I discuss later, I redo the estimation with different proxies to verify the results are robust to the choice of the skill proxy.

I use a nonlinear specification where total compensation is a power function of the worker's wage:

$$x_i = \eta + \gamma_1 \frac{w_i^{\gamma_2}}{\gamma_2} + \sum_{n \in \mathcal{N}} \pi_n A_{j(i),n} + \epsilon_i, \tag{1.8}$$

where $j(i)$ denotes the occupation where worker i is employed and w_i is the worker's

[32] Assumption 1 implies $\mathbb{E}[x|w = \overline{w}, \{A_n = \overline{A}_n\}] = \mathbb{E}[x|\{A_n = \overline{A}_n\}, \overline{\phi}_0]$, and the conditional independence then implies $\mathbb{E}[x|\{A_n = \overline{A}_n\}, \overline{\phi}_0] = \mathbb{E}[x|\overline{\phi}_0]$.

wage measured from the NLSY79.[33] I estimate this equation with nonlinear least squares. Under this specification, the amenity price is also a power function of the worker's wage:

$$\hat{\delta}_n(w) \equiv \frac{\frac{\partial x}{\partial A_{j,n}}}{\frac{\partial x}{\partial w}}|_{x=\hat{x}} = \frac{\hat{\pi}_n}{\hat{\gamma}_1} w^{1-\hat{\gamma}_2}. \tag{1.9}$$

Occupation-level amenity and total compensation measures. With these estimates, we can now compute occupation-level amenity and total compensation measures. Instead of using NLSY79 data, I compute average wages by occupation, w_j, using the 1980 census, which is the year we are interested in computing the total compensation measure. Accordingly, I inflate the 1980 wages to the corresponding price level in the year used for the estimation in the NLSY79. The estimated amenity measure for occupation j is

$$\hat{A}_j = \sum_{n \in \mathcal{N}} \hat{\pi}_n A_{j,n}, \tag{1.10}$$

and the total compensation measure

$$\hat{\bar{\phi}}_j = \hat{\gamma}_1 \frac{w_j^{\hat{\gamma}_2}}{\hat{\gamma}_2} + \hat{A}_j. \tag{1.11}$$

Age effects on amenity prices. The NLSY79 panel that I use for the estimation tracks the same group of workers whose ages ranged between 14 and 22 in the first year of the data. To take care of potential age effects, I estimate the coefficients $\hat{\gamma}_1$, $\hat{\gamma}_2$ and $\{\hat{\pi}_n\}$ in Equation (1.8) separately at four different points in time (1986, 1996, 2006, and 2016).[34] Then, to construct the total compensation measure, I inflate the 1980 wages used in Equation (1.11) to the corresponding year to be consistent with the estimated coefficients. Finally, I aggregate the estimated amenity and total compensation distributions of each year using the employment shares across age groups in 1980.[35]

Estimates. Table A.3 in Appendix A.2.2 contains the estimated coefficients of Equation (1.8) for each year. Most coefficients are right-sided in that they are positive if they are

[33] I estimate several alternative specifications and verify that the main results are robust (see Appendix A.2.3).

[34] In particular, we may be concerned that changes in skills over the lifecycle could lead to different wage amenity frontiers.

[35] More formally, let $Q_j^{\hat{A}}(a)$ and $Q_j^{\hat{\bar{\phi}}}(a)$ be the estimated amenity and total compensation quantiles of occupation j at age group a (i.e., at a particular year in the NLSY79). Then, for $y \in \{\hat{A}, \hat{\bar{\phi}}\}$, I aggregated the quantiles as $\overline{Q}_j^y = \sum_a S^{1980}(a) Q_j^y(a)$, where $S^{1980}(a)$ is the employment share of age group a in 1980.

characteristics that we would expect to be liked by workers and negative otherwise. Recall that the predicted proxy should grow with the value of an amenity if workers must pay a compensating differential for it. Moreover, in all years $\hat{\gamma}_2 \in (0.5, 1)$, so the local amenity prices are an increasing and concave function of the wage.

In Table 1.3, I report the amenity prices by one standard deviation of each amenity with a linear specification and for the amenities with the largest prices.[36,37] I do this for the linear specification so that compensating differentials are constant and independent of wages. Recall that the price should be positive if a characteristic is valued by workers and negative otherwise. The amenity prices for some amenities are large, especially for those of the O*NET interests file. The amenities with the largest prices (by one standard deviation) are *investigative*, *realistic*, and *artistic*. The largest compensating differential that I estimate is for the amenity *investigative* in 2006, where one standard deviation higher is equivalent to approximately 36 thousand dollars in 2012 prices. Except for a few cases (such as *no burns and cuts* in 2016), the coefficients have the sign that we would intuitively expect, and when that is not the case, the coefficients are not significantly different from zero.

Table A.12 contains the average percentile in the amenity and total compensation measures by occupation groups. The lowest amenity occupation groups are: *housekeeping and cleaning* (with an average percentile of 8.6% in the amenity distribution); *building, grounds cleaning, maintenance* (11.9%); *transportation and material moving* (16.3%); and *other agricultural and related* (20.7%). The highest amenity occupations groups are: *professional speciality* (85.1%); *executive, administrative and managerial* (77.2%); *management related* (76.5%); and *technicians and related support* (71.6%).

Although the standard errors of the coefficient estimates are often large, the resulting amenity and total compensation measures are estimated much more precisely. The reason is that the final measures aggregate over many amenities and the four different NLSY79 samples. I show this in Appendix A.2.2 using Montecarlo simulations. Moreover, the results of Section 1.5 rely on the ordinal ranking of occupations by these measures. I find that the correlations of these rankings across pairs of draws are close to one, suggesting that they are quite precisely estimated.

Finally, in Appendix A.2.2, I use the amenity measure to explore which workers sort into low-amenity occupations. Conceptually, sorting may be driven by tastes for amenities and by how much workers value getting paid extra compensating differentials. There-

[36]The amenities estimates with the linear specification are similar (see Appendix A.2.3).

[37]I compute the Anderson Rubin confidence intervals as recommended by Bell (2022).

Table 1.3: Amenity prices by one standard deviate of each amenity (amenities with largest prices)

	1986	1996	2006	2016
No conflict	6301.6	5645.5	12479.4	15212.2
	[3962.2; 8780.4]	[1922.4; 9450.9]	[5846.8; 19409.1]	[7781; 23047.9]
No contaminants	-87.77	9595.3	6052.5	16598.1
	[-4914.5; 4742.4]	[2776.8; 16548]	[5733.6; 18002]	[3685; 29979.9]
No burns and cuts	5584.2	6105.8	2467.0	-7525.4
	[2200.8; 9096.1]	[680.1; 11638]	[-7477.8; 12553.3]	[-18914 ; 3797.8]
Time standing	864.2	2008.7	-2828.2	-18417.1
	[-2948.2; 4687.4]	[-7523.9 ; 3475]	[-12129.8; 6503.7]	[-29072.5; -8131.4]
Artistic	5334.1	10547.4	17686.2	21450.7
	[2627.3; 8161.4]	[6191.9; 15049.1]	[9583.4 ; 26210]	[12318; 31139.8]
Conventional	-7470.7	-672.5	-3290.4	-6725.8
	[-11296; -3847.3]	[-5969.2; 4546.6]	[-5958.6; -12744.4]	[-17207 ; -3474.8]
Enterprising	10707.1	11679.5	14799.0	4623.2
	[-7028.3; 14698.7]	[6205.7; 17406.6]	[5280.3; 24851.8]	[-5961.7; 15493]
Investigative	24531.7	24345.2	36306.9	30407.8
	[20556.5; 29176]	[19381; 29818.4]	[27535; 46237]	[20878; 41167.9]
Realistic	-15498.4	-11652.2	-31646.8	-29995.0
	[-21162.3; -10230.5]	[-19383.9; -4180.4]	[-46557.9 ; -17815.6]	[-45847; -15296]
Social	-131.9	1012.3	-15984.1	-18739.6
	[-3788.2; 3406]	[-4782; 6686.4]	[-26865; -5782.8]	[-31055; -7315]
Observations	7919	6462	4961	3947

Note: Anderson-Rubin confidence intervals in brackets (95%). I select the amenities that have among the five largest compensating differentials in at least one year.

fore, we expect workers with high marginal utility (e.g., wealth-poor) to sort more into low-amenity occupations. Luo and Mongey (2019) and Boar and Lashkari (2021) have documented evidence of this type of sorting.[38] I also find consistent evidence: controlling for wages, workers with less wealth, more children, or poorer parents tend to be employed in lower-amenity occupations.

1.4.4 Potential biases and robustness checks

Validity of the skill proxy. The main concern comes from the validity of the skill proxy used. In particular, that it does not satisfy the conditional independence assumption. One

[38] Luo and Mongey (2019) find that graduates with higher student debt take jobs with higher wages and lower job satisfaction. Boar and Lashkari (2021) show that children of poorer parents select occupations with lower intrinsic quality.

may worry that the AFQT score could be more correlated with skills in cognitive/abstract occupations, usually high-amenity, than in manual occupations, which tend to have lower amenities. Bell (2022) proposes other proxies available in the NLSY79 that can be used: self-esteem, mastery, and height. I estimate the model and construct the amenity and total compensation measures with these alternative proxies. The resulting measures are similar with all the proxies and the main results are robust (see Appendix A.2.3).

Alternative specifications. The choice of specification determines how the amenity prices are allowed to depend on the wages and amenities, which can affect the amenity and total compensation measures. To assess the sensitivity to the specification, I estimate several alternative specifications, and I also find that the amenity and total compensation measures are very similar and the main results robust (see Appendix A.2.3).

Unionized workers. Another concern is that the presence of unionized workers earning rents could bias the estimates. As I explain in Appendix A.2.3, union rents could lead to a violation of Assumption 1 because they imply that the frontiers of similarly skilled workers cross. The NLSY79 contains information on whether workers are unionized or covered by a union. Hence, I can redo the estimation removing these workers from the sample. I find that the estimates are not very sensitive to the inclusion of unionized workers, the amenity and total compensation measures are similar, and the main results are robust.

Noise in income data and upward bias in amenity prices. As discussed in Bell (2022), if income is measured with noise, the coefficient on income $\hat{\gamma}_1$ is biased downward and the amenity price upward (see Equation (1.9)). This does not affect the result on the amenity-biased reallocation because it uses the amenity measure, which does not depend on $\hat{\gamma}_1$. However, it can affect the result of the absence of polarization by total compensation distribution. As I discuss in Section 1.5.2 and Appendix A.2.3, a direct approach to assess this concern is to analyze how sensitive this result is to a downward rescaling of the amenity price estimates. That is, to check how much we can arbitrarily increase the value of $\hat{\gamma}_1$ and still have the same qualitative result.

Gender differences in amenity prices and occupational composition. Labor economics research has documented sizable gender differences in demand for job amenities such as schedule flexibility (e.g., Wiswall and Zafar, 2018; Vattuone, 2023; Morchio and Moser, 2024). Therefore, it is important to understand how these gender differences affect the estimates and results. Since I estimate amenity prices, gender differences in preferences should not affect the estimates. However, the price estimates may differ if, for example, men have a comparative advantage in low-amenity occupations. To assess this, in Appendix A.3, I redo the estimation separately by gender. I find that the amenity values are similar (with slightly higher total prices for men) and that the resulting amenity and total compensation measures are also very similar. Additionally, in Appendix A.3, I study and compare the labor market changes across the wage, amenity, and total compensation distribution separately by gender.

Change in amenities within occupations. The O*NET and ATUS data used for the amenity estimation are not available for the 1980s. However, for the latter years (since 2006), I can use the different releases of the O*NET context file to measure changes in amenities within occupations. I generally find improvements in amenities between 2006 and 2022. In Section 1.6, I use these changes in the O*NET data to measure the increases in the value of amenities within occupations. Regarding the robustness of the results, I find that the amenity and total compensation measures are similar, and all the results go through whether I use the 2006 or 2022 data (see Appendix A.4). Additionally, in Appendix A.4, I analyze the evolution of injury rates within industries and occupations using data from the Survey of Occupational Injuries and Illnesses (SOII) and the NLSY79.

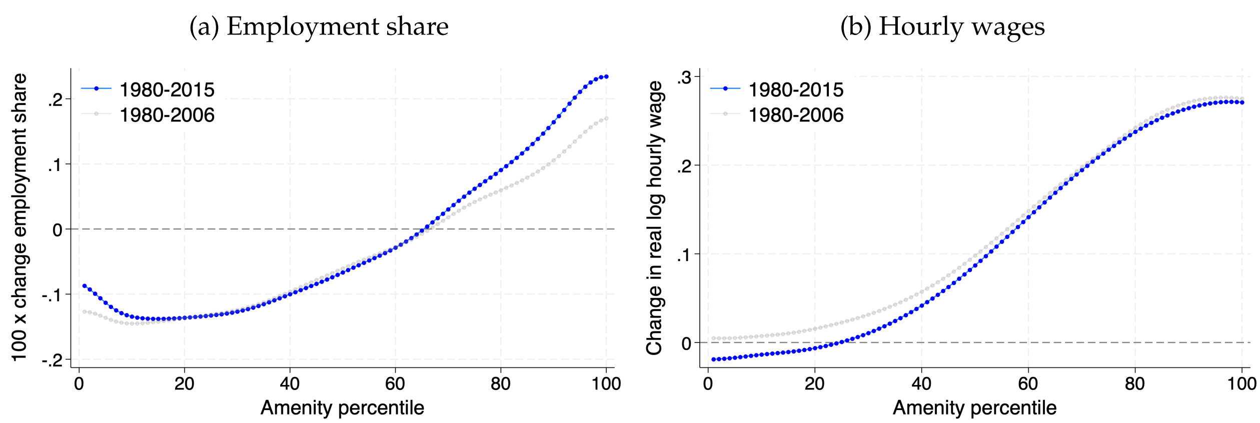
1.5 Amenity bias and no polarization by total compensation

In this section, I present the main empirical results of the paper. I start by documenting the reallocation of employment and wages across occupations (Section 1.5.1). Then, I show that, as a consequence, there is no polarization if we rank occupations by total compensation (Section 1.5.2).

1.5.1 Amenity-biased reallocation

The first result is an amenity-biased reallocation between 1980 and 2015 due to labor demand shifts across major occupation groups. Figure 1.5 plots the smoothed changes in employment and wages by amenity percentile between 1980 and 2006 and between 1980 and 2015. In both periods, there is a positive relationship between the amenity percentile and the increase in employment and wages. Employment decreased and (real) wages stagnated (or even decreased) in low-amenity occupations. Conversely, both wages and employment increased in the high-amenity ones. Moreover, most of the reallocation had already occurred by 2006.

Figure 1.5: Smoothed changes in employment and wages by amenity percentile, 1980-2006 & 1980-2015



Note: The plots are constructed with the same steps as Figure 1.2.

The second column of Table 1.4 shows the average amenity percentile by major occupation groups. Blue-collar occupations (specifically Transportation/construction/... and Machine operators/assemblers) have the lowest amenity levels and experienced decreases in employment and relative wages. By contrast, white-collar occupations (Managers/professionals...) have the highest amenities and saw large increases in employment and wages. In fact, the amenity bias documented in Figure 1.5 appears to be driven mainly by shifts across these broad occupation groups. In Figure A.23 in Appendix A.5, I compute the changes in employment and wages by amenity level within every major occupation group. Except for the wage changes within blue-collar occupations, there is no clear evidence of amenity bias in employment or wages within the other major occupation groups.

An amenity bias in both employment and wages suggests a demand-driven change, i.e., a decrease in labor demand in low-amenity occupations. Intuitively, if the drop in employment was caused by workers valuing amenities more in 2015, we should observe an increase in compensating differentials (and so wages) in low-amenity occupations. Conversely, improved working conditions in low-amenity occupations could have eroded compensating differentials and led to stagnant wages. However, without changes in labor demand, employment in these low-amenity occupations should increase as labor costs decrease.

The literature has argued that the decreases in labor demand for middle-wage—i.e., blue-collar (routine-manual) and clerical (routine-cognitive)—occupations were driven mainly by automation (Autor and Dorn, 2013), as well as offshoring (Goos *et al.*, 2014) or changes in consumer demand (Comin *et al.*, 2020). Therefore, these shifts in labor demand away from middle-wage routine occupations have not only led to a labor market polarization but also an amenity-biased reallocation because blue-collar occupations—which employed the largest share of workers—generally had the lowest amenities.

Table 1.4: Wages, amenity and total compensation by major occupation groups

	Wage percentile (1980)	Amenity percentile	Total compensation percentile	(100x) Change employment share (1980-2015)	Change log. hourly wage (1980-2015)
Managers/professionals/ technicians/ finance/ public safety	73.9	78	80.1	12.1	0.25
Production/craft	70.0	39	46	-2.35	-0.1
Transportation/construction/ mechanics/mining/ farm	52.4	23.1	26.7	-3.9	-0.02
Machine operators/assemblers	36.7	23.8	23.1	-6.2	0.0
Clerical/retail sales	30.6	59.2	55.8	-4.0	0.1
Service occupations	11.0	32.8	24	4.3	0.08

Note: The average percentiles in the table are computed as in Table 1.1.

Figure A.21 in Appendix A.5 plots the evolution of employment and wages by quartile of the amenity distribution since 1950. The reallocation from low- to high-amenity occupations has grown steadily since 1980. The amenity bias is less clear for the period 1950 to 1980 (see Figure A.22). The employment share increased in the highest-amenity occupations (the top 10%), but there is no clear bias in employment otherwise. For wages, we observe some positive relationship between the amenity and the increase in wages, but it is weaker.

Role of deunionization. The US labor market has experienced a large decrease in union membership rates in the last decades—from 20.1% in 1983 to 10.1% in 2022 (Macpherson and Hirsch, 2023). This fact raises the question of whether the low-amenity occupations also had high unionization rates in the 1980s and so if the deunionization can explain the amenity bias. I explore this using occupation-level data on union membership and coverage rates. Figure A.24 shows that low-amenity occupations had, on average, higher unionization rates in 1983 and experienced larger deunionizations between 1983 and

2015. Moreover, this decrease is mainly driven by blue-collar occupations (Table A.14), which previously had the highest unionization rents. Therefore, the erosion of rents from the deunionization could explain part of the decreases in relative wages at low-amenity occupations. However, the declines in both employment and relative wages in these occupations indicate again a demand-driven reallocation. The intuition is similar to the effect of improved working conditions in low-amenity occupations. The erosion of union rents should translate into an increase in employment in these low-amenity occupations as labor becomes cheaper for firms.

1.5.2 No polarization along the total compensation distribution

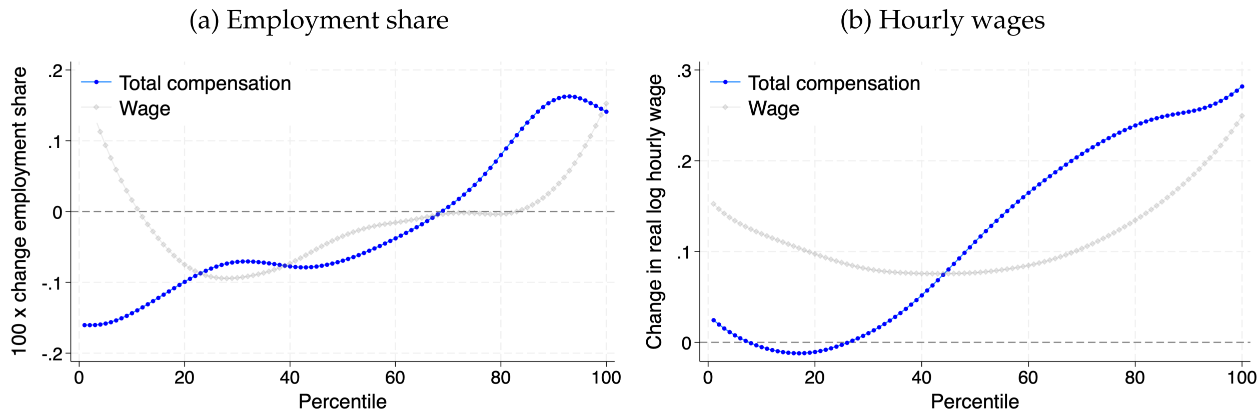
The previous section documented a decrease in employment and relative wages in low-amenity occupations due to declines in labor demand for blue-collar occupations. If workers in these occupations received a compensating differential for the low amenities, their total compensation would be (relatively) lower than their wages. Thus, a natural question is whether there is polarization if we rank occupations by total compensation. That is, did the jobs in the middle of the total compensation distribution suffer drops in employment and wages?

Figure 1.6 plots the smoothed changes in employment and wages ranking occupations by total compensation and wages. Despite the polarization along the wage distribution, the figure shows no polarization if occupations are ranked by total compensation. Instead, employment and relative wages decreased in the occupations with low total compensation.[39] Compared to the polarization along the wage distribution, we may view the changes in employment as more welfare improving because workers reallocated away from the worst (in terms of total compensation) occupations. By contrast, the wage declines would be a more negative phenomenon as they affected the workers who were worse off instead of those in the middle of the distribution.

The third column of Table 1.4 shows the average percentile in the total compensation distribution by major occupation groups. Indeed, since blue-collar occupations that populated the middle of the wage distribution had low amenities, they show up at the lower end of the total compensation distribution. By contrast, service occupations—that explained the increases in employment and wages at the bottom of the wage distribution (Autor and Dorn, 2013)—have relatively higher amenities, so they appear higher in

[39] At the bottom of the total compensation distribution, there is an uptick in the increase in wages, but it is very small (lower than the change in the middle of the distribution ranking by wages).

Figure 1.6: Smoothed changes in employment and wages by total compensation and wage percentile, 1980-2015



Note: The plots are constructed with the same steps as Figure 1.2.

the total compensation than in the wage distribution. As a result, many blue-collar and service occupations end up with a similar total compensation level, which makes the increase in employment and wages at the low end of the distribution disappear because the employment share of services was lower.[40]

It is important to stress that even if two occupations have the same total compensation level, workers may not be indifferent between them due to heterogeneous preferences. In particular, a blue-collar worker would generally be worse in a higher-amenity and lower-wage service occupation with the exactly same total compensation level. Hence, the blue-collar workers that (optimally) reallocate to services could be worse off than in the prior period. However, a proper welfare measurement needs to account for amenities. As long as these blue-collar workers that reallocate have some preference for amenities, we would underestimate utility changes by only measuring wage changes.

Figure 1.6 also has important implications for understanding the impact of polarization along the skill distribution. The polarization literature often refers to wages and skills interchangeably. Hence, it is often stated that polarization has impacted middle-skilled workers. As discussed in Section 1.4.2, total compensation can also be interpreted as a measure of the worker's skills that corrects for compensating differentials. Therefore, Figure 1.6 indicates that there is no polarization across the skill distribution after correcting for compensating differentials. That is, on aggregate, employment and relative wages decreased in low-skill occupations instead of middle-skill ones. Hence, on aggregate terms,

[40]By contrast, the clerical occupations, which were also in the middle of the wage distribution and experienced a decrease in employment, have relatively higher amenities and show up higher in the total compensation distribution.

this resembles more a skill-biased change in labor demand where low-skill workers are worse off and high-skill workers better off—as in the older skill-biased technical change literature (Bound and Johnson, 1992; Katz and Murphy, 1992).

Robustness to upwards biased compensating differentials. As discussed in Section 1.4.2, a downward bias in the coefficient on wages in Equation (1.8) leads to an upward bias in the estimated compensating differentials. This bias would overstate the importance of amenities relative to wages when ranking occupations by total compensation. Notice that if compensating differentials are sufficiently large, the total compensation ranking would approximate the ranking by amenities and conversely. Since there is polarization over wages but not over amenities, we must be careful that the absence of polarization in Figure 1.6 is not a byproduct of overestimating the compensating differentials. In Appendix A.2.3, I analyze the sensitivity of this result to the size of the estimated compensating differentials. I find even when reducing the estimated compensating differentials by half—that is, assuming we overestimated them by a factor of two—the absence of polarization remains robust. However, if we reduce the compensating differentials by a factor of four, we start to observe some polarization.

1.6 Growth accounting

The previous section documented a reallocation of workers from low- to high-amenity occupations between 1980 and 2015. However, the associated improvements in non-pecuniary compensation would not be accounted for in conventional income and output measures. In this section, I aim to answer the following questions. How does this amenity-biased reallocation affect conventional output and productivity estimates, and how should they be adapted to account for the presence of amenities?

To this end, I start in Section 1.6.1 by introducing the production side of the model. Then, in Section 1.6.2, I use a simple example to illustrate how the conventional TFP measure underestimates productivity growth and how measuring the value of amenities allows us to properly estimate it. Next, in Section 1.6.3, I generalize the previous intuitions, show that Hultens' theorem holds for an augmented output definition that includes the value of amenities, and derive growth accounting formulas. Finally, I explain how I map the amenity price estimates to the growth accounting formulas in Section 1.6.4, and I present the quantitative results in Section 1.6.5.

1.6.1 Production side

Workers' distribution, endowments, and budget constraints. I assume now that there is a unit mass continuum of workers indexed by $i \in [0,1]$ and I let θ_i denote the productivity type of i. Workers derive utility from a final consumption good c_i and amenities $\{A_n\}$ according to the utility function $U_i(c_i, \{A_n\})$, which is increasing in all arguments. Workers are also heterogeneous in their endowment of capital k_i, which pays a return of Q. Hence, the budget constraint of worker i employed in occupation j writes:

$$c_i = w_j(\theta_i) + Qk_i. \tag{1.12}$$

Each worker i chooses the occupation j that maximizes her utility subject to the budget constraint.

Production functions. There is a unit mass continuum of identical firms within each occupation. Output in each occupation, y_j, is produced by combining the labor input of each productivity type $\theta \in \mathbb{R}^L$, denoted by $h^j(\theta)$, according to the production function

$$y_j = z_j m^j(\{h^j(\theta)\}_\theta), \tag{1.13}$$

where m^j is weakly increasing in all arguments and has constant returns to scale, and z_j is an occupation-specific technology shifter. I introduce these technology shifters $\{z_j\}_j$ to be able to capture biased shifts in labor demand across occupations.

Let p_j denote the price of the output from occupation j. The problem of the representative firm in occupation j is

$$\Pi_j = \max_{\{w_j(\theta)\}_\theta} p_j m^j(\{h^j(\theta)\}_\theta) - \int w_j(\theta) h^j(\theta) d\theta \tag{1.14}$$

subject to workers' occupational choice. The optimality conditions are

$$w_j(\theta) = p_j \frac{\partial m^j(\{h^j(\theta)\}_\theta)}{\partial h^j(\theta)} \equiv MPL_j(\theta) \tag{1.15}$$

For ease of notation, I denote $m^j_\theta = \frac{\partial m^j(\{h^j(\theta)\}_\theta, r_j)}{\partial h^j(\theta)}$. Notice that wages equal the marginal product of each worker type, so there is no markdown. This is because I have assumed that amenities are occupation-specific and that there is a continuum of identical firms within each occupation. Hence, each firm faces a perfectly elastic labor supply schedule.[41] Combined with the constant returns assumption in the production function, this also implies that profits are zero.

A final good producer aggregates the output of each occupation according to the constant returns to scale production function

$$Y = z\mathcal{F}(\{y_j\}, K), \tag{1.16}$$

where K is the aggregate capital stock and z is a Hicks-neutral aggregate technology shifter. The function $\mathcal{F}$ is increasing and concave in all arguments and satisfies the standard Inada conditions: $\lim_{x\to 0} \frac{\partial \mathcal{F}}{\partial x} = +\infty$ and $\lim_{x\to +\infty} \frac{\partial \mathcal{F}}{\partial x} = 0$, for all $x \in (\{y_j\}, K)$. The price of the final good is normalized to one, so the price of each occupation's output is

$$p_j = z\frac{\partial \mathcal{F}(\{y_j\}, K)}{\partial y_j}, \tag{1.17}$$

and capital demand satisfies

$$Q = z\frac{\partial \mathcal{F}(\{y_j\}, K)}{\partial K}. \tag{1.18}$$

I denote $\mathcal{F} = \mathcal{F}(\{y_j\}, K)$ and $\mathcal{F}_x = \frac{\partial \mathcal{F}(\{y_j\}, K)}{\partial x}$ for $x \in (\{y_j\}, K)$. As is common, automation in this model is driven by the cost of the new technology q_j (e.g., Autor and Dorn, 2013). A reduction in the price q_j will induce an increase in r_j and decreases in the labor demand—and so lower employment and wages—in that occupation if r_j and labor are sufficiently substitutes.

Equilibrium. An equilibrium consists of allocations $\{c_i\}_i, \{h^j(\theta)\}_{\theta,j}, K$ and prices $\{w_j(\theta)\}_{\theta,j}, \{p_j\}_j, Q$ such that: (i) workers' occupational choices maximize their utility; (ii) firms' optimality conditions (1.15)-(1.18) hold; (iii) markets clear:

- Labor market: for all j and θ, $h^j(\theta)$ is consistent with workers choices
- Capital market: $K = \int k_i di$
- Final good: $Y = \int c_i di$.[42]

[41] This contrasts labor models of amenities, which are often used to explain labor market power (see, for example, Berger *et al.* (2022) or Lamadon *et al.*, 2022). If I had assumed that there is only one firm in each occupation or amenities were firm-specific, firms would face upward-sloping labor supply curves and would mark down wages.

[42] In Appendix A.6.1, I write a more detailed equilibrium definition with the workers' choice rules as in Mongey and Waugh (2024) that I use to prove the theoretical results.

1.6.2 An example on the role of amenities in growth accounting

I now use a minimal example to illustrate how conventional TFP growth underestimates productivity improvements after workers reallocate to high-amenity occupations and how we can adequately estimate productivity growth by measuring total compensation growth.

Consider a simplified version of the economy with two occupations $\mathcal{J} = \{H, L\}$, one amenity, no occupation-specific technology shocks, and no capital. Assume $A_H > A_L$ and that all workers are homogeneous in their productivity in each occupation but have heterogeneous preferences for the amenity. This allows us to define the price of the amenity as $\delta = -\frac{w_H - w_L}{A_H - A_L}$. Total production in this economy is

$$Y = z\mathcal{F}(h^H, h^L),$$

where h^H and h^L denote the share of workers employed in occupations H and L, respectively. Since aggregate labor supply is one, measured TFP equals aggregate production: $TFP = Y$. Constant returns to scale imply:

$$z\mathcal{F}(h^H, h^L) = W,$$

where $W = \sum_j w_j h^j$ denotes the aggregate wage bill. Then, differentiating and using $dh^H = -dh^L$, we get:

$$dTFP = \mathcal{F}dz + \underbrace{(MPL_H - MPL_L)}_{<0} dh^H = dW. \tag{1.19}$$

TFP growth is equal to the growth in measured output (first equality) and in measured income (second equality). The first term in the output growth captures the direct effect of an increase in the technology shifter dz. The second term is equal to the difference in the marginal product of labor between the high- and low-amenity occupations times the change in employment in the high-amenity one.

Without amenities, wages—and marginal products of labor (MPL)—are equalized across occupations, so we would have $MPL_H - MPL_L = 0$. With amenities, because the low amenity occupation must pay higher wages, its marginal product of labor is also higher. As a result, if workers reallocate to the high amenity occupation ($dh^H > 0$), output, income, and TFP decrease. Thus, even changes in preferences can affect measured productivity. If workers suddenly start valuing more amenities, we would measure a drop in TFP.

Although there appears to be misallocation in output, the equilibrium is efficient (as I prove in the following section). The intuition is simply that the amenity is a good whose (shadow) value is not measured in output. However, providing the amenity is costly. The shadow cost is equal to the difference in the MPL between the high- and low-amenity occupations, and it does (indirectly) affect the measured output. When workers reallocate to the high-amenity occupation ($dh^H > 0$), they increase their consumption of the (unmeasured) amenity and reduce their consumption of the (measured) final consumption good. Since the costs of producing the amenity are measured, but not their value, output and TFP decrease.

Because we can price the amenity, we can value the consumption of amenities in terms of the final good. Crucially, the differences in the MPL across occupations are exactly the amenity prices that the workers face. Notice that since $w_j = MPL_j$ for all j, the local price of the amenity satisfies $MPL_H - MPL_L = -\delta(A_H - A_L)$. That is, the (total) amenity price $\delta(A_H - A_L)$ is equal to the shadow cost of providing the amenity for an extra worker, which is also equal to the valuation for the amenity of the workers that switch occupations (after a first-order change in wages). Therefore, substituting into Equation (1.19) and rearranging, we can write

$$\mathcal{F} dz = \underbrace{dW + \delta(A_H - A_L) dh^H,}_{\text{Change total compensation}} \tag{1.20}$$

or dividing by output Y:

$$d \log z = \underbrace{\frac{W}{Y} d \log W + \frac{\delta(A_H - A_L) h^H}{Y} d \log h^H.}_{\text{Total compensation growth}} \tag{1.21}$$

By using amenity prices to measure the total compensation growth instead of income growth, we can achieve two things. First, we get a more welfare-relevant growth estimate as the amenity prices measure the valuation for the amenity of the workers that switch to first order. Second, we obtain an estimate of the growth in technology $d \log z$ that removes the effect of the reallocation of workers on TFP.

If workers reallocate to the high amenity occupation ($d \log h^H > 0$), as documented in Section 1.5, the increase in total compensation is larger than the increase in the wage bill. Hence, the growth in technology $d \log z$ is larger than what the change in measured TFP would suggest.

1.6.3 Efficiency, augmented output, and Hulten's theorem

I now generalize the intuitions from the previous example, show that a Hulten's theorem holds for an augmented output definition that includes the value of amenities, and derive growth accounting formulas to measure productivity growth with amenity prices.

First, it is useful to understand how output and TFP would be conventionally measured in this economy. Because there are no intermediate production inputs, Measured output is equal to $Y = \int c_i di$. Output can also be measured from national income, which, due to the constant returns assumption, is $Y = W + QK$, where $W = \sum_j \int w_j(\theta) h^j(\theta) d\theta$ is the aggregate wage bill. Then, the conventional way of measuring productivity growth would be

$$d \log TFP = d \log Y - \lambda_k d \log K, \tag{1.22}$$

where $\lambda_k = \frac{QK}{Y}$ is the capital expenditure share in output.[43] That is, TFP growth is equal to output growth minus the share-weighted growth in factor inputs (aggregate labor is fixed). Expanding $d \log Y$ we have:

$$d \log TFP = d \log z + \sum_j \frac{W_j}{Y} d \log W_j + \int \sum_j \frac{MPL_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta, \tag{1.23}$$

where $W_j = \int w_j(\theta) h^j(\theta) d\theta$ is the wage bill in occupation j. The first term captures the contribution to TFP growth of growth in the aggregate technology shifter. The second term measures the TFP gains from a reduction in the marginal costs of producing the new technologies, which can be interpreted as a productivity shock to the new technology producers. Again, the term

$$\begin{aligned} \int \sum_j \frac{MPL_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta &= \int \sum_j \frac{w_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta \\ &= \int cov_j \left(\frac{w_j(\theta)}{\widetilde{Y}}, d \log h^j(\theta) \right) d\theta, \end{aligned}$$

depends on the differences in marginal products/wages across occupations and the reallocation of workers. Without amenities, this term is zero because, to first order, the

[43] The term λ_k is the revenue-based Domar weight of capital. Baqaee and Farhi (2020) show that cost-based Domar weights are preferable because revenue-based weights have undesirable properties in inefficient economies. For ease of exposition and because this economy will turn out to be efficient, I use the revenue-based Domar weight.

workers who switch earn the same wage in the two occupations.[44] As in the example, if workers reallocate from low-wage to high-amenity occupations (so the covariance is negative), measured TFP decreases. Typically, this term would capture the change in allocative efficiency (also called misallocation) as a result of the reallocation of factor inputs, and it shows up in inefficient economies (see Baqaee and Farhi, 2020). However, the allocation is efficient and the first welfare theorem holds, as shown in the following proposition.

Proposition 1. *The equilibrium is a Pareto efficient allocation.*

The proposition is proven in Appendix A.6.1. Economies with differences in factors' marginal products are usually thought of as being inefficient. However, the differences in MPLs are driven solely by amenities. Closing differences in MPLs by reallocating workers to high MPL occupations would not yield any Pareto improvement because it would require a reduction in the utility from the amenity.

Equation (1.23) shows that Hulten's theorem (Hulten, 1978) does not hold for the conventional output and TFP growth measures. This is perhaps surprising because Hulten's theorem is often understood as a direct consequence of the first welfare theorem: since output is maximized with production efficiency, a macro envelope condition implies that only the changes in technology parameters ($d \log z$ and $\{d \log z_j\}$) have first-order effects on TFP. Equivalently, changes in the allocation of factor inputs (here $\{d \log h^j(\theta)\}$) do not have first-order effects on TFP. However, with amenities, efficiency does not imply that (measured) output is maximized because workers also derive utility from amenities.[45] Amenities are simply another good whose value is not included in measured output. With the appropriate definition of amenity prices, we can include the value of amenities in output and derive a Hulten's theorem.

Amenity prices. I now explain how I define local and total amenity prices in this economy. First, I assume from now on that there is a unidimensional amenity $\{A_j\}_j$. The definition of amenity prices with multiple amenities and a finite number of occupations is not straightforward and is left for future work. This simplification can also be rationalized by assuming that the equilibrium is such that all worker types face the same relative

[44] Without amenities, workers choose the occupation that pays the higher wage. To first order, only the workers that were previously indifferent switch, i.e., those that earn the same wage in the two occupations.

[45] This is a similar intuition why Hulten's theorem does not hold for output when households have a disutility of supply factors (Dávila and Schaab, 2023). However, in these cases, Hulten's theorem would generally still hold for TFP if computed by appropriately subtracting the changes in the factors' supplies (Baqaee and Rubbo, 2023).

prices across different amenities, which would allow us to aggregate: $A_j = \mathcal{A}(\{A_{jn}\}_n)$, for some function $\mathcal{A}$.

For every θ, let $\mathcal{J}(\theta) = \{\underline{j}(\theta), ..., J(\theta)\} \subseteq \mathcal{J}$ be the subset of occupations that are at the frontier of the offer set of type θ. Within this group, I order the index of these occupations by their amenity, from low to high. Then, we can recursively define the local amenity prices $\delta_j(\theta)$ as:

$$\delta_j(\theta) = -\frac{w_j(\theta) - w_{j-1}(\theta)}{A_j - A_{j-1}} > 0, \tag{1.24}$$

for all $j \in \{\underline{j}(\theta) + 1, ..., J(\theta)\}$. In some cases, we will need to compute the local amenity prices for amenities levels that are not in the frontier, i.e., $A < A_{\underline{j}(\theta)}$ and $A > A_{J(\theta)}$. In these cases, I will extrapolate using the local price $\delta_{\underline{j}(\theta)}(\theta) \equiv \delta_{\underline{j}(\theta)+1}(\theta)$ for $A < A_{\underline{j}(\theta)}$ and $\delta_{J(\theta)+1}(\theta) \equiv \delta_{J(\theta)}(\theta)$ for $A > A_{J(\theta)}$.

With this definition, we can link type-specific amenity prices to differences in the marginal product of labor across occupations, as we did in the simple example. That is, the difference in the MPL between occupations $j-1, j \in \mathcal{J}(\theta)$ is equal to

$$\mathcal{F}_{y_{j-1}} m_\theta^{j-1} - \mathcal{F}_{y_j} m_\theta^j = \delta_j(\theta)(A_j - A_{j-1}).$$

Finally, I define the total price of the amenity in each occupation as

$$\Delta_j(\theta) = \sum_{\substack{m \in \mathcal{J}(\theta) \\ \underline{j}(\theta)+1 \leq m \leq j}} \delta_m(\theta)(A_m - A_{m-1}), \tag{1.25}$$

for all $j \in \{\underline{j}(\theta) + 1, ..., J(\theta)\}$ and $\Delta_{\underline{j}(\theta)}(\theta) = 0$. Notice that this is simply the difference in wages relative to the lowest-amenity occupation in the frontier, i.e., $\Delta_j(\theta) = -(w_j(\theta) - w_{\underline{j}(\theta)}(\theta)) > 0$.

Augmented output and TFP. With amenity prices, we can now derive a Hulten's theorem with an augmented output definition. I define the augmented output as the output of the final good plus the total value of amenities:

$$Y^a = Y + \int \sum_{j \in \mathcal{J}(\theta)} \Delta_j(\theta) h^j(\theta) d\theta. \tag{1.26}$$

Output can be computed with firms' value added, consumers' expenditures, or income. The same equivalence applies to augmented output. With the income approach, we would include the monetary value of amenities in workers' income, i.e., we would measure their total compensation—as we did in the example of Section 1.6.2. For the other two approaches, we can think of a fictitious market where firms sell amenities to consumers. So, in the value added approach, we would include firms' sales of amenities, and in the expenditure approach, consumers' purchases.

Now, we can define augmented TFP ($ATFP$) growth as the growth in total factor productivity in the production of augmented output Y^a. Formally,

$$d\log ATFP = d\log Y^a - \frac{QK}{Y^a} d\log K, \tag{1.27}$$

where $d\log Y^a$ is defined as the growth in augmented output at constant prices. That is, holding the price of the final good normalized to one and the local amenity prices $\{\delta_j(\theta)\}_{j,\theta}$ fixed.[46] This term is equal to:

$$d\log Y^a = \frac{Y}{Y^a} d\log Y + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j(\theta)h^j(\theta)}{Y^a} d\log h^j(\theta) d\theta + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\delta_j(\theta;dA_j)A_j h^j(\theta)}{Y^a} d\log A_j d\theta, \tag{1.28}$$

where

$$\delta_j(\theta;dA_j) = \mathbb{1}\{dA_j \leq 0\}\delta_j(\theta) + \mathbb{1}\{dA_j > 0\}\delta_{j+1}(\theta).$$

Because the local amenity prices are nondifferentiable at every A_j, the price used to evaluate changes in amenities depends on the sign of dA_j. I illustrate this in Figure 1.7. Imagine that A_j increases by $dA_j > 0$. The increase in total amenity value is equivalent to reallocating all workers in j to an occupation j' with amenity $A_{j'} = A_j + dA_j$ and total amenity price

$$\Delta_{j'}(\theta) = \Delta_j(\theta) + \delta_{j+1}(\theta)(A_{j'} - A_j) = \Delta_j(\theta) + \delta_{j+1}(\theta)dA_j.$$

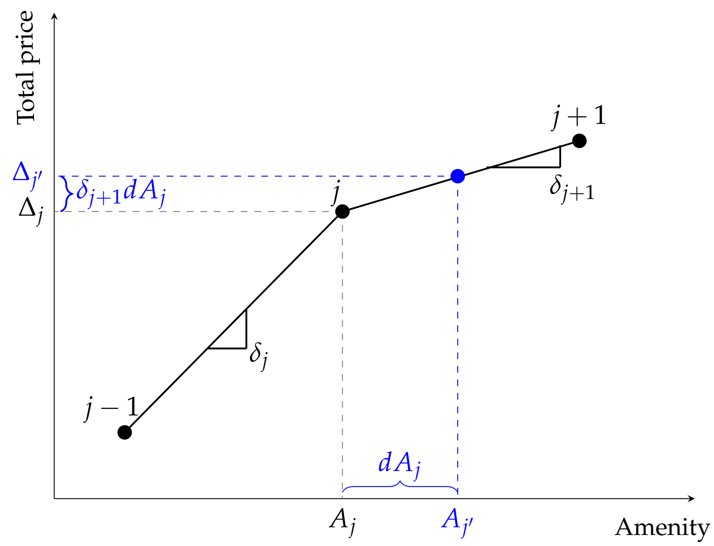
Hence, the change in the total value of the amenity is equal to:

$$\int [\Delta_{j'}(\theta)h^j(\theta) - \Delta_j(\theta)h^j(\theta)]d\theta = \int \delta_{j+1}(\theta)h^j(\theta)dA_j d\theta.$$

Hulten's theorem. The following proposition derives a Hulten's theorem for augmented TFP growth.

[46] This definition of output growth with constant prices is common in models with multiple final consumption goods (see Baqaee and Farhi, 2019 and Baqaee and Rubbo, 2023).

Figure 1.7: Local amenity prices and increase in amenity value



Proposition 2. *The augmented TFP growth satisfies:*

$$d\log ATFP = \frac{Y}{Y^a}d\log z + \sum_j \frac{W_j}{Y^a}d\log z_j + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\delta_j(\theta;dA_j)A_j h^j(\theta)}{Y^a}d\log A_j d\theta. \tag{1.29}$$

If we use amenity prices to measure augmented output and TFP, a Hulten's theorem holds so changes in the endogenous allocation (i.e., $\{d\log h^j(\theta)\}$) do not have first-order effects. As in Hulten's theorem, productivity growth only depends on changes in primitives (here $d\log z$, $\{d\log z_j\}$, and $\{d\log A_j\}$) weighted by their Domar weights. However, the Domar weights are now defined relative to augmented output $\widetilde{Y}^a$ instead of the conventional output measure $\widetilde{Y}$.

Additionally, Proposition 2 shows how to measure the contribution to aggregate productivity of changes in amenities within occupations. Think, for example, of a new machine that produces as much as its older version but is safer for the operator using it. The formula provides a way to measure the productivity gains from this safer machine. They are proportional to the share of workers that benefit from the higher safety $h^j(\theta)$ times the local price of amenities $\delta_j(\theta;dA_j)$.

Corrected TFP. Proposition 2 provides a formula to measure productivity growth for the production of augmented output, which includes the value of amenities. However, we can also use amenity prices to obtain an estimate for the conventional TFP growth—i.e., the total factor productivity in the production of the final consumption $\widetilde{Y}$—that corrects for the "measured misallocation" induced by amenities. To this end, I define the corrected TFP ($CTFP$) growth as:

$$d\log CTFP = d\log z + \sum_j \frac{W_j}{Y} d\log z_j, \tag{1.30}$$

which can be computed as

$$d\log CTFP = d\log TFP + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j(\theta)h^j(\theta)}{Y} d\log h^j(\theta)d\theta. \tag{1.31}$$

Note that apart from not measuring changes in amenities within occupations, the difference compared to the augmented TFP growth is that the corrected TFP growth is calculated with Domar weights based on the conventional output measure $\widetilde{Y}$.

1.6.4 Mapping to amenity price estimates

The growth accounting formulas derived in the previous section are nonparametric and only require occupation-level wage and employment data, amenity prices, and standard macroeconomic data. I now explain how we can map the amenity price estimates of Section 1.4 to these formulas and measure the change in total compensation, the augmented output, and the growth in corrected and augmented TFP.

First, to aggregate the amenities data to a unidimensional amenity, I use the estimates of relative amenity prices to aggregate $A_j = \sum_n \hat{\pi}_n A_{j,n}$ (i.e., I compute the amenity measure as in Section 1.4). If we assume that amenity prices are linear, we have $\Delta_j(\theta) = \delta(A_j - A_{\underline{j}(\theta)})$ for all θ, and we can compute the change in amenity value as:

$$\int \sum_{j\in\mathcal{J}(\theta)} \Delta_j(\theta)dh^j(\theta) = \sum_{j=1}^{J} \delta A_j dh^j, \tag{1.32}$$

where $dh^j = \int dh^j(\theta)d\theta$. So we only need to measure the employment share at each occupation.

The mapping is more involved when prices are nonlinear. First, notice that the amenity prices are the same for all types θ that have the same frontier of the offer set. Hence, imposing Assumption 1 used in the empirical part, we can apply a change of variables from the types θ to the frontiers of the offer set $\overline{\phi}$. Using the estimates of Section 1.4, I compute, for each census sample, the total compensation level of every worker as in Equation (1.11), which is given by: $\overline{\phi}_i = \hat{\gamma}_1 \frac{w_i^{\hat{\gamma}_2}}{\hat{\gamma}_2} + \hat{A}_{j(i)}$.

Next, at every census sample (i.e., each year), I discretize the total compensation distribution by $Q = 20$ quantiles. For every quantile $q \in \{1,...,Q\}$ and occupation j, we can then compute the employment share $h^j(q)$ and the amenity prices $\Delta_j(q)$ and $\delta_j(q)$.[47] To find the prices, we first need to assign a unique wage $\overline{w}_j(q)$ to every occupation and quantile pair (j,q). I do this by solving the root $\overline{\phi}_q = \hat{\gamma}_1 \frac{\overline{w}_j(q)^{\hat{\gamma}_2}}{\hat{\gamma}_2} + \hat{A}_j$, where $\overline{\phi}_q$ is the average total compensation at quantile q.[48] Then, the total amenity prices are:

$$\Delta_j(q) = -(\overline{w}_j(q) - \overline{w}_{\underline{j}(q)}(q)),$$

where $\underline{j}(q)$ is the lowest amenity occupation among those that have a positive employment share ($h^j(q) > 0$) at total compensation quantile q. Then, because $A_j = \sum_n \hat{\pi}_n A_{j,n}$, the local amenity price is

$$\delta_j(q) = \frac{\overline{w}_j(q)}{\hat{\gamma}_1}^{1-\hat{\gamma}_2}.$$

Finally, we can approximate the total amenity value to compute Y^a, and all the terms needed to estimate $d\log ATFP$ and $d\log CTFP$. The total amenity value is approximated by:

$$\int \sum_{j \in \mathcal{J}(\theta)} \Delta_j(\theta) h^j(\theta) d\theta \approx \sum_{q=1}^{Q} \sum_{j \in \mathcal{J}(q)} \Delta_j(q) h^j(q).$$ [49]

Similarly, we can approximate the changes in amenity value from the reallocation of workers between any periods t and $t+1$ as:

$$\int \sum_{j \in \mathcal{J}(\theta)} \Delta_j(\theta) dh^j(\theta) d\theta \approx \sum_{q=1}^{Q} \sum_{j \in \mathcal{J}(q)} \Delta_{t,j}(q) \left(h^{t+1,j}(q) - h^{t,j}(q) \right).$$

Lastly, we can approximate the change in amenity value from changes in amenities within

[47] All the parameters for the amenity prices are estimated with 2012 prices, and I also normalize all the wages to 2012 prices.

[48] Alternatively, we can obtain $\overline{w}_j(q)$ by computing the average or median wage at every (j,q) pair. The resulting values are similar.

[49] Note that in practice, we do not need to sum only over the subset of occupations in the frontier $\mathcal{J}(q)$ because the employment share is zero for the occupations that are not in the total compensation quantile q.

occupations as:

$$\int \sum_{j \in \mathcal{J}(\theta)} \delta_j(\theta) h^j(\theta) dA_j d\theta \approx \sum_{q=1}^{Q} \sum_{j \in \mathcal{J}(q)} \delta_{t,j}(q) h^j(q) (A_{j,t+1} - A_{j,t}).$$

Further details. For the macro variables, I use gross national income (GNI) for $\widetilde{Y}$. Then, I use capital stock data from Feenstra *et al.* (2015) and data on the capital share on national income from the World Inequality Database (Alvaredo *et al.*, 2020). Since in the model the mass of workers is normalized, I express all aggregates in per-worker units. The augmented TFP growth ($d \log ATFP$) is computed with the Domar weight on capital relative to augmented output. We can compute this weight as $\frac{Y}{Y^a}\alpha$, where α is the capital share.

Finally, the estimates of total amenity prices $\Delta_j(q)$ are sensitive to the wages in the lowest amenity occupation in every total compensation quantile $\overline{w}_{\underline{j}(q)}(q)$. In particular, at high total compensation levels, there may be some low-amenity occupations with a very small employment share $h^j(q)$ and very high wages, which leads to very large total amenity prices. This can significantly affect the total amenity value used to compute augmented output $\widetilde{Y}^a$. For this reason, I compute $\overline{w}_{\underline{j}(q)}(q)$ by pooling and averaging over the lowest-amenity occupations in each total compensation quantile. I select these low amenity occupations so that together they add up to less than 5% of the employment share in the quantile. Notice that this is not a problem when computing the changes in amenity value, as they are invariant to any additive rescaling of the $\Delta_j(q)$s, i.e.,

$$\sum_{j \in \mathcal{J}(q)} (\Delta_{t,j}(q) + \Delta) \left(h^{t+1,j}(q) - h^{t,j}(q) \right) = \sum_{j \in \mathcal{J}(q)} \Delta_{t,j}(q) \left(h^{t+1,j}(q) - h^{t,j}(q) \right).$$

1.6.5 Quantitative results on total compensation and productivity growth

I now present the main quantitative results on total compensation and productivity growth. The precise estimates have to be interpreted with caution because I imposed several assumptions to map the amenity price estimates to the growth accounting formulas. However, the exercise highlights the quantitative importance of measuring amenities to estimate productivity growth.

Wage vs total compensation growth. Figure 1.8 plots the cumulative changes in average wages and amenity value (with linear and nonlinear prices) for the period 1980 to 2015. During this period, I estimate that the average amenity value increased by 6424.73 dollars (in 2012 prices). Hence, this implies an increase in total compensation that is about 40% larger than the increase in average wages. Imposing linear amenity prices leads to an increase in amenity value of 11695.78 dollars and an increase in total compensation 80% larger than the increase in wages.

Figure 1.8: Changes in income and amenity value per worker (2012 prices)

plots_2006/change_nl_linear.png

Note: The values are expressed in per-worker units and 2012 prices. I commute the average wage from the Census and ACS samples, and the amenity value with linear prices is computed as in Equation (1.32) and with the price estimates from a linear specification (see Appendix A.2.3).

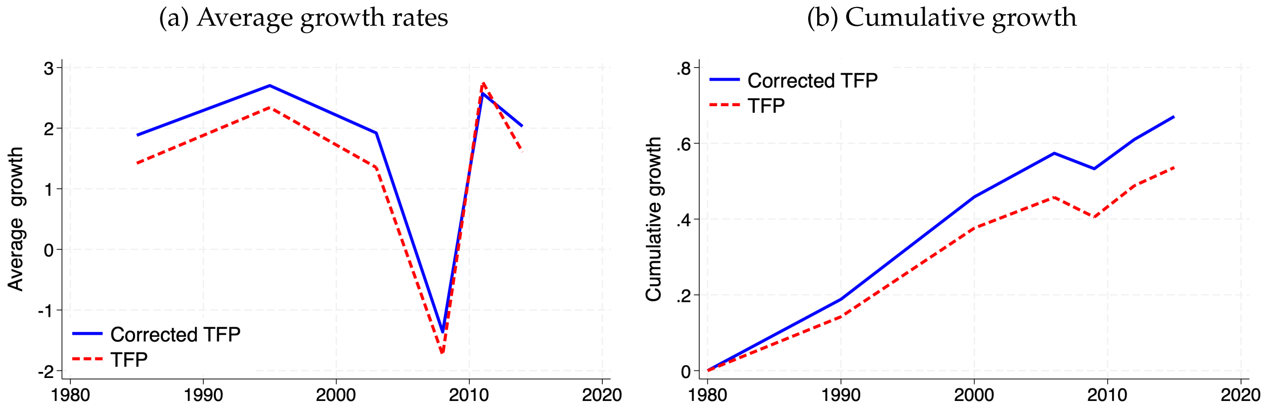
In Appendix A.6.2, I decompose the growth in total compensation by demographic groups. The increase in amenity value was relatively largest for the non-college educated and for men, which is consistent with this group being the most represented in blue-collar occupations.

Corrected TFP growth. Figure 1.9 plots the average growth rates (left panel) and the cumulative growth (right panel) of the conventional TFP measure and the corrected TFP. Recall that the corrected TFP growth is equal to the conventional TFP growth plus the growth in the amenity value, and it corrects for the "measured misallocation" in TFP due to compensating differentials. While TFP cumulatively grew by 54% between 1980 and 2015, the corrected TFP growth was equal to 67%. As a result, the total corrected TFP growth is 25% larger than the conventional TFP growth. Therefore, this estimate indicates that, due to amenity-biased of workers between 1980 and 2015, the conventional TFP measure has underestimated improvements in productivity by 25%. The average growth rates of corrected TFP and TFP are equal to 1.9% and 1.5%, respectively. Moreover, except for the period between 2009 and 2012, the corrected TFP growth is always larger.

Productivity slowdown. An extensive literature has documented and studied the causes of the US productivity slowdown between 2004 and 2015 (see Fernald, 2015; Syverson, 2017; Byrne *et al.*, 2016; Gordon, 2018). As explained by Syverson (2017), the evolution of productivity growth in the US is often divided into four periods: a period of high productivity growth from 1947 to 1973; a slowdown from 1974 to 1994; another large acceleration from 1995 to 2004; and another slowdown from 2004 to 2015 with the lowest average growth rates.

This raises the question: Can the amenity-biased reallocation explain the recent productivity slowdown? The estimates suggest that, on the contrary, accounting for amenities makes the productivity slowdown of the period 2004-2015 slightly worse. The reason

Figure 1.9: Corrected TFP and TFP



is that most of the amenity-biased reallocation of workers had already occurred by 2006 (see Figure 1.5).[50] Hence, correcting for amenities increases the productivity estimates relatively more in the earlier decades. Specifically, the average growth rate of corrected TFP between 1980 and 2006 was 2.22%, compared to 1.76% for conventional TFP. By contrast, between 2006 and 2015, the average growth rates for corrected and conventional TFP were 1.1% and 0.88%, respectively.[51] Therefore, the corrected TFP is 0.46 percentage points larger in the first period and 0.22 percentage points larger in the second, indicating an even larger slowdown than previously thought. It is important to note that these estimates are only suggestive because I only measure amenity improvements from the reallocation of workers across occupations. In particular, increases in expenditures on amenities within occupations or firms may have contributed to the slowdown.

Changes in amenities within occupations and augmented TFP growth. For the later periods (since 2006), we can use different releases of the O*NET to measure changes in amenities within occupations and so obtain an estimate for the augmented TFP growth.[52] Recall that the augmented TFP growth measures the growth in total factor productivity in the production of augmented output.

Table 1.5 contains estimates of the growth in augmented TFP and its decomposition.

[50] During the 1980s, the growth rate of corrected TFP is substantially larger than the conventional one (1.88% compared to 1.42%), implying that the productivity growth slowdown during the 1974-1994 period may have been smaller than otherwise thought.

[51] Note that in constructing the TFP, I do not adjust for factor utilization, which is known to be important for the business cycle (Basu *et al.*, 2006). However, this would have the same effect on the corrected TFP, so the percentage point difference between the two measures would not be affected.

[52] At any new release of the O*NET data, only the information for a subset of occupation is updated. For this reason, I use the data of 2022 despite doing the analysis for the period 2006-2015.

First, I find that augmented output per worker was equal to 131,020 in 2012 prices, compared to the measured value of 102,329. That is, the augmented output was 28% larger than the measured one. This is a substantial difference. The estimate has to be interpreted with caution because, as discussed in Section 1.6.4, it is sensitive to the wages of the lowest amenity occupation in each total compensation quantile. Nonetheless, it highlights the welfare importance of amenities.

The augmented TFP growth was equal to 8.5% dollars between 2006 and 2015. This compares to a growth in TFP of 7.9% and in corrected TFP of 9.7% during that period. The relative growth in productivity from changes in technology is now smaller than the corrected TFP due to dividing by Y^a instead of Y. The growth in augmented output resulting from changes in amenities within occupations is equal to 844 dollars or 0.6% percent of augmented output. This value should be interpreted as a lower bound as I cannot measure changes over time in all amenities.

Table 1.5: Decomposition growth in augmented TFP (2006-2015)

	Absolute value (Thousand of 2012 Dollars)	Relative to $\widetilde{Y}^a$
Augmented output ($\widetilde{Y}^a$) in 2006	131.02	1
Output ($\widetilde{Y}$) in 2006	102.32	78%
Change in productivity ($d\log z$ and $\{d\log z_j\}_j$ terms in (1.29))	10.32	7.9%
Change in amenity value within occupation ($\{d\log A_j\}_j$ terms in (1.29))	0.84	0.6%
Change in augmented TFP (ATFP)	$dATFP = 11.16$	$d\log ATFP = 8.5\%$

1.7 Extensions

I study growth accounting in two extensions. In the first one, I consider a model where firms endogenously produce amenities. In the second one, I extend the model to an inefficient economy with occupation-level wage markdowns. In this section, I briefly describe each, and I leave the detailed exposition for Appendix A.7.

Endogenous production of amenities by firms. Throughout the paper, I have assumed that amenities are fixed characteristics of occupations. However, some amenities are better modeled as being endogenously chosen by firms, e.g., office perks. I study growth accounting with endogenous amenities in Appendix A.7.1. I assume that firms incur costs to provide amenities to their workers. An increase in firms' amenity production shows up as an increase in production costs but is not reflected in either workers' compensation or firms' sales. Hence, an increase in endogenous amenities also decreases measured TFP. This effect on TFP is akin to an increase in the cost of a factor input. I show how we can also use the implicit prices of the endogenous amenities to define augmented output and derive a Hulten's theorem.

Inefficient economy with wage markdowns. In Section 1.6, I have studied an efficient economy where the first welfare theorem holds. This allowed us to isolate the role of amenities and to highlight how, due to compensating differentials, we can measure changes in allocative efficiency in an efficient economy. However, recent research—starting with Restuccia and Rogerson (2008) and Hsieh and Klenow (2009)—has quantified large aggregate effects on output, TFP, and welfare from misallocation in inefficient economics.

To understand how inefficiencies in the labor market interact with compensating differentials, I extend the model by introducing occupation-level exogenous wage markdowns in a similar way as Baqaee and Farhi (2020). In this case, the amenity prices are defined exactly as in Section 1.6.3 and can also be mapped to the amenity price estimates. Moreover, these prices still measure the valuations for amenities of workers that reallocate. However, due to the markdowns, they are not equal to the differences in MPLs across occupations—that is, they are not equal to the marginal costs of producing the amenities. I characterize the change in aggregate TFP in response to changes in occupation-specific productivities and markdowns. Changes in the allocation of workers affect TFP due to both markdowns and compensating differentials. Finally, I show how the corrected TFP growth measures the change in allocative efficiency due to the differences in markdowns net of the compensating differentials.

1.8 Conclusion

Smith (1776) (Chapter X part I) already introduced the idea of compensating differentials: *the wages of labour vary with the ease or hardship, the cleanliness or dirtiness, the honourableness or dishonourableness of the employment.* Since then, the theory and empirical evidence of compensating differentials have become well-established. However, the macro literature studying technical change, as well as traditional growth, welfare, and distributional accounting abstract from non-pecuniary aspects of jobs. In this paper, I introduced the notion of amenity-biased technical change, based on the idea that technical change may have also raised living standards by improving non-pecuniary aspects of jobs. By measuring the shadow value of job amenities (i.e., compensating differentials) as the labor literature, I could measure changes in non-pecuniary compensation, redo the traditional growth accounting, and revisit several macroeconomic changes of the last decades.

First, I documented an amenity-biased shift in labor demand between 1980 and 2015, which reallocated workers from low- to high-amenity occupations. On aggregate, I estimate that this reallocation implies that the growth in total compensation (wage plus the value of amenities) was 40% larger than the growth in wages.

Second, I theoretically showed how augmenting output to include the shadow value of amenities properly measures productivity improvements. Quantitatively, I found 25% larger productivity growth than suggested by the conventional TFP measure from 1980 to 2015. However, the corrected TFP estimates point to an even larger productivity growth slowdown between 2004 and 2015. Although the precise estimates have to be interpreted with caution, they highlight the importance of amenities for measuring productivity growth and welfare.

Third, I find that there is no labor market polarization if we measure workers' total compensation instead of wages. Moreover, both in the motivating exercises with skill proxies and interpreting total compensation as a skill measure that corrects for compensating differentials, I find evidence that there is no polarization across the skill distribution. Hence, this suggests that, on aggregate, there has been a shift in demand from low- to high-skill workers, as in the older skill-biased technical change literature (Bound and Johnson, 1992; Katz and Murphy, 1992).

Finally, this paper also makes a conceptual point on the definition of goods and national accounting. Some goods are not measured because there is no explicit monetary transaction, yet they are valued by people and costly to provide. This is the case for job amenities. The existence of compensating differentials—i.e., shadow prices—implies that job amenities are costly to provide and valued by workers. The costs can be explicit, like buying coffee for employees, or implicit, such as the differences in the usefulness of workers across occupations. Urban and spatial amenities are also such goods. I argue that because the costs of producing these goods are typically accounted for in national accounting, their value should also be (ideally) measured. Otherwise, changes in people's consumption between measured and unmeasured goods can lead to misleading inferences about productivity improvements.

Chapter 2

The Hidden Demand for Flexibility - A Theory of Gendered Employment Dynamics

Maria Frech[1]
Gerard Maideu-Morera

Abstract

Empirical evidence highlights women's demand for flexible working hours as a critical cause of the persistent gender disparities in the labor market. We propose a theory of how hidden demand for flexibility drives gendered employment dynamics. We develop a dynamic contracting model between an employer and an employee whose time availability is stochastic and unverifiable. We model men and women only to differ in their probability of having low time availability, which we measure in the ATUS. We explore contracts designed specifically for each gender (*gender-tailored*) and the polar case where a *male-tailored* contract is given to both men and women. For the latter, we show that contracting frictions endogenously give rise to well-documented gendered labor market outcomes: (i) the divergence and non-convergence of gender earnings differentials over the life-cycle, and (ii) women's shorter job duration and weaker labor force attachment.

[1]We are grateful to Christian Hellwig and Nicolas Werquin for their advice and guidance. We would also like to thank Charles Brendon, Ludovica Ciasullo, Fabrice Collard, Eugenia Gonzalez-Aguado, Alexander Guembel, Anna Sanktjohanser and participants at TSE Macro Workshops, the Vigo Macro Workshop, SYME 2023, RES 2024 and ESEM 2024.

2.1 Introduction

Women's need for flexible working hours has emerged as the primary source of the remaining gender differences in the labor market (Goldin, 2014). While men and women have converged on many employment dimensions, a persistent gender gap in wages and working hours remains (Blau and Kahn, 2017). In the US, 70% of the gender earnings gap can be explained by the child penalty (Cortés and Pan, 2020). The burden of unpaid care responsibilities such as childcare, mostly carried out by women, creates unpredictable schedule changes and calls for fewer working hours. Empirically, the relationship between temporal flexibility and labor market outcomes has been widely studied.[2] However, there is no theoretical understanding of the way in which women's higher flexibility needs drive gendered employment dynamics.

In this paper, we propose an explanation for gender differences in the life-cycle dynamics of wages and employment based on a hidden – external time commitments are unverifiable – demand for flexible working hours. We develop a theoretical framework that takes the unpredictable and unverifiable nature of flexibility needs seriously and studies them in dynamic employment relationships. We model the demand for flexibility as stochastic and unverifiable shocks to time availability. To study gender differences, we allow men and women to differ only in their probability of having limited time availability, $p_{\text{men}} < p_{\text{women}}$, as we will measure in the data. Hence, gender is not encoded as an ad-hoc difference in preferences. This allows for between gender similarities and within gender differences in working conditions over time depending on their exposure to flexibility shocks.[3] We study the effects of gendered flexibility needs on wage and employment dynamics through the lens of a dynamic contracting problem between an employer and an employee. We explore two types of contracts: (i) contracts designed for each gender's flexibility needs (*gender-tailored*) and (ii) the polar case where a *male-tailored* contract is given to both men and women. When contracts do not internalize women's flexibility needs (i.e. under *male-tailored* contracts), contractual frictions endogenously give rise to gendered employment dynamics that have been empirically documented. In particular, the model can account for the divergence and non-convergence of earnings differentials over the life-cycle, and women's shorter job duration and weaker labor force attachment. By contrast, under gender-tailored contracts, while individual differences may exist, the

[2] Mas and Pallais (2017), Mas and Pallais (2020),Wiswall and Zafar (2018).

[3] Our framework serves more generally to understand the consequences of flexibility needs on employment dynamics and is not per se gendered. It is also applicable to other sociodemographic groups, e.g. single fathers, that might experience unpredictable external time commitments.

systematic disparities between men's and women's labor market outcomes are negligible.

Job flexibility covers various temporal aspects like hours worked, specific times, and work hour predictability. Reduced-hour arrangements alone don't fully meet the demand for flexibility since care and family responsibilities can be unpredictable[4]: consider a woman heading to work, only to receive a call that her child is sick and requires immediate attention. Despite the employer allowing her to work from home, frequent emergencies will make it too costly for the employer to accommodate her needs while maintaining current working conditions and raise doubts about her genuine need for flexibility. Over time, this can have long-lasting consequences, e.g. not getting promoted.

We study a dynamic contracting problem (in the spirit of Clementi and Hopenhayn, 2006; Dovis, 2019) that aims to capture the mechanism of this story. A risk-neutral employer hires a risk-averse employee based on a contract that specifies a wage and working hours for every period. The employee's time availability is subject to i.i.d shocks. A low time availability shock corresponds to the case where e.g. a child unexpectedly needs to be picked up from school. Conversely, a high time availability shock coincides with the case where the employee can work as planned. Experiencing a low shock will make every hour worked more costly to that employee. Informational asymmetries arise as the employer can not observe the realization of the employee's time availability shock and thus does not know how costly it is for the employee to work (private information). Furthermore, the employee does not have to commit to staying in the contract (limited commitment) and can pursue an outside option. Likewise, the employer can also choose to terminate the contract.

We first characterize analytically the main properties of the optimal contract for compensation dynamics, working hours and termination probabilities as well as the region in the state space where termination may be optimal. The private information friction implies that in order to provide flexibility in hours, an employee with high (low) time availability must be rewarded (penalized). Moreover, because the employment relationship is dynamic and the employee is risk averse, it is optimal for the employer to smooth these rewards and penalties over time. Therefore, an employee who demands to work fewer hours due to low time availability will be penalized with lower wages in all future periods. When an employee experiences a sufficiently long sequence of low time availability, working conditions worsen. Pursuing an outside option becomes more attractive

[4]We focus on flexibility needs arising from unforeseeable family emergencies. A stable and predictable work schedule, i.e. not to be called in for work unexpectedly when they have scheduled care responsibilities, is equally important for women, see e.g. Ciasullo and Uccioli (2022). See Appendix B.6.2 on how our model can be extended to a setting where the employer seeks flexibility in working hours.

for the employee, which makes providing flexibility more costly. Eventually, terminating the employment relationship can become optimal.

To understand the gendered employment dynamics arising from these contracts we allow the probability of having low time availability to differ by gender, namely, $p_{men} < p_{women}$. This is the only parameter allowed to encode gender differences. We contrast two types of contracts. First, we study *gender-tailored* contracts. These contracts are designed for each gender specifically, meaning we solve the optimal contract for each p. Through a comparative statics exercise on p, we show that for a male and female employee that have the same compensation level, men experience larger wage penalties for demanding flexibility.[5] We find that when contracts fully internalize differences in flexibility needs, the average gender wage gap is constant over time. This is because the higher frequency of women's penalties is offset by men's higher penalties.

Second, we study *male-tailored* contracts. These are contracts initially designed for men but which then are also given to women. In practice, employers may not be able or allowed to offer gender-specific work arrangements. Hence, this exercise highlights the dynamic consequences of the incompatibility of women's flexibility needs in male-dominated work environments.[6] When employment relationships are designed to fit men's flexibility needs, women's average wages gradually start to diverge from men's. This is because women get both more frequent but also higher penalties. We also explore an intermediate case where the employer simultaneously employs men and women but is constrained to design a unique contract for both genders (*team-tailored* contract). For this case, we show that men's wages diverge upward and women's downward.

To directly compare wage paths and termination probabilities, the provision of temporal flexibility, and the flexibility penalties by gender, we also solve and simulate the model numerically. We combine the recursive Lagrangian method of Marcet and Marimon (2019) with a direct promised utility approach. We use the latter to solve the model in the termination region. We calibrate the model and use the American Time Use Survey (ATUS) to identify meaningful values for p_{men} and p_{women}. In the ATUS, we observe daily minutes spent on care activities during usual working hours. From this, we find that, for our baseline calibration, men have a frequency of limited time availability of $p_{\text{men}} = 6\%$, whereas women are more than twice as likely to be interrupted with a chance of $p_{\text{women}} = 15\%$. We also show substantial heterogeneity in these probabilities across

[5] This is indeed an empirical finding with respect to part-time work, sometimes referred to as "flexibility stigma"(Coltrane *et al.*, 2013; Golden, 2020; Aaronson and French, 2004; Dunn, 2018; Wolf, 2014; O'Dorchai *et al.*, 2007).

[6] Torre (2017), Mas and Pallais (2020), Patrick *et al.* (2016), Cha (2013).

different socioeconomic groups and quantify the effect on wage dynamics in our model.

Comparing the numerical results of *male-tailored* and *gender-tailored* contracts allows us to understand what drives differences in employment dynamics. Our results shed light on the mechanisms behind two well-documented gendered labor market outcomes:

1. *Women's wages start to diverge from men's after childbirth but do not fully converge back to men's after children grow up.*
   A large and internationally diverse literature has shown that women and men have divergent earnings growth trajectories after childbirth, even when they were previously on the same career paths (Barth *et al.*, 2021; Paul, 2016). This is partially explained by occupational sorting of women anticipating greater flexibility needs before having children (Kleven *et al.*, 2019; Mas and Pallais, 2020; Cortés and Pan, 2019). However, mother's wages also diverge from men's within the same firm and occupation, mainly through the lack of promotions (Lucifora *et al.*, 2021; Bronson and Thoursie, 2019). Over time, many women are being pushed out of current work arrangements due to parental demands and end up in predominantly lower-paying jobs (Patrick *et al.*, 2016).

   Importantly, despite children growing up and the gender wage gap narrowing, women's wages still never fully converge back to men's (Goldin *et al.*, 2022). Our model provides a potential underlying mechanism for both the divergence (within and across jobs) and non-convergence in wages. When contracts do not fully internalize differences in flexibility needs (i.e. *male-tailored* or *team-tailored* contracts), the average gender wage gap gradually grows over time and women that are particularly exposed to flexibility needs will be pushed out of the current employment relation. For employees that stay, in our main calibration, we are able to explain more than 40% of the within-firm divergence of wages. Lastly, in our model, demanding flexibility is penalized with both current and future wage cuts so that even when men's and women's flexibility needs are the same again after children grow up, wages will not converge back.

2. *Women's job duration is shorter and labor force attachment is weaker.*
   It is well documented that women's job spells are, on average, substantially shorter than men's (Hall, 1982; Molloy *et al.*, 2020; Munasinghe *et al.*, 2008). This is in part related to gender differences in job mobility patterns. A higher job turnover for women has been shown to contribute to the gender wage gap (Amano-Patiño *et al.*, 2020). Empirical evidence suggests that this is a result of women's care responsi-

bilities leading also to a weaker labor force attachment.[7] Unlike static occupational choice models, our dynamic contracting setting is able to speak to the impact of gender differences in external demands on job duration and turnover. Under *male-tailored* contracts, a larger share of women end up with depressed working conditions due to their higher flexibility needs. This leads to higher termination rates and shorter job duration for women.

Finally, comparing *gender-tailored* and *male-tailored* (and *team-tailored*) contracts highlights the potential adverse consequences of non-discriminatory contracts. This is a practical concern, as policymakers may overlook the unfavorable effects of not allowing work arrangements to be targeted towards one gender. For example, Antecol *et al.* (2018) show that gender-neutral tenure clock extensions reduced women's tenure probability while increasing men's.

Related Literature. Our theory provides a conceptually new way of thinking about the link between flexibility needs and working conditions. We propose that a hidden demand for flexible working hours drives gender differences in employment dynamics. That is, they are a result of information and contracting frictions. The seminal work in the literature providing microfoundations on how gendered flexibility needs can explain wage differentials is Goldin (2014). By analyzing the convexity of the hour-wage relationship in a static model, she can explain differences in the gender pay gap across and within occupations.[8] Erosa *et al.* (2022) conduct a quantitative analysis of the Goldin (2014) theory and find that it can account for a large share of the gender gaps in occupational choice, wages, and hours. By contrast, our dynamic model gives insights into the divergence (within and across occupations) of wages after childbirth and the non-convergence of wages after children are grown up – which cannot be rationalized with static models.

The across-occupation wage gap has been studied by Flabbi and Moro (2012) and Morchio and Moser (2024).[9] They quantify the effect of gender differences in preferences

[7] Married women are more likely to stay at home, find a new, more family-friendly job (Mas and Pallais, 2020; Mas and Pallais, 2017; Wiswall and Zafar, 2018) or pursue flexibility-oriented self-employment (Lim, 2019; Gurley-Calvez *et al.*, 2009; Bento *et al.*, 2021). Unmet needs for workplace flexibility push women into less profitable work arrangements or home production (Patrick *et al.*, 2016).

[8] Our theory disconnects wage from the marginal product of labor, which has important normative implications. If wage differentials are driven by a convex production function as in Goldin (2014), a reduction of the wage differential inside a family will entail a loss in production efficiency, which may be inefficient both socially and for the family. By contrast, if they are driven by contracting frictions, there may not be any production efficiency loss from a reduction in wage differentials.

[9] Relatedly, Le Barbanchon *et al.* (2021) show that women are more willing to tradeoff wages for a shorter commute time and study the consequences for the gender wage gap.

for occupational amenities, including job flexibility. In our model, differential flexibility needs create within-firm and within-occupation gender wage gaps and push women into lower-wage work arrangements that were suboptimal ex-ante.

Another strand of the literature uses models of human capital accumulation to explain the child penalty (Erosa *et al.*, 2016; Amano-Patiño *et al.*, 2020; and Barigozzi *et al.*, 2023). Due to maternity leave and lower working hours after returning to employment, women are able to accumulate less human capital on the job. These models hence generate (within-firm) divergence and non-convergences of wages. Our theory provides an alternative to generating these gendered wage dynamics that is based on contracting frictions where women can be paid less even if they are equally productive as men. In addition, human capital models remain silent about gender differences in job duration. Lastly, models of promotions (Lazear and Rosen, 1990; Lommerud *et al.*, 2015; and Bronson and Thoursie, 2019) can also give rise to within-firm gender wage gaps.[10]

Conceptually closest to our approach is Albanesi and Olivetti (2009).[11] They show how the combination of information frictions in the labor market (moral hazard and adverse selection) and intra-household home production decisions can generate self-fulfilling gendered equilibria. We abstract from home production decisions, take the gendered flexibility needs as given, and instead focus on its unpredictable nature and study the consequences for the life-cycle dynamics of wages and employment.

A growing empirical literature has tried to measure gender differences in external demands on time (Buzard *et al.*, 2023), including unexpected incidences during working hours (Cubas *et al.*, 2021; Schoonbroodt, 2018). We use our model to study the effects of these unpredictable and gendered demands on time. Moreover, we provide our own model-consistent estimates using the ATUS.

The dynamic contracting problem we study combines private information (from the flexibility shocks being unobservable to the employer) and limited commitment (from the employee being able to leave at any time) frictions. In this sense, the model is close to the sovereign debt model in Dovis (2019). As in Dovis (2019), the combination of the two frictions implies that the optimal contract may (temporarily) end up in a region with ex-post inefficiencies. In our case, this results from the fact that as the employee's value

[10] In our model, we could interpret increases in wages as – current or as changes in consumption in anticipation of future – promotions. Hence, our model may shed light on how differences in expected demands on time can create gender differences in promotion opportunities.

[11] Relatedly, Albanesi *et al.* (2015) study gender differences in top executives' compensation through a moral hazard model but find that a model of "managerial power" is more consistent with empirical evidence. Models that feature self-fulfilling gendered equilibria have also been studied by Dolado *et al.* (2013) and Lommerud *et al.* (2015).

approaches the outside option, it is impossible to induce an employee with low time availability to work positive hours. Then the employer's cost of increasing the compensation is smaller than the gain of inducing an employee with limited time to work positive hours, which implies that the Pareto frontier is increasing. In addition to Dovis (2019), in this region, we allow the principal the option to terminate the contract, which will be optimal if its outside option is high enough.

Outline. The paper is organized as follows. Section 2.2 lays down the environment. Section 2.3 sets up the dynamic contracting problem and characterizes the main properties of the optimal contract, including optimality of termination. In Section 2.4 we show how differences in time availability can generate gendered employment dynamics. We explore *gender-tailored*, *male-tailored*, and *team-tailored* contracts. In Section 2.5, we explain how we use the American Time Use Survey to calibrate the model, present the parametrization and discuss the solution method. Section 2.6 presents the numerical results and Section 2.7 connects them to the aforementioned gendered labor market outcomes. Section 2.8 concludes.

2.2 Environment

Time is discrete and indexed by $t = 0, 1, ..., \infty$. There is one employer, the principal, and one employee, the agent. They contract on a long-term employment relation that specifies hours and wages for each period. The employee does not have to commit to staying in the contract and can pursue an outside option. Conversely, the employer can also choose to terminate the employment relation. Thus, the contract features two-sided limited commitment. The source of uncertainty is a shock to the time availability of the employee, which is not observable to the employer. We model the contracting relation between the employer and the employee as a message game. At time 0, the employer makes a take-it-or-leave-it offer to the employee. The offer consists of a contract whose terms can be contingent on all public information.

2.2.1 Preferences

The employer and the employee are infinitely lived. At every period, the employee's time availability f_t can take two values $f_t \in \{f^L, f^H\}$ with $0 < f^L < f^H < 1$. Time availability is independently and identically distributed over time with $\mathbb{P}(f_t = f^L) = p$. This is the only parameter that will differ by gender in the entire model. The employee values a stochastic sequence of wages $\{w_t\}_{t=0}^{\infty}$ and working hours $\{h_t\}_{t=0}^{\infty}$ according to

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t U(w_t, h_t; f_t), \tag{2.1}$$

where $\beta \in (0,1)$ is the discount factor. The per-period utility is specified by

$$U(w,h;f) = u(w) - (1-f)\psi(h), \tag{2.2}$$

where $u : \mathbb{R}^+ \to \mathbb{R}$ is strictly increasing and concave and $\psi : \mathbb{R}^+ \to \mathbb{R}$ is strictly increasing and convex. The employee has an outside option that gives a life-time value $\overline{v} \in \mathbb{R}$.[12]

The employer is risk-neutral, also discounts the future with $\beta \in (0,1)$ and has a per period profit of

$$\pi(w,h) = g(h) - w, \tag{2.3}$$

where $g : \mathbb{R}^+ \to \mathbb{R}^+$ is increasing, concave and satisfies the condition $\lim_{h\to 0} g_h(h) = +\infty$. The employer has an outside option that gives a value of $\overline{\Pi} \in \mathbb{R}$.

2.2.2 Information

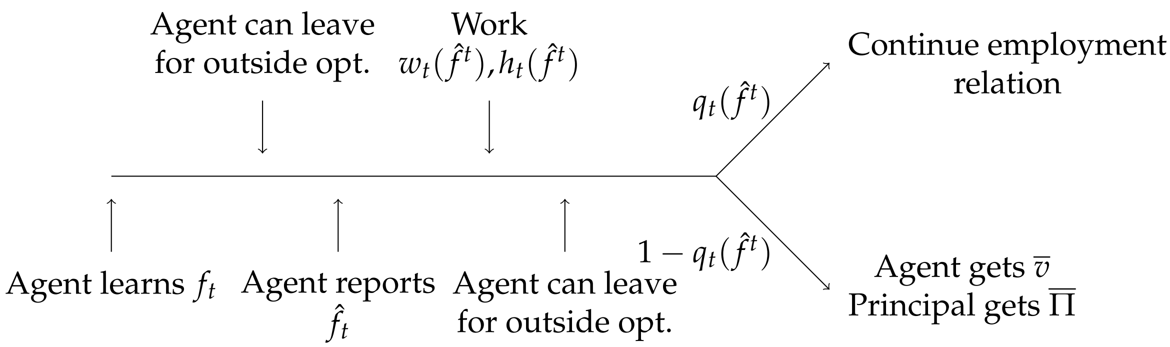
The employer cannot observe the realization of f_t, the time availability of the employee in period t. The employer bases the contract on the time availability reported by the employee in each period, $\hat{f}_t$. We invoke the Revelation Principle to, without loss of generality, restrict attention to truth-telling mechanisms. This reduces the message space to $\{f^L, f^H\}$. A reporting strategy for the employee is given by $\hat{f} = \{\hat{f}_t(f^t)\}_{t=0}^{\infty}$, where $f^t = (f_1, ..., f_t)$.

2.2.3 Timing

The timing of the contract within each period is as depicted in Figure 2.1. At the beginning of the period, the employee learns their time availability $f_t \in \{f^L, f^H\}$ and then has the possibility to pursue their outside option. If the employee stays, they then report $\hat{f}_t \in \{f^L, f^H\}$ to the employer. Based on the (history of) report(s) the employer offers the employee to work $h_t(\hat{f}^t)$ for a wage $w_t(\hat{f}^t)$. The employee again has the option to leave and pursue the outside option. At the end of the period, the employer proposes

[12] We leave the interpretation for the outside option open. It could, for example, include a more flexible work arrangement, home production, or self-employment.

Figure 2.1: Timing



the employment relation to continue with a probability $q_t(\hat{f}^t)$ and to terminate with a probability $1 - q_t(\hat{f}^t)$. In the latter case, the employee gets the outside option $\overline{v}$ and the employer gets $\overline{\Pi}$. The employee can choose to leave both after learning this period's time availability and at the end of the period.

2.2.4 Contract

The dynamic contract specifies employment policies which are contingent on all information provided by the employee. Letting $\hat{f}^t = (\hat{f}_1, ..., \hat{f}_t)$ denote the history of the employee's reports, the contract $\sigma = \{w_t(\hat{f}^t), h_t(\hat{f}^t), q_t(\hat{f}^t)\}$ specifies a contingent policy of termination probabilities q_t, hours h_t and wages w_t.

2.2.5 Feasible Contract

Definition 1. *A contract σ is feasible if $\forall t \geq 1$ and $\forall \hat{f}^t \in \{f^L, f^H\}^t$*

(i) $q(\hat{f}^t) \in [0,1]$

(ii) $w(\hat{f}^t), h(\hat{f}^t) \geq 0$

We collect the assumptions made so far:

Assumption 3. *(i) The utility function $U(w,h;f) = u(w) - (1-f)\psi(h)$ is strictly increasing and concave in both arguments and satisfies the single crossing property $U_{hf} > 0$. The production function g is increasing, continuously differentiable and satisfies $\lim_{h \to 0} g_h(h) = +\infty$.*
(ii) The time availability shock can take on two values $f \in \{f^L, f^H\}$ and is independent and identically distributed over time. Each period time availability can be low with probability $p \in (0,1)$ and high with probability $1 - p \in (0,1)$.

The following convention for notation is useful: a superscript H indicates an element in the contract designed for an employee that reports to be of high time availability, e.g. the wage specified for an employee that reports to have high time availability at time t is denoted by $w_t^H = w(f^{t-1}, f^H)$. Equivalently, the superscript L denotes an element in the contract designed for an employee that reports to be of low time availability in the current period.

2.2.6 A Benchmark: Contracts under Symmetric Information

As a first step, we consider the case of symmetric information, where the employer observes the employee's realization of time availability and the employee has full commitment.

Proposition 3. *In the first best, at every history f^t, the contract satisfies:*

1. *No distortions in hours worked, that is,*

$$g'(h_t(f^t)) = \frac{(1-f_t)\psi'(h_t(f^t))}{u'(w_t(f^t))}. \tag{2.4}$$

2. *Full insurance and perfect intertemporal wage smoothing, that is $w = w_t(f^t)$.*

3. *No termination, that is $q_t(f^t) = 1$.*

In the first best, the employer is able to set hours and wage optimally for both high and low time availability. In particular, this means hours are not distorted. The employee is perfectly insured against flexibility shocks and is thus allowed to reduce hours worked without a penalty on total compensation. In particular, wages for both employees with high and low time availability are the same and constant over all periods. Lastly, given that both parties agreed to enter the employment relation at time zero, it is never optimal for the employer to lay off the employee.

2.2.7 Model Discussion

We study a particular type of contract where wages are adjusted as frequently as flexibility needs arise. While such contracts may not have a direct counterpart in reality, employers may implicitly track attendance and adjust wages in the long run through, for example, reduced promotion opportunities. In this case, we can interpret the changes in current compensation as a result of the employees adjusting their private borrowing or saving in anticipation of future pay raises or wage cuts.

We purposely avoid putting structure on the outside option. We keep it the same for both genders throughout since we do not want our termination rates to be driven mechanically, but it is reasonable to expect this variable to be gendered. The outside option could capture, for example, the value of going into a higher flexibility and lower-wage occupation, dropping out of the labor force, or pursuing flexibility-oriented self-employment.

To focus on gender differentials in wage dynamics resulting from flexibility needs, we deliberately abstract from other drivers of wage dynamics over the life-cycle, such as human capital accumulation. For this reason, when women's wages diverge downwards in the model, we do not interpret it as women's wages decreasing over time. Instead, we view this as women's wage divergence relative to men's wage path.

Lastly, our framework focuses on flexibility needs that arise when employees need to adjust their working hours due to unpredictable family emergencies. As highlighted by Ciasullo and Uccioli (2022), a stable and predictable work schedule is equally important for women to be able to, for example, schedule care responsibilities. In Appendix B.6.2, we lay out a simple extension of our model where the employer can also ask to work unpredictable extra hours. We assume that the employee's cost of working these extra hours is also stochastic and unverifiable. All the main intuitions and results go through in that setting.

2.3 Optimal Contract

In this section, we define the recursive formulation of the (constrained) optimal contract with information asymmetry in f and limited commitment induced by the outside option. We characterize the properties of the optimal contract upon continuation and then show that termination may be optimal in some states. These results hold for a given probability p of having low time availability.

2.3.1 Recursive Formulation

The employee's continuation value from staying in the contract, denoted v_t, evolves according to

$$v_t = (1-p)\left(U(w^H, h^H; f^H) + \beta v^H_{t+1}\right) + p\left(U(w^L, h^L; f^L) + \beta v^L_{t+1}\right), \tag{2.5}$$

which is composed of the expected per period utility and the expected continuation value. The quantities v^H_{t+1} and v^L_{t+1} are the continuation values contingent on high and low time-availability report, respectively. Following Sargent and Ljungqvist (2000) and Clementi and Hopenhayn (2006), we present the problem in a recursive form, where the state variable of the problem is the employee's continuation value at the beginning of a period. In particular, the contract σ, introduced in Section 2.2.4, can now be decomposed into current allocations $(w_t(f^j), h_t(f^j))$ and a continuation value v^j_{t+1} contingent on reported f^j for $j \in \{H, L\}$.[13]

First, define the total value to the employee if she stays in the contract. The flow equation (2.5) now becomes the promise-keeping constraint, as the initial value v must be delivered to the employee by the continuation contract. Suppressing time subscript this gives:

$$v = (1-p)\left(U(w^H, h^H; f^H) + \beta v^H\right) + p\left(U(w^L, h^L; f^L) + \beta v^L\right). \tag{PK}$$

As usual, the only relevant *incentive constraint* is the one imposing truthful reporting in the high state:[14]

$$U(w^H, h^H; f^H) + \beta v^H \geq U(w^L, h^L; f^H) + \beta v^L. \tag{IC}$$

The incentive compatibility constraint (IC) captures the informational frictions in setting the contract. It ensures that the employee has no incentive to misreport. In particular, an employee with high time availability does not want to take the contract designed for someone with low time availability.

The lack of commitment of the employee imposes two constraints that are of different

[13] In this section we focus on contracts without termination, thus continuation probabilities are one for both reported time availabilities, i.e. $q(f^H) = q(f^L) = 1$.

[14] This is shown formally in the proof of Proposition 4 in the Appendix.

timing. First consider the standard *limited commitment constraints*:

$$v^H, v^L \geq \overline{v}. \tag{LC}$$

These constraints impose that the employee's continuation value promised in the current contract must exceed their outside option, i.e. after working in a period they will still prefer to stay in the contract. Additionally, a contract should be sustainable in the following way:

$$\begin{aligned} U(w^H, h^H; f^H) + \beta v^H &\geq \overline{v}, \\ U(w^L, h^L; f^L) + \beta v^L &\geq \overline{v}. \end{aligned} \tag{SUST}$$

These *sustainability constraints* ensure that even before working and regardless of their time availability, employees will want to continue in the employment relationship. Hence, (LC) and (SUST) guarantee that the employee prefers staying in the employment relation to the outside option at different times of the contract.

We denote the employer's value contingent upon continuation as $\hat{\Pi}$. This is different from the value of the employer prior to the firing decision, which is denoted as Π. The choice variables are the wages and hours for the current period contingent on the reported time availability as well as the contingent continuation values v^H and v^L. The employer then solves:

$$\widehat{\Pi}(v) = \max_{\substack{w^H, h^H, v^H \\ w^L, h^L, v^L}} (1-p)\left[\pi(w^H, h^H) + \beta\Pi(v^H)\right] + p\left[\pi(w^L, h^L) + \beta\Pi(v^L)\right] \tag{2.6}$$

subject to (PK), (IC), (SUST), (LC). To recover the time zero problem, we can solve $\widehat{\Pi}(v_0)$ for a given starting promised utility v_0.

Contingent on the high or low time availability, the first term of the maximand corresponds to the current period's expected profits and the second term indicates the expected discounted total value of the employer prior to the firing the decision.

2.3.2 Termination Decision

We now turn to the termination decision of the employer. In some states, a stochastic firing decision may be optimal. Technically speaking, this is the case when the outside option for the employer implies that the Pareto frontier is not a convex set. Allowing for randomization over the termination decision is equivalent to assuming that the employer offers a lottery to the employee at the end of every period. The employment relation ends with probability $1-q$, in which case the employer receives $\overline{\Pi}$, and it continues with probability q. In the latter case the employee receives continuation value v_c. Then the function $\Pi(v)$ solves the following functional equation:

$$\Pi(v) = \max_{q\in[0,1],v_c} (1-q)\overline{\Pi} + q\widehat{\Pi}(v_c)$$
$$\text{subject to } (1-q)\overline{v} + qv_c = v.$$

2.3.3 Time Zero Contracting

We are agnostic about how the starting promised utility v_0 is determined and how it differs by gender. For employees to engage in the employment relation we need $v_0 \geq \overline{v}$. The choice of v_0 does not affect the results we present in the following sections.

2.3.4 Properties

We now present the main properties of the optimal contract in case of continuation. The results on optimal termination are discussed in Section 2.3.5.

Under symmetric information, the employer is able to fully insure the employee against having low time availability. As shown in the following proposition, under private information and lack of commitment, the optimal contract no longer provides full insurance against flexibility shocks.

Proposition 4. *At every history f^t, the optimal allocation satisfies*

1. *Hours worked of an employee with high time availability are undistorted and satisfy equation (2.4). The hours of an employee with low time availability are distorted downwards and satisfy*
$$g'(h^L) > \frac{(1-f^L)\psi'(h^L)}{u'(w^L)} \tag{2.7}$$
2. *For $v > \overline{v}$, the employer compensates high time availability with a higher wage ($w^H > w^L$) and higher continuation utility $v^H > v^L$. Moreover, when the (SUST) and (LC) constraints do not bind, $v^H > v > v^L$.*

3. If the (SUST) and (LC) constraints do not bind, the following Inverse Euler equation holds

$$\frac{1}{u'(w_{t-1})} = p\frac{1}{u'(w_t^L)} + (1-p)\frac{1}{u'(w_t^H)}. \tag{2.8}$$

The presence of private information induces a higher compensation for working more hours for an employee with high time availability. Having low time availability will be penalized with a lower wage and accommodated with working less hours. Low time availability employees are offered contracts with working hours lower than predicted in the first best setting. This is directly evident by comparing equation (2.7) with (2.4). When an employee demands flexibility, hours are distorted below the first best optimum. Hence, the contract under-works the employees with low time availability. This is because, to correctly screen time availability, lying about f needs to be less attractive for the high time availability employee. Therefore, the optimal contract prescribes fewer hours than at the first best after low time availability to discourage the type with high time availability from claiming otherwise.

Because the employee is risk averse and the employment relationship dynamic, it is optimal for the employer to smooth the employee's compensation intertemporally. This means that high time availability not only is rewarded with a higher wage at t, but is also compensated with a higher continuation value, which implies that wages increase in all future periods. Wage premia are thus permanent. Similarly, the provision of flexibility comes at the cost of lower wages, which, due to the optimality of smoothing compensation, implies a lower continuation value and, therefore, lower wages in all future periods.

The third result of the proposition shows that when the (SUST) and (LC) constraints do not bind, the inverse marginal utilities follow a martingale, implying that the cross-sectional average along a large population of workers would be constant over time. This typically implies that average wages are also approximately constant over time, as we will show in the numerical simulations. With log utility ($u(c) = \log(c)$), the average wages are exactly constant, i.e.

$$w_{t-1} = pw_t^L + (1-p)w_t^H.$$

In Section 2.4, we explore the implications of this result for the dynamics of the gender wage gap.

2.3.5 Optimality of Termination

We now characterize when terminating the employment relation is optimal. Intuitively, this occurs when the employee has accumulated sufficiently many incidences of low time availability and the cost of providing flexibility is too high for the employer. If the outside option for the employer is high enough, termination can be optimal.

Formally, our approach to show this property of the optimal contract is the following. First, we study a constrained problem where the principal is not allowed to terminate the contract. We show that the principal's value in the constrained problem $\Pi^c(v)$ is increasing in some region around $\overline{v}$. This implies that there must be some range of values for the outside option $\overline{\Pi}$ such that the (constrained) Pareto frontier $\{\overline{\Pi}, \Pi^c(v)\}$ is not a convex set. Then, Π^c and Π do not coincide, and terminating the contract with a positive probability is optimal. Intuitively, in the inefficient region where $\Pi^c(v)$ is increasing, both the employee and the employer would gain by increasing v, but this would violate the (PK) constraint. If $\overline{\Pi}$ is high enough, the principal can offer a lottery between terminating and continuing at a higher v such that the (PK) constraint is satisfied and the principal obtains a strictly higher value.

The value of the principal in the constrained problem Π^c satisfies the following Bellman equation:

$$\Pi^c(v) = \max_{\substack{w^H,h^H,v^H \\ w^L,h^L,v^L}} (1-p)[\pi(w^H,h^H) + \beta\Pi^c(v^H)] + p[\pi(w^L,h^L) + \beta\Pi^c(v^L)]$$

$$\text{subject to (PK), (IC), (SUST), (LC).}$$

The contract becomes inefficient as v approaches $\overline{v}$ because it is not possible to induce an employee with low time availability to work positive hours while satisfying the (IC) and (SUST) constraints. Intuitively, when $v = \overline{v}$ the (SUST) and (LC) constraints together imply that the continuation utility of both high and low time availability employees must be the same. Full insurance can only be incentive-compatible if the information rent given to the time-affluent employee is 0, which requires $h^L = 0$. The following lemma formalizes this result.

Lemma 2. *For an allocation satisfying (IC), (SUST) and (PK), h^L converge to 0 as v converges to $\overline{v}$.*

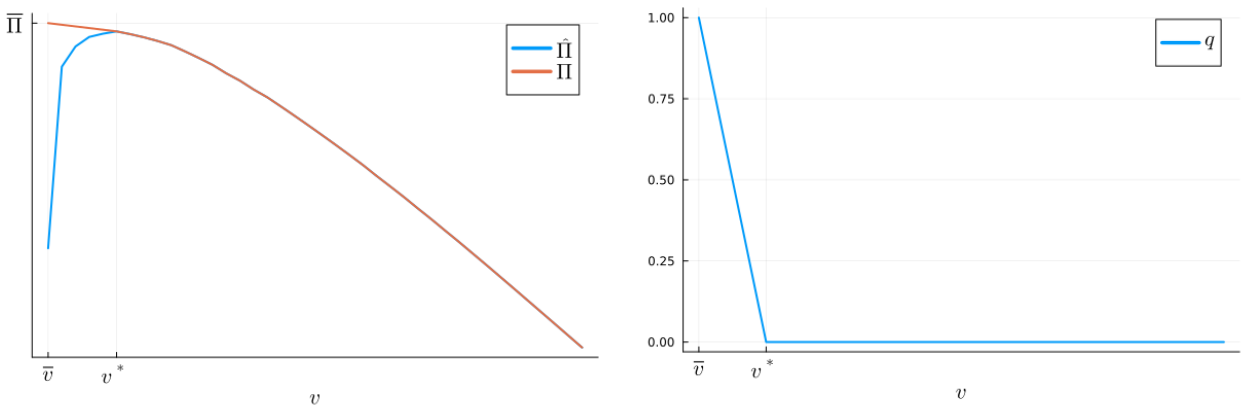
Relying on Lemma 2, we can now show that Π^c must be increasing in a region around $\overline{v}$. The result and proof are akin to Dovis (2019).

Proposition 5. *There exists a $\widetilde{v} > \overline{v}$ such that the (constrained) profit function $\Pi^c(v)$ is increasing over $[\overline{v}, \widetilde{v})$ and decreasing for $v > \widetilde{v}$.*

Figure 2.2: Pareto frontier and termination probabilities

(a) Principal's values. In red Π, in blue $\widehat{\Pi}$.

(b) Termination probabilities



The proof relies on the fact that the condition $g'(0) = \infty$ implies that the gains from slightly increasing h^L must be larger than the cost of providing a higher expected utility to the employee. Finally, we characterize the optimality of termination in the following corollary.

Corollary 1. *If $\overline{\Pi} > \Pi^c(\overline{v})$, there exists a set of values $(\overline{v}, v^*)$ where a positive termination probability, i.e. $q < 1$, is optimal.*

Figure 2.2 depicts the results of Proposition 5 and Corollary 1. On the left, we see that there exist values v for which the principal's values are increasing. Hence, both parties would improve from increasing promised utilities v. On the right, we see that for promised utilities close enough to the outside option $\overline{v}$ termination occurs, i.e. termination probabilities are positive. Typically, v^* will be the point where the line going from $(\overline{v}, \overline{\Pi})$ to $(v^*, \Pi(v^*))$ is the tangent at $(v^*, \widehat{\Pi}(v^*))$. Below this point, increasing v would be beneficial for both the employer and employee. However, this would violate the (PK) constraint. This can be bypassed by offering a lottery between terminating and continuing at a higher v. Then, the promised utility upon continuation will be v^* and the termination probabilities will be given by

$$q = \frac{v - \overline{v}}{v^* - \overline{v}}. \tag{2.9}$$

This mechanism has an intuitive explanation in our setting. The employees that end up in the termination region are the ones that suffered from a long sequence of low time availability, permanently depressing continuation values. The closer the promised continuation value gets to the outside option, the costlier it is for the employer to provide flexibility. At $v = \overline{v}$, demanding flexibility would lead to a contract with zero hours $h^L = 0$ (Lemma 2) and correspondingly a low wage. From the employer's perspective it makes sense to either terminate the employment relationship or to increase the hours to a point where the gains from increased production are higher than the cost of providing a higher continuation value.

2.4 Generating Gendered Employment Dynamics

Up to this point we have studied the characteristics of the constrained optimal contract for a fixed probability of having low time availability. First, we have shown that low time availability is penalized with both current ($w_L < w_H$) and future ($v_L < v_H$) wage cuts. Second, a sufficiently long sequence of low time availability can drive employees into the termination region, making the outside option attractive. Hidden flexibility needs encoded as private information on time availability thus allow the model to generate employment dynamics.

We now introduce gender by comparing contracts with two different probabilities of low time availability. This will be the only parameter that will encode differences by gender with $p_{\text{men}} < p_{\text{women}}$. First, we consider *gender-tailored* contracts. These contracts are designed with the flexibility needs of each gender in mind. In practice, that means we solve the optimal contract twice, once for each p. Second, we consider *male-tailored* contracts. These contracts are designed with men's need for flexibility in mind but then given to both men and women. Third, we explore the wage dynamics in an intermediate case where the employer has to give the same contract to a team of men and women (*team-tailored* contract).

When contracts fully internalize flexibility needs (*gender-tailored* contracts), there are no systematic differences in employment dynamics by gender. In particular, this implies that the average wage gap is constant over time. We also show that men receive higher wage cuts when demanding flexibility which is consistent with the empirically documented flexibility stigma (Aaronson and French (2004)). Under *male-tailored* and *team-tailored* contracts, the wage gap gradually grows over time.

Finally, we show that men's and women's wages will not fully converge back after women's need for flexibility converges back to that of men (e.g. children are grown up). This is the case for each of the three types of contracts studied.

2.4.1 Comparative Statics in Gender-tailored Contracts

We now derive comparative statics results on how the optimal contract varies with p for a fixed continuation utility v. This allows us to understand how the optimal contract differs for a man and a woman who were promised the same compensation. The following proposition shows that for a male employee –i.e., lower p– the contract features lower hours when time-limited and lower current and future compensation for both high time availability (w^H and v^H) and low time availability (w^L and v^L).

Proposition 6. *For a fixed current promised utility, v, hours, h^L, wages, w^H and w^L, and promised utilities, v^H, and v^L, are all increasing in p. In particular, if $p_{men} < p_{women}$, men*

1. *earn smaller rewards for high time availability, i.e.*

$$w^{H,men} < w^{H,women} \text{ and } v^{H,men} < v^{H,women},$$

2. *experience larger penalties for low time availability, i.e.*

$$w^{L,men} < w^{L,women} \text{ and } v^{L,men} < v^{L,women},$$

3. *and work fewer hours after low time availability, i.e. $h^{L,men} < h^{L,women}$.*

The result is proven by considering an admissible perturbation of the optimal allocation that lowers h^L and wages (w^H and w^L) or promised utilities (v^H and v^L). Then, we show that the gains from this perturbation decrease with p, implying that in an optimal allocation hours h^L, wages and promised utilities are increasing in p for a fixed v.

The parameter p captures the employee's probability of experiencing low time availability and this is known by the employer. For example, this means that the employer understands that, on average, a young mother is more likely to ask for lower working hours than a single male employee. From the employer's perspective, a higher p increases the average costs of providing flexible working hours and increases the rewards and/or decreases the penalties required to deliver the same average compensation.

First, to understand why wages and promised utilities are higher for women, notice that a higher p increases the frequency of penalties (w^L and v^L) and decreases the frequency of rewards (w^H and v^H). Consequently, to deliver the same average compensation, the contract must offer an incentive-compatible combination of larger rewards and lower penalties. Conversely, since men have to be penalized less often, the employer can achieve the same average compensation with lower rewards and higher penalties. This result is in line with the so-called flexibility stigma (Aaronson and French, 2004; Golden, 2020): when men reduce work hours for family reasons, they are punished with higher wage cuts than women.[15]

To understand why h^L is lower for men than for women, recall from Proposition 4 that the employer provides flexibility but underworks the employee below the first best level. By doing so, the employer discourages an employee with high time availability from claiming low time availability, which, in turn, allows to provide more insurance against flexibility shocks.[16] For women, the expected direct cost (lower production) from reducing h^L is higher than for men due to the higher frequency of low time availability. Therefore, it is optimal to underwork the employee less. While this may appear counterintuitive, note that in Proposition 6, we study how optimal contracts differ by gender assuming that both were promised the same compensation level. However, in our dynamic model, the distribution of promised compensation levels is endogenous and, in particular, depends on the frequency of experiencing limited time availability.

The key insight from studying *gender-tailored* contracts follows from Proposition 4, where we showed that optimal contracts display approximately constant average wages over time. Recall that for *gender-tailored* contracts, we solve for the optimal contract twice, once for each p. This means, in particular, that if the (SUST) and (LC) constraints do not bind, an Inverse Euler equation holds for each $p \in \{p_{\text{women}}, p_{\text{men}}\}$:

$$\frac{1}{u'(w_{t-1})} = p\frac{1}{u'(w_t^L)} + (1-p)\frac{1}{u'(w_t^H)}. \tag{2.10}$$

Note that, with log utility, this implies that wages follow a martingale $w_{t-1} = pw_t^H + (1-p)w_t^L$ and averages wages would be constant over time. Therefore, if men and women entered the contract with the same initial wage, the gender gap in average wages would be approximately zero and constant over time. Intuitively, following from Proposition 6, the frequency offsets the size of the penalties. Women have lower penalties and higher rewards than men but at a higher frequency and vice versa. Hence, contracts that fully internalize the flexibility needs of each employee do not display systematic differences in wage dynamics.

[15]It also follows that h^H is decreasing in p because the hours of the time affluent employee are decreasing in wages due to income effects as shown in the optimality condition (2.4).

[16]Formally, by lowering h^L, the employer reduces information rent given to an employee with high time availability at the direct cost of reducing production. Lower information rents allow the employer to provide more insurance to employees with low time availability, which is profitable as the same v can now be delivered with lower average wages.

Finally, although men experience larger drops in wages for demanding flexibility, a direct implication of Proposition 6 is that the total utility penalty is actually larger for women. To understand this, notice that from the incentive constraints we can write:

$$\Big(U(w^H, h^H; f^H) + \beta v^H\Big) - \Big(U(w^L, h^L; f^L) + \beta v^L\Big) = (f^H - f^L)\psi(h^L).$$

Since h^L is larger for women (as shown in Propositon 6) the gap in utility between high and low time availability is also larger for women. Intuitively, because the women are less underworked with low time availability, the employer is not able to provide as much insurance against flexibility needs.

2.4.2 Male-tailored Contracts

So far we have considered optimal contracts catering to the individual's flexibility needs. In practice, it may not be feasible –legally or due to complexity reasons– for the employer to design multiple types of contracts for various employees. Our framework can be easily used to study employment dynamics with contracts designed only for one type of employee. In particular, we consider the case where both men and women are under a contract initially designed for men. We believe that this a useful polar case to study, because actual contracts may not be designed specifically to cater each employees' needs. Specifically, in male dominated working environments, women's higher need for temporal flexibility may not be taken into account.

In Section 2.4.1, we showed that there are no systematic differences in wage dynamics for *gender-tailored* contracts. Conversely, if a woman took a contract designed for men (*male-tailored* contract), she would experience the same penalty and reward structure as her male colleague, but she would still be penalized more often. By Proposition 6, she now suffers higher penalties and lower rewards than in her optimal contract. Equation (2.8) and $p_{\text{women}} > p_{\text{men}}$ imply

$$\frac{1}{u'(w_{t-1}^{\text{men}}(f^{t-1}))} > p_{\text{women}} \frac{1}{u'(w_t^{\text{men}}(f^{t-1}, f^L))} + (1 - p_{\text{women}}) \frac{1}{u'(w_t^{\text{men}}(f^{t-1}, f^H))}, \quad (2.11)$$

where superscript *men* denotes allocations under men's optimal contract. Women's wages diverge downwards under *male-tailored* contracts and the average wage gap grows gradually over time. Intuitively, women now bear both disadvantages, higher penalties/lower rewards as well as a higher chance of being time-limited. For the same reason, women's promised utility also diverges downwards under a *male-tailored* contract. This will generally result in larger termination rates for women.

Women's wage dynamics depend on the characteristics of men's contract and their probability of low time availability. The following proposition uses the Inverse Euler equation for men's contracts to characterize the growth of the gender wage gap under *male-tailored* contracts.

Proposition 7. *With log utility ($u(c) = \log(c)$), the expected growth rate of women's wages under a male-tailored contract is equal to:*

$$\mathbb{E}_{p_{women}}\left(\frac{w_t^{men}(f^t)}{w_{t-1}^{men}(f^{t-1})}\right) - 1 = \underbrace{\left(\frac{p_{women} - p_{men}}{1 - p_{men}}\right)}_{>0,\ \textit{Difference in ps}} \underbrace{\left(\frac{w_t^{men}(f^{t-1}, f^L)}{w_{t-1}^{men}(f^{t-1})} - 1\right)}_{<0,\ \textit{Men's penalty}} < 0, \tag{2.12}$$

where $\mathbb{E}_{p_{women}}$ denotes the expectation under the probability p_{women}.[17]

The growth rate of women's wages – and so of the gender wage gap– depends on two terms: the difference in the probability of low-time availability and the change in men's wages in case of low time availability. Hence, the wage gap will grow faster when the differences between p_{women} and p_{men} are larger and/or men are penalized more for low-time availability.

Finally, since continuation utilities for women under contracts designed for men do not coincide with men's continuation utility, it is not guaranteed that the contract remains incentive-compatible for women. In Appendix B.5, we provide conditions for incentive compatibility (for both types f^H and f^L) in this case and verify them numerically.

2.4.3 Team-tailored Contracts

A less polar case than the *male-tailored* contract is a scenario in which an employer hires a fraction $s \in (0,1)$ of men and a fraction $(1-s)$ of women but has to give the same contract to both genders. We refer to this intermediate case as *team-tailored* contracts. Again, we can show that the optimal contract satisfies an Inverse Euler equation but with the

[17]With the general utility function, we have a similar expression for the growth rate of the inverse marginal utilities:

$$\mathbb{E}_{p_{\text{women}}}\left(\frac{u'(w_{t-1}^{\text{men}}(f^{t-1}))}{u'(w_t^{\text{men}}(f^t))}\right) - 1 = \left(\frac{p_{\text{women}} - p_{\text{men}}}{1 - p_{\text{men}}}\right)\left(\frac{u'(w_{t-1}^{\text{men}}(f^{t-1}))}{u'(w_t^{\text{men}}(f^{t-1}, f^L))} - 1\right).$$

(history-contingent) average probability of low time availability. To this end, we denote by $P^{\text{men}}(f^t)$ and $P^{\text{women}}(f^t)$ the probability measure over histories induced by the probabilities of low time availability p_{men} and p_{women}, respectively.

Proposition 8. *If an incentive-compatible team-tailored contract exists, at every history f^t, it satisfies the following Inverse Euler equation:*[18]

$$\frac{1}{u'(w_{t-1}^{avg}(f^t))} = p^{avg}(f^t)\frac{1}{u'(w_t^{avg,L}(f^t,f^L))} + (1-p^{avg}(f^t))\frac{1}{u'(w_t^{avg,H}(f^t,f^H))},$$

where

$$p^{avg}(f^t) \equiv \frac{sP^{men}(f^t)p_{men} + (1-s)P^{women}(f^t)p_{women}}{sP^{men}(f^t) + (1-s)P^{women}(f^t)},$$

and $p_{men} < p_{avg}(f^t) < p_{women}$ for all f^t.

Because $p_{\text{men}} < p_{\text{avg}}(f^t) < p_{\text{women}}$, in a *team-tailored* contract, men's wages always diverge upwards and women's wages downwards. If $p^{\text{avg}}(f^t)$ is close to p_{men}, i.e. when s is close to one and/or in histories f^t that are more likely for men, the contract approaches the *male-tailored* contract. Therefore, men's wages will be approximately constant over time, and women's wages diverge downwards faster. Conversely, if $p_{\text{avg}}(f^t)$ is close to p_{women}, women's wages are approximately constant, and men's wages diverge upwards faster.

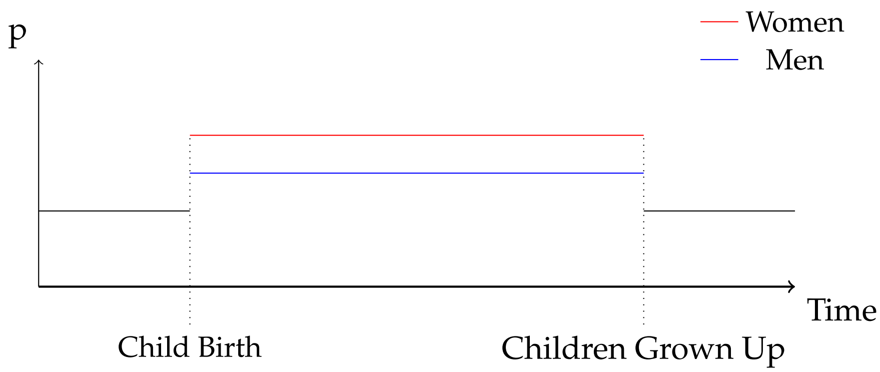
We can generally not show that an incentive-compatible *team-tailored* contract exists. However, if the *male-tailored* contract studied in the previous section is incentive-compatible for both men and women, there must also exist a team-tailored contract as the employer can always use the *male-tailored*.

2.4.4 Non-convergence of Wages

Goldin *et al.* (2022) document that women's wages do not converge back to the level of men after children grow up. In our model, we can interpret the children growing up as p_{women} converging to p_{men}. Figure 2.3 simplifies how we think our model fits into the life-cycle. Before childbirth, men and women have a similar external demand on time. After

[18] In the proof of the proposition in Appendix B.2 we also lay out the employer's problem. Solving for the optimal contract is challenging because the continuation utilities of men and women do not coincide, and, a priori, we do not know which incentive constraints bind. However, we can characterize the Inverse Euler equation in the proposition with a variational argument because uniform changes in utilities preserve the incentive and participation constraints of men and women.

Figure 2.3: Differences in p over time by gender.



childbirth but before children grow up, women are more exposed to flexibility needs than men so $p_{\text{women}} > p_{\text{men}}$. Once children have grown up, men and women have a similar probability of limited time availability again.[19]

In this case, the wage dynamics in our model are consistent with the documented non-convergence of wages. Under *gender-tailored*, *male-tailored* as well as *team-tailored contracts*, the wages of women who have been penalized do not converge back to the level of men.

First, consider the *male-tailored* contract. If p_{women} converges to p_{men}, women's wage process converges to the same process as men's. By Proposition 4, inverse marginal utilities then follow a martingale, so average wages are approximately constant over time. Therefore, the wages would stop diverging downwards, but they do not converge to the same level as men's. The same logic applies for *team-tailored* contracts.

For *gender-tailored* contracts, Appendix B.6.3 studies an extension in which we allow the probability p to follow a time-varying and possibly stochastic process. Because p is observable, the employee is perfectly insured against changes in p. In particular, this implies that an employee who was penalized with low wages will not experience a sudden increase in wages if p increases.

2.5 Numerical Simulations

To connect our theoretical results with empirical regularities on gendered employment dynamics we solve the model numerically. We use the American Time Use Survey (ATUS) to calibrate the probabilities of being time limited by gender. We use these values to

[19]There may also be gender differentials before and after childbirth, e.g., caring for elderly parents.

understand a reasonable range of values that our model is able to generate with respect to wage gaps and exit rates.

2.5.1 Parametrization

We use the following isoelastic parametrization of the utility function

$$U(w,h;f) = \frac{w^{1-\sigma}}{1-\sigma} - (1-f)\frac{h^{1+\eta}}{1+\eta}, \tag{2.13}$$

where we set $\sigma = 2$ and $\eta = 2$. The Frisch elasticity, $1/\eta$, is thus equal to 0.5 following Chetty *et al.* (2011). For the numerical exercise, we consider a period to be equal to a quarter. Hence, we set the discount factor equal to $\beta = 0.987$. We believe this to be the best compromise between the high frequency of interruptions during a working week and the low frequency of wage adjustments observed in employment contracts.[20] We parameterize the production by

$$g(h) = \frac{h^{\alpha}}{\alpha},$$

and set $\alpha = 0.7$, which is the value commonly used to match the labor share.[21] In our model, α captures how costly it is for the employer to allow for volatility in the employee's working hours. We would thus expect this parameter to vary across occupations. If $\alpha = 1$, the production function is linear, and the employer only cares about expected hours. As α decreases, the employer prefers smoother working hours as the cost of providing flexibility increases. Qualitatively, our results do not rely on the specific value of α, except that decreasing returns ($\alpha < 1$) are needed to generate termination.

2.5.2 Identifying p, f^H and f^L

Our main parameter of interest is the probability of having low time availability p. This is the only parameter we use to differentiate men and women. Thus, we need to determine p_{men} and p_{women}. To identify meaningful values we turn to the American Time Use Survey (ATUS). The ATUS is a nationally representative U.S. time diary survey with detailed information on how many minutes at a certain time of the day respondents spent on different activities including work, care and leisure. We restrict our sample to full-time

[20]More generally, we may think that wages are adjusted at a lower frequency (e.g. yearly) but the employee can adjust its borrowing or saving in anticipation of future pay raises or wage cuts.

[21]For robustness, we do comparative statics on the value of α for our main result, see Figure B.3 in Appendix B.1.

workers (excluding self-employed), aged between 20 and 65, and having at least one child below the age 12. We gather information on when and how many minutes a respondent spent on care activities, as well as the typical working hours in their current job.

Our goal is to determine gendered values for p that align with the reduction in hours implied by f^L. In the data, care activities are given in minutes, but in our model time availability is binary. To map the data to the model, we need to find a threshold of minutes spent on care to define an interruption. First, we construct the *Care-Work-Ratio*, which indicates the share of minutes of care activities during usual working time:

$$\text{Care-Work-Ratio} = \frac{\text{Minutes of Care Activities between 9am and 5pm}}{\text{Usual Minutes of Work}}.$$

We put the usual working minutes in the denominator so as not to inflate the care-work ratio by reducing the working time due to care activities. For more details on the sample selection and robustness of the following results, see Appendix B.4.

Second, we define the cutoff X such that, among all individuals with a Care-Work-Ratios above X, the average interruption comprises 25% of the working day:

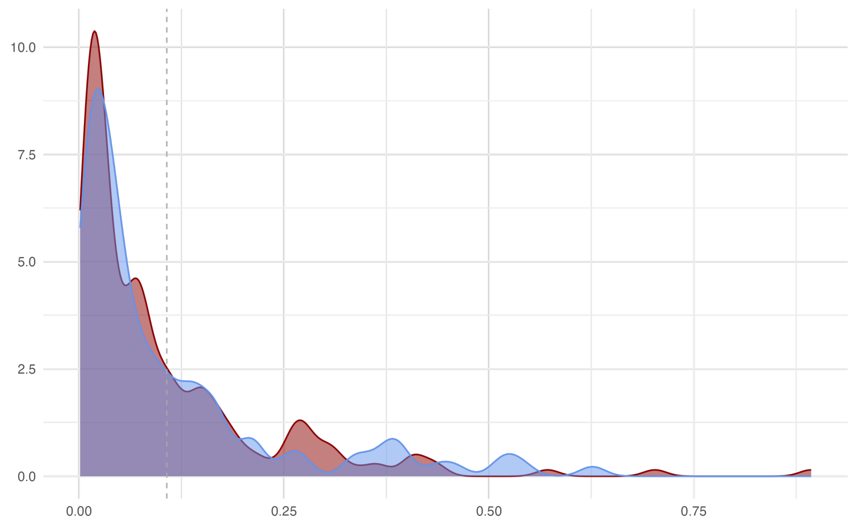
$$\mathbb{E}\left(\text{Care-Work-Ratio} \mid \text{Care-Work-Ratio} \geq X\right) = 0.25.$$

We do so to then choose f^L accordingly such that, in a fixed per-hour wage contract, an employee would reduce work time by 25%. We find a cutoff of X close to 11%. In our sample, 18% of men have a positive care-work ratio compared to 43% of women. The average care-work ratio in the whole sample is 2% for men and 4% for women. However, conditional on being positive, the average care-work ratio is similar: 11% for men and 10% for women. Figure 2.4 shows the distribution of all positive care-work ratios by gender. Figure 2.4 highlights that conditional on being positive, the distribution of care-work ratios for men and women is very similar. The size of most care interruptions during working hours is small, i.e. close to zero and below the cutoff X we find.

Hence, when we then translate the continuous minutes of care work during working hours into binary interruptions, we classify all care activities whose share of the working day is above 11% as an interruption. The probability of being hit with low time availability is then simply:

$$p_{\text{men}} = \frac{\text{Number of men with Care-Work-Ratio} \geq X}{\text{Number of men}}$$

Figure 2.4: Distribution of positive Care-Work-Ratios by gender



Note: Dashed line indicates cutoff X. Women in red, men in blue.

and

$$p_{\text{women}} = \frac{\text{Number of women with Care-Work-Ratio} \geq X}{\text{Number of women}}.$$

Using this approach we obtain $p_{\text{men}} = 0.06$ and $p_{\text{women}} = 0.15$. This means that while men have a 6% chance of having low time availability, women experience low time availability with a probability of 15%. Women are more than twice as likely to experience low time availability than men.

The cutoff X is arbitrary in the sense that we choose the size of the average interruption. In Table B.1 in Appendix B.1, we present results for different cutoffs. As expected p_{men} and p_{women} increase when we reduce the average reduction in hours to 20% and decrease when we increase the average reduction in hours to 30%.

Moreover, in Table 2.1, we show further probabilities of limited time availability by gender across socioeconomic groups. While p_{men} remains fairly stable across groups, p_{women} varies substantially but always remains larger than p_{men}. Women without a college degree, below median family or weekly income are the most likely to be interrupted during working hours.

We are not the first to use the ATUS to document gender differences in external demands on time during working hours. Both Cubas *et al.* (2021) and Schoonbroodt (2018) analyze the effects on wages of parental childcare during working hours. In comparison to them, we document relatively small incidences of limited time availability, suggest-

Table 2.1: Probabilities of low time availability across different socioeconomic groups.

	p_{men}	p_{women}
Baseline	0.06 (0.012)	0.15 (0.022)
Non-College	0.07 (0.017)	0.16 (0.033)
College	0.06 (0.015)	0.11 (0.025)
Below Median Family Income	0.06 (0.020)	0.17 (0.041)
Above Median Family Income	0.05 (0.001)	0.11 (0.023)
Below Median Weekly Earnings	0.10 (0.027)	0.18 (0.031)
Above Median Weekly Earnings	0.05 (0.012)	0.09 (0.027)

Note: Standard errors in parenthesis.

ing that our results can be seen as a lower bound. Moreover our results are not directly comparable, since we map (continuous) minutes of care activities into our binary framework of high/low time availability. The most precise measure of differences in external demand on time can be found in Buzard *et al.* (2023). They conduct an experiment to measure how frequently schools contact mothers compared to fathers. They find that mothers are contacted 1.4 times more often. Since this addresses external time demands in a narrow context, their results can be considered a conservative estimate, as Buzard *et al.* (2023) argue.

To determine values for f^H and f^L, we start by normalizing the disutility of working when having high time availability such that the hours of the time affluent employee in the first best would equal 1/2.[22] The calibration of this parameter depends on p; using $p = p_{men}$ for both genders, we get $f^H = -5.2$. The value of f^L deserves some attention, as it determines how much a flexibility shock affects the disutility of working. We calibrate it by assuming that after a flexibility shock, if the employee was offered a fixed per-hour wage contract, she would want to work 25% less hours. Therefore, the size of an interruption in the model coincides with the one we used in the data to calibrate p. Under this assumption we obtain a value of $f^L = 1 - (4/3)^{\sigma+\eta}(1 - f^H) = -18.7$.[23]

[22]Note that in the first best, we have $v_0 = \frac{1}{1-\beta}\left(u(w^{FB}) - (1-p)(1-f^H)\psi(h^{H,FB}) - p(1-f^L)\psi(h^{L,FB})\right)$, which combined with (2.4) allows us to solve for the wage and hours.

[23]To derive this, notice that under a wage per-hour contract, for any $f \in \{f^L, f^H\}$, we have $w(wh)^{-\sigma} = (1-f)h^{\eta}$. So the ratio of hours is $\frac{h^H}{h^L} = \left(\frac{1-f^L}{1-f^H}\right)^{\frac{1}{\sigma+\eta}}$, then setting $\frac{h^H}{h^L} = \frac{4}{3}$ and solving for f^L we derive the

2.5.3 Solution Method

We need to deal with two technical challenges to solve our dynamic contracting problem numerically. First, common to all recursive contracting problems, because the constraints are forward-looking, the transversality conditions may not hold. Marcet and Marimon (2019) provide a recursive Lagrangian formulation to solve this problem. However, a direct application of this method is known to fail when the Pareto frontier is not strictly concave (Cole and Kubler (2012)), which is the case in our model as shown in Section 2.3.5.

Our approach consists of using two different methods at different parts of the state space. From Section 2.3.5, we know that the constrained Pareto frontier Π^c is only concave when v is close enough to $\overline{v}$. Thus, for $v > v_{min}^{MM}$ with v_{min}^{MM} large enough, we first solve the model with the recursive Lagrangian following (Marcet and Marimon (2019)), which guarantees that the transversality condition holds. Then, for $v \in [\overline{v}, v_{min}^{MM}]$ we solve the model using a direct promised utility approach. A more detailed description of the solution method and algorithm can be found in Appendix B.3.

2.5.4 Outside Options

Finally, we also need to set values for the outside options of the employer and the employee, which, for the numerical exercise, we allow to be stochastic, as we explain next.

Stochastic Employee's Outside Option. To have a positive termination probability, the continuation value v must eventually reach the region with ex-post inefficiencies. Hence, we need $v^L < v$ at all points outside the region with positive termination probability. When the (SUST) constraint binds, a high cost of reaching the inefficient region can render $v^L > v$ optimal.[24] In our parametrization, with both p_{men} and p_{women}, we get $v^L > v$ as v approaches the termination region, so there is no exit in the optimal contract.

To circumvent this issue, we extend the model by letting the employee's outside option be stochastic. That is, we assume $\overline{v}$ can take values on a grid $\{\overline{v}_1, .., \overline{v}_i, ..., \overline{v}_I\}$ with corresponding probabilities $\{\overline{p}_i\}_{i=1}^{I}$. However, we maintain the assumption that this outside option is observable by the employer. Intuitively, this stochasticity smooths the employer's cost of lowering v^L as the (SUST) constraint may not bind at $t+1$ when $\overline{v}_i$ is

expression above.

[24] Dovis (2019) derives conditions to have $v^L < v$ everywhere, but we find numerically that they typically are not sufficient in our model.

small. More details on the employer's problem with the stochastic outside option can be found in Appendix B.6.1.

Employer's Outside Options. Next, we need to assign a value to the employer's outside options in a contract with a man ($\overline{\Pi}^{\text{men}}$) and a woman ($\overline{\Pi}^{\text{women}}$). For every v, the employer value in a contract with a man $Q^{\text{men}}(v) \equiv \sum_{i=1}^{I} \overline{p}_i \Pi^{\text{men}}(v, \overline{v}_i)$ is generally much larger than $Q^{\text{women}}(v)$. Hence, if we set a common outside option $\overline{\Pi}^{\text{men}} = \overline{\Pi}^{\text{women}}$, the contract will typically deliver either no termination for men or termination at all v for women. For this reason, we pin down the outside options based on the highest value that the employer can attain by replacing the current employee with another one of the same gender.[25] We approximate the value of replacing the employee with the value functions in the constrained problem where the employer cannot terminate Π^c defined in Section 2.3.5.[26] That is, we set $\overline{\Pi}^{\text{men}} = \max_v \sum_{i=1}^{I} \overline{p}_i \Pi^{c,\text{men}}(v, \overline{v}_i)$ and $\overline{\Pi}^{\text{women}} = \max_v \sum_{i=1}^{I} \overline{p}_i \Pi^{c,\text{women}}(v, \overline{v}_i)$.

2.6 Numerical Results

Solving for the optimal contract numerically allows us to get a better picture on how gendered labor market outcomes arise from our theoretical framework of hidden demand for flexibility. We first consider wage dynamics and gender differences in penalties. Then, we assess differences in employment dynamics captured by job duration and termination probabilities. Comparing *gender-tailored* and *male-tailored* contracts allows us to emphasize the importance of employment contracts that account for differences in flexibility needs.

2.6.1 Wage Dynamics

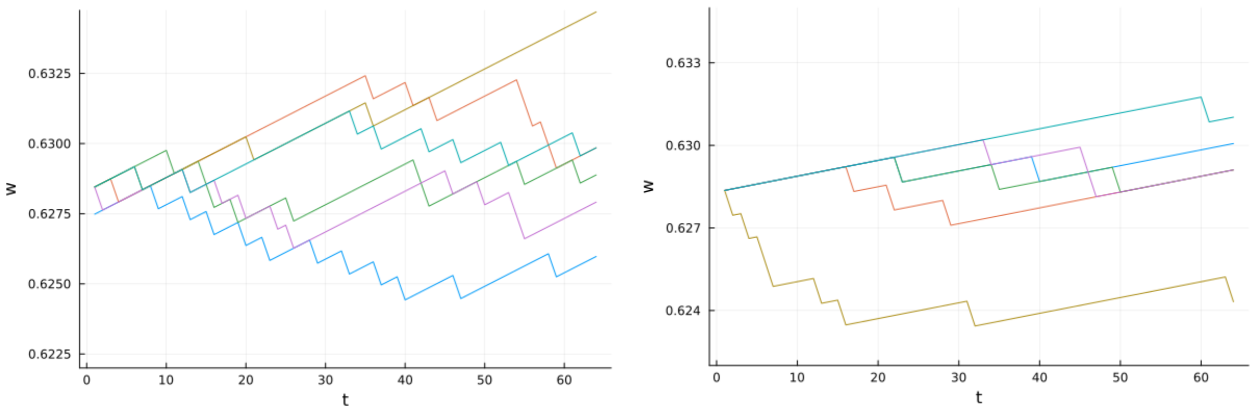
Gender-tailored Contracts. Figure 2.5 shows five (randomly selected) sample paths of wages for women (left panel) and men (right panel) under *gender-tailored* contracts over 64 periods. To focus on the gender differences in wage dynamics, in this figure and throughout the section we set v_0^{men} and v_0^{women} such that men and women have the same average initial wages. Moreover, we solve the model at a v_0 far away from the termination region to focus on the effects of the private information friction. Each colored line represents

[25] This is a conservative value. If we assume, for example, that the employer may sometimes also hire an employee of the other gender, we would get higher termination probabilities for women.

[26] Otherwise, finding the outside options would require us to solve the fixed point problem $\overline{\Pi}^{g} = \max_v \sum_{i=1}^{I} \overline{p}_i \Pi^{c,g}(v, \overline{v}_i; \overline{\Pi}^{g})$ for each $g \in \{\text{men}, \text{women}\}$.

Figure 2.5: Wage paths gender-tailored contracts

(a) Wage Paths for Women (b) Wage Paths for Men



Note: Each colored line represents a randomly selected sampled path for women (left panel) and men (right panel).

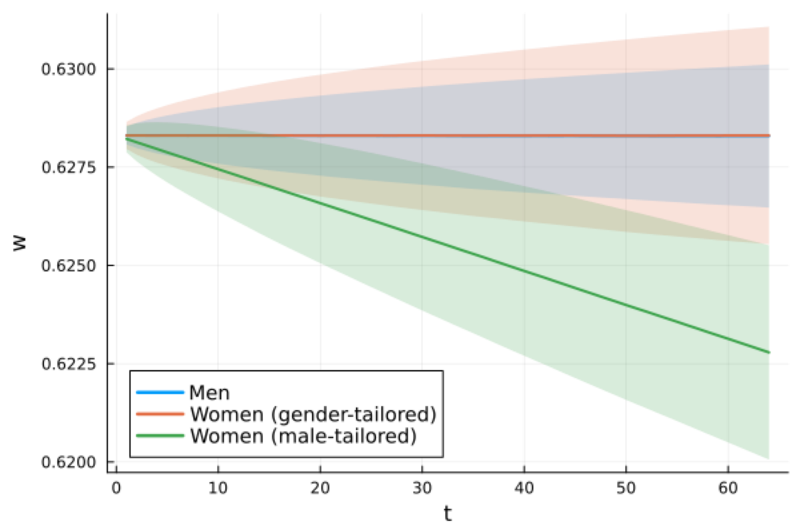
the wage path of an individual employee characterized by their respective sequence of time availability. Since, in our framework, gender is not mechanically encoded in the employee's preferences, both men and women could experience long sequences of low time availability. However, it is more likely for a woman to have such an extended series of low time availability. Hence, it is not the difference in p directly that the employer penalizes but the realizations of low time availability.

We observe that women suffer frequent and small wage penalties ($w^L_{\text{women}}/w_{t-1} - 1 = -0.11\%$), whereas men's wages feature slightly larger ($w^L_{\text{men}}/w_{t-1} - 1 = -0.12\%$) but less frequent penalties. In contrast, women experience larger rewards ($w^H_{\text{women}}/w_{t-1} - 1 = 0.019\%$) than men ($w^H_{\text{men}}/w_{t-1} - 1 = 0.007\%$), but less frequently.

Male-tailored Contracts. Figure 2.6 shows the average wage dynamics of men and women under *gender-tailored* and *male-tailored* contracts. The line indicates the average wage and the shaded area one standard deviation along the cross-section in each period. In blue (men) and red (women) we show wage dynamics under *gender-tailored* contracts, i.e with p_{women} and p_{men}. Average wages of men and women are the same – the red and the blue lines overlap – but women's wages feature an increasingly larger variance over time. For women under *male-tailored* contracts, the gender wage gap widens. Average wages between men and women gradually diverge, accompanied by an increasing volatility. After 16 years (64 periods), the wage gap grows to be 0.88%.

Given our normalization of a 25% reduction in working hours from an interruption, it is important to check the sensitivity of the resulting wage gap to this value. We recalibrate the parameters f^L, p^{women} and p^{men} by setting an interruption to involve 20% and 30% reductions in working hours (see Table B.1 in Appendix B.1). We find that the ratio between p^{women} and p^{men} is larger with the 20% reduction, leading to a wage gap of 1.12%. However, this ratio is smaller with the 30% reduction, where we obtain a wage gap of 0.64%.

Figure 2.6: Average wage dynamics



Note: The lines denote the average wages along the cross-section each period and the shaded areas one standard deviation.

To put these numbers in perspective, note that in Figure 2.6, we are comparing the wage trajectory of equally productive men and women in the same occupation and firm that started with the same wages and that made no adjustments to their initial working arrangement. Moreover, by setting the same initial compensation the employer fully internalizes the higher cost of childcare responsibilities of women compared to men. In Section 2.7, we discuss how our mechanism explains a sizeable amount of the corresponding gender wage gap the empirical literature has found.

Figure 3.2 shows the distribution of wages after 16 years (64 periods), with men's wages displayed in blue and women's in red. The left panel shows the wage distribution for *gender-tailored* contracts. There is a higher mass of women with lower wages corresponding to female employees with a longer sequence of low time availability. However, we also see a larger mass of women with higher wages than men, resulting from the "lucky" women who suffered almost no low time availability and benefited from the higher rewards.

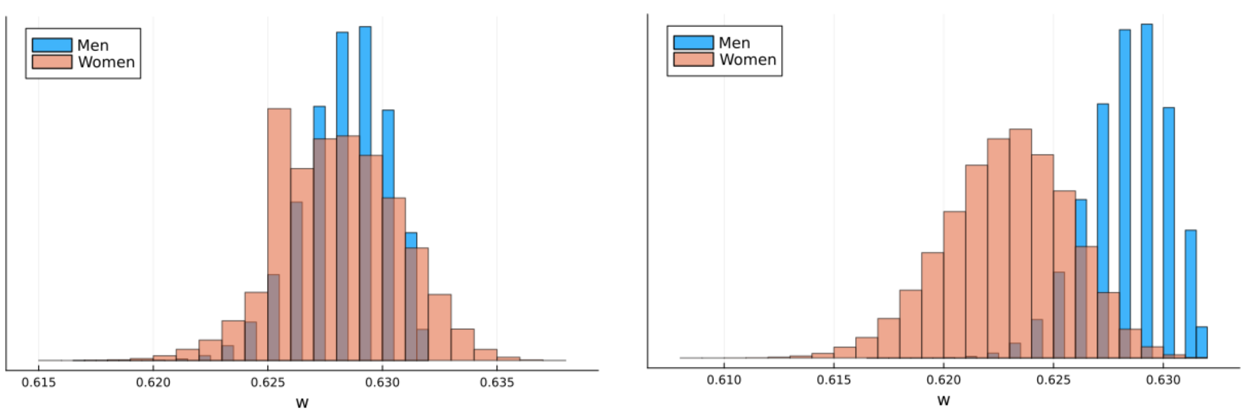
On the right panel, we compare the wage distributions of men with women under *male-tailored* contracts. Women's wage distribution is shifted to the left. Conversely, the distribution of men's wages is more skewed and shifted to the right. While men and women now experience the same penalty and reward structure, women get penalized more frequently than men.

Finally, we may also consider the case where men are under a women-tailored contract. In this scenario, men would be exposed to less frequent and smaller penalties, so their wages would diverge upwards (see Figure B.1 in Appendix B.1). However, this result must be interpreted with caution because, as shown in Appendix B.5, the contract may not remain incentive-compatible.

Figure 2.7: Wage Distributions (at $t = 64$)

(a) Gender-tailored contract

(b) Male-tailored contract



Comparison across Socioeconomic Groups. Table 2.2 shows the average growth rate of women's wages as well as the average wage gap after 16 years across different socioeconomic groups. Differences in p between men and women across groups directly translate into differences in average growth rates and wage gaps. Our model predicts the smallest wage gap between men and women who earn more than the median weekly income, while the largest can be found among those whose family income is below the median. The wage gap depends on both men's and women's probability of limited time availability within each group. As discussed in Section 2.4.2, the average growth rate of women's wages under *male-tailored* contracts depends on the size of the penalties for men and the ratio of p_{men} and p_{women}. Compare, for example, college-educated and respondents with family income above the median: while p_{women} is the same, p_{men} is one percentage point higher for college-educated men. This relative difference translates into sizeable differences in wages after 16 years.

Table 2.2: Comparison of the wage gaps in male-tailored contracts across socioeconomic groups

	p_{men}	p_{women}	Avg. growth rate of womens' wages	Avg. wage gap after 16 years
Baseline	0.06	0.15	-0.014%	0.88%
Non-college	0.07	0.16	-0.014%	0.88%
College	0.06	0.11	-0.008%	0.49%
Below median weekly earnings	0.10	0.18	-0.012%	0.78%
Above median weekly earnings	0.05	0.09	-0.006%	0.39%
Below median family income	0.06	0.17	-0.018%	1.07%
Above median family income	0.05	0.11	-0.009%	0.58%

2.6.2 Employment Dynamics

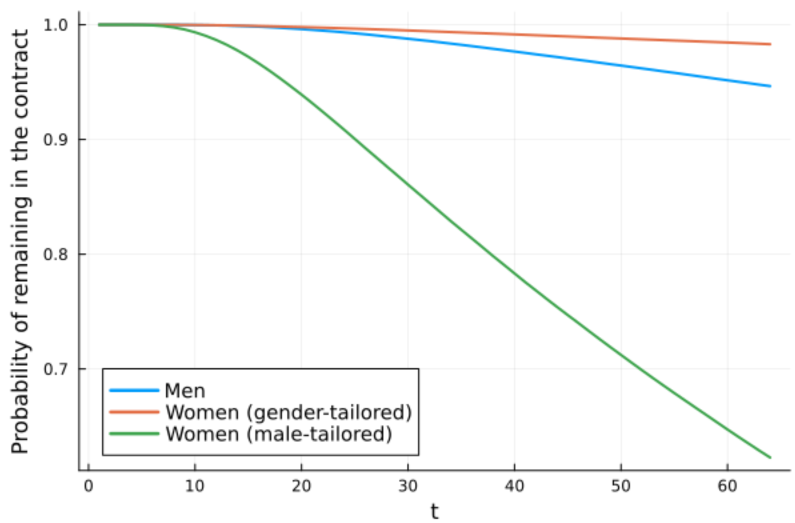
Figure 2.8 plots the probability of remaining on the contract over time for men and women under the *gender-tailored* contract and for women under the *male-tailored* contract. We set the same starting promised utility for men and women. Under *gender-tailored* contracts, the termination rates are slightly higher for men despite a higher mass of women being penalized. The intuition is the following. As compensation levels get pushed down, the contract requires that h^L goes to zero, which is costly for the employer. For men, on average, this cost is lower due to the lower probability of limited time availability. Hence, the cost of decreasing future compensation v^L as the contract approaches the termination region is lower for men. As a result, the penalized men end up reaching the termination region faster than women.

Under the *male-tailored* contract, similar to the wage dynamics, low time availability is now penalized with even lower future compensation pushing the contract faster to the termination region. Combined with the higher frequency of low time availability, this leads to larger termination rates for women.

After 16 years (64 periods), more than a third of women would have exited the contract, while under the *gender-tailored* contract, the share would be less than 2%. The share of women who terminate the contract is more than six times larger than for men. To put this result into context, note again that our model only captures ending employment relations due to the incompatibility of unpredictable flexibility needs and the current work arrangement. There are many other reasons why job duration rates vary by gender. For example, men are more likely to switch to other jobs for opportunity than for flexibility reasons, which is outside the scope of our model. In Section 2.7, we discuss to what extent our termination rates might even be smaller than their empirical counterpart.

Notice also that the termination probabilities are zero for some initial periods as it takes some time for the employees to reach the inefficient region. Whether termination is highest in the initial or later periods depends on how close the starting promised utility v_0 is to the outside option (see Figure B.2 in Appendix B.1). However, regardless of the chosen starting value, it is always the case that the termination rates of men and women are similar under the *gender-tailored* contract but much larger for women under the *male-tailored* contract.

Figure 2.8: Termination dynamics



Note: Each line represents the fraction of men or women that remain in the contract at every period.

2.7 Recovering Stylized Facts

Our model simultaneously accounts for both wage divergence and non-convergence, as well as gender differences in labor force attachment. In this section, we contextualize our numerical results with empirical estimates and summarize the key mechanisms of our model.

2.7.1 Women's wages start to diverge from men's after childbirth but do not fully converge after children grow up

The bulk of the gender wage gap can be traced back to the child penalty. After childbirth, men and women's wages diverge even if they were on the same career paths before (Kleven *et al.*, 2019; Cortés and Pan, 2023, Barth *et al.*, 2021).

Divergence. The empirical literature has found various estimates for the child penalty, comparing mothers to non-mothers (Yu and Hara, 2021), women to men (Blau and Kahn, 2017), mothers to fathers (Yu and Hara, 2021), and within couples (Angelov *et al.*, 2016), revealing a range of wage gaps between 10% to more than 30%. Different studies emphasize different drivers of this penalty, but they consistently identify the most important factors to be reductions in hours (Angelov *et al.*, 2016; Adda *et al.*, 2017; Goldin, 2014; Kleven *et al.*, 2019; Wasserman, 2023; Erosa *et al.*, 2016), workforce interruptions (Adda *et al.*, 2017; Bertrand *et al.*, 2010), pre- and post-selection into more flexible and lower-paying jobs (Blau and Kahn, 2017; Adda *et al.*, 2017, Yu and Hara, 2021; Goldin *et al.*, 2017; Morchio and Moser, 2024; Felfe, 2012; Wiswall and Zafar, 2018; Card *et al.*, 2016) or leaving the labor force (Kleven *et al.*, 2019; Harkness *et al.*, 2019; Adda *et al.*, 2017). Our focus, however, lies in understanding how women fare in the same job compared to male coworkers and how, over time, they get pushed out of an employment relation and into an ex-ante inefficient outside option.

We thus study wage trajectories of equally productive men and women in the same occupation and firm who started with the same wages and made no adjustments to their initial working arrangement. When ps are different, under *male-tailored* and *team-tailored* contracts, women's average wages gradually diverge downwards. In work arrangements that are designed with men's flexibility needs in mind, women bear both the higher penalties as well as the higher frequency of being time limited.

The closest empirical counterpart to the gender wage gap we are considering is thus within-firm, within-occupation differences. Here, estimates of the average wage gap range from 2% to 4% (Budig and England, 2001; Yu and Hara, 2021; Morchio and Moser, 2024; Felfe, 2012; Lucifora *et al.*, 2021; Bronson and Thoursie, 2019). The main driver of this gap is predominantly forgone promotion opportunities. In our main calibration, we find a wage gap of 0.88% after 16 years, which is sizeable for the narrow context we are considering. In particular, this suggests that our mechanism can explain between 20% and 40% of the gap. This result does not take into account the wage cuts of women that end the employment relation. Our model can capture this too. Women's wages also diverge because some women - those with a sufficiently long sequence of limited time - are being pushed out of their current work employment. In particular, they leave for an outside option that has a lower continuation value than their current employment. While we are agnostic about what the outside option entails, we can interpret this as any type of lower wage work arrangement. Hence, the outside option could entail dropping out of the labor force, more flexible work arrangements or even self-employment. The literature has focused on modeling the wage gap across occupations as a result of an optimal trade-off between wages and flexibility/amenities (Goldin (2014), Morchio and Moser (2024)). By contrast, our analysis shows how women can be ex-post pushed into lower-paying arrangements.

Non-convergence. Having children has long-lasting effects. Even when children are grown up women's wages do not fully converge back (Goldin *et al.*, 2022; Angelov *et al.*, 2016).

Our model is able to capture the non-convergence of wages after women's flexibility needs converge back to men's.[27] As discussed in Section 2.4.4, because the penalties for demanding flexibility are permanent, the wages of women who experienced low time availability more often are permanently depressed and do not converge. This is true under any of the three types of contracts studied. One could interpret these long-lasting effects on wages as forgone pay raise opportunities. An example of this could be promotions, which are the predominant drivers of within-firm differences in wage trajectories (Lucifora *et al.*, 2021; Bronson and Thoursie, 2019).

2.7.2 Women's job duration is shorter and labor force attachment is weaker

Women have been shown to have both shorter job duration and weaker labor force attachment than men, often stemming from child and elderly care responsibilities (Hall, 1982; Molloy *et al.*, 2020; Cortés and Pan, 2023; Lundborg *et al.*, 2017). This is reflected in women being more likely to stay at home, find a new family-friendly job (Mas and Pallais, 2020; Wiswall and Zafar, 2018; Aaronson *et al.*, 2021), or pursue flexibility-oriented self-employment (Bento *et al.*, 2021; Gurley-Calvez *et al.*, 2009). Unmet needs for workplace

[27] In our model, having a child can be interpreted as a change in external demands on time and, therefore, differences in p. As discussed in Section 2.4.4, we can also interpret the children growing up as p_{women} converging to p_{men}.

flexibility push women into less profitable work arrangements, including home production (Patrick *et al.*, 2016). Moreover, in male-dominated working environments, women's termination rates are relatively higher (Torre, 2017; Cha, 2013).

Studies show a wide range in the employment rates of new mothers, from 20.9% to 70%, depending on e.g. the age of the child, the occupation or the mother's education level (Felfe, 2012; Harkness *et al.*, 2019; Erosa *et al.*, 2016; Bertrand *et al.*, 2010). Moreover, while many women choose their occupations based on anticipated childcare needs, almost all change or adjust jobs (through hours) after returning to the labor force post-maternity leave (Felfe, 2012; Adda *et al.*, 2017, Bertrand *et al.*, 2010; Hotz *et al.*, 2018).

In our model, we find that after 16 years over a third of women have exited their current employment relation. Since we are agnostic about the outside option it can entail both leaving the labor force or finding a new job. Taken together, while sizeable, our result may be in the lower end of the empirical estimates.

Our model highlights how working conditions that do not internalize differences in external demands on time can push women out of an employment relationship. As highlighted in Figure 2.8, when men and women are given contracts initially designed for them, both termination rates are similar. However, when contracts do not internalize differences in flexibility needs (*male-tailored*), a larger share of women exit the contract. Moreover, we are agnostic about the meaning of the outside option and keep it the same for both men and women. This might be a conservative choice as the value of outside options, such as home production, could significantly vary by gender.

2.8 Concluding Remarks

This paper proposes a theoretical framework to explain how gender differences in external demands on time can drive gendered employment dynamics. At the core of our theory is modeling the unpredictability and unverifiability of flexibility needs and studying its consequences in dynamic employment relationships. To study gender differences, we allow men and women to differ only in their probability of having limited time availability, which we find to be more than twice as large for women using survey data. This allows for between gender similarities and within gender differences in working conditions over time depending on their exposure to flexibility shocks. When contracts do not internalize women's flexibility needs (*male-tailored*), contractual frictions give rise to meaningful gendered labor market outcomes, including the divergence and non-convergence of gender earnings differentials over the life-cycle and women's shorter job duration and weaker labor force attachment. However, when contracts internalize women's flexibility needs (*gender-tailored*), there are no systematic gender differences in labor market outcomes.

The framework we propose is a useful starting point for understanding how gender differences in external demands (e.g., for parental involvement), can have permanent consequences on employees' working conditions. Inflexible workplaces can drive women out of breadwinner roles and simultaneously men out of caregiver roles, solely as a result of how unmet flexibility needs are penalized. Optimal contracting models like ours are also well suited to study policy implications as employment relations can endogenously respond to them.

Finally, the contrast between *male-tailored* and *gender-tailored* contracts illustrates how non-discriminatory contracts can have adverse consequences. In our model, gender differences in labor market outcomes result from the employer's inability to tailor contracts to each gender's flexibility needs. Well-intended policies that aim to reduce gender differences by imposing gender-neutral rules or contracts may have unintended consequences (see, for example, Antecol *et al.* (2018) on the effects of gender-neutral tenure clock extensions).

Chapter 3

Firm Size and Compensation Dynamics with Risk Aversion and Persistent Private Information

Gerard Maideu-Morera[1]

Abstract

I study a dynamic cash flow diversion model between a risk neutral lender and a risk averse entrepreneur who has persistent private information about the firm's productivity. In the optimal contract, the firm's size is always distorted downwards and its distortions inherit the autoregressive properties of the type process. The entrepreneur's compensation is smoothed and decoupled from the firm size dynamics. These results contrast those of equivalent models with risk neutrality. I use numerical simulations to study a quasi-implementation with simpler contracts, which highlights that this class of models is unable to generate realistic firm size and equity share dynamics simultaneously.

[1] I am indebted to Christian Hellwig for his advice and guidance throughout the project. I have benefited very much from advice and discussions from Charles Brendon, Fabrice Collard, Alessandro Pavan and, especially, Nicolas Werquin. I would also like to thank George-Marios Angeletos, Matteo Broso, Eugenia Gonzalez-Aguado, Johannes Hörner, David Martimort, B. Ravikumar, Jean-Charles Rochet, Andreas Schaab, Stéphane Villeneuve and seminar participants at the TSE Macro Workshop, EEA 2022, NSEF PhD Workshop 2022, EWMES 2022, ESEM 2023, Northwestern Macro Lunch Seminar and Midwest Macro Fall 2023.

3.1 Introduction

Financing constraints slow down growth over the lifecycle of a firm. Dynamic contracting models have proved useful in understanding the underlying agency frictions that generate financing constraints and can account for several stylized facts on the lifecycle of firms. The canonical setting in this literature is the dynamic contracting model with cash flow diversion: at each period, an entrepreneur needs funds from a lender to operate a project, but only the entrepreneur observes the project's returns (i.e. cash flows) and can secretly divert them for consumption. These contracting problems have typically been analyzed assuming a risk-neutral entrepreneur and i.i.d productivity shocks (Clementi and Hopenhayn, 2006; DeMarzo and Sannikov, 2006; DeMarzo and Fishman, 2007; Biais *et al.*, 2007).[2] Under these assumptions, in the optimal contract, the firm size (i.e. working capital invested) drifts upwards, and the entrepreneur is compensated once the undistorted first best size is reached. Notably, the optimal contract generates a one-to-one link between firm size and compensation dynamics, and implementations typically link the entrepreneur's promised compensation to her equity share in the firm (Clementi and Hopenhayn, 2006).

In this paper, I revisit the predictions of the optimal contracting solution of dynamic cash flow diversion models. I depart from the previous literature by allowing the entrepreneur to be risk averse and to have persistent private information about the firm's productivity.[3] Risk aversion creates a consumption smoothing motive that makes it costly to backload the entrepreneur's compensation. Persistence allows the entrepreneur to have more information about the firm's future profitability, thus making her preferences for future contract arrangements depend on the current profitability.

I show that introducing risk aversion and persistent private information fundamentally changes the nature of the optimal contract and the resulting firm size and compensation dynamics. First, the firm size and compensation dynamics are decoupled. The entrepreneur is compensated for her current and past returns, but the firm size dynamics are driven by the costs of increasing the entrepreneur's ability to divert funds at a particular period or history. Hence, the dynamics of the two variables can be essentially characterized separately. Second, the firm's size never converges to the first best and its distortions

[2] A notable exception is Fu and Krishna (2019), who study a similar cash-flow diversion model with risk neutrality but with persistent shocks. However, as I will show, the role of persistence on firm dynamics depends on the entrepreneur's risk aversion.

[3] With risk neutrality, the contracting problem amounts to maximizing the value of the firm, which can be justified under complete markets. However, this assumption may be harder to justify in an entrepreneurship context as it involves significant non-diversifiable risks.

inherit the autoregressive properties of the type process. Their drift depends on the initial uncertainty about the entrepreneur's productivity. In particular, if the initial type is known – as assumed in the literature – the distortions drift upwards, so firm size tends to decrease over time. Third, the entrepreneur's compensation is smoothed intertemporally, but the variance of consumption grows over time (as in Thomas and Worrall (1990) and Atkeson and Lucas (1992)). Fourth, I show that implementing the optimal contract with risk aversion requires separately keeping track of the entrepreneur's wealth and equity share in the firm. Finally, I argue that cash flow diversion models always generate a tight link between the firm's size and the entrepreneur's equity share, posing a challenge for this class of models to generate realistic dynamics for the two variables simultaneously. In particular, they cannot account for the dilution of the entrepreneur's equity share as the firm grows.

I first derive a recursive characterization of the optimal contracting problem building on recent advances in dynamic mechanism design (Pavan *et al.*, 2014; Kapička, 2013; Farhi and Werning, 2013). With i.i.d shocks, the literature has characterized the optimal contract using the entrepreneur's promised continuation utility as a state variable. I can also derive a recursive representation of the problem with risk aversion and persistent private information by adding dynamic information rents as an extra state variable. These two state variables break the tight link between firm size and compensation dynamics. Moreover, risk aversion creates a consumption smoothing motive that qualitatively and quantitatively modifies the relation between promised utility and firm size. As a result, in the optimal contract, the promised utility drives the compensation dynamics, and the dynamic information rents drive the firm size dynamics with little interaction between the two.

I characterize the firm size dynamics with return-dependent investment wedges, which reduce capital below its first best level. Investment wedges are positive because more productive entrepreneurs have a relatively higher ability to divert funds as capital increases. The size of the investment wedges depends on the upper Pareto coefficient of the distribution of the marginal product of capital and a normalized shadow cost of information rents. With risk aversion and persistent private information, the lender reduces the cost of screening types at period t by promising to lower future expected information rents. However, this increases the shadow costs of information rents at $t+1$ and onwards.[4] Consequently, when the initial productivity of the entrepreneur is known, the

[4]With persistent private information, more productive types at t prefer contracts with higher expected information rents at period $t+1$. Therefore, committing to lower information rents at $t+1$ reduces the cost of screening types at t. This is the same reason why the labor wedges tend to increase over time in dynamic

investment wedges tend to increase over time and firm size tends to decrease because the firm starts operating without any promise to lower information rents. By contrast, with sufficiently high uncertainty about the initial type, the wedges decrease over time as the distortions from the initial screening problem gradually vanish. Finally, with i.i.d productivity shocks, there is no gain of promising to lower future information rents, so wedges and firm size are (approximately) stationary.

Next, I show that the entrepreneur's consumption process satisfies a Generalized Inverse Euler Equation (GIEE) similar to Hellwig (2021). With risk aversion, the lender smooths the entrepreneur's compensation intertemporally. After a history of high (low) productivity shocks, the entrepreneur is rewarded with high (low) consumption. Hence, the cross-sectional variance of the entrepreneur's consumption grows over time (Thomas and Worrall, 1990; Atkeson and Lucas, 1992). However, the dispersion in compensation does not translate to firm size distortions. Numerically, I find that the investment wedges are essentially uncorrelated with compensation. For instance, after several periods, the firm's size can be distorted downwards, but the entrepreneur receives a high compensation.

To further understand the compensation dynamics, I use numerical simulations and analyze a (quasi-)implementation with simpler contracts. With i.i.d shocks, the following contract gets very close to the optimal allocation. The lender gives the entrepreneur a constant equity share in the firm's reported returns. Then, the entrepreneur can pledge her shares as collateral to borrow and smooth consumption given his implied wealth. Pledging shares is a common practice (Fabisik, 2019); this implementation shows how it can be rationalized as part of a nearly optimal contract. Moreover, the implementation is independent of dividend payout policies, which are typically used to implement the compensation with risk neutrality.

With persistent private information, we need an extra instrument to replicate the dynamic information rents. The entrepreneur's equity share is a natural candidate because it controls the sensitivity of the entrepreneur's compensation to productivity shocks. Intuitively, varying the equity share is informative for the lender because a more productive entrepreneur expects higher returns in the future and is less inclined to give up equity. Hence, the entrepreneur's equity share tracks the dynamics of the expected information rents. The equity share should be high when the expected information rents are high, and conversely. In particular, this implies that the equity share decreases over time if there is no uncertainty about the initial productivity.

Mirrlees models (see Farhi and Werning (2013) and Makris and Pavan (2020)).

The equity share dynamics help understand the distinct firm size dynamics with risk neutrality and risk aversion. Regardless of the entrepreneur's preferences, the lender provides more capital when the equity share is high because of the lower incentives to divert funds –implying that there is always a positive link between the entrepreneur's equity share and the firm's size. Hence, the opposite drift in the firm's size across different models can be understood from the opposite drift in the equity share. With risk neutrality, the equity share instead maps to the promised utility, which drifts upwards (Clementi and Hopenhayn, 2006). As a result, the equity share and the firm's size also drift upwards.

The equity share dynamics of the risk neutral model may be at odds with what we observe in the data. For example, in the venture capital industry, the founder's equity share gets diluted over the financing rounds as the firm grows (Sahlman, 1990). By introducing screening about the initial productivity, the model with risk aversion and persistence can generate more realistic firm dynamics, but this again implies an increasing equity share. Therefore, it appears to be challenging for this class of cash flow diversion models to break the embedded link between the firm's size and the entrepreneur's equity share and, consequently, generate realistic dynamics of these two variables simultaneously.

I explore three extensions (see Online Appendix C.3): (i) limited commitment of the entrepreneur, which can generate dynamics where firm size increases over time as in Albuquerque and Hopenhayn (2004); (ii) a model where the entrepreneur can choose the fraction of funds invested and diverted, which delivers the same characterizations of the firm dynamics and the GIEE; and (iii) allowing the lender to terminate the contract, which may be optimal but does not affect the equations characterizing the optimal contract. Moreover, in a simplified version of the model, I show that if termination is optimal, termination probabilities increase with the persistence of the process. The intuition is similar to that of the equity share dynamics.

Related literature. This paper contributes to the dynamic financial contracting literature. Important early work on this class of models includes Clementi and Hopenhayn (2006), Albuquerque and Hopenhayn (2004), Biais *et al.* (2007), Biais *et al.* (2010), DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007), DeMarzo and Fishman (2007) and DeMarzo *et al.* (2012).[5] In particular, I contribute to the literature by studying a workhorse dynamic cash flow diversion model with risk aversion and persistent private information.

[5] That firm size decreasing over time can be the outcome of an optimal contract has also been shown in Clementi *et al.* (2010). They study a dynamic moral hazard model where the firm's productivity distribution depends on the entrepreneur's costly effort exerted. In their model, the entrepreneur becomes wealthier over time, which lowers the effort and, consequently, firm size.

Models with persistence have been recently analyzed in DeMarzo and Sannikov (2016), Fu and Krishna (2019) and Krasikov and Lamba (2021), but all these papers assume a risk neutral entrepreneur. To my knowledge, this is the first paper in the dynamic financial contracting literature with both persistent private information and risk aversion. I show that the entrepreneur's preferences determine the effects of persistent private information on the firm dynamics. In particular, the distortions to firm size inherit the autoregressive properties of the type process only with risk aversion. Fu and Krishna (2019) and Krasikov and Lamba (2021) show that distortions gradually vanish as with i.i.d types, but the speed of convergence to the first best decreases with the persistence of the process.

Khan *et al.* (2020) analyze the efficient allocation of capital across a continuum of risk-averse agents subject to i.i.d productivity shocks. They show that it is optimal to allocate more capital to agents with higher promised utility. Numerically, this link is weak in my model, so the dynamic information rents determine the allocation of capital. He (2012) and Di Tella and Sannikov (2021) also study dynamic contracting models with risk aversion. Both papers study a hidden savings problem, so the entrepreneur has persistent private information about his savings. I do not allow for hidden savings but allow for persistent private information about the firm's productivity.

Throughout the paper, I use tools and insights from the dynamic Mirrlees literature. I use the first-order approach (FOA) (Pavan *et al.*, 2014) and set up the principal's problem recursively using dynamic information rents as state variables as in Kapička (2013), Farhi and Werning (2013) or Golosov *et al.* (2016). The FOA consists of solving a relaxed problem with the local incentive compatibility constraints and allows for solving the model with persistent private information.[6] Applying the FOA to a cash flow diversion model with risk aversion, an extra challenge emerges because the marginal information rents depend on consumption. So if the principal adjusts the consumption of some type θ, the slope of the profile of information rents for types $\theta' > \theta$ changes. This problem does not arise in commonly studied Dynamic Mirrlees problems with additively separable preferences between income and consumption, but it does when preferences are nonseparable. Following Hellwig (2021), I use incentive-adjusted probability measures to derive analytical characterizations of the optimal contract with risk aversion.[7] The incentive-adjusted measures reweight the density of types such that the lender's evaluation of allocations

[6] A priori, global incentive compatibility constraints may bind. Following the procedure in Kapička (2013) and Farhi and Werning (2013), I verify ex-post that this is not the case in all the numerical simulations.

[7] The idea of using incentive-adjusted measures was developed in Hellwig (2021) to study a dynamic Mirrlees taxation problem with non-separable preferences between consumption, leisure, and productivity.

accounts for the changes in information rents, and therefore, incentive compatibility is preserved. Moreover, the finding that firm size can drift downwards follows from the insight of the Dynamic Mirrlees literature that labor wedges tend to increase over time (Farhi and Werning, 2013; Makris and Pavan (2020)).

Finally, this paper is also related to the literature on insurance with persistent private information (Williams, 2011; Bloedel *et al.*, 2023; and Bloedel *et al.*, 2023). With fixed capital, the cash flow diversion model studied in this paper is equivalent to the hidden endowment model used in this literature. Their focus is on the role of persistent private information for the long-run distribution of consumption and whether or not it features immiseration (Thomas and Worrall, 1990; Atkeson and Lucas, 1992). In the paper, I present some results and discussion on the long-run consumption dynamics. Nevertheless, numerical simulations show that in this model, immiseration is a very long-run phenomenon so that it may be irrelevant for the usual lifespan of a firm.

Outline. The rest of the paper is organized as follows. Section 3.2 describes the model, sets up the relaxed planning problem, and describes the first best allocation. Section 3.3 presents the main results on the firm size and consumption dynamics, and Section 3.4 illustrates them with numerical simulations. Section 3.5 studies the quasi-implementation. Section 3.6 discusses the differences in models with risk neutrality and risk aversion and their implications. Finally, Section 3.7 concludes.

3.2 Model

Time is discrete and indexed by $t = 0, 1, ..., \infty$. Every period, an entrepreneur (the agent, "he") needs funds k_t from a lender (the principal, "she") to operate a project. Both the entrepreneur and the lender are long-lived. At period t, the project generates returns equal to $f(k_t, \theta_t)$, where $\theta_t \in [\underline{\theta}, \overline{\theta}]$ is the entrepreneur's productivity type. The agent's type history is denoted by $\theta^t = \{\theta_0, ..., \theta_t\}$ and is the agent's private information. θ_t follows a first-order Markov process with conditional density $\varphi_t(\theta_t|\theta_{t-1})$ (and CDF $\Phi_t(\theta_t|\theta_{t-1})$), and the initial type θ_0 is drawn from the density $h(\theta_0)$ (and CDF $H(\theta_0)$).

The lender cannot observe the returns and instead relies on the entrepreneur's report. The entrepreneur can misreport and divert a fraction of the returns for his consumption. There is a deadweight loss $(1 - \iota) \in [0, 1)$ on diverted funds. That is, for every dollar of funds diverted, the entrepreneur only gets to consume a fraction ι. Capital k_t fully depreciates at the end of every period. After the entrepreneur reports returns $f(k_t, \widetilde{\theta}_t)$, the lender asks for a repayment $b_t(\widetilde{\theta}_t)$ and advances funds $k_{t+1}(\widetilde{\theta}_t)$ for the next period. The entrepreneur cannot privately save, so the entrepreneur's period t consumption if the true returns are $f(k_t,\theta_t)$ but he reports $f(k_t,\widetilde{\theta}_t)$ is

$$c_t = f(k_t,\theta_t) - (1-\iota)\left(f(k_t,\theta_t) - f(k_t,\widetilde{\theta}_t)\right) - b_t(\widetilde{\theta}_t). \tag{3.1}$$

In particular, if the entrepreneur does not misreport returns, he consumes $c_t = f(k_t,\theta_t) - b_t(\theta_t)$. As is common, I further assume that the agent cannot overreport his returns. That is, reports are restricted to $\widetilde{\theta}_t \leq \theta_t$. This assumption is motivated by the restriction that the entrepreneur cannot save outside the contract with the lender. The entrepreneur also has limited liability, so his consumption must always be non-negative $c_t \geq 0$. The entrepreneur is risk averse, derives utility $u(c_t)$ from consumption, and discounts the future with factor $\beta \in (0,1)$. Throughout the paper, I will use the following notation for the derivatives of the production function: $f_k(k_t,\theta_t) \equiv \frac{\partial f(k_t,\theta_t)}{\partial k_t}$, $f_\theta(k_t,\theta_t) \equiv \frac{\partial f(k_t,\theta_t)}{\partial \theta_t}$, and $f_{\theta k}(k_t,\theta_t) \equiv \frac{\partial^2 f(k_t,\theta_t)}{\partial \theta_t \partial k_t}$.

Below, I summarize all the assumptions on the functions f and u and the productivity process.

Assumptions:

A1: The utility function satisfies $u'' < 0 < u'$, and the Inada conditions $\lim_{c\to 0} u'(c) = \infty$ and $\lim_{c\to\infty} u'(c) = 0$.

A2: The production function is twice differentiable and satisfies $f_{kk} < 0 < f_k$, $f_\theta > 0$, the Inada conditions $\lim_{k\to 0} f_k(k,\theta) = \infty$ and $\lim_{k\to\infty} f_k(k,\theta) = 0$, and $f_{\theta k} > 0$.

A3: The conditional density $\varphi_t(\theta_t|\theta^{t-1})$ is differentiable with respect to the second argument and persistent, i.e.

$$\mathcal{E}(\theta_t,\theta_{t-1}) \equiv \frac{\frac{\partial \varphi_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}}}{\varphi_t(\theta_t|\theta_{t-1})}$$

is non-decreasing in θ_t.

Assumption (A1) implies that the agent is risk averse and the optimal allocation is generally interior. Assumption (A2) states that there is decreasing marginal product of investment, higher types obtain higher returns and have a higher marginal product. This last assumption ($f_{\theta k} > 0$) is key as it will imply that higher capital increases information rents. Finally, assumption (A3) imposes that the type process has either positive persistence or is independent over time, in which case $\frac{\partial \varphi_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}} = 0$. The process is allowed to be time-dependent. Differentiability will be needed to use the envelope condition for the local incentive constraint. For future use, it is useful to define:

$$\rho_t(\theta^t) \equiv \frac{1-\Phi_t(\theta_t|\theta_{t-1})}{\varphi_t(\theta_t|\theta_{t-1})}\mathbb{E}\left[\mathcal{E}(\theta',\theta_{t-1})|\theta' \geq \theta_t,\theta_{t-1}\right] = \frac{\frac{\partial}{\partial\theta_{t-1}}\left(1-\Phi_t(\theta_t|\theta_{t-1})\right)}{\varphi_t(\theta_t|\theta_{t-1})}. \quad (3.2)$$

This is the impulse response of θ_t to θ_{t-1} as defined in Pavan, Segal and Toikka (2014). It is a measure of the persistence of the process. If the type process follows an AR(1) with autoregressive parameter ρ, then $\mathcal{E}(\theta_t,\theta_{t-1}) = -\rho\frac{\partial\varphi_t(\theta_t|\theta_{t-1})}{\partial\theta_t}/\varphi_t(\theta_t|\theta_{t-1})$ and $\rho_t(\theta^t) = \rho$.

3.2.1 Lender's problem

The lender is risk neutral and discounts the future with factor $q \in (0,1)$. At an ex-ante stage ($t = 0$) before the firm starts operating, the lender must screen over the entrepreneur's initial type (θ_0) with continuation contracts for periods $t = 1$ and onwards. In this initial screening stage, the entrepreneur does not consume and there is no production. By the revelation principle, it is without loss to focus on direct revelation mechanisms. At any history, the entrepreneur sends a report $r \in [\underline{\theta},\theta_t]$ about θ_t to the lender. Define a reporting strategy by $\sigma = \{\sigma_t(\theta^t)\}$, it implies a history of reports $\sigma^t(\theta^t) = \{\sigma_1(\theta_0),...,\sigma_t(\theta^t)\}$. Let $\Sigma = \{\sigma|\sigma_t(\theta^t) \leq \theta_t \quad \forall\theta^t \in [\underline{\theta},\overline{\theta}]^t\}$ be the set of feasible reporting strategies. For $t \geq 1$, the entrepreneur's continuation utility with truth-telling can be written recursively as

$$w_t(\theta^t) = u(c(\theta^t)) + \beta\int w_{t+1}(\theta^t,\theta_{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}, \quad (3.3)$$

where $c(\theta^t) = f(k_t(\theta^{t-1}),\theta_t) - b_t(\theta^t)$. Similarly, the continuation utility of type θ^t with reporting strategy σ is

$$w_t^\sigma(\theta^t) = u(c(\theta_t,\sigma^t(\theta^t))) + \beta\int w_{t+1}^\sigma(\theta^t,\theta_{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}, \quad (3.4)$$

where

$$c(\theta_t,\sigma^t(\theta^t)) = \iota f(k_t(\sigma^{t-1}(\theta^{t-1})),\theta_t) + (1-\iota)f(k_t(\sigma^{t-1}(\theta^{t-1})),\sigma_t(\theta_t)) - b_t(\sigma^t(\theta^t)). \quad (3.5)$$

At $t = 0$, the continuation utility with truth-telling writes

$$w_0(\theta_0) = \beta\int w_1(\theta^1)\varphi(\theta_1|\theta_0)d\theta_1, \quad (3.6)$$

and with reporting strategy σ

$$w_0^\sigma(\theta_0) = \beta \int w_1^\sigma(\theta^1)\varphi(\theta_1|\theta_0)d\theta_1. \tag{3.7}$$

Finally, at the start of the contract, the lender must deliver a minimum level of compensation to the entrepreneur equal to

$$(1-\kappa)w_0(\theta_0) + \kappa \int w_0(\theta_0')h(\theta_0')d\theta_0' \geq v_- \tag{3.8}$$

for all $\theta_0 \in [\underline{\theta},\overline{\theta}]$ and where $\kappa \in \{0,1\}$.[8] I set this up as Makris and Pavan (2020), which allows us to consider both ex-ante and ex-post participation constraints. If $\kappa = 1$, we have an ex-ante constraint that requires the principal to deliver expected utility v_- to the agent. Conversely, if $\kappa = 0$, we have an ex-post constraint so the agent's utility must be at least v_- for all θ_0 realizations.

The lender's problem consists of choosing an allocation $\{k_{t+1}(\theta^t), b_t(\theta^t)\}$ to minimize its expected discounted cost subject to the participation, incentive compatibility, and limited liability constraints:

$$\begin{aligned} K_0(v_-) \quad = \quad & \min_{\{k_{t+1}(\theta^t), b_t(\theta^t)\}} \mathbb{E}\left[k_1(\theta_0) + \sum_{t=1}^{\infty} q^t \left(k_{t+1}(\theta^t) - b_t(\theta^t)\right)\right] \\ & s.t \qquad (1-\kappa)w_0(\theta_0) + \kappa\mathbb{E}[w_0(\theta)] \geq v_- \quad \forall \theta_0 \in [\underline{\theta},\overline{\theta}] \quad (PK) \\ & w_t(\theta^t) \geq w_t^\sigma(\theta^t) \quad \forall \theta^t \in [\underline{\theta},\overline{\theta}]^t \text{ and } \sigma \in \Sigma. \quad (IC) \\ & c_t(\theta^t) \geq 0 \quad \forall \theta^t \in [\underline{\theta},\overline{\theta}]^t \quad (LL) \end{aligned} \tag{3.9}$$

Relaxed problem at $t \geq 1$

I start by deriving a – relaxed – recursive representation of the problem for periods $t \geq 1$, and then set up the time-0 screening problem in Section 3.2.1. With Markov shocks, it is sufficient to consider only the temporary incentive compatibility constraint (Fernandes and Phelan, 2000; Kapička, 2013)

$$w_t(\theta^t) = \max_{r\in[\underline{\theta},\theta_t]} u(c(\theta_t,(\theta^{t-1},r))) + \beta \int w_{t+1}(\theta^{t-1},r,\theta_{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}, \tag{3.10}$$

[8]The constant v_- may correspond to the entrepreneur's outside option, or it can be pinned down by a break-even condition for the lender.

where type's θ_t consumption if he reports r, $c(\theta_t, (\theta^{t-1}, r))$, is given by equation (3.5). This allows us to solve a recursive problem. Write entrepreneur's continuation utility under truth-telling as

$$w_t(\theta^t) = u(c(\theta^t)) + \beta v_t(\theta^t) \tag{3.11}$$

$$v_t(\theta^t) = \int w_{t+1}(\theta^{t+1}) \varphi_{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}. \tag{3.12}$$

Following Kapička (2013), Farhi and Werning (2013) and Pavan, Segal and Toikka (2014), I use a first-order approach. That is, I solve a relaxed problem with the local IC constraint.[9] The envelope condition of the temporary IC (3.10) gives

$$\frac{\partial}{\partial \theta_t} w_t(\theta^t) = \underbrace{u'(c(\theta^t)) \iota f_\theta(k_t(\theta^{t-1}), \theta_t)}_{\text{Static marginal info rent}} + \underbrace{\beta \Delta_t(\theta^t)}_{\text{Dynamic marginal info rent}} \tag{3.13}$$

$$\Delta_t(\theta^t) = \int w_{t+1}(\theta^{t+1}) \frac{\partial \varphi_{t+1}(\theta_{t+1}|\theta_t)}{\partial \theta_t} d\theta_{t+1}. \tag{3.14}$$

With persistent private information, the marginal information rents depend on two terms. The static component captures how much the agent can gain by marginally misreporting returns in the current period. The dynamic marginal information rent, which can be rewritten as $\Delta_t(\theta^t) = \mathbb{E}\left[\rho(\theta^{t+1}) \frac{\partial w(\theta^{t+1})}{\partial \theta_{t+1}} | \theta^t\right]$, captures the rent that the agent obtains by having more information about future types than the principal. If types are i.i.d we have $\Delta_t(\theta^t) = 0$.

If the entrepreneur is risk averse, the static marginal information rent $(u'(c(\theta^t)) \iota f_\theta(k_t(\theta^{t-1}), \theta_t))$ depends on the entrepreneur's consumption. Intuitively, if the entrepreneur's productivity increases by $d\theta_t$, he generates an extra return of $f_\theta(k_t(\theta^{t-1}), \theta_t) d\theta_t$. The entrepreneur can then decide to mimic the returns of the type right below him and divert the extra funds, he can then obtain $\iota f_\theta(k_t(\theta^{t-1}), \theta_t) d\theta_t$ extra consumption units. This extra information rent has to be transformed into utils by multiplying by $u'(c(\theta^t))$. The fact that information rents depend on the entrepreneur's consumption poses a challenge for characterizing the solution to this problem. If the principal increases the type's θ_t consumption, then this type's information marginal rent changes. But then the information rents of all types $\theta' > \theta_t$ must be adjusted non-linearly in order to preserve incentive compatibility. In Section 3.3.2, I will show how the incentive-adjusted probability measures developed in Hellwig (2021) can be used to take into account these changes in information rents.

The principal solves a dynamic programming problem where, within every period,

[9] Following Kapička (2013) and Farhi and Werning (2013), I verify numerically that the global IC constraints do not bind. More details can be found in Section 3.4 and Appendix C.4.

there is an optimal control problem. I drop the limited liability constraints from the problem and verify ex-post that they do not bind.[10] For $t \geq 1$, the relaxed problem is

$$
\begin{aligned}
& K_t(v_{t-1}, \Delta_{t-1}, \theta_{t-1}, k_t) = \\
\min_{\substack{\{k_{t+1}(\theta^t), b_t(\theta^t), \\ w_t(\theta^t), v_t(\theta^t), \Delta_t(\theta^t)\}}} \quad & \int \left[k_{t+1}(\theta^t) - b_t(\theta^t) + qK_{t+1}(v_t(\theta^t), \Delta_t(\theta^t), \theta_t, k_{t+1}(\theta^t)) \right] \varphi_t(\theta_t|\theta_{t-1}) d\theta_t \\
s.t \quad (PK) \quad & w_t(\theta^t) = u(c(\theta^t)) + \beta v_t(\theta^t) \qquad [\varphi_t(\theta_t|\theta_{t-1}) \xi_t(\theta^t)] \\
& v_{t-1} = \int w_t(\theta^t) \varphi_t(\theta_t|\theta_{t-1}) d\theta_t \qquad [\varphi_t(\theta_t|\theta_{t-1}) \lambda_t] \\
(IC) \quad & \dot{w}_t(\theta^t) = u'(c(\theta^t)) \iota f_\theta(k_t, \theta_t) + \beta \Delta_t(\theta^t) \qquad [\mu_t(\theta^t)] \\
& \Delta_{t-1} = \int w_t(\theta^t) \frac{\partial \varphi_t(\theta_t|\theta_{t-1})}{\partial \theta_{t-1}} d\theta_t \qquad [\varphi_t(\theta_t|\theta_{t-1}) \gamma_t] \\
(Feasibility) \quad & c(\theta^t) = f(k_t, \theta_t) - b_t(\theta^t)
\end{aligned}
\tag{3.15}
$$

Note that I write the multipliers associated with each constraint inside square brackets. To economize notation, I will write directly $u(\theta^t)$ and $f(\theta^t)$ instead of $u(c(\theta^t))$ and $f(k_t(\theta^{t-1}), \theta_t)$. Along with the promised utility, v_{t-1}, the previous period's type, θ_{t-1}, and the funds advanced at $t-1$, k_t, the past dynamic information rents, Δ_{t-1}, become an extra state variable of the problem. The principal can lower current dynamic information rents by promising to reduce information rents in future periods. Intuitively, she does so by reducing the expected sensitivity of the entrepreneur's value to his productivity, i.e. by reducing his exposure to returns. Because the past promises must be satisfied, Δ_{t-1} has to be added as an extra state variable of the problem. Throughout the paper, I will refer to this state variable as the promised information rent. The co-state variable of the within period Hamiltonian is $\mu_t(\theta^t)$. This co-state variable will become key for the dynamics later. We will refer to it as the shadow cost of information rents, as it captures the principal's resource gain from reducing information rents.

[10]The allocation is generally interior because of the Inada condition on u. However, without extra assumptions on the utility function, understanding the behavior of the contract around the boundary $c_t = 0$ is more complex with persistent private information (see Bloedel *et al.*, 2023). In any case, this is only a concern in the immiseration limit, which is not the focus of the paper, and numerically, I find that consumption is always strictly positive.

Time-0 problem

At $t = 0$, the principal screens over θ_0 with the continuation contracts for periods $t = 1$ and onwards. I also employ the FOA and solve the following relaxed problem[11]

$$
\begin{aligned}
K_0(v_-) = & \min_{\substack{\{k_1(\theta_0), w_0(\theta_0),\\ v_0(\theta_0), \Delta_0(\theta_0)\}}} \int \left(k_1(\theta_0) + qK_1(v_0(\theta_0), \Delta_0(\theta_0), \theta_0, k_1(\theta_0))\right) h(\theta_0) d\theta_0 \\
s.t \;\; & w_0(\theta_0) = \beta v_0(\theta_0) \\
& v_- = (1-\kappa) w_0(\underline{\theta}) + \kappa \int w_0(\theta_0) h(\theta_0) d\theta_0 \\
& \dot{w}_0(\theta_0) = \beta \Delta_0(\theta_0)
\end{aligned}
$$

When there is no uncertainty about the initial type (i.e. θ_0 is fixed), the problem can be directly solved by treating Δ_0 and k_1 as free variables and setting $v_- = v_0$: $K_0(v_-) = \min_{\Delta_0, k_1} k_1 + qK_1(v_-, \Delta_0, \theta_0, k_1)$.

3.2.2 First Best

To gain intuition on the model, it is useful to first look at the first best allocation, i.e. with no private information. The results are summarized in the following proposition.

Proposition 9. *In the First Best, at any history θ^t, there is*

1. *No diversion of funds: $f(k_t, \widetilde{\theta}_t) = f(k_t, \theta_t)$.*
2. *No distortion of the firm's size: $\frac{1}{q} = \mathbb{E}\left[f_k(k_{t+1}(\theta^t), \theta_{t+1}) | \theta_t\right]$.*
3. *Full insurance and intertemporal consumption smoothing: $u'(c(\theta^t)) = \frac{\beta}{q} u'(c(\theta^{t+1}))$.*

By the revelation principle and because diverting funds is inefficient, there will also be no diversion of funds in the second-best. However, points 2. and 3. of the proposition do not hold in the second best allocation. In particular, firm size is distorted downwards, and the entrepreneur is exposed to risk. In the following section, we will study how, with private information, the firm size and compensation dynamics differ from the first best benchmark and the risk neutral and i.i.d cases.

[11] Notice that in the ex-post participation constraint, I have used the fact that if $w_0(\underline{\theta}) = v_-$, the constraint must also hold for $\theta_0 > \underline{\theta}$ due to the incentive constraint.

3.3 Optimal allocation

In this section, I present the main results on the dynamics of the optimal allocation. I begin with a brief overview of the optimal contract in the benchmark with risk neutrality and discuss why risk aversion breaks the tight link between compensation and firm size dynamics (Section 3.3.1). As a result, the firm size and compensation dynamics can be characterized separately. I start with the firm size dynamics (Section (3.3.2)). First, I show that they are driven by the dynamics of the normalized shadow cost of information rents ($\widetilde{\mu}_t$). Second, I introduce the incentive-adjusted probability measures as in Hellwig (2021) to characterize $\widetilde{\mu}_t$ and its dynamics. Then, I turn to the compensation dynamics (Section (3.3.3)). I again use incentive-adjusted measures to characterize the entrepreneur's consumption process and discuss the implications.

3.3.1 Risk neutral benchmark and breaking the size-compensation link

To facilitate the comparison, I now simplify the model as in Clementi and Hopenhayn (2006), but I allow the entrepreneur to be risk averse. That is, I assume binary i.i.d shocks $\theta \in \{0,1\}$, a return function of the form $\theta f(k)$ and that there is no deadweight loss on diverted funds. Dropping the time subscripts, the incentive constraint of the high type ($\theta = 1$) now writes

$$u(f(k) - b^H) + \beta v^H \geq u(f(k) - b^L) + \beta v^L,$$

or

$$\beta(v^H - v^L) \geq u(f(k) - b^L) - u(f(k) - b^H),$$

where superscripts H and L denote allocations for types $\theta = 1$ and $\theta = 0$, respectively. With risk-neutrality, Clementi and Hopenhayn (2006) show that the limited liability constraint binds for both types (i.e. $b^H = f(k)$ and $b^L = 0$) outside the region with no distortions. Intuitively, the principal wants as high-powered incentives as possible with risk neutrality, so the agent receives zero compensation until the first best is reached. In this case, the incentive constraint, which always binds, writes

$$u(f(k)) = \beta(v^H - v^L).$$

So, the current firm size uniquely pins down the required spread in continuation utilities. The cost of spreading continuation utilities increases with the concavity of the lender's value function. With risk neutrality, the value function is increasing and concave, and it becomes flat for a high enough v. As a result, firm size increases with v until it reaches the first best.

When the entrepreneur is risk averse, the consumption smoothing motive implies that it is never optimal (except, possibly, in the immiseration limit) to set $b^H = f(k)$ or $b^L = 0$. Hence, firm size does not directly pin down the required spread in continuation utilities because the incentive constraints also depend on the repayments b^H and b^L. Therefore, consumption smoothing breaks the tight link between the promised utility and firm size. It may still be the case that, in the optimal contract, capital increases with the promised utility as in Khan *et al.* (2020).[12] However, in the numerical simulations (Section 3.4), I find that firm size is approximately constant with i.i.d shocks. As I show in the following section, persistent private information generates time-varying dynamic information rents that drive the firm size dynamics with risk aversion.

3.3.2 Firm size dynamics

The firm dynamics implied by this cash flow diversion model with risk neutrality are well understood. On average, firm size tends to increase over time until it converges to the first best (Clementi and Hopenhayn, 2006). This is true regardless of whether the shocks are i.i.d or persistent (Fu and Krishna, 2019). The firm dynamics are remarkably different when we allow the entrepreneur to be risk averse. First, the firm's size is always below its first best level. With persistent private information, the firm size distortions inherit the autoregressive properties of the type process, and the drift depends on the uncertainty about the initial type. In particular, if θ_0 is fixed – as typically assumed in the literature – the firm's size tends to decrease over time.

Following the Public Finance tradition, it is helpful to describe the optimal allocation in terms of implicit wedges, i.e. distortions in the second best allocation relative to the first best. I define the lending wedges as distortion to the cost of capital faced by the lender

$$\frac{1}{q(1-\tilde{\tau}^k(\theta^t))} = \mathbb{E}\left[f_k(k_{t+1}(\theta^t),\theta_{t+1})|\theta_t\right]. \tag{3.16}$$

Besides the direct effect of the productivity process $\{\theta^t\}$, now the dynamics of the firm's size ($k_{t+1}(\theta^t)$) also depend on the dynamics of the lending wedge. Therefore, to

[12] Higher capital increases information rents and requires higher sensitivity of consumption. Khan *et al.* (2020) show that, with decreasing absolute risk aversion, this implies that it is optimal to allocate more capital to agents with higher promised utility.

characterize the firm size dynamics in the second best, it is sufficient to focus on the dynamics of the lending wedge. The following proposition shows that this wedge can be characterized with return-dependent investment wedges.

Proposition 10. *At any history θ^t, the lending wedge $\tilde{\tau}^k(\theta^t)$ satisfies*

$$\tilde{\tau}^k(\theta^t) = \frac{\mathbb{E}[f_k(\theta^{t+1})\tau^k(\theta^{t+1})|\theta^t]}{\mathbb{E}[f_k(\theta^{t+1})|\theta^t]} \tag{3.17}$$

where the return-dependent investment wedges satisfy

$$\tau^k(\theta^{t+1}) \equiv \iota\Psi^{f_k}(\theta^{t+1}) \times \tilde{\mu}_{t+1}(\theta^{t+1}) \geq 0$$

where

$$\Psi^{f_k}(\theta^{t+1}) = \frac{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)}{f_k(\theta^{t+1})\varphi_{t+1}(\theta_{t+1}|\theta_t)} f_{\theta,k}(\theta^{t+1}) \geq 0.$$

$$\tilde{\mu}_{t+1}(\theta^{t+1}) = \frac{\mu_{t+1}(\theta^{t+1})}{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)} u'(\theta^{t+1}) \geq 0$$

Because $\tau^k(\theta^{t+1}) \geq 0$ we have $\tilde{\tau}^k(\theta^t) \geq 0$, so the lending wedge lowers capital below its first best level, i.e. $k^{SB}_{t+1}(\theta^t) \leq k^{FB}_{t+1}(\theta^t)$. From now on, I will focus on the return-dependent investment wedges $\tau^k(\theta^{t+1})$. The first term of $\tau^k(\theta^{t+1})$ equals the upper Pareto coefficient of the distribution of the marginal product of capital, $\Psi^{f_k}(\theta^{t+1})$, times the ability to consume diverted funds, ι.[13] Intuitively, because $f_{\theta k} > 0$, increasing capital increases the returns of higher types relatively more. Therefore, their ability to divert funds increases, i.e. the higher types' information rents (in consumption units) increase by more, which is costly for the lender. Hence, $\Psi^{f_k}(\theta^{t+1})$ measures the total increase in information rents above θ_{t+1}. Finally, the cost of increasing the information rents is proportional to the normalized shadow cost $\tilde{\mu}(\theta^{t+1})$. This term increases when the lender wants (or has promised) to lower information rents. So when $\tilde{\mu}(\theta^{t+1})$ is high, increasing information rents is more costly.

For log-additive production functions, i.e. $f(k_{t+1},\theta_t) = g(\theta_{t+1})\tilde{f}(k_{t+1})$, $\Psi^{f_k}(\theta^{t+1})$ is only a function of θ_{t+1}. Consequently, this term usually only depends on the (exogenous) type process, and is stationary as long as the process is also stationary. Therefore,

[13]The upper Pareto coefficient of the distribution of the marginal product is defined as $\Psi^{f_k}(\theta^{t+1}) \equiv \frac{1-G^{f_k}(f_k(\theta^{t+1})|\theta^t)}{f_k(\theta^{t+1})g^{f_k}(f_k(\theta^{t+1})|\theta^t)}$, where g^{f_k} (G^{f_k}) are the density (CDF) of the distribution of the marginal product of capital. Then, using $G^{f_k}(f_k(\theta^{t+1})|\theta^t) = \Phi(\theta_t|\theta_{t-1})$ and $g^{f_k}(f_k(\theta^{t+1})|\theta^t) = (1-\varphi(\theta_t|\theta_{t-1}))f_{\theta,k}(\theta^{t+1})$ we get the equation in the proposition.

the normalized shadow cost of information rents ($\widetilde{\mu}(\theta^{t+1})$) typically drives all the wedge dynamics.

It has been shown in dynamic screening models that with risk aversion and persistent private information, these shadow costs (and so wedges) are persistent, and - if the initial type is fixed- they tend to increase over time (Farhi and Werning, 2013; Makris and Pavan, 2020). In what follows, we first characterize the shadow costs at periods $t \geq 2$ and $t = 1$ and then discuss how the dynamics depend on the time-0 uncertainty.

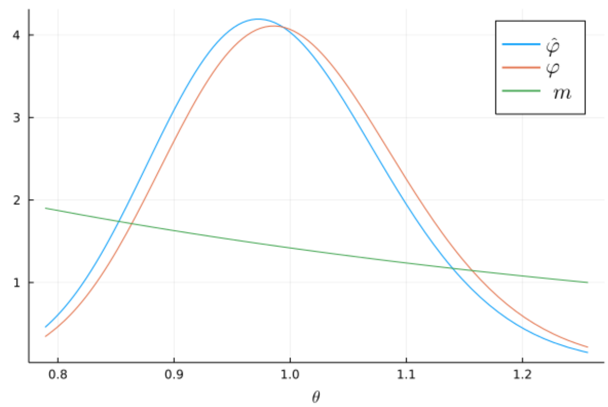
Characterization of $\widetilde{\mu}$. As discussed, the main challenge for characterizing the optimal allocation in this problem is that the static marginal information rents, $u'(c(\theta^t))\iota f_\theta(k_t(\theta^{t-1}), \theta_t)$, depend on consumption. This implies that any perturbation of the allocation of some type θ_t – either through a change in capital or consumption – induces a non-uniform change in the information rents of types $\theta'_t > \theta_t$. To understand this, fix a history θ^{t-1}, and consider a perturbation in the consumption of type θ_t that changes its utility by $\Delta u(\theta^t) > 0$. This changes its marginal information rent by $\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}\Delta u(\theta^t)$. For types $\theta'_t > \theta_t$, due to the change in the slope of the profile of information rents, incentive compatibility requires a change in utility equal to:

$$\Delta u(\theta^{t-1}, \theta'_t) = \exp\left(\int_{\theta_t}^{\theta'_t} \frac{u''(\theta^{t-1}, \theta'')\iota f_\theta(\theta^{t-1}, \theta'')}{u'(\theta^{t-1}, \theta'')} d\theta''\right) \Delta u(\theta^t).$$

This non-separability between consumption and information rents is also a feature of dynamic Mirrlees taxation problems with arbitrary non-separable preferences between consumption, income, and productivity $U(c, y, \theta)$ (see Hellwig, 2021). Hellwig (2021) shows that accounting for this non-separability amounts to evaluating the changes in utility of each type θ_t according to $m(\theta^t) \equiv e^{-\int_{\theta_t}^{\bar{\theta}} \frac{u''(\theta^{t-1}, \theta')\iota f_\theta(\theta^{t-1}, \theta')}{u'(\theta^{t-1}, \theta')} d\theta'}$ (and the changes in consumption according to $M(\theta^t) \equiv \frac{1}{u'(\theta^t)} m(\theta^t)$). Crucially, the factor $m(\theta^t)$ can be interpreted as a reweighting of the type distribution. Accordingly, we define the incentive-adjusted probability measures as

$$\hat{\varphi}_t(\theta_t|\theta^{t-1}) \equiv \frac{\varphi_t(\theta_t|\theta_{t-1}) m(\theta^t)}{\mathbb{E}[m(\theta^t)|\theta_{t-1}]}. \tag{3.18}$$

Figure 3.1: Incentive-adjusted probability measure



Note: The plot is computed with the same calibration as the main simulations in Section 3.4 for i.i.d types. Observe that m is monotonically decreasing, and the incentive-adjusted measure $\hat{\varphi}$ puts more weight on the lower type realizations.

Therefore, the new measure $\hat{\varphi}$ reweights the density of types such that these perturbations preserve incentive compatibility. Because $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} < 0$, the function $m(\theta^t)$ is decreasing in θ_t. So, $\Phi_t(\cdot|\theta_{t-1})$ first-order stochastically dominates $\hat{\Phi}_t(\cdot|\theta^{t-1})$. That is, incentive compatibility requires evaluating allocations as if the principal puts more weight on lower types, see Figure 3.1. Intuitively, because lower types have a higher marginal utility, their information rents are more sensitive to changes in consumption. Therefore, the incentive-adjusted measure that guarantees incentive compatibility has to put more weight on lower types. The following proposition uses the incentive-adjusted measure to characterize $\widetilde{\mu}_t$ for periods $t \geq 2$.

Proposition 11. *(Hellwig (2021)) The normalized shadow cost of information rents $\widetilde{\mu}_{t+1}(\theta^{t+1})$ satisfies, for $t \geq 1$,*

$$\widetilde{\mu}_{t+1}(\theta^{t+1}) = \hat{MB}(\theta^{t+1}) + \hat{\rho}(\theta^{t+1}) \frac{\beta}{q} \frac{1 - \Phi_t(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi_t(\theta_t|\theta_{t-1})} \widetilde{\mu}_t(\theta^t), \tag{3.19}$$

with

$$\hat{MB}(\theta^{t+1}) = \frac{\mathbb{E}\left(m(\theta^t,\theta')|\theta' \geq \theta_{t+1},\theta_t\right)}{M(\theta^{t+1})} \left\{ \hat{\mathbb{E}}\left[\frac{1}{u'(\theta',\theta^t)} \mid \theta' \geq \theta_{t+1},\theta^t\right] - \hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] \right\} \geq 0 \tag{3.20}$$

$$\hat{\rho}(\theta^{t+1}) \equiv \frac{\mathbb{E}\left(m(\theta^t,\theta')|\theta' \geq \theta_{t+1},\theta_t\right)}{M(\theta^{t+1})} \left\{ \hat{\mathbb{E}}\left[\mathcal{E}(\theta',\theta^t) \mid \theta' \geq \theta_{t+1},\theta^t\right] - \hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta^t)|\theta^t\right] \right\} \geq 0. \tag{3.21}$$

Note that the operator $\hat{\mathbb{E}}$ denotes expectations under the measure $\hat{\varphi}$. The proposition shows that the normalized shadow cost of information rents is a function of two terms. The first, $\hat{MB}(\theta^{t+1})$, is the current marginal benefit of redistributing consumption from types $\theta' > \theta_{t+1}$ to $\theta'' < \theta_{t+1}$ while preserving the promise-keeping constraint.[14] The incentive-adjusted measure accounts for the non-uniform utility adjustments resulting from the changes in marginal information rents. Because lower types have higher marginal utility, this perturbation carries a cost reduction for the lender, implying that this term is always (weakly) positive. The second is a backward-looking term that accounts for how changes in information rents at $t+1$ affect information rents at t. Intuitively, it measures the pass-through of information rents from periods $t+1$ to t (captured by $\hat{\rho}(\theta^{t+1})$), times the cost of the resulting change in information rents at t. Hence, these shadow costs inherit the autoregressive properties of the type process.

Proposition 11 shows that the shadow costs (and wedges) inherit the autoregressive properties of the type process. The time-0 screening problem determines the starting value of this process, i.e. $\widetilde{\mu}_1(\theta^1)$. The following proposition shows that, at $t=1$, a similar backward-looking term accounts for the promises to lower information rents in the time-0 screening problem.

Proposition 12. *At $t=1$, the normalized shadow cost of information rents satisfies*

$$\widetilde{\mu}_1(\theta^1) = \hat{MB}(\theta^1) + \hat{\rho}(\theta^1) MB_0(\theta^0) \tag{3.22}$$

where $\hat{MB}(\theta^1)$ and $\hat{\rho}(\theta^1)$ are given by equations (3.20) and (3.21), respectively, and $MB_0(\theta^0)$ solves

$$MB_0(\theta^0) = \frac{1-H(\theta_0)}{h(\theta_0)} \left\{ \mathbb{E}_h \left[\lambda_1(\theta_0') | \theta_0' > \theta_0 \right] - \kappa \mathbb{E}_h \left[\lambda_1(\theta_0) \right] \right\}, \tag{3.23}$$

where

$$\lambda_1(\theta_0) = \hat{\mathbb{E}} \left[\frac{1}{u'(\theta^1)} | \theta_0 \right] + MB_0(\theta^0) \hat{\mathbb{E}} \left[\mathcal{E}(\theta_1, \theta_0) | \theta_0 \right]. \tag{3.24}$$

In particular, if the initial type θ_0 is fixed, $MB_0(\theta^0) = 0$.

When $\kappa = 1$, the backward-looking term in $\widetilde{\mu}_1$ measures the marginal benefit of redistributing expected continuation utilities $v_0(\theta_0)$ from types $\theta' > \theta_0$ to $\theta'' < \theta_0$. With the

[14] The literature on dynamic mechanism design with risk aversion typically analyses hidden effort models with separable preferences of the form $U(\theta, e, c) = u(c) - \psi(e, \theta)$, where e is the (unobservable) agent's effort. This includes dynamic taxation models with separable preferences, but also models of managerial compensation, among others. With these preferences, the static marginal information rents $\psi_\theta(e, \theta)$ are independent of consumption. So, if the principal increases the utility of type θ_t, it is sufficient to increase utility uniformly to all types $\theta_t' > \theta_t$ to preserve incentive compatibility, so consumption has to be redistributed in proportion to $\frac{1}{u'(\theta')}$. As a result, in these settings, one can derive the same characterization but under the original measure φ and with $m(\theta^t) = 1$ (Makris and Pavan, 2020; Brendon, 2013; Hellwig, 2021).

ex-post constraint ($\kappa = 0$), this is simply the marginal benefit of lowering the continuation utilities of types $\theta' > \theta_0$. In both cases, this term is zero when θ_0 is fixed, so there is no backward-looking term in $\widetilde{\mu}_1$. More generally, as I find numerically, this term should typically increase with the variance of θ_0. Moreover, because $\mathbb{E}_h[\lambda_1(\theta_0)] > 0$, this marginal benefit will usually be much larger with the ex-post constraint, especially if the constraint is tight so that the multipliers $\lambda_1(\theta_0)$ are large.[15]

Dynamics of $\widetilde{\mu}$. Proposition 11 shows that the shadow costs $\widetilde{\mu}_{t+1}$ are persistent. Iterating backward on equation (3.19) and using (3.22) we get

$$\widetilde{\mu}_{t+1}(\theta^{t+1}) = \sum_{\tau=0}^{t}\left(\frac{\beta}{q}\right)^{\tau}\prod_{s=0}^{\tau-1}\left(\hat{\rho}_{t+1-s}(\theta^{t+1-s})\frac{1-\Phi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}{u'(\theta^{t-s})\varphi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}\right)\hat{MB}(\theta^{t+1-\tau})$$
$$+\left(\frac{\beta}{q}\right)^{t}\prod_{s=0}^{t-1}\left(\hat{\rho}_{t+1-s}(\theta^{t+1-s})\frac{1-\Phi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}{u'(\theta^{t-s})\varphi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}\right)\hat{\rho}(\theta^{1})MB_0(\theta^0) \tag{3.25}$$

The formula shows that the shadow costs of information rents are a function of current and past marginal benefits of redistribution $\{\hat{MB}(\theta^{t+1-\tau})\}$ and $MB_0(\theta^0)$. In particular, because the passthrough terms $\hat{\rho}_{t+1-s}(\theta^{t+1-s})\frac{1-\Phi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}{u'(\theta^{t-s})\varphi_{t-s}(\theta_{t-s}|\theta_{t-s-1})}$ are always positive, the shadow costs are always increasing in the past marginal benefits. Moreover, the drift will depend on whether the time-0 marginal benefits $MB_0(\theta^0)$ are larger than the marginal benefits in subsequent periods $\{\hat{MB}(\theta^{t+1-\tau})\}$.

Consider first the case where θ_0 is fixed so that $MB_0(\theta_0) = 0$. Then, $\widetilde{\mu}_{t+1}(\theta^{t+1})$ and $\tau^k(\theta^{t+1})$ will tend to grow with the distance from the starting period. The intuition is the following. With persistent private information, different types θ_t have different preferences for period $t+1$ contracts. In particular, higher types value relatively less contracts with low information rents at $t+1$, as they know they are expected to be more productive then and so collect higher information rents. The principal can use this to lower the resource cost of screening types at every period. More concretely, if the principal promises to lower the future information rents to type (θ^{t-1}, θ') (i.e lowers $\Delta_t(\theta^{t-1}, \theta')$) this relaxes the incentive constraints of types (θ^{t-1}, θ') with $\theta'' > \theta'$. Because every period the principal can gain by promising to lower future information rents, the shadow costs $\widetilde{\mu}_t$ will tend to increase over time. More concretely, the gain of relaxing type (θ^{t-1}, θ')'s incentives constraint equals the marginal benefit of redistributing consumption around him. Hence, the

[15]Notice also that $MB_0(\theta_0)$ is typically inverse U-shaped in θ_0 with $\kappa = 1$ but decreasing with $\kappa = 0$. Thus, the firm size distortions will be the largest for the intermediate (lowest) types with ex-ante (ex-post) constraints.

increases in $\widetilde{\mu}_{t+1}$ are proportional to the intertemporal passthroughs of information rents times the marginal benefits $\{\hat{MB}(\theta^{t+1-\tau})\}$. However, as will be shown in the numerical simulations, the wedges may, over time, converge to a stationary distribution. I use this intuition to explain why the lender may want to use equity purchases in the implementation (see Section 3.5.2).[16] It is important to stress that, in this case, for every type θ_t firm size ($k_{t+1}(\theta_t)$) is never larger than in the initial period. The reason is that the principal initializes the contract by setting Δ_0 freely. So Δ_0 is set to not have any "promises" to lower dynamic information rents. Consequently, for every $\theta_t \in [\underline{\theta}, \overline{\theta}]$, the wedges will not be smaller than in the initial period.

In the time-0 problem, by the same logic, the lender also lowers dynamic information rents ($\Delta_0(\theta_0)$) to reduce the cost of screening over θ_0. Hence, with initial uncertainty, the shadow costs at $t = 1$ can already be high, which can make wedges decrease over time. Consider the extreme case where all the uncertainty about the entrepreneur's productivity is realized in the initial period, i.e. $\theta_t = \theta_0$ for all $t \geq 1$. Then, the marginal benefits of redistribution would be zero for all periods following $t = 0$, i.e. $\hat{MB}(\theta^t) = 0$ for all $t \geq 1$. If the passthrough term times $\frac{\beta}{q}$ is smaller than one – as will typically be the case with a mean-reverting process– the effect of the time-0 screening problem on the shadow costs will gradually vanish and the investment wedges go to zero. More generally, the drift in the wedges will depend on the relative magnitudes of $MB_0(\theta^0)$ and $\{\hat{MB}(\theta^t)\}$. With sufficiently high variance in θ_0, $MB_0(\theta^0)$ can be high enough such that the shadow cost and investment wedges decrease over time. As I will show numerically, if the initial variance is not as high, the investment wedges can increase during a few initial periods and then gradually decrease.

Moreover, both risk aversion and persistence are necessary to have these investment wedge dynamics. If the agent is risk neutral we have $\hat{MB}(\theta^{t+1}) = 0$, which implies $\widetilde{\mu}_{t+1}(\theta^{t+1}) = 0$. If the type process is not persistent we have $\hat{\rho}_t(\theta^t) = \rho_t(\theta^t) = 0$ and

$$\widetilde{\mu}_{t+1}(\theta^{t+1}) = \hat{MB}(\theta^{t+1}) \tag{3.26}$$

so past marginal benefits of lowering information rents do not affect the current shadow costs. However, it is still the case that wedges are always positive and so firm size is below

[16] Alternatively, imagine the principal increases consumption of all types $(\theta^{t-1}, \widetilde{\theta}_t)$ with $\widetilde{\theta}_t > \theta_t$. To preserve incentive compatibility, the principal needs to adjust the information rent of all types $(\theta^{t-2}, \theta'_{t-1})$ with $\theta'_{t-1} > \theta_{t-1}$. Because if types are persistent (i.e $\rho_t(\theta^t) > 0$), types $\theta'_{t-1} > \theta_{t-1}$ have a higher probability of being type $\widetilde{\theta}_t$ at period t. This adjustment has to be done for all types $(\theta^{\tau-1}, \theta'_{\tau-1})$ with $\theta'_{\tau-1} > \theta_{\tau-1}$ at all periods $\tau < t$. Therefore, these costs will tend to increase over time if types are persistent. For a clearer and more detailed intuition on this, see Makris and Pavan (2020).

the first best. As I will show in the numerical simulations, wedges are approximately stationary with i.i.d types.

The persistence of the wedges can be amplified or dampened with the incentive-adjusted measure relative to the impulse responses under the original type measure $\rho_t(\theta^t)$. The (unnormalized) persistence is

$$\frac{1-\Phi(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi(\theta_t|\theta_{t-1})}\hat{\rho}_t(\theta^t) \gtreqless \rho_t(\theta^t)$$

if $\rho_t(\theta,\theta^{t-1})\frac{u''(\theta,\theta^{t-1})f_\theta(\theta,\theta^{t-1})}{u'(\theta,\theta^{t-1})}$ is increasing/constant/decreasing in θ (see proposition 3 in Hellwig (2021)). This condition depends, in particular, on the properties of the utility function. For instance, assume that the type process is (log) AR(1) with autoregressive parameter ρ (i.e $\frac{\partial\varphi_t(\theta_t|\theta_{t-1})}{\partial\theta_{t-1}} = -\rho\frac{\partial\varphi_t(\theta_t|\theta_{t-1})}{\partial\theta_t}$ and $\rho_t(\theta^t) = \rho$) and that the production function is linear in the type (i.e $f_{\theta\theta} = 0$). Then, the persistence of the wedges is amplified, i.e. $\frac{1-\Phi(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi(\theta_t|\theta_{t-1})}\hat{\rho}_t(\theta^t) > \rho$, if the utility features decreasing absolute risk aversion (DARA), but $\frac{1-\Phi(\theta_t|\theta_{t-1})}{u'(\theta^t)\varphi(\theta_t|\theta_{t-1})}\hat{\rho}_t(\theta^t) = \rho$ with CARA utility.

In the data, we consistently observe a strong lifecycle component in firm dynamics (Evans, 1987). Young firms are usually small and face strong financing constraints. Over time, the firm size tends to increase, and financing constraints are relaxed. A cash flow diversion model with a risk neutral agent and limited liability (Clementi and Hopenhayn, 2006, Fu and Krishna, 2019) can qualitatively replicate the dynamics observed in the data. However, this is no longer the case once we introduce risk aversion and persistent private information in the benchmark without time-0 uncertainty. The opposite dynamics emerge: the firm size tends to decrease over time, and it never reaches the first best. In Section 3.6, I discuss in more detail why models with risk neutrality generate different firm dynamics using intuitions from the implementation. Introducing uncertainty about the starting productivity can solve this and generate dynamics where firm size increases over time.

3.3.3 Compensation dynamics

We now turn to the compensation dynamics. As in all dynamic insurance models, at the optimum, the principal equalizes the cost of increasing the agent's utility at periods t and $t+1$ in an incentive-compatible manner, i.e.

$$\lambda_{t+1}(\theta^t) = \frac{\beta}{q}\xi(\theta^t), \tag{3.27}$$

where $\lambda_{t+1}(\theta^t)$ is the multiplier on the period $t+1$ promise keeping constraint and $\xi(\theta^t)$ is the multiplier on the type's θ^t period t continuation utility constraint. Again, when the principal promises to increase utilities at period $t+1$, this changes all marginal information rents (as they depend on consumption). So, utility has to be distributed non-uniformly to preserve incentive compatibility. Hellwig (2021) shows how incentive-adjusted measures can be used to derive a Generalized Inverse Euler Equation (GIEE). I derive a similar characterization in this model.

Proposition 13. *In the optimal allocation, the following Generalized Inverse Euler Equation holds at any history θ^t*

$$\frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \frac{1}{u'(\theta^t)}(1+s(\theta^t)) \tag{3.28}$$

where

$$s(\theta^t) = \left(\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} - \hat{\mathbb{E}}\left[\rho_{t+1}(\theta^{t+1})\frac{u''(\theta^{t+1})\iota f(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right]\right)\frac{f_k(\theta^t)}{f_{k\theta}(\theta^t)}\tau^k(\theta^t). \tag{3.29}$$

The GIEE provides an intuitive representation that clarifies what effects drive consumption dynamics and allows to perform direct comparative statics. As in the standard Inverse Euler Equation, the costs of increasing utility at period t and $t+1$ are proportional to $\frac{1}{u'(\theta^t)}$ and $\frac{1}{u'(\theta^{t+1})}$.[17] However, expectations are taken with respect to the incentive-adjusted probability measure because utility at $t+1$ has to be redistributed non-uniformly to preserve incentive compatibility. Moreover, an extra wedge emerges that captures how savings decisions affect marginal information rents at periods t and $t+1$. Changes in marginal information rents at $t+1$ lower information rents in period t at rate $\rho_{t+1}(\theta^{t+1})$. Therefore, the size and sign of the savings wedge depend on the persistence of the process. Intuitively, when the persistence is higher, increasing consumption at $t+1$ lowers the cost of incentive provision at t by more, and so the principal wants relatively higher savings. In general, if persistence (i.e. $\rho_{t+1}(\theta^{t+1})$) is not too high, we will have $s(\theta^t) < 0$. The savings wedge is then scaled by the cost of information rents at period t.

The savings wedge takes a particularly simple form with CARA utility $u(c) = -e^{-\sigma c}$

[17] As in the characterization of $\widetilde{\mu}$, in hidden effort models with separable preferences it is sufficient to increase utility uniformly across all realizations of θ_{t+1} to preserve incentive compatibility. So, equation (3.27) leads to the well know Inverse Euler Equation $\frac{1}{u'(c(\theta^t))} = \frac{\beta}{q}\mathbb{E}\left[\frac{1}{u'(c(\theta^{t+1}))}|\theta_t\right]$. One cannot derive this tight characterization in all other settings studied in the literature: this includes models with taste shocks (as in Atkeson and Lucas, 1992), hidden endowment (as in Thomas and Worrall, 1990), Mirrlees with non-separable preferences, and also this model. For this reason, results on the agent's consumption process are usually derived from the principal's marginal cost martingale (Golosov *et al.*, 2016).

with $\sigma > 0$. Assume also an autoregressive process $\rho_t(\theta^t) = \rho$ and $f(k,\theta) = \theta\tilde{f}(k)$, then

$$\begin{aligned} s(\theta^t) &= -\sigma\iota\theta_t \times \tau^k(\theta^t) \times \left(\tilde{f}(k_t) - \rho\tilde{f}(k_{t+1}(\theta^t))\right) \\ &= -\sigma\iota \times \tau^k(\theta^t) \times \left(f(k_t,\theta_t) - \mathbb{E}\left[f(k_{t+1}(\theta^t),\theta_{t+1})|\theta_t\right]\right) \end{aligned}$$

Because $-\sigma\iota\theta_t \times \tau^k(\theta^t) \leq 0$, we have $s(\theta^t) \leq 0$ if $\rho \leq \frac{\tilde{f}(k_t)}{\tilde{f}(k_{t+1}(\theta^t))}$. Moreover, savings are, on the margin, more discouraged when the agent is more risk-averse (higher σ), the costs of diverting funds are small (high ι), and the costs of incentive provision are high (high $\tau^k(\theta^t)$). With fixed capital ($k_t = k$) and $\iota = 1$, this model nests a hidden endowment model.[18] In this case, $s(\theta^t) = 0$ if $\rho = 1$ and we can use the following result.

Proposition 14. *Assume $\frac{q}{\beta} \leq 1$ and $\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right) \geq 0$, if $s(\theta^t) \geq 0$ marginal utility follows a super-martingale*

$$u'(\theta^t) \geq \mathbb{E}\left[u'(\theta^{t+1})|\theta_t\right].$$ [19]

Moreover, if $s(\theta^t) \geq 0$ for all θ^t, $u' \to 0$ almost surely.

The proposition shows that the marginal utility dynamics are preserved under the original measure when $s(\theta^t) \geq 0$. Therefore, in a hidden endowment model with a unit root process ($\rho = 1$) there is no immiseration (Thomas and Worrall, 1990; Atkeson and Lucas, 1992), and the contract sends the agent to bliss, which is consistent with the results in Bloedel *et al.* (2023) and Bloedel *et al.* (2023).[20] When $s(\theta^t) < 0$, we do not have direct implications for the dynamics under the original measure. The numerical simulations indicate, as expected, that consumption converges to zero, and so there is immiseration. However, the convergence is very slow, so these results may be irrelevant for the usual lifespan of a firm.

Compared to a hidden endowment model, time-varying capital generates an extra motive to increase the variance in compensation over time. For the parametric specification above, as long as $\frac{\tilde{f}(k_t)}{\tilde{f}(k_{t+1}(\theta^t))}$ is decreasing in θ_t, given some high enough ρ, there can exist a $\widetilde{\theta}_t$ such that $s(\theta^t) < 0$ if $\theta_t \leq \widetilde{\theta}_t$ and $s(\theta^t) \geq 0$ otherwise.[21] So, savings are on

[18] With CARA utility, it is also equivalent to a taste shocks model as in Atkeson and Lucas (1992).

[19] With i.i.d shocks, it is easy to verify that the inequality $\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right) \geq 0$ holds. However, with persistent private information, the principal can provide incentive by lowering the dynamic information rents, and, a priori, this inequality may not hold (see Bloedel *et al.*, 2023). Hence, I include it as an assumption in the proposition and verify that it holds in all numerical simulations.

[20] Bloedel *et al.* (2023) and Bloedel *et al.* (2023) have corrected the findings in Williams (2011) and shown (with more general utility functions and processes) that there is immiseration whenever there is some mean-reversion in the type process.

[21] The condition that $\frac{\tilde{f}(k_t)}{\tilde{f}(k_{t+1}(\theta^t))}$ is decreasing in θ_t would not be satisfied if, for some types $\theta_t' > \theta_t''$, the

the margin more discouraged for lower types. Intuitively, because $f_{\theta k} > 0$, higher capital increases information rents. If lower types will have less capital at $t+1$, their incentive constraints will be less tight. Hence, the benefit of increasing consumption at $t+1$ to lower information rents is smaller for lower types.

In sum, the lender minimizes the cost of compensating the agent across periods in an incentive-compatible manner. For this reason, it is optimal to smooth the entrepreneur's compensation over time. Moreover, because the entrepreneur always needs to be compensated for reporting high returns, the cross-sectional variance of consumption grows over time.

3.4 Numerical simulations

In this section, I numerically solve and simulate the model. This will help us better understand the results in the previous section and allow us to quantify the effect of persistent private information on firm size and compensation dynamics. The numerical simulations will also be used to guide the implementation in Section 3.5.

I assume the agent has CRRA utility $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$, and the production function is given by $f(k,\theta) = z\theta k^{\alpha}$, where $\alpha \in (0,1)$ and z is a positive constant used to scale up the problem. The agent's productivity follows a geometric AR(1) process $\theta_t = \theta_{t-1}^{\rho}\varepsilon_t$, where $\log(\varepsilon_t) \sim N(\mu,\sigma_\varepsilon^2)$. I set $\alpha = 3/4$, $1-\iota = 0.05$, $\sigma = 2$ and assume the lender and the entrepreneur have the same discount rate $\beta = q = 0.95$. For the productivity process, I set $\mu = 1$ and $\sigma_\varepsilon^2 = 0.01$. The comparative statics of this section focus on the effect of the persistence ρ. The model is solved with $\rho = 0$ (i.i.d types) and $\rho = 0.7$.[22] Details on the solution method, algorithm and the procedure to check global incentive compatibility can be found in Appendix C.4. After solving for the value functions (K, v and Δ), the policy functions (c_t, λ_{t+1}, γ_{t+1} and k_{t+1}), and the costate (μ_t), I run a Monte Carlo simulation with 10^6 draws over 25 periods each. I start with the benchmark case where θ_0 is known, and at the end of the section, introduce the screening over θ_0.

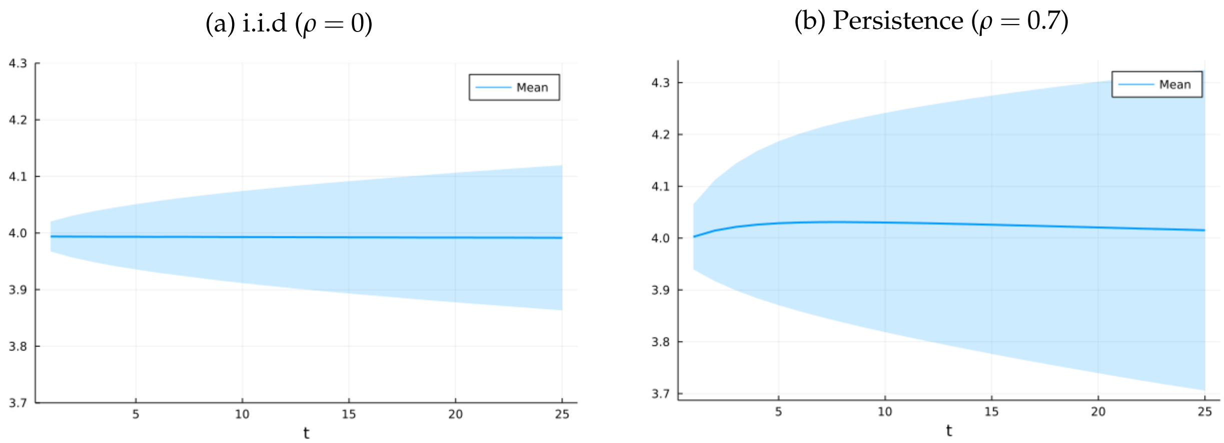
Consumption dynamics. Figure 3.2 illustrates the evolution of the mean and standard deviation of consumption along the cross-section over time with $\rho = 0$ and $\rho = 0.7$. As expected, the variance of consumption is permanently increasing in both cases. With i.i.d

effect of higher wedges at $t+1$ for type θ_t' is stronger than from the higher expected productivity. Numerically I find that $k_{t+1}(\theta^t)$ is indeed increasing.

[22] I also solve the model with different parametrizations of the utility function (log utility ($\sigma = 1$) and CARA), qualitatively, the results are the same (see Appendix C.2).

types, average consumption is approximately constant. With persistence, there is also a slight increase in average consumption in the initial periods. Since the savings wedge $s(\theta^t)$ is proportional to the investment wedge $\tau^k(\theta^t)$, this is consistent with the initial increase in the investment wedge that we will observe (see Figure 3.4). Moreover, because the agent is risk averse, the average marginal utility tends to increase over time.[23]

Figure 3.2: Consumption dynamics



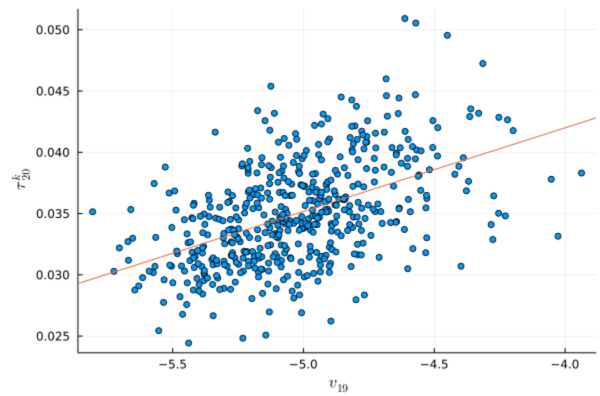
Note: For each period, the blue line is the mean consumption along the cross-section, and the shaded blue area is one standard deviation.

Separating compensation and firm size distortions. The separation between compensation and firm size (or wedge) dynamics can be illustrated very clearly with the numerical simulations. Figure 3.3 shows the relation between the promised utility and the investment wedge at age 20. There appears to be some positive association between the two variables, but they are not linked one to one. We can observe that there is some probability that at age 20, the entrepreneur receives a high compensation (high v_t) but that the firm is financially constrained (high τ^k). The converse is also possible: the compensation is low, but the financing constraints are also low. As discussed, this is not the case in a model with risk neutrality (Clementi and Hopenhayn, 2006), where the promised utility is linked one-to-one with the distortions to firm size.

Firm and wedge dynamics without time-0 screening. Figure 3.4 plots the firm size and investment wedge dynamics. In both cases, the firm size closely follows the dynamics of the investment wedge. With i.i.d shocks, the wedges are approximately stationary, so

[23]To visualize the immiseration dynamics, in Figure C.1 in Appendix C.2 I plot the median and quantiles of the distribution of consumption over a long time horizon. The median consumption monotonically decreases, indicating that consumption will converge to its lower bound. However, the decrease is very slow, so it may be irrelevant for the usual lifespan of a firm.

Figure 3.3: Investment wedge and promised utility at $t = 20$ ($\rho = 0.7$)



Note: Each dot is a random realization of the investment wedge and promised utility at period 20. The red line is a linear regression line on the 500 draws ploted.

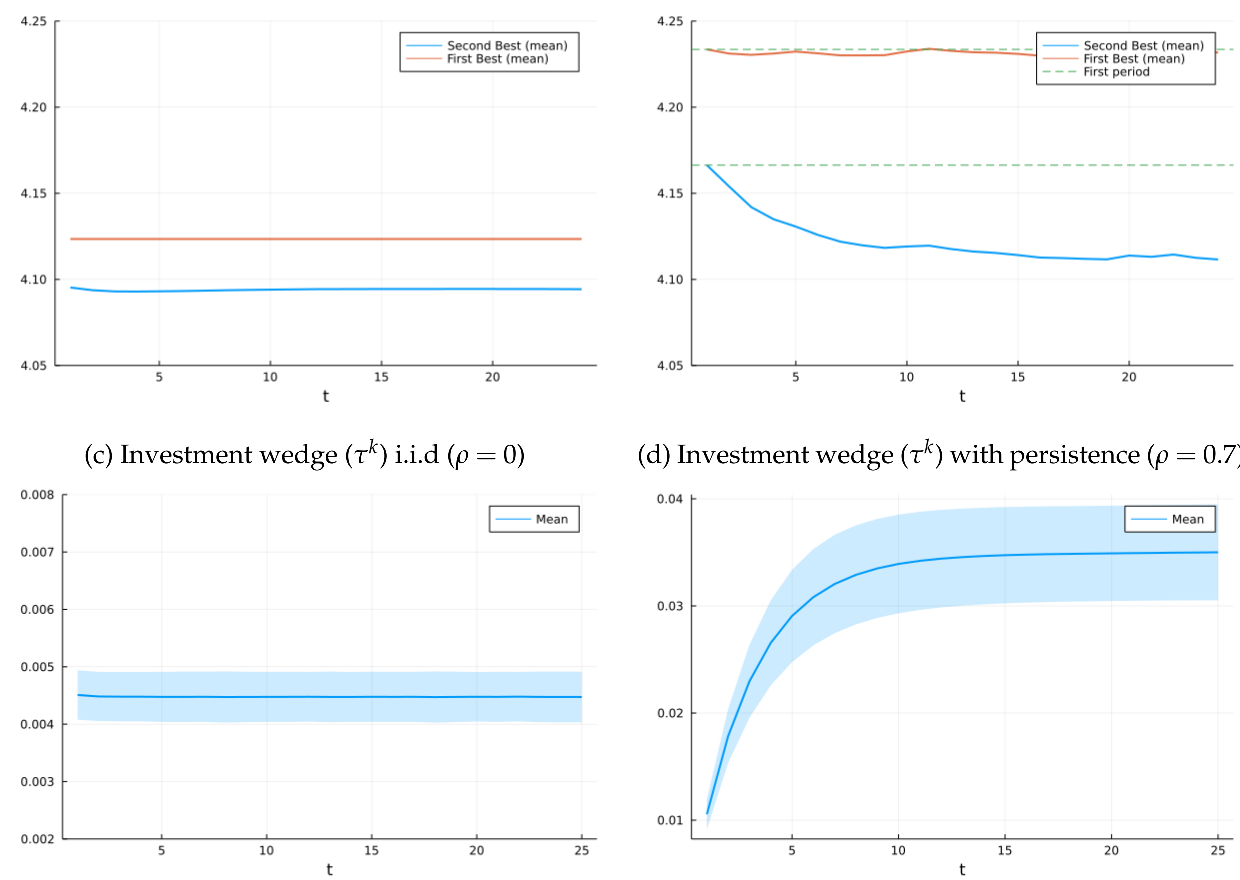
firm size is constant (Panels 3.4a and 3.4c).[24] Hence, the firm size dynamics are essentially independent of the compensation dynamics. Intuitively, the lender compensates the entrepreneur by permanently increasing his consumption, not by lending more capital to the firm. Moreover, the wedges are small, so firm size is also very close to – but always below – the first best level.

For the persistent case, at the first best, the variation in firm size is driven only by differences in expected returns. Moreover, because the type process is mean-reverting, firm size is stationary. At the second best, on average, the wedges tend to increase over time and firm size tends to decrease (Panels 3.4b and 3.4d). However, the wedges do not increase indefinitely. Over time, they converge to a stationary distribution, and so does firm size. With log utility (lower risk aversion), the wedges and the decrease in firm size are smaller (see Figure C.6 in Appendix C.2). Overall, the decrease in firm size will be larger the higher the risk aversion and persistence.

Wedge dynamics with time-0 screening. I now introduce the time-0 screening problem. For this, I assume an ex-ante participation constraint and consider two parametrizations of h. In the benchmark, I set it equal to the ergodic distribution of θ_t but double the standard deviation, and I increase the standard deviation by five times in the second one. Panel 3.5a in Figure 3.5 plots the average investment wedge over the initial type θ_0. As usual, the average wedges are inverse U-shaped in θ_0 due to the no distortions at the top and bottom in the time-0 screening problem. Moreover, as expected, the average wedges are larger in the calibration with higher variance. Panel 3.5b plots the invest-

[24] This is also the case with the other parametrizations of the utility function.

Figure 3.4: Firm size and investment wedge dynamics (θ_0 fixed)



(a) Size (k) i.i.d ($\rho = 0$)

(b) Size (k) with persistence ($\rho = 0.7$)

(c) Investment wedge (τ^k) i.i.d ($\rho = 0$)

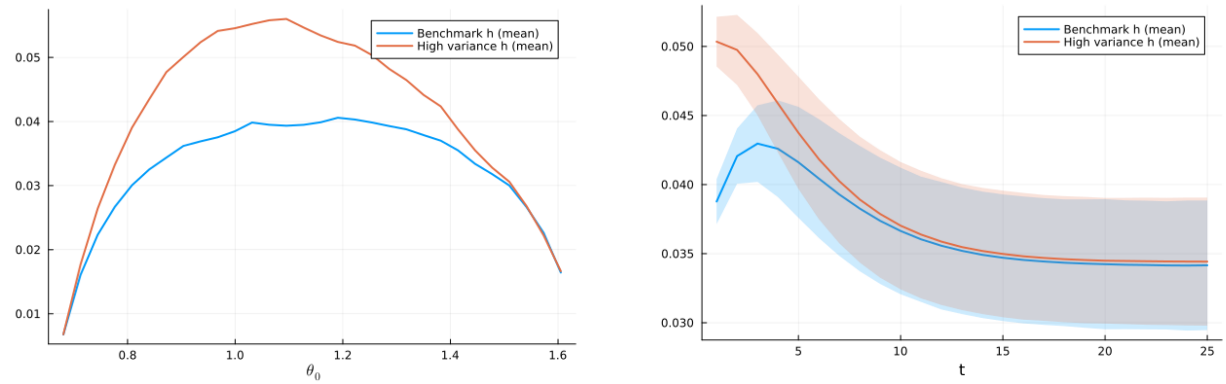
(d) Investment wedge (τ^k) with persistence ($\rho = 0.7$)

Note: Panel (a): The red line is the size at the first best (constant). The blue line is the average size at the second best; it is the same for almost all realizations as expected wedges are approximately constant. Panel (b): The red (blue) line is the average size in the first (second) best. The dashed green lines are the average size in the first period. Panels (c) and (d): The blue lines are the average investment wedges, and the shaded blue areas are one standard deviation.

ment wedge dynamics starting with the θ_0 with the lowest promised information rent (i.e. lowest $\Delta_0(\theta_0)$). In the high variance calibration, we see now that the investment wedge gradually decreases over time. Interestingly, with a lower variance, the wedges initially increase for a few periods and then gradually decrease. Consistent with equation (3.25), in the first periods, the marginal benefits of redistribution from $t \geq 1$ onwards ($\hat{MB}(\theta^1)$, $\hat{MB}(\theta^2)$...) are summed with the initial marginal benefit $MB_0(\theta^0)$, making the initial wedges increase. Over time, the distortions from the time-0 screening start to vanish because the passthrough term is below one, and so wedges start to decrease. Accordingly, the wedges appear to converge to the same stationary distribution with both calibrations. As discussed, the wedges will typically be much larger with an ex-post participation constraint, so they could decrease over time even with a much smaller variance. Moreover, in that case, the average τ_1^K would be decreasing in θ_0 instead of inverse U-shaped.

Figure 3.5: Investment wedge with time-0 screening

(a) Average τ_1^K (b) Wedge dynamics (highest starting distortion)



Note: Panel (a): On a grid for θ_0, I compute the policy functions $(\lambda_1(\theta_0), \gamma_1(\theta_0), k_1(\theta_0))$ from the time-0 problem. Then, I use these policies as starting values to compute the average τ_1^K with a Montecarlo simulation. Panel (b): For both paths, I start with the θ_0 with the lowest $\gamma_1(\theta_0)$. The shaded areas are one standard deviation.

3.5 Quasi-implementation

The optimal contract studied thus far may a priori be complex, which limits the insights we can derive from the problem. Therefore, it is helpful to study implementations of the optimal contract. This will also provide intuition on what drives the different firm size dynamics in models with risk neutrality and risk aversion. A full implementation of the optimal contract is challenging and left for future work. In this section, I use numerical simulations to study a quasi-implementation with simpler contracts that approximate the optimal allocation.

The approach to deriving the quasi-implementation with the numerical simulations will be the following. First, I use regressions with the model simulated data to better understand the compensation dynamics. Then, I propose a simple contract and use the simulated data and regression estimates to calibrate the parameters of the contract. Finally, I solve the entrepreneur's problem under the simple contract and compare the induced consumption dynamics with the optimal contract. With i.i.d types, firm size is constant, so I also fix capital to be constant in the implementation. For simplicity, I will also keep capital fixed for the persistent case. Therefore, we will focus solely on the compensation dynamics.

3.5.1 i.i.d types

I use the simulated data from Section 3.4 to run regressions of consumption on returns and promised utility. The regression results are in Table C.1 in Appendix C.2; we make three observations:

1. Returns at any period $t-k$ have the same effect as returns at t on consumption at t (column 2). Relatedly, consumption follows a random walk (column 5). This suggests that compensation is perfectly smoothed across periods.

2. The sensitivity of compensation to returns does not depend on the current promised utility. Note the interaction $returns_t \times v_{t-1}$ is close to 0 in column 3.

3. The sensitivity of compensation to returns is close to linear. See the linear relation between consumption and returns in Figure C.3 in Appendix C.2 or note that $returns_t^2$ is close to 0 in column 4.

Points 2. and 3. suggest that a constant equity share can be a good approximation. If the promised utility (v_{t-1}) were related to the equity share, we would observe that it affects the sensitivity of consumption to returns, even if the entrepreneur is smoothing consumption intertemporally. Point 1. indicates that in the implementation, the entrepreneur's implicit wealth can be used to perfectly smooth consumption intertemporally. As is known, the promised utility can be naturally mapped to the agent's wealth (Atkeson and Lucas, 1992). Let W_t denote the agent's wealth and χ the (inside) equity share, i.e. the portion of the returns accruing to the entrepreneur. Let $\overline{f}(k_t) = \mathbb{E}\left[f(k_t, \theta_t)\right]$ denote the expected returns if capital is k_t. I fix capital to the optimum in the second best k_{SB}. The entrepreneur also receives initial cash W_0.[25] Therefore, at period 1, the entrepreneur's wealth is: $W_1 = W_0 + \frac{\chi \overline{f}(k_{SB})}{1-q}$. At every period, after returns are realized, if the entrepreneur does not misreport, his wealth changes by $\chi\left(f(k_{SB}, \theta_t) - \overline{f}(k_{SB})\right)$. So, the law of motion of the entrepreneur's wealth follows

$$c_t + W_{t+1} = \frac{1}{q} W_t + \chi\left(f(k_{SB}, \theta_t) - \overline{f}(k_{SB})\right) \equiv C(W_t, \theta_t). \tag{3.30}$$

[25] This is just a free variable used to match the chosen initial promised utility in the second best, so we may also have $W_0 < 0$ if the entrepreneur initially transfers funds to the lender.

Given the entrepreneur's wealth, savings can be chosen to smooth consumption. Therefore, this contract is equivalent to allowing the entrepreneur to pledge his shares as collateral and borrow to consume. This practice is prevalent; Fabisik (2019) reports that between 2007 and 2016, 7.6% of CEOs of US public companies had pledged shares. Moreover, she estimates that 90.5% of CEOs use it to obtain liquidity while maintaining ownership. This motive is consistent with this implementation. Pledging shares aligns the entrepreneur's consumption with the firm's value but without having to sell shares, which is costly as it reduces the entrepreneur's incentives. Moreover, the implementation is independent of dividend payout policies. Notice that it is equivalent if the extra returns $\left(f(k_{SB},\theta_t) - \overline{f}(k_{SB})\right)$ are paid as dividends or are kept as savings inside the firm, and the entrepreneur and the firm face the same interest rate $\frac{1}{q} - 1$.

The next step for the numerical implementation is to obtain a value for χ. I back out this value from the regressions on model simulated data. For an entrepreneur that does not misreport and is allowed to save by himself, to a first-order approximation, we have: $\frac{dc_t}{df(k_t,\theta_t)} \approx (1-q)\chi$. So χ can be identified from the regressions as $\hat{\chi} = \frac{\beta_{returns}}{(1-q)} = \frac{0.0488}{0.05} \approx \iota$, where $\beta_{returns}$ is the regression coefficient on returns in column (1) of Table C.1. So I set directly $\hat{\chi} = \iota$.[26] Then, given $\hat{\chi}$, the entrepreneur's recursive problem with wealth W_t and productivity θ_t is

$$\begin{aligned} \mathcal{W}(W_t,\theta_t) &= \max_{\widetilde{\theta}\leq\theta} u(\widetilde{c}_t) + \beta\mathcal{V}(W_{t+1}) \\ s.t \qquad W_{t+1} &= qC(W_t,\widetilde{\theta}_t) \\ c_t &= (1-q)C(W_t,\widetilde{\theta}_t) \\ \widetilde{c}_t &= c_t + \iota(f(k_{SB},\theta) - f(k_{SB},\widetilde{\theta})) \end{aligned} \tag{3.31}$$

where $\mathcal{V}(W_{t+1}) = \mathbb{E}\left[\mathcal{W}(W_{t+1},\theta_{t+1})\right]$ and W_0 is chosen such that $\mathcal{V}(W_1) = v_1$, i.e. the promised utility under the direct mechanism. Throughout the paper, I have assumed that the entrepreneur cannot secretly save. So in the implementation, there is a double deviation problem if the entrepreneur is allowed to save freely. That is, the entrepreneur deviates by misreporting funds and saving more. For this reason, I assume that the lender directly assigns a consumption/savings level given the entrepreneur's report and wealth $(W_t,\widetilde{\theta}_t)$.[27] Equivalently, we can imagine that the entrepreneur is penalized if the lender

[26] It is a regular result in cash flow diversion models (especially in static versions) that the equity share is linked to the deadweight loss of diverting funds.

[27] I assign the consumption to be $c_t = (1-q)C(W_t,\widetilde{\theta}_t)$ because I observe that average consumption is approximately constant in the numerical simulations. But this is not the optimal savings level of the entrepreneur, as he would save more for precautionary motives. To relax this restriction, we could introduce

Table 3.1: Welfare comparisons i.i.d

	Total Welfare	Deadweight loss diversion of funds	Risk premium (relative to SB)
Optimal contract (SB)	-55.88	0	0
Quasi-Implementation	-56.11 (-0.4% loss)	5.7e-8	0.22

observes that his savings choices are not optimal given the reported type and wealth.

I solve numerically for the policy functions $\widetilde{\theta}(W_t, \theta_t)$ in the entrepreneur's problem (3.31). Then, I run the same Monte Carlo simulation as for the optimal allocation and compare the results.[28] Figure C.4 in Appendix C.2 shows that the consumption paths are very close to the optimal allocation and that this contract induces minimal diversion of funds. Not surprisingly, this simple contract also reaps most of the benefits of the optimal allocation (see Table 3.1). Given a fixed initial promised utility (v_0), we can decompose the lender's loss from using the simple contract

$$K^I(v_0, \theta_0) - K^{SB}(v_0, \theta_0) =$$

$$\underbrace{(1-\iota)\mathbb{E}\left[\sum_{t=0}^{\infty} q^t \left(f(k_{SB}, \theta^t) - f(k_{SB}, \widetilde{\theta}_t(\theta^t))\right) |\theta_0\right]}_{\geq 0, \text{Deadweight loss diversion of funds}} + \underbrace{\mathbb{E}\left[\sum_{t=0}^{\infty} q^t \left(c^I(\theta^t) - c^{SB}(\theta^t)\right) |\theta_0\right]}_{\text{Risk premium}, >0 \text{ if less risk in SB}},$$

where the superscript I is used to denote allocations under the implementation. As shown in Table 3.1, most of the losses from the simple contract result from exposing the entrepreneur to more risk, but the differences are negligible. The implementation performs even better with log utility, see Figure C.5 in Appendix C.2.

3.5.2 Persistent types

With persistent private information, the dynamic information rents (Δ_{t-1}) must be added as an extra state variable in the problem. Intuitively, this variable captures the expected sensitivity of the entrepreneur's utility to the productivity realization (i.e. his exposure),

an extra wedge (or tax) on the entrepreneur's returns on savings to exactly counteract the precautionary motive.

[28]To have accurate comparisons, in the Monte Carlo simulation, for each realization of the shock process $\{\varepsilon_t\}_{t=1}^{25}$ I compute consumption for both the optimal allocation and the implementation. Then for each realization and period, I compute the distance and average across all draws. That is, I compute for every period $\bar{c}_t^{dist} = \sum_i \sqrt{\left(c_t^{SB}(\{\varepsilon_{i.\tau}\}_{\tau=1}^t) - c_t^I(\{\varepsilon_{i.\tau}\}_{\tau=1}^t)\right)^2}$, where c^{SB} is the consumption under the optimal allocation and c^I under the implementation.

as equation (3.14) can be written as

$$\Delta_{t-1} = \mathbb{E}\left[\rho(\theta^t)\frac{\partial w(\theta^t)}{\partial \theta_t}|\theta_{t-1}\right]. \tag{3.32}$$

We can also verify this in the regressions with model simulated data, where we obtain

$$c_t = -0.2^{***}\theta_t - 3.327^{***}\Delta_{t-1} + 1.011^{***}\theta_t \times \Delta_{t-1} + 0.652^{***}\theta_{t-1} + 0.382^{***}v_{t-1}.$$

The coefficient on the interaction term $\Delta_{t-1} \times \theta_t$ is positive. So, for a given level of persistence, when the lender has promised low information rents (i.e low Δ_{t-1}), the entrepreneur's exposure to returns decreases. In this implementation, the exposure of the entrepreneur is controlled by the equity share. Notice that for the i.i.d case (problem (3.31)), if it is optimal for the entrepreneur to not divert funds, we have

$$\frac{\partial \mathcal{W}(W_t, \theta_t)}{\partial \theta_t} = \chi \times u'(c_t) f_\theta(k_{SB}, \theta_t).$$

Thus, a full implementation of the optimal contract would need to allow for a time-varying equity share. In general, lowering the entrepreneur's equity is beneficial as it increases the entrepreneur's insurance against productivity shocks, but it also comes at the cost of increasing the incentives to misreport funds. If types are persistent, there is an extra gain of lowering the equity share at period $t+1$ because it helps screen types.

Why does buying equity help screen types? Imagine that, at period t, the lender offers to buy some equity from type (θ^{t-1}, θ'). Assume also that the lender offers to pay him a price $P_{\Delta\chi}((\theta^{t-1}, \theta'))$ such that he is indifferent between accepting the offer or rejecting it. If returns are persistent, types (θ^{t-1}, θ'') with $\theta'' > \theta'$ have higher expected returns at period $t+1$. So it is not attractive for them to sell equity at price $P_{\Delta\chi}((\theta^{t-1}, \theta'))$. That is, if the entrepreneur is not willing to sell equity at a fair price relative to the projected returns – which both parties agree upon given the current reported returns – it signals that he is misreporting funds. Therefore, the lender can use equity purchases, which inefficiently lower the equity share, to better screen types.

An implementation with a time-varying equity share is substantially more challenging.[29] However, I find that the contract with a constant equity share still delivers small welfare losses relative to the optimal contract.[30] Compared to the i.i.d case, we now only

[29] Now it is more challenging to infer the equity share from the regressions directly. Moreover, it may follow a complicated stochastic process. As it would be persistent but also because there is no distortion at the top ($\bar{\theta}$) and bottom ($\underline{\theta}$) in the promised information rents.

[30] I have experimented with contracts where the equity share is uniformly decreased over time for all

Table 3.2: Welfare comparison with persistence

	Total Welfare	Deadweight loss diversion of funds	Risk premium (relative to SB)
Optimal contract (SB)	-58.49	0	0
Quasi-Implementation	-59.12 (-1.07% loss)	3e-3	0.572

have to make one modification. Notice that when the entrepreneur's period t returns increase, the net present value of the firm's future returns also increases. So, the firm's value increases and the entrepreneur experiences a capital gain. Define the value of the firm by: $\overline{f}_{t+1}(k_{SB},\theta_t) \equiv \mathbb{E}\left[\sum_{\tau=1}^{\infty} q^{\tau-1} f(k_{SB},\theta_{t+\tau})|\theta_t\right]$. Recall capital is fixed to the same level k_{SB} as in the i.i.d case. Then, the entrepreneur's cash on hand at period t if he reports type $\widetilde{\theta}_t$ and his past type report was $\widetilde{\theta}_{t-1}$ is

$$C(W_t,\widetilde{\theta}_t,\widetilde{\theta}_{t-1}) = \frac{1}{q}W_t + \chi\left(f(k_{SB},\widetilde{\theta}_t) + q\overline{f}_{t+1}(k_{SB},\widetilde{\theta}_t) - \overline{f_t}(k_{SB},\widetilde{\theta}_{t-1})\right) \tag{3.33}$$

$$= \frac{1}{q}W_t + \chi\left(\underbrace{f(k_{SB},\widetilde{\theta}_t) - \mathbb{E}\left[f(k_{SB},\widetilde{\theta}_t)|\widetilde{\theta}_{t-1}\right]}_{\text{Innovation returns}} + \underbrace{q\left(\overline{f}_{t+1}(k_{SB},\widetilde{\theta}_t) - \overline{f_{t+1}}(k_{SB},\widetilde{\theta}_{t-1})\right)}_{\text{Capital gain}}\right).$$

The entrepreneur's problem is the same as in (3.31) but with the cash on hand given by (3.33). Because the entrepreneur can borrow using his shares as collateral, the capital gains also increase the entrepreneur's consumption. Moreover, we also need to keep track of the past report $\widetilde{\theta}_{t-1}$ as an extra state variable because it affects the expected returns. Table 3.2 contains the welfare comparison and decompositions with the optimal contract, assuming that the initial type is fixed in both cases. The risk premium is higher than for the i.i.d case, but the welfare losses from the simple contract continue to be small.

Fu and Krishna (2019) study a similar model with persistent private information and a risk neutral entrepreneur. They show that as persistence increases, the sensitivity of the entrepreneur's compensation to returns also increases. In their implementation, this implies that the entrepreneur is compensated more with stock options and less with equity. A priori, a full implementation of the optimal contract could require using stock options. But only with equity the entrepreneur already has a higher exposure to returns. Because the entrepreneur experiences a capital gain and can borrow using his shares as collateral, his compensation increases by more than his equity share times the reported

types. The idea is that when the agent underreports at t, he experiences a capital loss but expects to recover it at $t+1$ with returns that are higher than expected. However, if his equity share is lower at $t+1$, he cannot fully recover the capital loss. However, I have not found any gains from these types of contracts.

returns. Intuitively, simply accounting for capital gains may allow the implementation to approximate the optimal contract because it (approximately) generates the extra sensitivity in compensation required by the dynamic information rents. To understand this, notice that by expanding the dynamic information rent term (Equation (3.32)), we can write

$$\Delta_t = \iota \times \mathbb{E}\left[\sum_{j=1}^{\infty} I_t^{t+j}(\theta^{t+j})\beta^{j-1}u'(c(\theta^{t+j}))f_\theta(k_{t+j},\theta_{t+j})|\theta^t\right],$$

where $I_t^{t+j}(\theta^{t+j}) \equiv \rho(\theta^t) \times ... \times \rho(\theta^{t+j})$ are the impulse response functions as defined in Pavan *et al.* (2014). With $\chi = \iota$ and fixed capital, this is actually the capital gain of the entrepreneur from an increase in productivity $d\theta_t$ if the firm is priced with the entrepreneur's stochastic discount factor.

Finally, this implementation bears some resemblance with the "Dynamic Incentive Account" (DIA) implementation in Edmans *et al.* (2012) for a CEO compensation model. There, the DIA tracks the agent's NPV of future pay (i.e., his wealth) and invests a fraction in the company stock and a fraction in interest-bearing cash. This portfolio is then constantly rebalanced as the firm value changes to maintain incentives. By contrast, here, because the agent's compensation comes solely from his ownership of the firm, letting him borrow against his stock is (approximately) sufficient to implement the required compensation. Moreover, the lender uses equity purchases to lower dynamic information rents, but in this cash flow diversion model, there is no need to rebalance the agent's portfolio every period to maintain incentives.

3.6 Comparison with risk neutral and equity dynamics

The quasi-implementation helps understand the different firm size dynamics with risk neutrality and risk aversion. In Table 3.3, I summarize the main features of the optimal contract and implementations with the different assumptions about the agent's utility and shock process. With risk neutrality, as long as the limited liability constraint ($c \geq 0$) is satisfied, increasing the agent's exposure to risk bears no cost. After high returns, it is optimal to compensate the entrepreneur with a higher stake in the project, i.e. by increasing his equity share. Therefore, with risk neutrality, the entrepreneur's promised utility maps to the value of equity, as shown in Clementi and Hopenhayn (2006).

If the entrepreneur is risk averse, increasing his exposure to risk through a higher equity share is costly. In the numerical simulations, we have seen that the entrepreneur's

Table 3.3: Comparisons optimal contract and implementation across models

	Risk neutral & i.i.d (CH 2006)	Risk neutral & persistent (FK 2019)	Risk averse & i.i.d	Risk averse & persistent
Convergence to FB	Yes	Yes	No	No
Deferred compensation	Yes	Yes	No	No
Link firm size & compensation	Yes (strong)	Yes	Weak	No
Firm size drift w/out time-0 uncertainty	Increasing	Increasing	(Approximately) Constant	Decreasing
Implementation	Equity & cons. = div.	Equity + stock options & cons. = div. + option	Wealth + Equity (fixed) & cons. ≠ div.	Wealth + Equity (varying) & cons. ≠ div.

Notes: CH 2006 stands for Clementi and Hopenhayn (2006), FK 2019 for Fu and Krishna (2019), cons. for consumption and div. for dividends. Implementations of the optimal contract are generally not unique. So, in the table, I just describe the implementation in each of the corresponding papers.

exposure to returns is independent of his promised utility. So with i.i.d types, a constant equity share and mapping the entrepreneur's promised utility to his private wealth gives a good approximation to the optimal allocation. With persistent types, the equity share should also be time-varying as in the risk neutral model, but the driving forces are different. With persistence, the lender has an incentive to lower equity below the efficient level at $t+1$ as it helps screen types at period t. Hence, when θ_0 is fixed and so the dynamic information rents decrease over time, the equity share of the entrepreneur also tends to decrease. Then, when the equity share is low, the entrepreneur has more incentives to divert funds, so the lender is less willing to provide capital.

In both models there is a positive relation between the entrepreneur's equity share and firm size. A lower equity share always increases the implicit lending costs because the incentives to divert funds are higher. However, without time-0 uncertainty, the equity share drifts in opposite directions: upwards with risk neutrality and downwards with risk aversion and persistence. With risk neutrality and i.i.d types, firm size converges to the first best level only because the entrepreneur's equity share goes to one (Clementi and Hopenhayn, 2006). That is, he becomes the sole owner of the firm, and the value of debt and outside equity go to zero. With persistent types and risk neutrality, the equity share does not necessarily have to converge to one for the firm's size to reach the first best (Fu and Krishna, 2019). However, the combination of equity and stock options also increases once the firm becomes unconstrained, and so it also tends to increase over time. These equity dynamics may be inconsistent with what is observed in the data. For example, in the venture capital industry, the founder's ownership is typically diluted over time as the firm's capital grows through multiple financing rounds (Sahlman, 1990).

By introducing uncertainty about the entrepreneur's initial type, we have seen that it is possible to have an upward drift in firm size even with risk aversion and persistent private information. However, this is achieved by making the information rents increase over time, which implies that the equity share should also increase in an implementation. Accordingly, to simultaneously explain firm size and equity dynamics, it may be necessary to break the tight link between equity and firm size that these models generate.

3.7 Conclusion

In this paper, I revisited the firm size and compensation dynamics predicted by the optimal contracting solution of dynamic cash flow diversion models. I departed from the previous literature by allowing the entrepreneur to be risk averse and to have persistent private information about the firm's productivity.

Relaxing these assumptions leads to remarkably different dynamics than those of models with a risk neutral entrepreneur. First, the interaction between risk aversion and persistent private information decouples the dynamics of the firm's size and the entrepreneur's compensation. Second, the firm's size never converges to the first best, and its distortions inherit the autoregressive properties of the type process. Moreover, if there is no initial uncertainty about the entrepreneur's productivity –as assumed in the literature– the distortions tend to increase over time, so the firm's size tends to decrease. Third, the entrepreneur's compensation is smoothed intertemporally, but the variance of consumption increases over time. Finally, implementing the optimal contract requires separately keeping track of the entrepreneur's wealth and equity share in the firm.

I argue that canonical cash flow diversion models cannot simultaneously generate realistic firm size and equity share dynamics due to the embedded link between the two variables. Accordingly, an important avenue for future work is to study departures from this model that can break the tight between the firm's size and the entrepreneur's equity share. In particular, an empirical regularity that these models should be able to rationalize is the dilution of the entrepreneur's equity share as the firm grows.

Bibliography

[1] AARONSON, D., DEHEJIA, R., JORDAN, A., POP-ELECHES, C., SAMII, C. and SCHULZE, K. (2021). The effect of fertility on mothers' labor supply over the last two centuries. *The Economic Journal*, **131** (633), 1–32.

[2] — and FRENCH, E. (2004). The effect of part-time work on wages: Evidence from the social security rules. *Journal of Labor Economics*, **22** (2), 329–252.

[3] ACEMOGLU, D. (1999). Changes in unemployment and wage inequality: An alternative theory and some evidence. *American economic review*, **89** (5), 1259–1278.

[4] — (2024). *The Simple Macroeconomics of AI*. Tech. rep., National Bureau of Economic Research.

[5] — and AUTOR, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, vol. 4, Elsevier, pp. 1043–1171.

[6] ADDA, J., DUSTMANN, C. and STEVENS, K. (2017). The career costs of children. *Journal of Political Economy*, **125** (2), 293–337.

[7] ALBANESI, S. and OLIVETTI, C. (2009). Home production, market production and the gender wage gap: Incentives and expectations. *Review of Economic dynamics*, **12** (1), 80–107.

[8] —, — and PRADOS, M. J. (2015). Gender and dynamic agency: Theory and evidence on the compensation of top executives. In *Gender in the Labor Market*, vol. 42, Emerald Group Publishing Limited, pp. 1–59.

[9] ALBUQUERQUE, R. and HOPENHAYN, H. A. (2004). Optimal lending contracts and firm dynamics. *The Review of Economic Studies*, **71** (2), 285–315.

[10] ALTONJI, J. G. and PIERRET, C. R. (2001). Employer learning and statistical discrimination. *The quarterly journal of economics*, **116** (1), 313–350.

[11] ALVAREDO, F., ATKINSON, A. B., BLANCHET, T., CHANCEL, L. and LUIS BAULUZ, E. A. (2020). Distributional national accounts: Guidelines, methods and concepts used in the world inequality database. [Research Report] PSE (Paris School of Economics). hal-03307274.

[12] AMANO-PATIÑO, N., BARON, T. and XIAO, P. (2020). Equilibrium wage-setting and the life-cycle gender pay gap.

[13] ANGELOV, N., JOHANSSON, P. and LINDAHL, E. (2016). Parenthood and the gender gap in pay. *Journal of Labor Economics*, **34** (3), 545–579.

[14] ANTECOL, H., BEDARD, K. and STEARNS, J. (2018). Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies? *American Economic Review*, **108** (9), 2420–2441.

[15] ATKESON, A. and LUCAS, R. E. (1992). On efficient distribution with private information. *The Review of Economic Studies*, **59** (3), 427–453.

[16] AUTOR, D. H. and DORN, D. (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, **103** (5), 1553–1597.

[17] —, KATZ, L. F. and KEARNEY, M. S. (2006). The polarization of the us labor market. *American economic review*, **96** (2), 189–194.

[18] —, LEVY, F. and MURNANE, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, **118** (4), 1279–1333.

[19] BAJARI, P. and BENKARD, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of political economy*, **113** (6), 1239–1276.

[20] BAQAEE, D. and FARHI, E. (2019). *A short note on aggregating productivity*. Tech. rep., National Bureau of Economic Research.

[21] — and RUBBO, E. (2023). Micro propagation and macro aggregation. *Annual Review of Economics*, **15** (1), 91–123.

[22] BAQAEE, D. R. and FARHI, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, **135** (1), 105–163.

[23] BÁRÁNY, Z. L. and SIEGEL, C. (2018). Job polarization and structural change. *American Economic Journal: Macroeconomics*, **10** (1), 57–89.

[24] BARIGOZZI, F., CREMER, H. and THIBAULT, E. (2023). The motherhood wage and income traps.

[25] BARTH, E., KERR, S. P. and OLIVETTI, C. (2021). The dynamics of gender earnings differentials: Evidence from establishment data. *European Economic Review*, **134**, 103713.

[26] BASU, S. and FERNALD, J. G. (2002). Aggregate productivity and aggregate technology. *European Economic Review*, **46** (6), 963–991.

[27] —, — and KIMBALL, M. S. (2006). Are technology improvements contractionary? *American Economic Review*, **96** (5), 1418–1448.

[28] —, PASCALI, L., SCHIANTARELLI, F. and SERVEN, L. (2022). Productivity and the welfare of nations. *Journal of the European Economic Association*, **20** (4), 1647–1682.

[29] BECKER, G. S., PHILIPSON, T. J. and SOARES, R. R. (2005). The quantity and quality of life and the evolution of world inequality. *American economic review*, **95** (1), 277–291.

[30] BELL, A. (2022). Job amenities and earnings inequality. *Available at SSRN 4173522*.

[31] —, BILLINGS, S. B., CALDER-WANG, S. and ZHONG, S. (2024). An anti-iv approach for pricing residential amenities: Applications to flood risk. *Available at SSRN*.

[32] BENTO, P., SHAO, L. and SOHAIL, F. (2021). Gender gaps in time use and entrepreneurship. *Working Paper*.

[33] BERGER, D., HERKENHOFF, K. and MONGEY, S. (2022). Labor market power. *American Economic Review*, **112** (4), 1147–1193.

[34] BERTRAND, M., GOLDIN, C. and KATZ, L. F. (2010). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American economic journal: applied economics*, **2** (3), 228–255.

[35] BIAIS, B., MARIOTTI, T., PLANTIN, G. and ROCHET, J.-C. (2007). Dynamic security design: Convergence to continuous time and asset pricing implications. *The Review of Economic Studies*, **74** (2), 345–390.

[36] —, —, ROCHET, J.-C. and VILLENEUVE, S. (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica*, **78** (1), 73–118.

[37] BLAU, F. D. and KAHN, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, **55** (3), 789–865.

[38] BLOEDEL, A., KRISHNA, R. and STRULOVICI, B. (2023). *Persistent private information revisited*. Tech. rep., Technical report, Stanford University.

[39] BLOEDEL, A. W., KRISHNA, R. V. and LEUKHINA, O. (2023). Insurance and inequality with persistent private information. *Available at SSRN 3091457*.

[40] BOAR, C. and LASHKARI, D. (2021). *Occupational choice and the intergenerational mobility of welfare*. Tech. rep., National Bureau of Economic Research.

[41] BOARINI, R., JOHANSSON, A. and D'ERCOLE, M. M. (2006). Alternative measures of well-being.

[42] BÖHM, M. J. (2020). The price of polarization: Estimating task prices under routine-biased technical change. *Quantitative Economics*, **11** (2), 761–799.

[43] BOUND, J. and JOHNSON, G. (1992). Changes in the structure of wages in the 1980's: an evaluation of alternative explanations. *The American economic review*, **82** (3), 371–392.

[44] BRENDON, C. (2013). Efficiency, equity, and optimal income taxation.

[45] BRONSON, M. A. and THOURSIE, P. S. (2019). The wage growth and within-firm mobility of men and women: New evidence and theory.

[46] BUDIG, M. J. and ENGLAND, P. (2001). The wage penalty for motherhood. *American sociological review*, **66** (2), 204–225.

[47] BURBANO, V. C., FOLKE, O., MEIER, S. and RICKNE, J. (2023). The gender gap in meaningful work. *Management Science*.

[48] BUZARD, K., GEE, L. K. and STODDARD, O. (2023). Who you gonna call? gender inequality in external demands for parental involvement. *Working Paper*.

[49] BYRNE, D. M., FERNALD, J. G. and REINSDORF, M. B. (2016). Does the united states have a productivity slowdown or a measurement problem? *Brookings Papers on Economic Activity*, **2016** (1), 109–182.

[50] CARD, D., CARDOSO, A. R. and KLINE, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly journal of economics*, **131** (2), 633–686.

[51] CHA, Y. (2013). Overwork and the persistence of gender segregation in occupations. *Gender & society*, **27** (2), 158–184.

[52] CHETTY, R., GUREN, A., MANOLI, D. and WEBER, A. (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *American Economic Review*, **101** (3), 471–475.

[53] CIASULLO, L. and UCCIOLI, M. (2022). What works for working mothers? a regular schedule lowers the child penalty.

[54] CLEMENTI, G. L., COOLEY, T. F. and DI GIANNATALE, S. (2010). A theory of firm decline. *Review of Economic Dynamics*, **13** (4), 861–885.

[55] — and HOPENHAYN, H. A. (2006). A Theory of Financing Constraints and Firm Dynamics. *The Quarterly Journal of Economics*, **121** (1), 229–265.

[56] — and — (2006). A theory of financing constraints and firm dynamics. *The Quarterly Journal of Economics*, **121** (1), 229–265.

[57] COLE, H. and KUBLER, F. (2012). Recursive contracts, lotteries and weakly concave pareto sets. *Review of Economic Dynamics*, **15** (4), 479–500.

[58] COLTRANE, S., MILLER, E. C., DEHAAN, T. and STEWART, L. (2013). Fathers and the flexibility stigma. *Journal of Social Issues*, **69** (2), 279–302.

[59] COMIN, D. A., DANIELI, A. and MESTIERI, M. (2020). *Income-driven labor market polarization*. Tech. rep., National Bureau of Economic Research.

[60] CORDOBA, J. C. and VERDIER, G. (2008). Inequality and growth: Some welfare calculations. *Journal of Economic Dynamics and Control*, **32** (6), 1812–1829.

[61] CORTES, G. M., JAIMOVICH, N. and SIU, H. E. (2017). Disappearing routine jobs: Who, how, and why? *Journal of Monetary Economics*, **91**, 69–87.

[62] CORTÉS, P. and PAN, J. (2019). When time binds: Substitutes for household production, returns to working long hours, and the skilled gender wage gap. *Journal of Labor Economics*, **37** (2), 351–398.

[63] — and Pan, J. (2023). Children and the remaining gender gaps in the labor market. *Journal of Economic Literature*, **61** (4), 1359–1409.

[64] Cortés, P. and Pan, J. (2020). *Children and the Remaining Gender Gaps in the Labor Market*. Working Paper 27980, National Bureau of Economic Research.

[65] Cubas, G., Juhn, C. and Silos, P. (2021). Work-care balance over the day and the gender wage gap. *AEA Papers and Proceedings*, **111**, 149–53.

[66] Dávila, E. and Schaab, A. (2023). *Welfare Accounting*. Tech. rep., National Bureau of Economic Research.

[67] De Schouwer, T. and Kesternich, I. (2023). *Work Meaning and the Flexibility Puzzle*. Tech. rep., Working Paper, KU Leuven. 3, 22.

[68] DeMarzo, P. M. and Fishman, M. J. (2007). Agency and optimal investment dynamics. *The Review of Financial Studies*, **20** (1), 151–188.

[69] — and — (2007). Optimal long-term financial contracting. *The Review of Financial Studies*, **20** (6), 2079–2128.

[70] —, —, He, Z. and Wang, N. (2012). Dynamic agency and the q theory of investment. *The Journal of Finance*, **67** (6), 2295–2340.

[71] — and Sannikov, Y. (2006). Optimal security design and dynamic capital structure in a continuous-time agency model. *The Journal of Finance*, **61** (6), 2681–2724.

[72] — and — (2016). Learning, termination, and payout policy in dynamic incentive contracts. *The Review of Economic Studies*, **84** (1), 182–236.

[73] Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, **132** (4), 1593–1640.

[74] Di Tella, S. and Sannikov, Y. (2021). Optimal asset management contracts with hidden savings. *Econometrica*, **89** (3), 1099–1139.

[75] Dolado, J. J., García-Peñalosa, C. and De La Rica, S. (2013). On gender gaps and self-fulfilling expectations: Alternative implications of paid-for training. *Economic Inquiry*, **51** (3), 1829–1848.

[76] Dorn, D. (2009). *Essays on Inequality, Spatial Interaction, and the Demand for Skills*. Dissertation, University of St. Gallen.

[77] DOVIS, A. (2019). Efficient Sovereign Default. *Review of Economic Studies*, **86** (1), 282–312.

[78] — (2019). Efficient sovereign default. *The Review of Economic Studies*, **86** (1), 282–312.

[79] DUNN, M. (2018). Who chooses part-time work and why?

[80] DURÁN, J. and LICANDRO, O. (2024). Is the output growth rate in nipa a welfare measure? *The Economic Journal*, p. ueae064.

[81] EDMANS, A., GABAIX, X., SADZIK, T. and SANNIKOV, Y. (2012). Dynamic ceo compensation. *The Journal of Finance*, **67** (5), 1603–1647.

[82] EROSA, A., FUSTER, L., KAMBOUROV, G. and ROGERSON, R. (2022). Hours, occupations, and gender differences in labor market outcomes. *American Economic Journal: Macroeconomics*, **14** (3), 543–590.

[83] —, — and RESTUCCIA, D. (2016). A quantitative theory of the gender gap in wages. *European Economic Review*, **85**, 165–187.

[84] EVANS, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *The journal of industrial economics*, pp. 567–581.

[85] FABISIK, K. (2019). Why do us ceos pledge their own company's stock? *Swiss Finance Institute Research Paper*, pp. 19–60.

[86] FARBER, H. S. and GIBBONS, R. (1996). Learning and wage dynamics. *The Quarterly Journal of Economics*, **111** (4), 1007–1047.

[87] FARHI, E. and WERNING, I. (2013). Insurance and taxation over the life cycle. *Review of Economic Studies*, **80** (2), 596–635.

[88] FEENSTRA, R. C., INKLAAR, R. and TIMMER, M. P. (2015). The next generation of the penn world table. *American economic review*, **105** (10), 3150–3182.

[89] FELFE, C. (2012). The motherhood wage gap: What about job amenities? *Labour Economics*, **19** (1), 59–67.

[90] FERNALD, J. G. (2015). Productivity and potential output before, during, and after the great recession. *NBER macroeconomics annual*, **29** (1), 1–51.

[91] FERNANDES, A. and PHELAN, C. (2000). A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory*, **91** (2), 223–247.

[92] FLABBI, L. and MORO, A. (2012). The effect of job flexibility on female labor market outcomes: Estimates from a search and bargaining model. *Journal of Econometrics*, **168** (1), 81–95.

[93] FLEURBAEY, M. (2009). Beyond gdp: The quest for a measure of social welfare. *Journal of Economic literature*, **47** (4), 1029–1075.

[94] — and GAULIER, G. (2009). International comparisons of living standards by equivalent incomes. *Scandinavian Journal of Economics*, **111** (3), 597–624.

[95] FOLKE, O. and RICKNE, J. (2022). Sexual harassment and gender inequality in the labor market. *The Quarterly Journal of Economics*, **137** (4), 2163–2212.

[96] FU, S. and KRISHNA, R. V. (2019). Dynamic financial contracting with persistent private information. *The RAND Journal of Economics*, **50** (2), 418–452.

[97] GOLDEN, L. (2020). Part-time workers pay a big-time penalty. *Economics Policy Institute Report*.

[98] GOLDIN, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, **104** (4), 1091–1119.

[99] —, KERR, S. P. and OLIVETTI, C. (2022). *When the Kids Grow Up: Women's Employment and Earnings across the Family Cycle*. Tech. rep., National Bureau of Economic Research.

[100] —, —, — and BARTH, E. (2017). The expanding gender earnings gap: Evidence from the lehd-2000 census. *American Economic Review*, **107** (5), 110–114.

[101] GOLOSOV, M., TROSHKIN, M. and TSYVINSKI, A. (2016). Redistribution and social insurance. *American Economic Review*, **106** (2), 359–86.

[102] —, TSYVINSKI, A. and WERQUIN, N. (2016). Recursive contracts and endogenously incomplete markets. In *Handbook of Macroeconomics*, vol. 2, Elsevier, pp. 725–841.

[103] GOOS, M. and MANNING, A. (2007). Lousy and lovely jobs: The rising polarization of work in britain. *The review of economics and statistics*, **89** (1), 118–133.

[104] —, — and SALOMONS, A. (2009). Job polarization in europe. *American economic review*, **99** (2), 58–63.

[105] —, — and — (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American economic review*, **104** (8), 2509–2526.

[106] GORDON, R. J. (2018). *Why has economic growth slowed when innovation appears to be accelerating?* Tech. rep., National Bureau of Economic Research.

[107] GURLEY-CALVEZ, T., BIEHL, A. and HARPER, K. (2009). Time-use patterns and women entrepreneurs. *American Economic Review*, **99** (2), 139–144.

[108] HALL, R. E. (1982). The importance of lifetime jobs in the u.s. economy. *The American Economic Review*, **72** (4), 716–724.

[109] — (1990). Invariance properties of solow's productivity residual. *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solow's Birthday*.

[110] HAMERMESH, D. S. (1999). Changing inequality in markets for workplace amenities. *The Quarterly Journal of Economics*, **114** (4), 1085–1123.

[111] — (2001). The changing distribution of job satisfaction. *Journal of Human Resources*, **36** (1), 1–30.

[112] HARKNESS, S., BORKOWSKA, M. and PELIKH, A. (2019). Employment pathways and occupational change after childbirth.

[113] HE, Z. (2012). Dynamic compensation contracts with private savings. *The Review of Financial Studies*, **25** (5), 1494–1549.

[114] HELLWIG, C. (2021). Static and dynamic mirrleesian taxation with non-separable preferences: A unified approach.

[115] HIRSCH, B. T. and MACPHERSON, D. A. (2003). Union membership and coverage database from the current population survey: Note. *ILR Review*, **56** (2), 349–354.

[116] HOTZ, V. J., JOHANSSON, P. and KARIMI, A. (2018). *Parenthood, family friendly workplaces, and the gender gaps in early work careers*. Tech. rep., National Bureau of Economic Research.

[117] HSIEH, C.-T., HURST, E., JONES, C. I. and KLENOW, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, **87** (5), 1439–1474.

[118] — and KLENOW, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, **124** (4), 1403–1448.

[119] HULTEN, C. R. (1978). Growth accounting with intermediate inputs. *The Review of Economic Studies*, **45** (3), 511–518.

[120] HWANG, H.-S., REED, W. R. and HUBBARD, C. (1992). Compensating wage differentials and unobserved productivity. *Journal of Political Economy*, **100** (4), 835–858.

[121] JONES, C. I. and KLENOW, P. J. (2016). Beyond gdp? welfare across countries and time. *American Economic Review*, **106** (9), 2426–2457.

[122] KAPIČKA, M. (2013). Efficient allocations in dynamic private information economies with persistent shocks: A first-order approach. *Review of Economic Studies*, **80** (3), 1027–1054.

[123] KAPLAN, G. and SCHULHOFER-WOHL, S. (2018). The changing (dis-) utility of work. *Journal of Economic Perspectives*, **32** (3), 239–258.

[124] KATZ, L. F. and MURPHY, K. M. (1992). Changes in relative wages, 1963–1987: supply and demand factors. *The quarterly journal of economics*, **107** (1), 35–78.

[125] KHAN, A., POPOV, L. and RAVIKUMAR, B. (2020). Enduring relationships in an economy with capital and private information. *FRB St. Louis Working Paper*, (2020-34).

[126] KLEVEN, H., LANDAIS, C., POSCH, J., STEINHAUER, A. and ZWEIMULLER, J. (2019). Child penalties across countries: Evidence and explanations. *AEA Papers and Proceedings*, **109**, 122–26.

[127] —, — and SØGAARD, J. E. (2019). Children and gender inequality: Evidence from denmark. *American Economic Journal: Applied Economics*, **11** (4), 181–209.

[128] KRASIKOV, I. and LAMBA, R. (2021). A theory of dynamic contracting with financial constraints. *Journal of Economic Theory*, **193**, 105196.

[129] LAMADON, T., MOGSTAD, M. and SETZLER, B. (2022). Imperfect competition, compensating differentials, and rent sharing in the us labor market. *American Economic Review*, **112** (1), 169–212.

[130] LAVETTI, K. (2023). Compensating wage differentials in labor markets: Empirical challenges and applications. *Journal of Economic Perspectives*, **37** (3), 189–212.

[131] LAZEAR, E. P. and ROSEN, S. (1990). Male-female wage differentials in job ladders. *Journal of Labor Economics*, **8** (1, Part 2), S106–S123.

[132] LE BARBANCHON, T., RATHELOT, R. and ROULET, A. (2021). Gender differences in job search: Trading off commute against wage. *The Quarterly Journal of Economics*, **136** (1), 381–426.

[133] LIM, K. (2019). Do american mothers use self-employment as a flexible work alternative? *Review of Economics of the Household*, **17** (3), 805–842.

[134] LOMMERUD, K. E., STRAUME, O. R. and VAGSTAD, S. (2015). Mommy tracks and public policy: On self-fulfilling prophecies and gender gaps in hiring and promotion. *Journal of Economic Behavior & Organization*, **116**, 540–554.

[135] LUCAS, R. E. (1977). Hedonic wage equations and psychic wages in the returns to schooling. *The American Economic Review*, pp. 549–558.

[136] LUCIFORA, C., MEURS, D. and VILLAR, E. (2021). The "mommy track" in the workplace. evidence from a large french firm. *Labour Economics*, **72**, 102035.

[137] LUNDBORG, P., PLUG, E. and RASMUSSEN, A. W. (2017). Can women have children and a career? iv evidence from ivf treatments. *American Economic Review*, **107** (6), 1611–1637.

[138] LUO, M. and MONGEY, S. (2019). *Assets and job choice: Student debt, wages and amenities*. Tech. rep., National Bureau of Economic Research.

[139] MACPHERSON, D. A. and HIRSCH, B. T. (2023). Five decades of cps wages, methods, and union-nonunion wage gaps at unionstats. com. *Industrial Relations: A Journal of Economy and Society*.

[140] MAESTAS, N., MULLEN, K. J., POWELL, D., VON WACHTER, T. and WENGER, J. B. (2023). The value of working conditions in the united states and implications for the structure of wages. *American Economic Review*, **113** (7), 2007–2047.

[141] MAKRIS, M. and PAVAN, A. (2020). Wedge dynamics with evolving private information. *Working Paper*.

[142] MARCET, A. and MARIMON, R. (2019). Recursive contracts. *Econometrica*, **87** (5), 1589–1631.

[143] MAS, A. and PALLAIS, A. (2017). Valuing alternative work arrangements. *American Economic Review*, **107** (12), 3722–3759.

[144] — and — (2020). Alternative work arrangements. *Annual Review of Economics*, **12**, 631–658.

[145] MOLLOY, R., SMITH, C. and WOZNIAK, A. K. (2020). *Changing stability in us employment relationships: A tale of two tails*. Tech. rep., National Bureau of Economic Research.

[146] MONGEY, S. and WAUGH, M. E. (2024). *Discrete choice, complete markets, and equilibrium*. Tech. rep., National Bureau of Economic Research.

[147] MORCHIO, I. and MOSER, C. (2024). *The gender pay gap: Micro sources and macro consequences*. Tech. rep., National Bureau of Economic Research.

[148] MUNASINGHE, L., REIF, T. and HENRIQUES, A. (2008). Gender gap in wage returns to job tenure and experience. *Labour economics*, **15** (6), 1296–1316.

[149] NDIAYE, A. (2020). *Flexible Retirement and Optimal Taxation*. Tech. rep.

[150] NEAL, D. A. and JOHNSON, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of political Economy*, **104** (5), 869–895.

[151] NORDHAUS, W. D. and TOBIN, J. (1972). Is growth obsolete? In *Economic Research: Retrospect and Prospect, Volume 5: Economic Growth*, 1–80. Cambridge, MA: National Bureau of Economic Research, Inc.

[152] O'DORCHAI, S., PLASMAN, R. and RYCX, F. (2007). The part-time wage penalty in european countries: how large is it for men? *International Journal of Manpower*, **28** (7), 571–603.

[153] PATRICK, C., STEPHENS, H. and WEINSTEIN, A. (2016). Where are all the self-employed women? push and pull factors influencing female labor market decisions. *Small Business Economics*, **46** (3), 365–390.

[154] PAUL, M. (2016). Is there a causal effect of working part-time on current and future wages? *The Scandinavian Journal of Economics*, **118** (3), 494–523.

[155] PAVAN, A., SEGAL, I. and TOIKKA, J. (2014). Dynamic mechanism design: A myersonian approach. *Econometrica*, **82** (2), 601–653.

[156] PETRIN, A. and LEVINSOHN, J. (2012). Measuring aggregate productivity growth using plant-level data. *The Rand journal of economics*, **43** (4), 705–725.

[157] RACHEL, L. (2024). Leisure-enhancing technological change. *Working Paper*.

[158] RESTUCCIA, D. and ROGERSON, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, **11** (4), 707–720.

[159] ROSEN, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of political economy*, **82** (1), 34–55.

[160] — (1986). The theory of equalizing differences. *Handbook of labor economics*, **1**, 641–692.

[161] SAHLMAN, W. A. (1990). The structure and governance of venture-capital organizations. *Journal of Financial Economics*, **27** (2), 473–521.

[162] SARGENT, T. J. and LJUNGQVIST, L. (2000). Recursive macroeconomic theory. *Massachusetss Institute of Technology*.

[163] SCHOONBROODT, A. (2018). Parental child care during and outside of typical work hours. *Review of Economics of the Household*, **16**, 453–476.

[164] SMITH, A. (1776). *The wealth of nations*.

[165] SORKIN, I. (2018). Ranking firms using revealed preference. *The quarterly journal of economics*, **133** (3), 1331–1393.

[166] STANTCHEVA, S. (2017). Optimal taxation and human capital policies over the life cycle. *Journal of Political Economy*, **125** (6).

[167] SYVERSON, C. (2017). Challenges to mismeasurement explanations for the us productivity slowdown. *Journal of Economic Perspectives*, **31** (2), 165–186.

[168] THOMAS, J. and WORRALL, T. (1990). Income fluctuation and asymmetric information: An example of a repeated principal-agent problem. *Journal of Economic Theory*, **51** (2), 367–390.

[169] TORRE, M. (2017). Attrition from male-dominated occupations: Variation among occupations and women. *Sociological Perspectives*, **60** (4), 665–684.

[170] VATTUONE, G. (2023). Worker sorting and the gender wage gap.

[171] WASSERMAN, M. (2023). Hours constraints, occupational choice, and gender: Evidence from medical residents. *The Review of Economic Studies*, **90** (3), 1535–1568.

[172] WILLIAMS, N. (2011). Persistent private information. *Econometrica*, **79** (4), 1233–1275.

[173] WISWALL, M. and ZAFAR, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, **133** (1), 457–507.

[174] WOLF, E. (2014). The german part-time wage gap: Bad news for men?

[175] YU, W.-H. and HARA, Y. (2021). Motherhood penalties and fatherhood premiums: Effects of parenthood on earnings growth within and across firms. *Demography*, **58** (1), 247–272.

Appendix A

Appendix of Chapter 1: Amenity-Biased Technical Change

A.1 Data details

A.1.1 Amenities data

O*NET work context. The O*NET context file provides detailed information about job characteristics across occupations. The data contains a total of 57 characteristics from which I manually select a subset of 17. First, not all the characteristics in the context may be considered amenities, as we can reasonably assume that they do not affect the workers' utility. For example, I exclude characteristics such as how often an occupation requires *writing letters and memos*, or *electronic mail*. Moreover, due to the limited sample size of the NLSY79, we have to be cautious with the dimensionality of the amenities in the estimation of amenity prices. For this reason, I further exclude some amenities that measure characteristics that are very similar to others or more specific. For instance, I exclude characteristics like the time spent *keeping or regaining balance*, or the time spent *in an enclosed vehicle or equipment*.

For every characteristic, there are five categories typically representing the frequency with which a worker experiences it. For example, for the *exposure to contaminants*, the categories range from "never" (category 1), "once a year or more but not every month" (category 2), up to "every day" (category 5). For every occupation and characteristic, the data contains a weight on each of these categories, which is based on the input from occupational experts. Table A.1 lists all the selected characteristics and explains how I define each amenity based on these categories.

The O*NET occupation data uses the O*NET-SOC codes. First, I crosswalk the data to the standard SOC codes and then to the census codes. Finally, I use the crosswalks constructed by (76) to convert the census codes of the corresponding year to the codes of (16) (which are based on the 1990 census codes). Finally, some of the characteristics are missing for a few occupations. Whenever this is the case, I use the average of the amenity among the occupation group (using the same occupation groups as Table A.12) of the corresponding occupation to impute the value.

Table A.1: Amenity definitions in context file

Amenity	Question	Response defining amenity
Contact	How much **contact with others** (by telephone, face-to-face, or otherwise) is required to perform your current job?	Every day (Category 5)
No discussions	How often does your current job require **face-to-face discussions** with individuals and within teams?	1-Every day (1-Category 5)
Teamwork	How important are interactions that require you to **work with or contribute to a work group or team** to perform your current job?	Extremely important (Category 5)
Responsibility	How responsible are you for **work outcomes and results** of other workers on your current job?	High responsibility and Very High responsibility (Categories 4 and 5)
No conflict	How often are **conflict situations** a part of your current job?	1-Every day (1-Category 5)
Indoors	How often does your current job require you to work **indoors in an environmentally controlled environment** (like a warehouse with air conditioning)?	Every day (Category 5)
No extreme temperatures	In your current job, how often are you exposed to **very hot** (above 90° F) or **very cold** (under 32° F) temperatures?	1-Every day (1-Category 5)
No contaminants	In your current job, how often are you exposed to **contaminants** (such as pollutants, gases, dust, or odors)?	1-Every day (1-Category 5)
No hazardous conditions	How often does your current job require that you be **exposed to hazardous conditions**?	1-Every day (1-Category 5)
No burns and cuts	How often does your current job require that you be **exposed to minor burns, cuts, bites, or stings**?	1-Every day (1-Category 5)
Time sitting	How much time in your current job do you spend **sitting**?	Continually or almost continually (Category 5)
Time standing	How much time in your current job do you spend **standing**?	Continually or almost continually (Category 5)
No consequence error	**How serious a mistake** can you make on your current job (one you can't easily correct)?	1-"Extremely serious" (1-"Extremely serious")
Decision making	In your current job, how often do your **decisions affect** other people or the image or reputation or financial resources of your employer?	Everyday (Category 5)
Freedom decisions	In your current job, how much **freedom** do you have to make decisions without supervision?	A lot of freedom (Category 5)
No repetition	How important to your current job are **continuous, repetitious physical activities** (like key entry) or **mental activities** (like checking entries in a ledger)?	1- "Extremely important" (1- Category 5)
No time pressure	How often does your current job require you to meet **strict deadlines**?	1- "Every day" (1- Category 5)

O*NET Interests. The O*NET interest file data is based on the RIASEC model. The RIASEC categories are: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (see Table A.2 for a description of each from the O*NET). The data assigns, for every occupation, a score to each of these characteristics based on input from analysts. I use these scores directly as amenities. I follow the same procedure as with context file data for the crosswalks and imputation of missing data.

Table A.2: RIASEC categories interest file

Title	Description
Realistic	Work involves designing, building, or repairing of equipment, materials, or structures, engaging in physical activity, or working outdoors.
Investigative	Work involves studying and researching non-living objects, living organisms, disease or other forms of impairment, or human behavior.
Artistic	Work involves creating original visual artwork, performances, written works, food, or music for a variety of media, or applying artistic principles to the design of various objects and materials.
Social	Work involves helping, teaching, advising, assisting, or providing service to others.
Enterprising	Work involves managing, negotiating, marketing, or selling, typically in a business setting, or leading or advising people in political and legal situations.
Conventional	Work involves following procedures and regulations to organize information or data, typically in a business setting.

ATUS (123). I use the data on the disutility of work across occupations from (123), which they construct from survey responses in the ATUS in 2010, 2012, and 2013. They use survey responses about workers' feelings during working hours. They have responses on: how happy, how sad, how stressed, how tired, how much pain, and how meaningful. I do not use the responses on happiness or sadness with the concern that these are more directly influenced by the workers' wages. Their data already uses the same (16) occupation codes, and I do the same imputation as with the other datasets.

A.1.2 NLSY79

The NLSY79 is a long panel that tracks labor market activities and life events of a sample of Americans born between 1957 and 1964. I restrict attention to employed workers who worked at least ten hours a week and nine weeks since the last interview (as in (30)). I also drop those employed in the military or armed forces and the observations without the AFQT score. The AFQT test was conducted in 1980 on respondents who were at least 17 years old. So, I only use cohorts born between 1957 and 1962. This leaves me with 7919 observations in 1986, 6462 in 1996, 4961 in 2006, and 3947 in 2016. The data uses the census occupation codes, so I also use the (76) crosswalks.

A.1.3 Census and ACS data

I use data on employment, wages, occupations, and years of schooling from the census and the American Community Survey (ACS). I obtain the data from IPUMS and use the census samples from 1980, 1990, and 2000, and the ACS samples from 2006, 2009, 2012, and 2015. I closely follow (73) for sample selection and aggregation to occupation-level wage and employment measures. I restrict attention to employed individuals aged 18 to 64. I drop those in the armed forces, unpaid family workers, or those in institutional group quarters. For the years 2009 to 2015, the weeks worked need to be imputed based on the data from 2006.

For the aggregation, each individual's labor supply weight is computed with the person's sampling weight times the usual hours worked per week and times the weeks worked. With this weight, I compute the occupation's employment shares and the average hourly wages, yearly wages, and years of schooling. Finally, I also use the (76) crosswalks to crosswalk the data from each year's census codes to the *occ1990dd* codes of (16).

A.2 Details amenity pricing method

In Section A.2.1, I discuss in more detail Assumption 1. Section A.2.2 contains the estimates of the amenity prices. In Section A.2.3, I explain the different robustness exercises conducted.

A.2.1 Assumption of non-crossing frontiers

Assumption 1 essentially requires that the workers' frontiers of the offer sets do not cross. That is if $\overline{\phi}_i < \overline{\phi}_{i'}$, worker i' must earn more at every amenity level. This assumption is hard to test directly. In this section, I discuss intuitively under what conditions on the distributions of the θ_is and the wage functions $\{w_j(\theta)\}$ this assumption will or will not be satisfied.

Consider workers i and i' with $\overline{\phi}_i < \overline{\phi}_{i'}$ but close. As in the growth accounting, let $\mathcal{J}(\theta)$ denote the subset of occupations in the frontier of the offer set for type θ. Since i and i' have similar total compensation levels, I assume further that these sets coincide $\mathcal{J}(\theta_i) = \mathcal{J}(\theta_{i'})$. In this case, the frontiers do not intersect if $w_j(\theta_i) < w_j(\theta_{i'})$ for all $j \in \mathcal{J}(\theta_i)$. Typical specifications of the production imply that the wage function is equal to a common price w_j times the effective labor input of each type, that is, $w_j(\theta) = w_j g_j(\theta)$ for some function $g_j : \mathbb{R}^L \to \mathbb{R}_+$ with $\frac{\partial g_j}{\partial \theta_l} \geq 0$ for all $l \in \mathcal{L}$.[1] Then, to first order, the frontiers do not cross if

$$\sum_l \frac{\partial g_j(\theta)}{\partial \theta_l}|_{\theta=\theta_{i'}}(\theta_{l,i'} - \theta_{l,i}) > 0.$$

This condition helps understand under which conditions the frontiers will not cross. First, if the $\theta_l s$ are highly correlated for every worker. That is, if $\theta_{l',i'} > \theta_{l',i}$ for some l', we should also have $\theta_{l,i'} > \theta_{l,i}$ for (most) of the other l. For example, this would be the case if workers with cognitive skills, also tend to do better in manual or social tasks. Second, if there are a few $\theta_l s$ that are very important (i.e if $\frac{\partial g_j(\theta)}{\partial \theta_l}$ is large) in all occupations. For example, this would be satisfied if cognitive skills were important determinants of wages in all occupations.

One clear example where Assumption 1 would not be satisfied is in the basic Roy model. In the simplest case, there are two occupations $\mathcal{J} = \{1,2\}$ and two task-specific productivities $\mathcal{L} = \{1,2\}$, and each occupation uses only one of the two tasks. That is, the wage in occupation $j = 1$ depends only on the productivity in task $l = 1$, θ_1, and the wage in occupation $j = 2$ depends only on θ_2. By construction, the frontiers will cross in this case.

It is important to note that the similar slopes of the frontiers are only needed for workers with a similar total compensation level. Hence, these are, in a sense, local conditions. For example, if workers with high cognitive skills earn sufficiently more in some occupations. Their frontiers may never cross with the ones of low cognitive skill workers even if they have different slopes.

[1]This will be the case in equilibrium if all types θ are perfect substitutes in an occupations production function. That is, if $m^j(\{h^j(\theta)\}_\theta, r_j) = \widetilde{m}^j(H^j, r_j)$, where

$$H^j = \int g_j(\theta) h^j(\theta) d\theta.$$

A.2.2 Estimates

Table A.3: Coefficients of the main specification

	1986		1996		2006		2016	
γ_1	0.00185	(0.00165)	0.0151*	(0.00915)	0.0154	(0.0111)	0.0191	(0.0149)
γ_2	0.825***	(0.0870)	0.617***	(0.0562)	0.593***	(0.0645)	0.576***	(0.0699)
Contact	8.749***	(3.254)	1.872	(3.327)	5.943	(3.679)	7.089*	(4.090)
No discussions	-9.685***	(3.525)	-2.052	(3.712)	-5.588	(4.400)	-3.549	(4.952)
Teamwork	-4.230	(3.292)	-6.210*	(3.511)	-7.891**	(3.852)	-11.85***	(4.246)
Responsibility	8.265***	(2.575)	4.546*	(2.634)	1.489	(2.983)	5.367	(3.271)
No conflict	20.47***	(3.805)	12.25***	(3.882)	16.26***	(4.364)	19.30***	(4.916)
Indoors	0.392	(1.940)	1.313	(2.209)	2.718	(2.638)	3.749	(2.953)
No extreme temp.	-0.579	(3.554)	4.835	(3.882)	5.155	(4.745)	12.92**	(5.491)
No hazardous cond.	1.058	(3.620)	-4.302	(3.673)	2.646	(4.403)	-6.786	(5.124)
No contaminants	-0.131	(3.161)	9.177***	(3.295)	4.540	(3.739)	11.66***	(4.243)
No burns and cuts	11.15***	(3.381)	10.07**	(4.083)	4.184	(4.912)	-5.709	(5.766)
Time sitting	7.081**	(2.890)	3.725	(2.827)	-0.705	(3.215)	-8.317**	(3.562)
Time standing	0.999	(2.054)	2.569	(2.314)	-0.0572	(2.705)	-8.984***	(3.138)
No consequence error	5.145	(3.177)	-4.674	(3.270)	-3.225	(3.523)	0.160	(3.831)
Decision making	5.674*	(3.165)	-1.566	(3.299)	1.543	(3.696)	-1.203	(4.112)
Freedom decisions	-9.005***	(2.311)	-1.088	(2.621)	6.732**	(3.170)	-0.0516	(3.556)
No repetition	-9.961***	(2.999)	-3.649	(3.340)	-5.014	(3.766)	-16.05***	(4.246)
No time pressure	8.179***	(2.392)	4.842*	(2.598)	6.554**	(2.922)	1.508	(3.178)
Artistic	1.568***	(0.405)	2.063***	(0.420)	2.328***	(0.533)	2.874***	(0.601)
Conventional	-1.623***	(0.392)	-0.172	(0.440)	-0.312	(0.511)	-0.751	(0.566)
Enterprising	1.952***	(0.325)	1.368***	(0.325)	0.927**	(0.365)	0.241	(0.414)
Investigative	5.653***	(0.346)	3.605***	(0.358)	3.326***	(0.413)	2.632***	(0.456)
Realistic	-2.389***	(0.408)	-1.146***	(0.415)	-2.151***	(0.477)	-1.974***	(0.512)
Social	0.0136	(0.335)	0.261	(0.364)	-1.204***	(0.408)	-1.377***	(0.450)
Pain (ATUS)	-1.794***	(0.602)	-1.993***	(0.639)	-2.242***	(0.776)	-2.440***	(0.843)
Tired (ATUS)	0.836	(0.526)	-0.589	(0.572)	0.467	(0.661)	0.253	(0.739)
Stress (ATUS)	0.000706	(0.533)	0.433	(0.578)	-0.640	(0.698)	0.642	(0.784)
Meaning (ATUS)	0.107	(0.548)	-0.581	(0.597)	0.824	(0.686)	1.772**	(0.790)
Observations	7919		6462		4961		3947	
R^2	0.259		0.272		0.275		0.287	
Adjusted R^2	0.256		0.268		0.271		0.282	

Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Montecarlo simulations for standard errors of amenity and total compensation measures

In this section, I explain how I use Montecarlo simulations to estimate the standard deviation of the final amenity and total compensation measures. To do this, for every year (i.e., NLSY sample), I first draw $R = 10.000$ times all the coefficients from normal distributions with each coefficient's point estimate and standard error: $\{\gamma_1(r), \gamma_2(r), \{\pi_n(r)\}_n\}_r$. Note that the draws from the normal distributions are independent. For a few draws, the linear coefficient on wages $\gamma_1(r)$ can become negative, which implies that total compensation is decreasing in wages. For this reason, I also redo the analysis assuming $\gamma_1(r)$ follows a lognormal distribution.

Then, for every draw r, I compute the amenity and total compensation of each occupation as usual:

$$\hat{A}_j(r) = \sum_n \pi_n(r) A_{j,n}$$

$$\hat{\overline{\phi}}_j(r) = \gamma_1(r) \frac{w_j^{\gamma_2(r)}}{\gamma_1(r)} + \hat{A}_j(r).$$

Next, for every draw r, I compute the percentile of each occupation in the amenity and total compensation distributions, denoted by $Q_j^A(r)$ and $Q_j^{\overline{\phi}}(r)$. Finally, I aggregate each year's estimates with the age shares in 1980 as in the main measures. That is, letting $Q_j^{\hat{A}}(a,r)$ and $Q_j^{\overline{\phi}}(a,r)$ be the amenity and total compensation quantiles at draw r and occupation j, and at age group a (i.e., at a particular year in the NLSY79), compute for every r and j:

$$Q_j^{\hat{A},agg}(r) = \sum_a S^{1980}(a) Q_j^{\hat{A}}(a,r),$$

and similarly for $Q_j^{\overline{\phi},agg}(r)$.

For every occupation, I compute the standard deviations of the percentiles in the amenity and total compensation measures across the draws. Figure A.1 plots the resulting distributions of standard deviations. The figure shows that the standard deviations of the occupation's percentiles are quite small. The average standard deviation for the amenity percentile is 3.18, and for the total compensation percentile is 2.92 (8.66 if γ_1 is normal).

The main empirical results of Section 1.5 are based on the (ordinal) rankings of occupation by amenity and total compensation. Hence, ultimately, we want to assess the sensitivity of these rankings. To this end, for random pairs of draws r and r', I compute the correlation between the corresponding amenity and total compensation percentiles—i.e., I calculate $corr_j(Q_j^{A,agg}(r), Q_j^{\overline{\phi},agg}(r'))$ and $corr_j(Q_j^{A,agg}(r), Q_j^{\overline{\phi},agg}(r'))$. Figure A.2 plots the resulting distributions of correlations. The correlations are very high; the average correlation for the amenity ranking is 98.5%, and of the total compensation 98.7% (84.3% if γ_1 is normal). Therefore, even if the standard errors on the coefficients of some amenities can be large, the amenity and total compensation rankings—used for the main results of Section 1.5—are much more precisely estimated.

Figure A.1: Distributions of standard deviations of amenities and total compensation percentiles

(a) Amenity Percentile

(b) Total compensation percentile

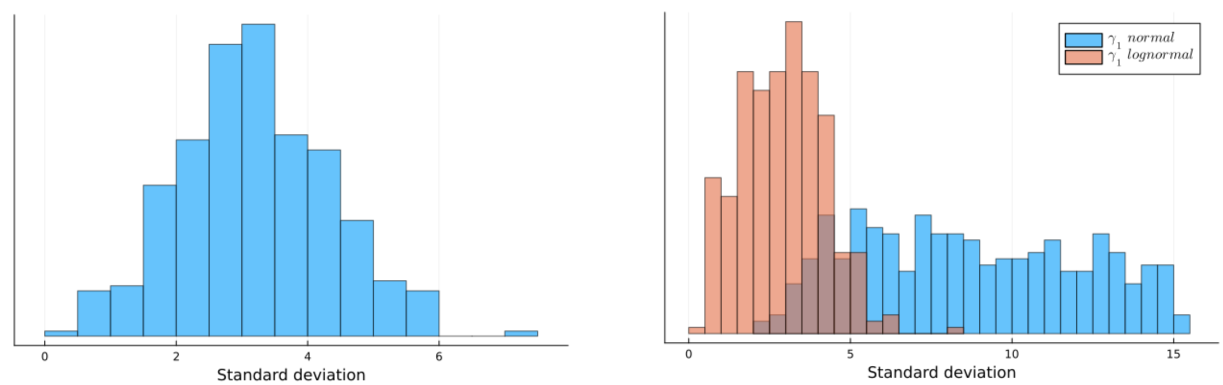
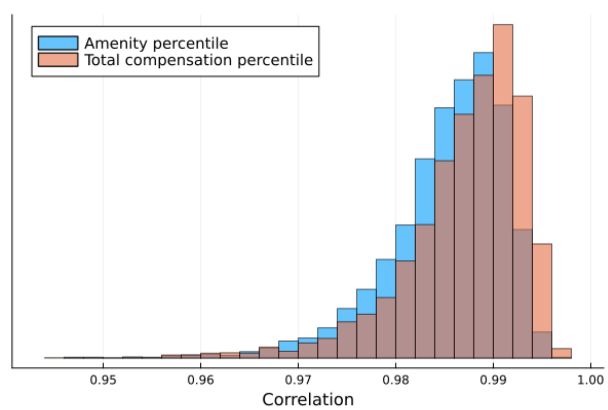


Figure A.2: Correlations amenity and total compensation percentiles (with γ_1 lognormal)



Sorting into low-amenity occupations based on consumption needs

In this section, I provide evidence that workers with high consumption needs (i.e., high marginal utility) tend to sort into low amenity occupations. Intuitively, this sorting is important from a welfare perspective because it implies that, all else equal, wage decreases in low-amenity occupations are relatively worse as they affect workers with high marginal utility.

Generally, given her offer set, a worker may prefer to go into the lower-amenity occupations for two main reasons. First, the worker may have a lower intrinsic preference for some amenities. Consider, for example, a young worker with a lower distaste for a physically demanding job. Second, she may choose a low amenity occupation because she values getting paid the compensating differential more due to higher consumption needs. Higher consumption needs can come, for example, from lower wealth or a larger number of children. (40) construct a measure of an occupation's intrinsic quality and document that children of richer parents are more likely to be employed in occupations with higher intrinsic quality. Similarly, (138) document that graduates with higher student debt accept jobs with higher wages and lower job satisfaction.

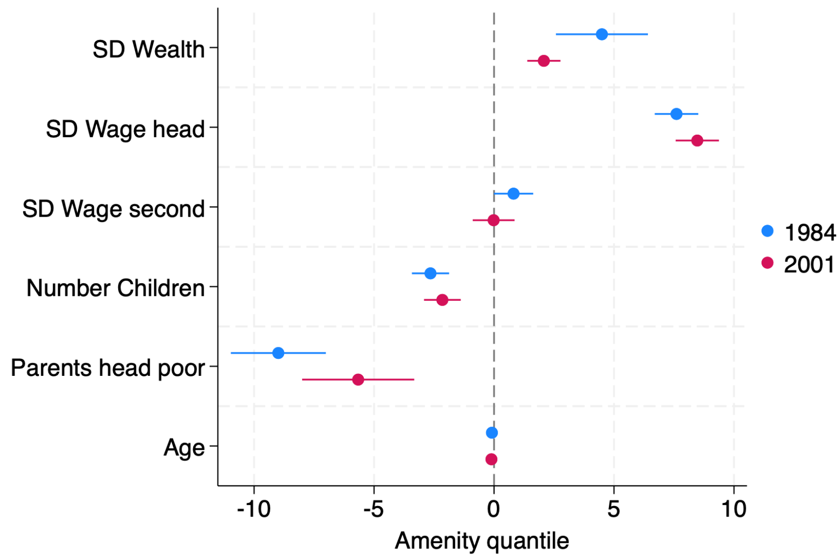
Using the constructed amenity measure, I provide further evidence of sorting by consumption needs. I use data from the Panel Study of Income Dynamics (PSID) as it provides better wealth data than the NLSY79 and covers a broader range of ages each year. Figure A.3 plots the coefficients of a regression of the amenity quantile of a worker's occupation on income, wealth, and controls for 1984 and 2001 data. Wealthier workers tend to be employed in occupations with higher amenities. Increasing wealth by one standard deviation leads, on average, to being employed in an occupation 4.5 (2.1 in 2001) percentage points higher in the amenity distribution. Moreover, a larger number of children or poor parents also predict sorting into low amenity occupations. The wage of the second earner also has a small positive effect in 1984 but not in 2001.

In Figure A.4, I repeat the same regressions but for the amenities of the occupation of the second earner. The results are similar, but now the wage of the first earner has larger effects on the amenity.

A.2.3 Alternative measures and robustness checks

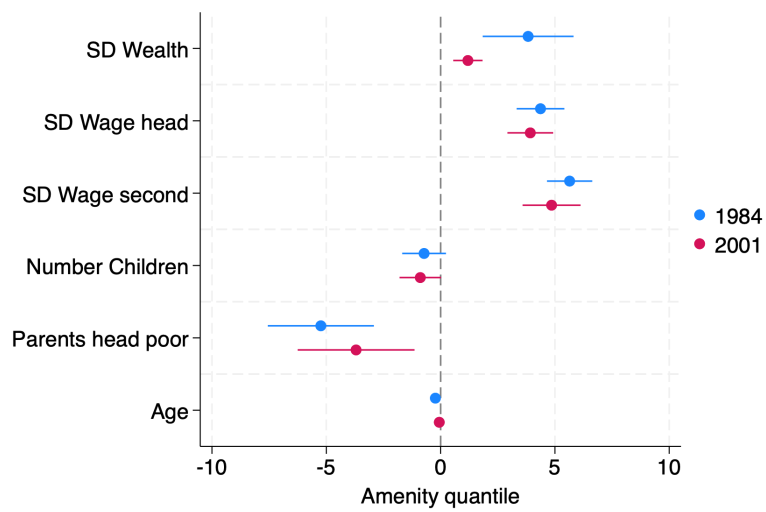
In this Section, I present the robustness checks with alternative ways of constructing the amenity and total compensation measures. In Subsection A.2.3, I explain the estimation with alternative specifications and in Subsection A.2.3 with the alternative proxies. The main results are robust in all cases. Moreover, the resulting amenity and total compensation measures are very similar, as we can observe in the correlation tables A.4 and A.5. In Section A.2.3, I explain how the presence of unionized workers could affect Assumption 1, and show that the results are similar if these workers are removed in the estimation.

Figure A.3: Regressions coefficients on worker's amenity



Note: The dots represent the point estimates and the lines the 95% confidence intervals.

Figure A.4: Regressions sorting into low amenity occupations for second earner



Finally, in Subsection A.2.3, I explore the robustness of the no polarization over total compensation to the size of the estimated compensating differentials.

Estimation with alternative skill proxies

In this section, I verify that the amenity and total compensation measures and the main results are robust to the choice of skill proxy used in the estimation. The AFQT is a good

Table A.4: Correlations table amenity measures

	Baseline	Linear	Polynomial	Interaction	Mastery	Self-esteem	Height	No union	No coverage
Baseline	1								
Linear	0.9993	1							
Polynomial	0.9997	0.9983	1						
Interaction	0.9994	0.9995	0.9987	1					
Mastery	0.8604	0.8662	0.8571	0.8587	1				
Self-esteem	0.9037	0.9060	0.9025	0.9008	0.9381	1			
Height	0.7989	0.7953	0.7989	0.7989	0.6740	0.7718	1		
No union	0.9992	0.9983	0.999	0.9989	0.8405	0.9140	0.8105	1	
No coverage	0.9988	0.9984	0.9984	0.9989	0.8438	0.9144	0.8118	0.9996	1

Table A.5: Correlations table total compensation measures

	Baseline	Linear	Polynomial	Interaction	Mastery	Self-esteem	Height	No union	No coverage
Baseline	1								
Linear	0.9994	1							
Polynomial	0.9988	0.9989	1						
Interaction	0.9996	0.9995	0.9985	1					
Mastery	0.9001	0.902	0.8971	0.8979	1				
Self-esteem	0.933	0.9336	0.9323	0.9308	0.9553	1			
Height	0.8499	0.8457	0.8437	0.8486	0.7528	0.8315	1		
No union	0.9992	0.9988	0.9988	0.999	0.8790	0.9367	0.8609	1	
No coverage	0.9989	0.9986	0.9981	0.9989	0.8817	0.9372	0.8627	0.9997	1

candidate for a proxy because it measures general aptitude skills and is predetermined, as the subjects took the test before they entered the labor market. However, a concern still remains because the AFQT score is a cognitive measure, and cognitive skills may be relatively more important in high-amenity occupations.

Hence, to assess the validity of the estimates, it is useful to examine how sensitive the results are to the choice of proxy. Specifically, whether they are robust if we use non-cognitive proxies. If two proxies satisfy Assumption 2, they should deliver similar results even if they measure different abilities.

I redo the estimation with three other proxies that I obtain from the NLSY79: mastery, self-esteem, and height (see https://www.nlsinfo.org/content/cohorts/nlsy79/topical-guide/attitudes). The Pearlin mastery scale is a measure of self-concept and references the extent to which individuals perceive themselves in control of forces that significantly impact their lives. The self-esteem proxy is based on the Rosenberg self-esteem scale which measures the degree of approval toward oneself. Finally, the height (which I first residualize on gender) can also be used as a proxy because it has been shown to correlate with earnings and is also predetermined (see (30)). For all the proxies, I use a linear specification as in (A.1) because the correlation between these proxies and wages is lower than for the AFQT.

As we can observe in Tables A.4 and A.5, the correlation between the resulting amenity and total compensation measures with each of the proxies is quite high, indicating that the different proxies deliver similar estimates.

Table A.6 shows that the average percentiles in the amenity distribution by the major occupation groups with each proxy. The values are quite similar with all the skill proxies. Moreover, the blue-collar occupations groups (Transportation/construction... and machine operators/assemblers) show up at the bottom of the skill distribution with all the proxies. This is also evidence that the AFQT score is a valid proxy. One specific concern would be that the cognitive skills measured in the AFQT score are relatively less important in these blue-collar occupations and that this leads to estimating that these are the lowest amenity occupations. The fact that these occupations are also the lowest amenity ones with the alternative proxies suggests that the estimates are robust to this concern.

Finally, figures A.5 and A.6 show the changes in employment and wages along the amenity and total compensation distributions, respectively, with each of the alternative proxies. The results are qualitatively robust. The only exception is the increase in employment at the bottom of the total compensation distribution for height, but the increase is very small.

Table A.6: Amenity percentiles by major occupation groups with alternative skill proxies

	AFQT (Benchmark)	Height	Self-esteem	Mastery
Managers/professionals/ technicians/finance/public safety	78.1	74	80.7	89.1
Production/craft	40	56.8	43.7	48.6
Transportation/construction/ mechanics/mining/ farm	23.7	29.7	25.9	29.3
Machine operators/assemblers	23.9	29.7	20.6	21.9
Clerical/retail sales	58.9	54.4	62.1	58.6
Service occupations	31.8	39.5	36.6	35.5

Note: Notice that the values with AFQT score do not coincide exactly with those in Table 1.4 because here I use a linear specification.

Figure A.5: Smoothed changes in employment and wages by amenity percentile with alternative skill proxies, 1980-2015

(a) Employment share (b) Hourly wages

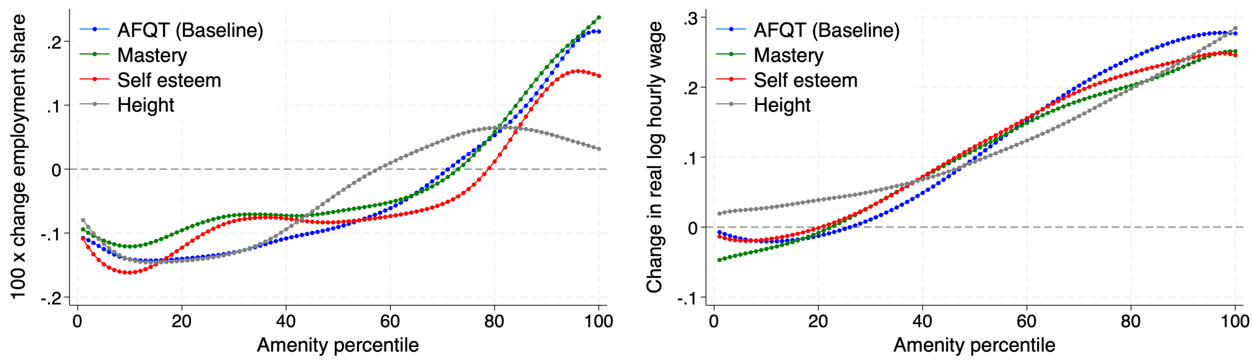
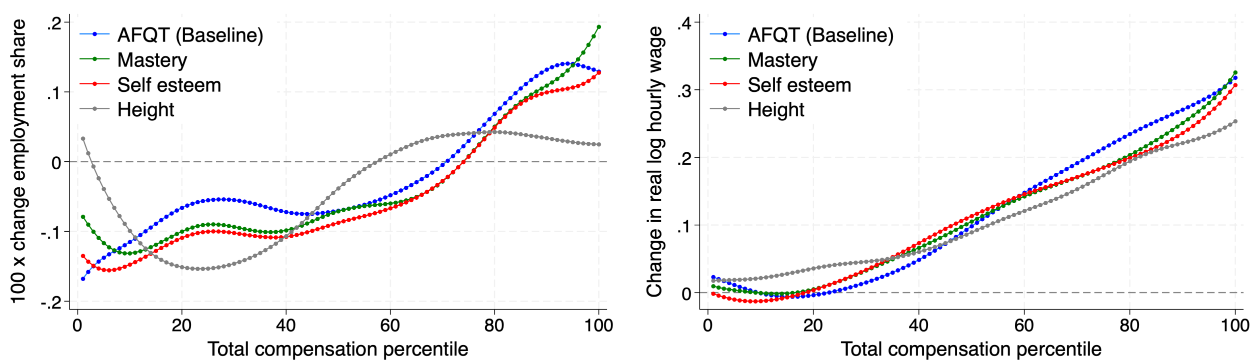


Figure A.6: Smoothed changes in employment and wages by total compensation percentile with alternative skill proxies, 1980-2015

(a) Employment share (b) Hourly wages



Alternative specifications

For robustness, I estimate several alternative specifications. I estimate a linear model

$$x_i = \eta + \gamma w_i + \sum_{n \in \mathcal{N}} \pi_n A_{j(i),n} + \epsilon_i, \tag{A.1}$$

a model with a third degree polynomial in the wage

$$x_i = \eta + \gamma_1 w_i + \gamma_2 w_i^2 + \gamma_3 w_i^3 + \sum_{n \in \mathcal{N}} \pi_n A_{j(i),n} + \epsilon_i, \tag{A.2}$$

and a model with an interaction between wages and the (aggregated) amenity

$$x_i = \eta + \gamma_1 w_i + (1 + \gamma_2 w_i) \sum_{n \in \mathcal{N}} \pi_n A_{j(i),n} + \epsilon_i. \tag{A.3}$$

I also estimate this last model with nonlinear least squares. I obtain similar results with all specifications as shown in Figure A.7 (for the amenity ranking) and in Figure A.8 (for the total compensation ranking). The correlation of the amenity and total compensation measures is also very high.

Figure A.7: Smoothed changes in employment and wages by amenity percentile with alternative specifications, 1980-2015

(a) Employment share (b) Hourly wages

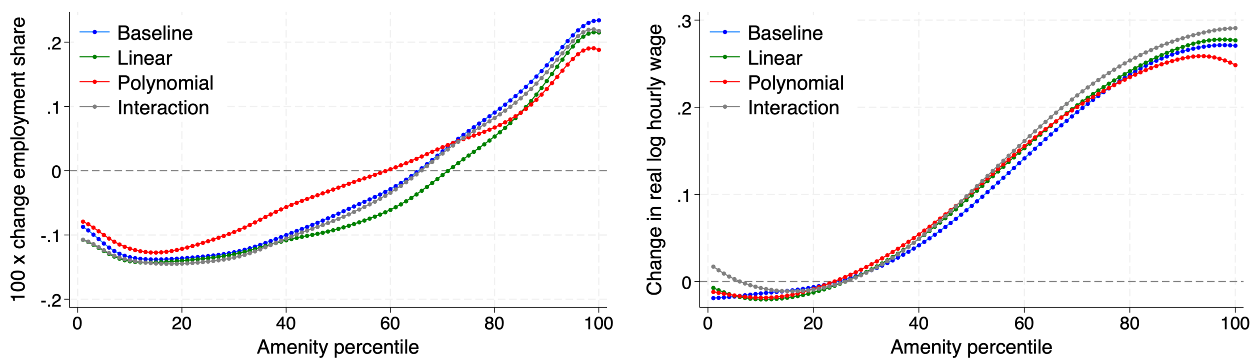
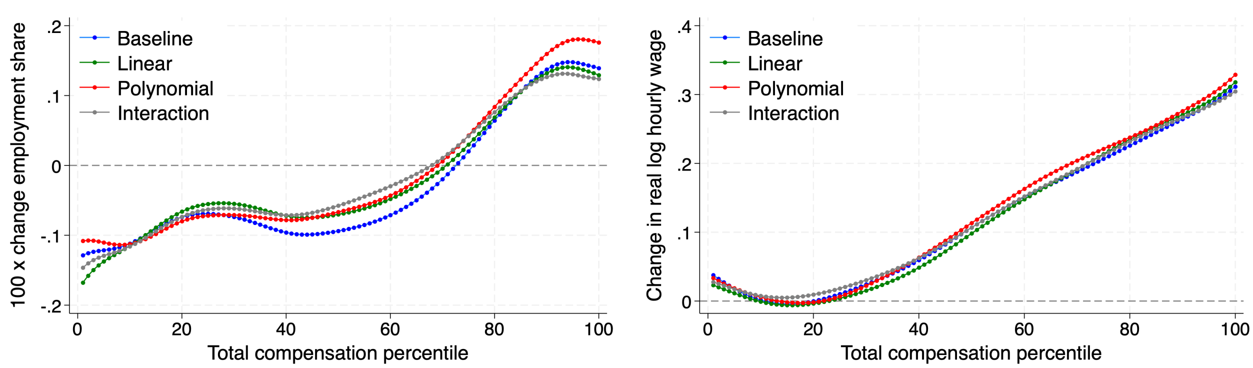


Figure A.8: Smoothed changes in employment and wages by total compensation with alternative specifications, 1980-2015

(a) Employment share (b) Hourly wages

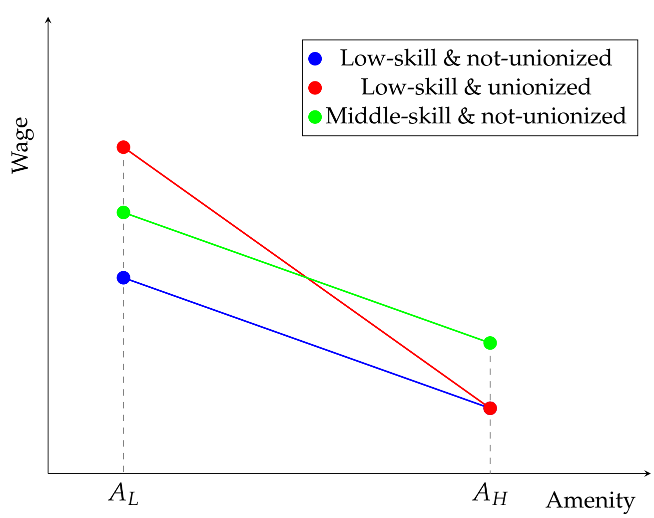


Removing unionized workers

Unionized workers earn a rent relative to their outside wages. Hence, we may worry that we could be measuring union rents with compensating differentials. The way in which unionization could affect the estimates is by violating Assumption 1 on the non-crossing frontiers. Unionized workers will earn higher wages at their current employment than in their outside options, which can imply that the slope of their frontiers is different from those of similarly skilled non-unionized workers. Moreover, a particular concern is that the blue-collar occupations, which I find to have the lowest amenities, also had the highest unionization rates.

Figure A.9 illustrates how unions could imply that Assumption 1 is not satisfied. There are two occupations, L and H, with $A_L < A_H$. Some workers in occupation L may be unionized, in which case they earn a higher wage than equally skilled workers. I assume that the frontiers are shifted up in parallel as the worker's skill increases, so Assumption 1 would always be satisfied without unions. In blue, I represent the frontier of a low-skill worker who is not unionized, and in red, the one of an equally skilled worker who is unionized in occupation L. In occupation L, the unionized worker earns a higher wage, but in occupation H, they earn the same wage. Hence, these two equally skilled workers face different amenity prices, and Assumption 1 is not satisfied. Moreover, the frontier of the unionized worker would also cross with those of slightly higher-skilled workers who are not unionized (as depicted by the green frontier).

Figure A.9: Unionization and crossing frontiers



The NLSY79 data contains information on whether workers are unionized that I can use to assess the sensitivity of the estimates. I re-estimate the amenity prices with the subset of workers that are not unionized and not covered.[2] Table A.7 compares the membership and coverage rates by major occupation groups in the NLSY79 with the aggregate data. The rates are of a similar magnitude, but they become relatively larger in the later periods for the NLSY79 because the sample contains older workers. I find that the estimates are not very sensitive to the exclusion of the unionized workers and, as a result, the amenity and total compensation measures are similar (see the correlation tables A.4 and A.5). Moreover, the main results on the amenity bias (Figure A.10) and the no polarization along the total compensation distribution (Figure A.11) are robust.

Table A.7: Comparison union membership and coverage (in parenthesis) rates in NLSY79 and aggregate data

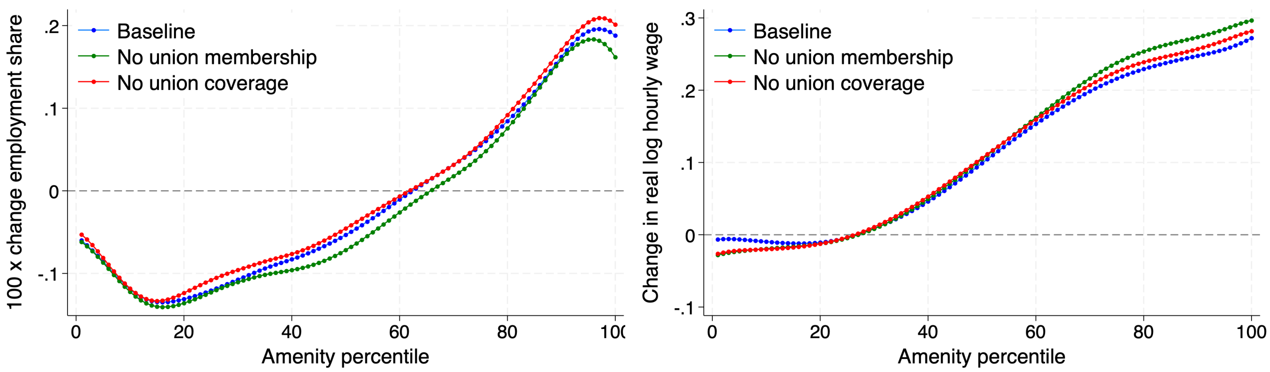
	1986		1996		2006		2016	
	NLSY79	Agg.	NLSY79	Agg.	NLSY79	Agg.	NLSY79	Agg.
Managers/professionals/ technicians/ finance/ public safety	0.08	0.15 (0.19)	0.13 (0.17)	0.13 (0.15)	0.16 (0.24)	0.13 (0.15)	0.17 (0.24)	0.11 (0.12)
Production/craft	0.19	0.24 (0.25)	0.18 (0.21)	0.13 (0.15)	0.2 (0.22)	0.15 (0.18)	0.22 (0.28)	0.11 (0.12)
Transportation /construction/ mechanics/mining/ farm	0.13	0.28 (0.3)	0.2 (0.22)	0.25 (0.27)	0.24 (0.28)	0.25 (0.26)	0.25 (0.3)	0.17 (0.18)
Machine operators/ assemblers	0.15	0.33 (0.35)	0.25 (0.27)	0.22 (0.23)	0.27 (0.31)	0.28 (0.29)	0.27 (0.31)	0.08 (0.09)
Clerical/retail sales	0.08	0.14 (0.17)	0.13 (0.17)	0.1 (0.11)	0.15 (0.2)	0.1 (0.12)	0.17 (0.21)	0.08 (0.09)
Service occupations	0.09	0.11 (0.13)	0.12 (0.14)	0.09 (0.1)	0.16 (0.2)	0.08 (0.09)	0.18 (0.23)	0.06 (0.07)

Note: The aggregate data on union membership and coverage rates by occupation is from https://www.unionstats.com/ (data described in 115 and 139).

[2] Covered workers are those who are under a collective bargaining agreement but who are not members of a union.

Figure A.10: Smoothed changes in employment and wages by amenity percentile with prices estimated without unionized workers, 1980-2015

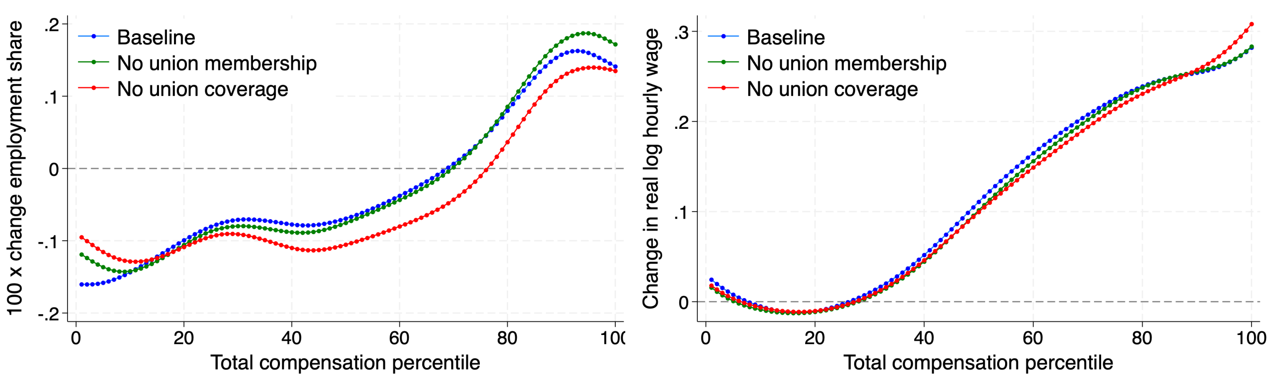
(a) Employment share (b) Hourly wages



Note: The measure *no union membership* (in green) is constructed by dropping all the union members in the estimation. The *no union coverage* (in red) drops all the union members as well as the workers covered by a union but who are not members.

Figure A.11: Smoothed changes in employment and wages by total compensation percentile with prices estimated without unionized workers, 1980-2015

(a) Employment share (b) Hourly wages



Rescaled compensating differentials

As discussed in Section 1.5.2, the total compensation measure is a combination of the wages and the amenity measure, where the coefficients on income γ_1 and γ_2 determine the importance of each. If the amenity prices/compensating differentials are high, the total compensation will be close to the amenity measure, and conversely. As a result it could be that the no polarization along the total compensation distribution documented is a result of overestimating the compensating differentials.

A direct approach to tackle this concern is to reconstruct the total compensation measure with re-scaled compensating differentials and check if the no-polarization of Figure 1.6 is robust. More concretely, we can re-compute total compensation as

$$\hat{\hat{\phi}}_j = \hat{\gamma}_1 \frac{w_j^{\hat{\gamma}_2}}{\hat{\gamma}_2} + \alpha \hat{A}_j,$$

for some factor $\alpha \in (0,1)$. I do this in Figure A.12 with $\alpha \in \{0.25, 0.5, 0.75\}$. The result is robust for $\alpha = 0.75$ and $\alpha = 0.5$, meaning that we can reduce the estimated compensating differentials by half, and there would still be no polarization. The picture is slightly less clear if we reduce the compensating differential four times ($\alpha = 0.25$). There is still a mostly monotone relationship for wages, but we start to observe some polarization in employment.

Table A.8 shows the resulting average total compensation percentile for each α by major occupation groups. As expected, as α decreases, the average total compensation percentile of the blue-collar occupations increases and of the services decreases. However, because there is still sufficient overlap and the employment share of the blue-collar occupations was much larger, we can still obtain that there is no polarization on aggregate with $\alpha = 0.5$, as shown in Figure A.12.

Figure A.12: Smoothed changes with rescaled compensating differentials, 1980-2015

(a) Employment share

(b) Hourly wages

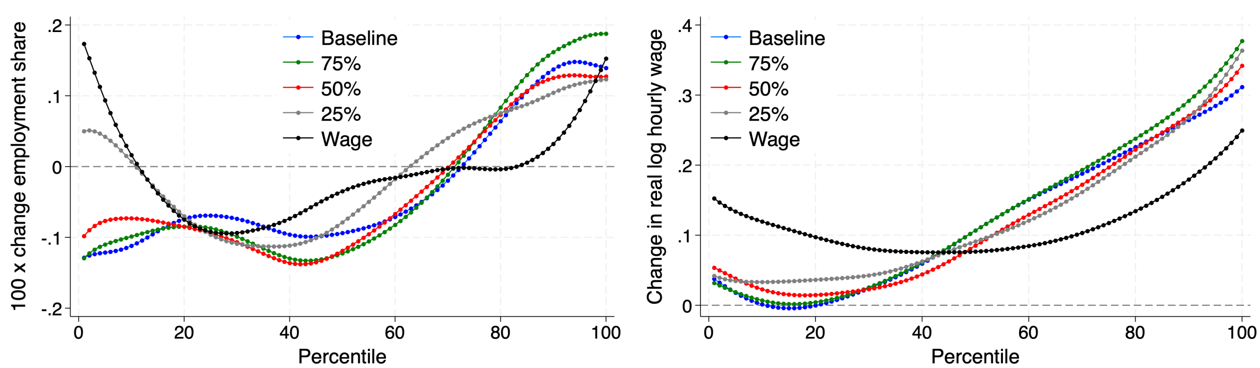


Table A.8: Total compensation with rescaled compensating differentials by major occupation groups

	Wage percentile (1980)	Amenity percentile	Total compensation percentile			
			Baseline ($\alpha = 1$)	$\alpha = 0.75$	$\alpha = 0.5$	$\alpha = 0.25$
Managers/professionals/ technicians/ finance/ public safety	73.9	78	80.1	80.5	80.8	80.5
Production/craft	70.0	39	46	48	51.4	57.8
Transportation/construction/ mechanics/mining/ farm	52.4	23.1	26.7	28	30.6	36.4
Machine operators/assemblers	36.7	23.8	23.1	23.2	24	26.4
Clerical/retail sales	30.6	59.2	55.8	54.4	51.7	44
Service occupations	11.0	32.8	24	21.6	17.6	12.2

A.3 Gender differences in amenity prices and sorting

In this subsection, I explore the gender differences in the amenity price estimates and how they affect the results. As discussed in the main text, since the estimation consists of measuring amenity prices, they should not be affected by gender differences in preferences for amenities. That is, even though women may have a higher preference for amenities than men, they may face the same market prices. With a hedonic pricing approach, preferences can be recovered in a second step using choice data (i.e., with employment shares), as was shown by (159). However, it may still be the case that men and women tend to face different amenity prices. For instance, men may be better at physical tasks—that tend to have low amenities—which gives them a comparative advantage in low-amenity occupations. Hence, in this case, men could face larger amenity prices despite valuing the amenities less. Then, the sorting into different occupations by gender will depend on both the differences in amenity prices and preferences.

To assess the sensitivity of the results to gender differences in amenity prices, I redo the estimation and the construction of the amenity and total compensation measures separately by each gender. Since the samples are reduced, I use a linear specification as in Equation (A.1). Figure A.13 plots the amenity value of every occupation for men against the value for women, both normalized by the lowest amenity value of each gender. Note that this normalization is consistent with the definition of total amenity prices used in the growth accounting (Section 1.6.3). The figure shows that the gender differences in the amenity values are small. The correlation across occupations is equal to 95.8%. The amenity values are indeed slightly higher for men, which is consistent with men earning higher wages, but also with men having a comparative advantage in low-amenity occupations.

Table A.9 computes the average log hourly wages, the average employment share, and the average percentiles in the amenity and total compensation distributions by gender across the major occupation groups. Men's wages are higher in all occupation groups. The largest relative difference is in *Machine operators/assemblers* and *Production/craft*, and the smallest in *Service occupations*, which is consistent with low-skill men having a comparative advantage in low-amenity occupations.[3] As is well known, blue-collar occupations predominantly employ men, and women tend to sort into clerical and service occupations. This sorting is consistent with both different comparative advantages and differences in preferences for amenities. The last columns report the average amenity and total compensation (computed with the occupation's average wage for the corresponding gender) percentiles. The amenity and total compensation percentiles of men and women are similar in all occupations.

Given the large differences in employment across major occupation groups, it is important to understand how labor market changes differ within each gender. Figure A.14 plots the changes in employment and wages across the wage, amenity, and total compensation distributions, separately for men and women. That is, I redo all the steps for each gender so, for example, the changes in employment share are computed with the employment shares within each gender. Qualitatively, all the main changes—wage polarization, amenity-biased reallocation, and no polarization by total compensation—go through for each gender. Perhaps the only exception is the smaller or negligible declines in employment at the lowest-amenity occupations for men. This is consistent with the fact that the aggregate employment share in blue-collar occupations decreased partly because more women entered the labor market and went into other occupations. Hence, the decline in the employment share in blue-collar occupations among men is smaller than in aggregate. Moreover, as shown in Table A.9, the average amenity percentiles of blue-collar and service occupations are much closer for men.

[3]Notice that the relative wage difference in *Transportation/construction...* is actually small. However, women's employment share in these occupations is also very small, suggesting that only a few women with relatively high wages sort into these jobs.

Figure A.13: Differences in amenity value for men and women (relative to lowest amenity occupation) 2006-2022

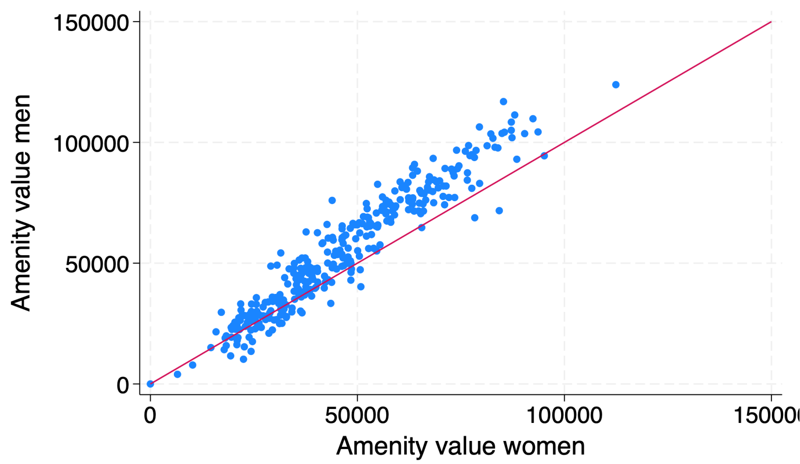


Table A.9: Wages, employment, and amenity and total compensation percentiles by gender across major occupation groups

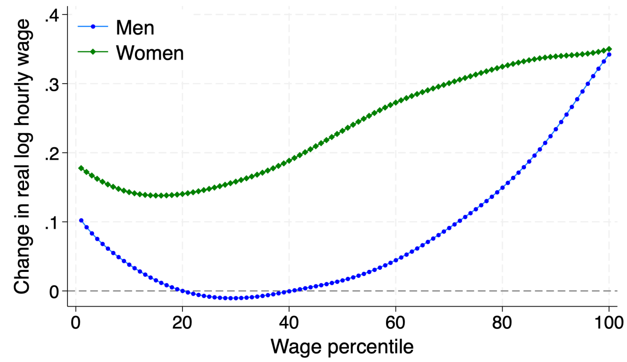
	Log hourly wages (1980)		Employment share (1980, in %)		Amenity percentile		Total compensation percentile	
	Men	Women	Men	Women	Men	Women	Men	Women
Managers/professionals/ technicians/finance/public safety	3.2	2.8	35	29.7	77.3	77.8	80.3	78
Production/craft	3	2.5	7.2	1.8	34.2	40.1	39.3	40.1
Transportation/construction/ mechanics/mining/ farm	2.8	2.5	28.9	3.7	23.2	19.5	24.7	19.3
Machine operators/assemblers	2.9	2.4	10.3	10	24.9	20.5	25.3	20.0
Clerical/retail sales	2.9	2.5	12.2	40.9	56.7	58.7	57.1	57.1
Service occupations	2.5	2.2	6.3	13.9	25.5	33.4	21.7	28.2

Note: To compute the total compensation, I use the occupation's average wages in 1980 of the corresponding gender. For all the variables (hourly wages, amenity percentile, and total compensation percentile) I compute the averages weighting by the employment shares of the corresponding gender.

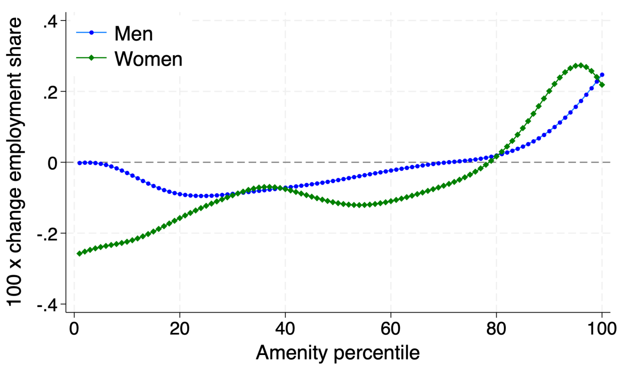
Figure A.14: Smoothed changes in employment and wages by wage, amenity and total compensation percentile separately by gender, 1980-2015

(a) Employment share by wage percentile

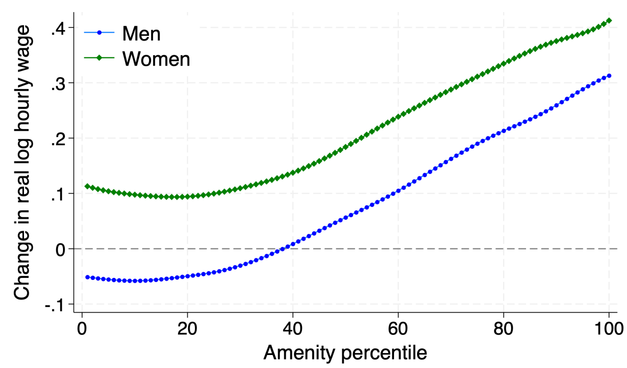
(b) Hourly wages by wage percentile



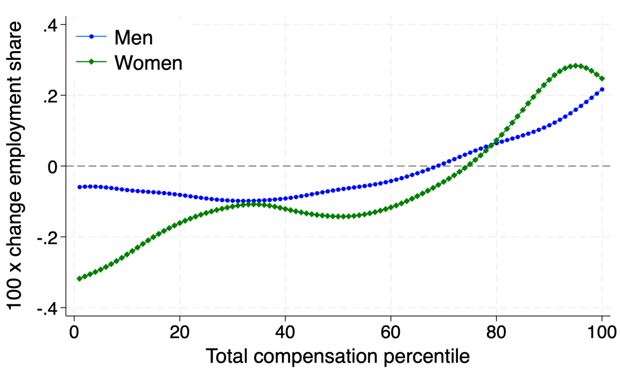
(c) Employment share by amenity percentile



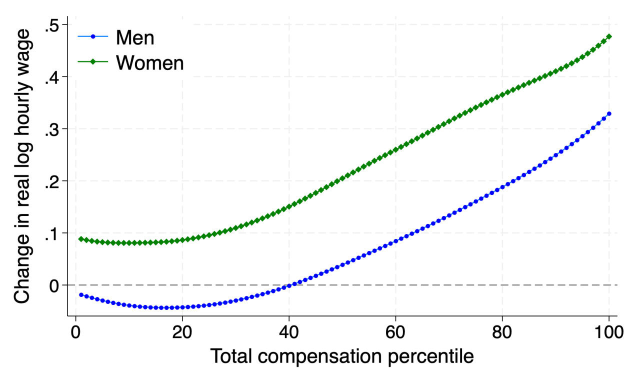
(d) Hourly wages by amenity percentile



(e) Employment share by total compensation percentile



(f) Hourly wages by total compensation percentile



A.4 Evolution of amenities within occupations

A caveat of the measurement exercise is that I am not able to measure the evolution within occupations over time and, in particular, that I do not have amenities data in 1980. For some amenities, this is a reasonable assumption as they are inherent to the task involved in the occupation. This may be the case for the RIASEC characteristics of the O*NET interests file. For example, whether an occupation is artistic or not should not change much over time. For other amenities, such as those related to job safety, we should expect some within-occupation improvements over time.

Change in amenities in O*NET context file. Starting from 2006, I can use different releases of the O*NET context file to measure changes in amenities within occupation (as I do for the growth accounting exercise) and assess the sensitivity and robustness of the results to changes in amenities. Table A.10 shows the absolute and relative changes in the amenities of the context file between 2006 and 2022. As expected, we observe improvements in most amenities.

Next, I use this data to assess the sensitivity of the results to changes in amenities. I re-estimate the amenity prices and re-construct the amenity and total compensation measures with the amenities data from 2022. Figure A.15 plots the percentiles in the amenity and total compensation distribution of each occupation with the 2006 amenities data against the percentiles with the 2022 data. In both cases, the percentiles are similar (the correlation is 0.97 in both), indicating that the rankings are not very sensitive to these changes. Figure A.16 compares the changes in the amenity value (with linear prices) with the 2006 and 2022 data. The amenity value is generally lower with the 2022 data. Hence, although the rankings used for the main empirical results appear quite robust, the monetary values—used for the growth accounting—may be more sensitive. Notice that if amenities improve uniformly across occupations, the (ordinal) amenity measure should, a priori, not change much. However, extrapolating, we may expect that the value of amenities would be larger with amenity data from 1980 instead of 2006, which would lead to larger growth in total compensation and productivity in the growth accounting part. Finally, Figure A.17 plots the smoothed changes in employment and wages by amenity and total compensation percentile estimated with the 2006 and 2022 data. All the results are similar and robust.

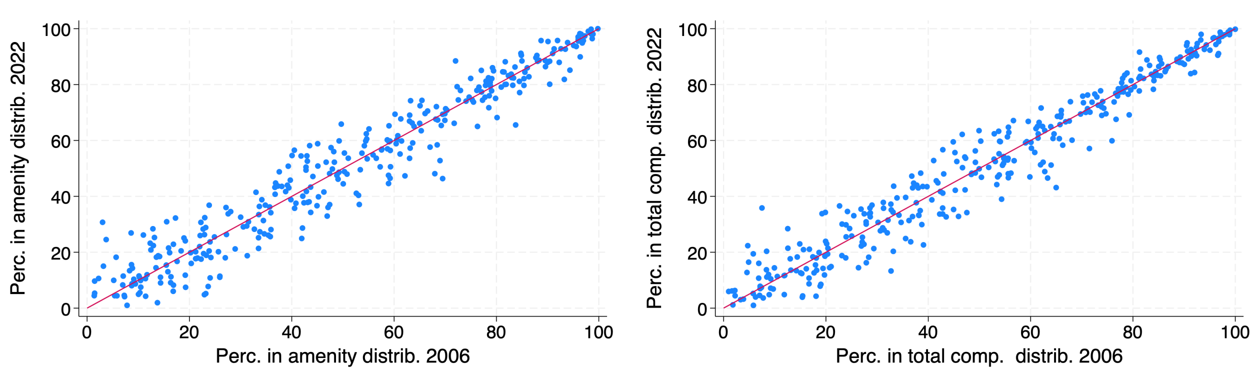
Table A.10: Changes in amenities of the O*NET context file 2006-2022

	Correlation	Avg. absolute change (100x and normalized)	Avg. percentage change
No burns and cuts	0.79	0.98	2.2%
No conflict	0.51	0.12	2.2%
No consequence error	0.7	0.13	3.75%
Contact	0.64	-1.96	4.7%
Contaminants	0.83	2.9	11.1%
Decision making	0.5	-3.8	4.9%
Discussions	0.36	3.9	115.8%
No extreme temperatures	0.78	0.1	1.5%
Freedom decisions	0.49	-13.61	1.8%
No hazardous conditions	0.8	1.04	4.3%
Indoors	0.84	2.74	17.2%
No repetition	0.6	-1.4	1.5%
Responsibility	0.48	4.85	20.6%
Teamwork	0.44	0.64	18.1%
No time pressure	0.54	-3.84	3.7%
Time sitting	0.85	3.92	194.1%
Time standing	0.86	4.23	131.2%

Note: The absolute change is normalized by the mean of each amenity across occupations. More concretely, for every amenity n and occupation j, I compute, $100 \times (\frac{A_{n,j,2022} - A_{n,j,2006}}{J^{-1} \sum_j A_{n,j,2006}})$, and then I average over occupations, that is $100 \times (\frac{\sum_j A_{n,j,2022}}{\sum_j A_{n,j,2006}} - 1)$. The average percentage change is, for every n, $J^{-1} \sum_j (\frac{A_{n,j,2022}}{A_{n,j,2006}} - 1) \times 100$.

Figure A.15: Changes in amenity and total compensation percentiles 2006-2022

(a) Amenity percentile (b) Total compensation percentile



Note: The correlation is equal to 0.97 for both the amenity and total compensation percentiles.

Figure A.16: Changes in amenity value (with linear specification and relative to lowest amenity occupation) 2006-2022

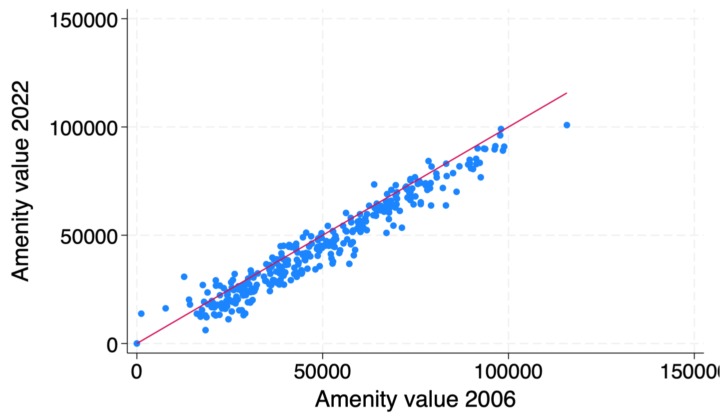
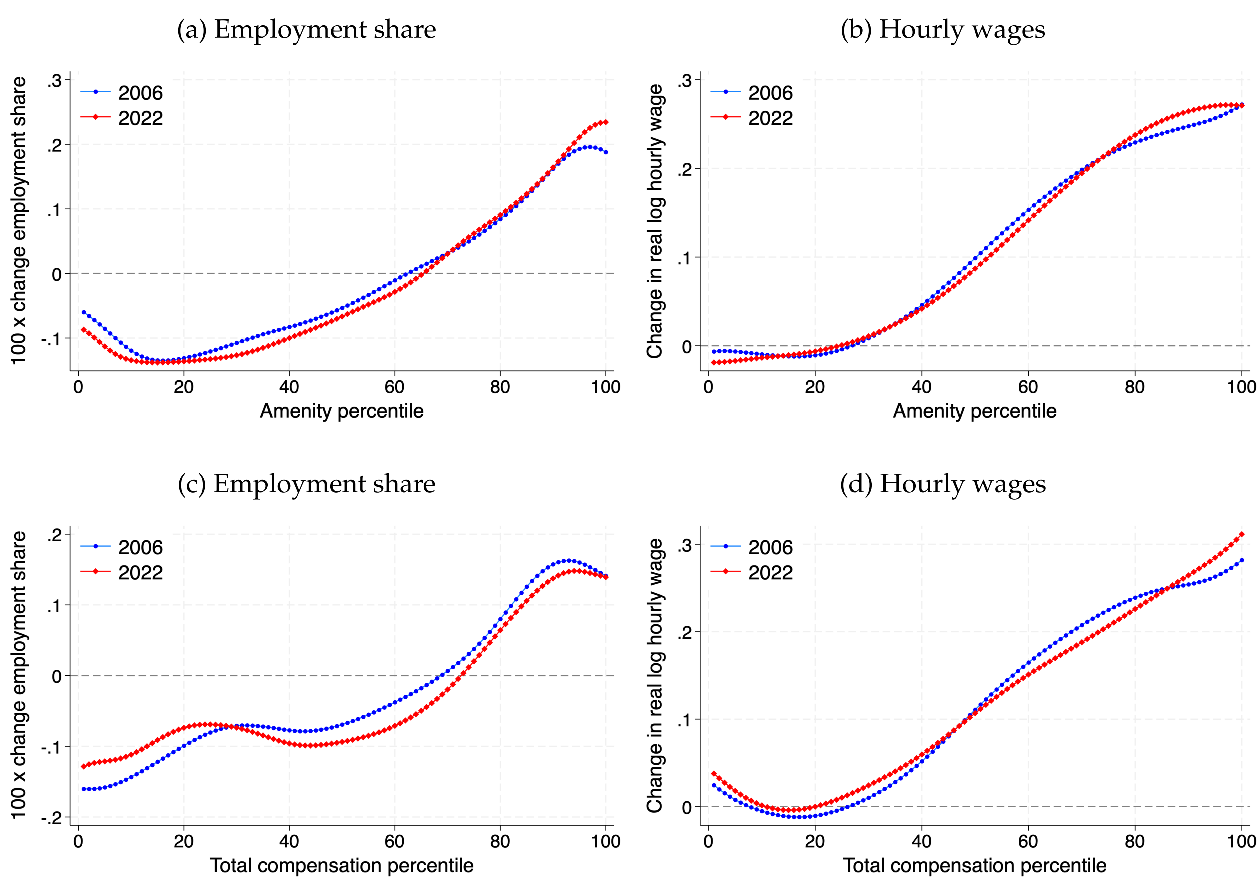


Figure A.17: Smoothed changes with 2006 and 2022 amenities data, 1980-2015



Evolution of job safety and injury rates. Data on injury rates can also help infer improvements within occupations in earlier periods. First, I use survey responses from the

NLSY79 to compute average injury rates by occupation groups between 1988-1990 and 1998-2000. This is the same data used by (110). Figure A.18 plots the average injury rates by occupation groups for both periods. The figure shows that, indeed, the injury rates for most occupation groups decreased (though it increased slightly for some). Notice, however, that because these occupation groups are quite aggregated, the decreases in injury rates can also be due to workers reallocating across different occupations within these groups.

In Figure A.19, I use data on average injury and illness rates by sector from the Survey of Occupational Injuries and Illnesses (SOII). This is the same data as I used in Figure 1.1, and is constructed from historical news releases of the SOII. Panel a) shows the changes in injury rate by sector between 1992 and 2002, and Panel b) between 2003-2015. Injury rates decreased in both periods and across all sectors. Again, this is only illustrative as these declines may be (partly) explained by reallocations across occupations within sectors.

Figure A.18: Changes in injury rates by occupation group (NLSY79)

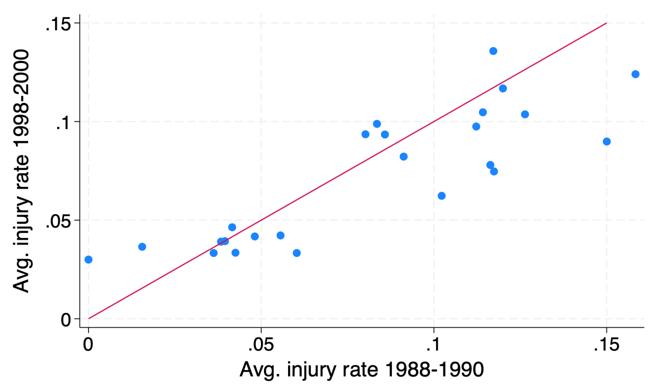
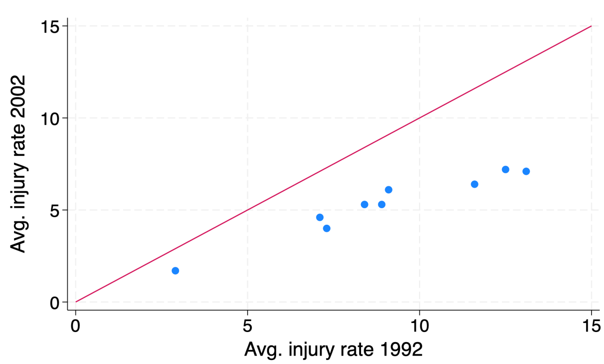
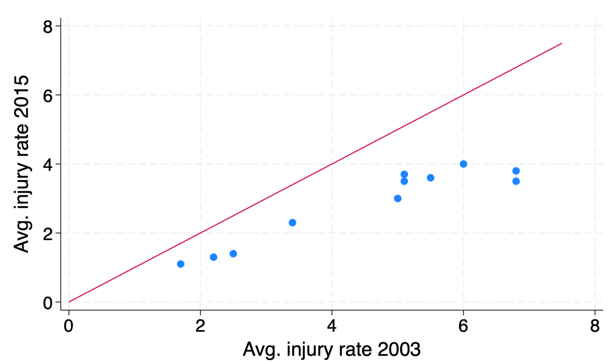


Figure A.19: Changes in average injury rates by sector

(a) 1992-2002



(b) 2003-2015



A.5 Extra figures

Table A.11: Alternative (non-cognitive) skill proxies by major occupation groups

	Wage (1980)	AFQT	Years schooling (1980)	Self-esteem	Mastery	Height (Residualized by gender)
Managers/professionals/ technicians/finance/public safety	73.9	79	80.2	71.5	61.9	61.5
Production/craft	70.0	41.2	38.6	51.3	51.2	67.1
Transportation/construction/ mechanics/mining/ farm	52.4	29.7	25.5	29.9	40.9	39.9
Machine operators/assemblers	36.7	16.9	17.3	28.1	29.2	31.6
Clerical/retail sales	30.6	55.5	57	53.3	56.2	47.6
Service occupations	11.0	23	24.7	34.2	34.5	30

Figure A.20: Changes in employed and wages by wages and skill proxies, 1980-1990, 1990-2000 and 2000-2015

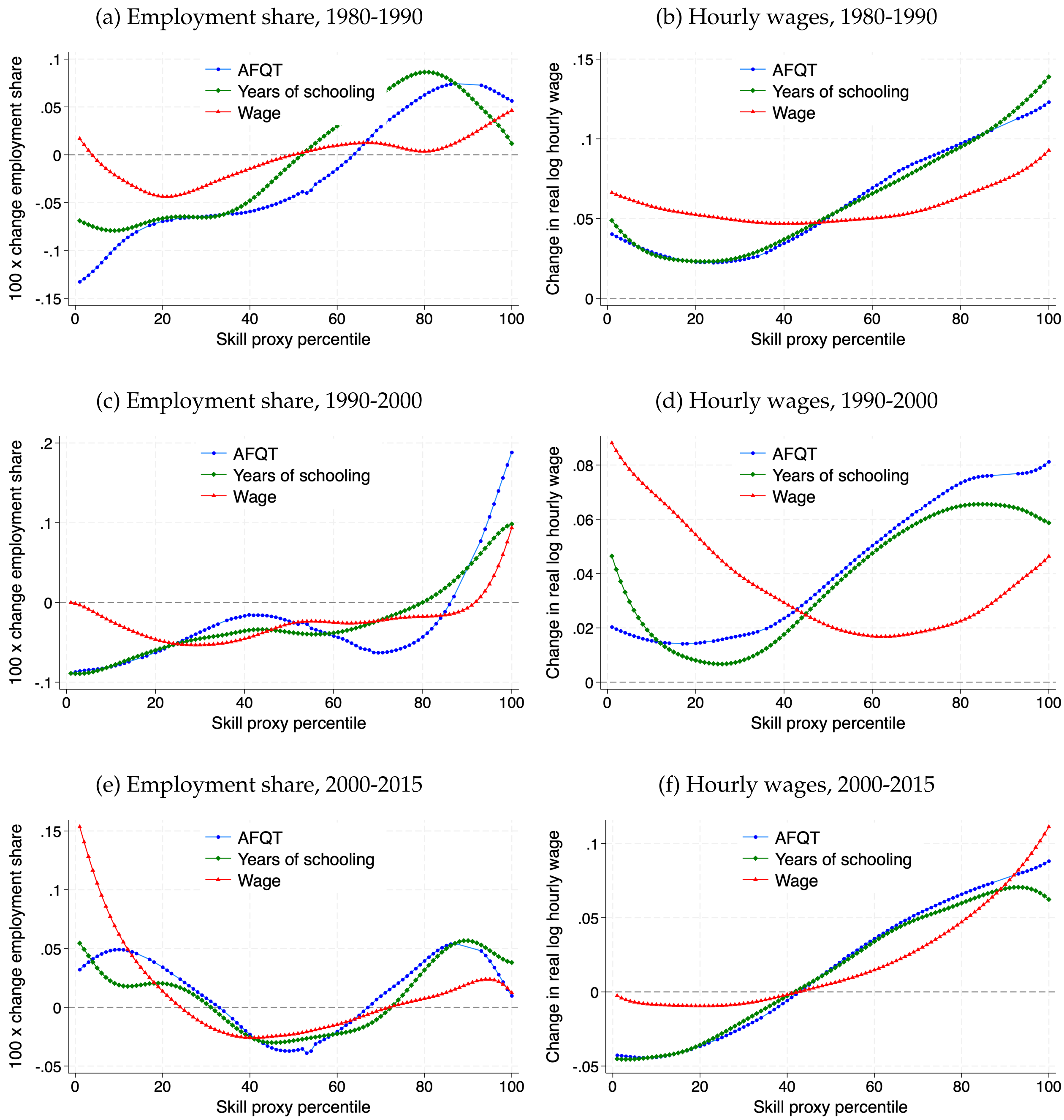
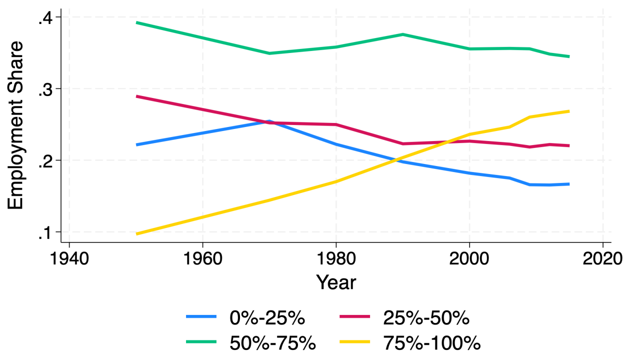


Table A.12: Employment and wage changes by occupation groups

Occupation group	Wage percentile (1980)	Amenity percentile	Total compensation percentile	(100x) Change employment share (1980-2015)	Change log. hourly wage (1980-2015)
Managerial and Professional Specialty Occupations:					
Executive, administrative and managerial	91	77.2	83.4	2.2	0.28
Management Related	75.9	76.5	78.6	1.6	0.3
Professional Specialty	65	85.1	83.9	5.9	0.24
Technical, Sales, and Administrative Support Occupations:					
Technicians and Related Support	54.2	71.6	70.7	0.19	0.33
Retail Sales Occupations	41.4	60.6	58.6	-0.3	0.06
Financial Sales and Related Occupations	77.3	70.8	75.3	1.3	0.14
Administrative Support (clerical)	26.9	58.8	54.9	-3.68	0.11
Service occupations:					
Housekeeping and Cleaning	2.7	8.6	2.7	-0.09	0.16
Protective Service	59.4	50.8	53.9	0.54	0.16
Healthcare Support	6.2	44	37.4	1.4	0.13
Building, Grounds Cleaning, Maintenance	20.4	11.9	9.8	0.27	-0.03
Personal Appearance	9.5	61.6	52.6	0.19	-0.03
Recreation and Hospitality	19	48.5	43	0.3	0.03
Child Care Workers	0.3	50.8	32.8	0.13	0.2
Misc. Personal Care and Service Occupations	23.4	43.5	37.4	0.1	0.08
Farming, Forestry, and Fishing Occupations:					
Farm Operators and Managers	39	58.5	57.4	-0.24	0.8
Other Agricultural and Related	13.5	20.7	15	-0.18	0.14
Precision Production, Craft, and Repair Occupations:					
Mechanics and Repairers	62.8	26.9	32.8	-1.2	0.02
Construction Trades	63.3	29.8	36.3	-0.9	-0.06
Extractive Occupations	71.0	25.6	33.3	-0.09	0.05
Precision Production	70	39	46	-2.5	-0.11
Operators, Fabricators, and Laborers:					
Machine Operators, Assemblers, and Inspectors	36.6	23.8	23.1	-6.2	0.0
Transportation and Material Moving	45.8	16.3	18.4	-1.4	-0.07

Figure A.21: Evolution of employment and wages by quartiles of amenity distribution

(a) Employment share



(b) Hourly wages

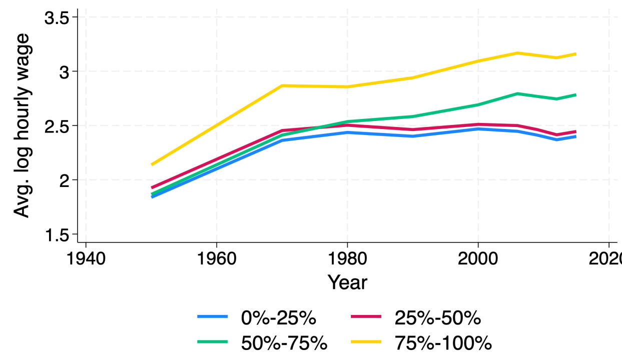
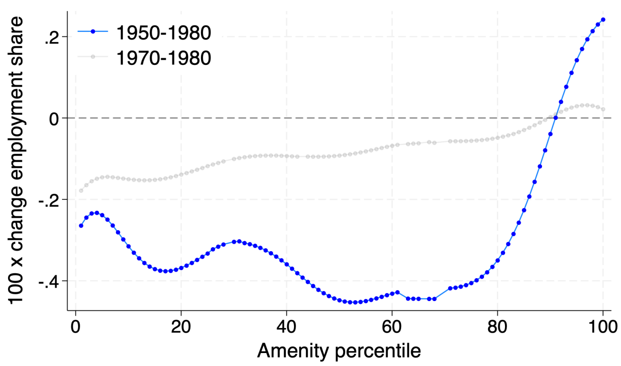


Figure A.22: Amenity bias 1950-1980

(a) Employment share



(b) Hourly wages

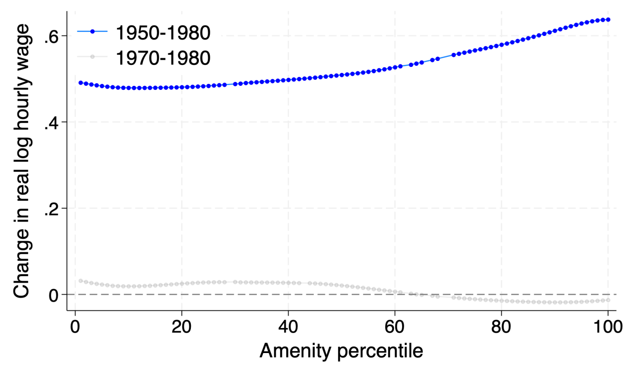
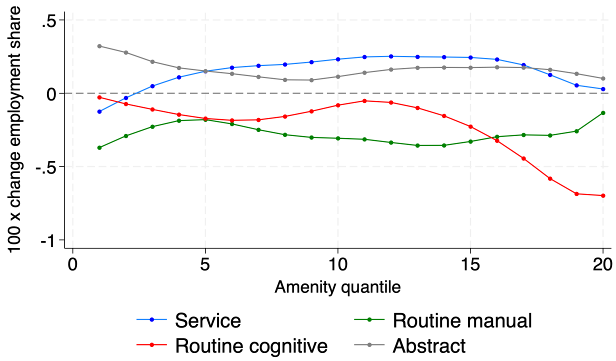
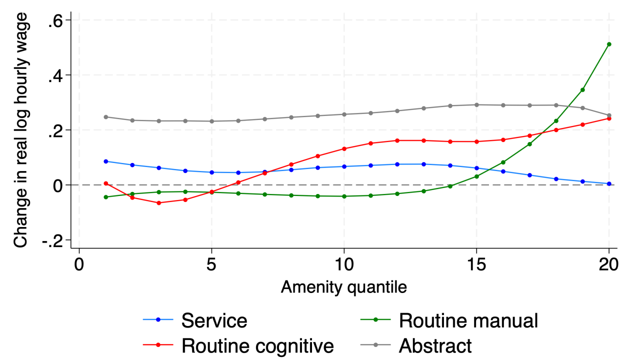


Figure A.23: Amenity bias within major occupation groups

(a) Employment share



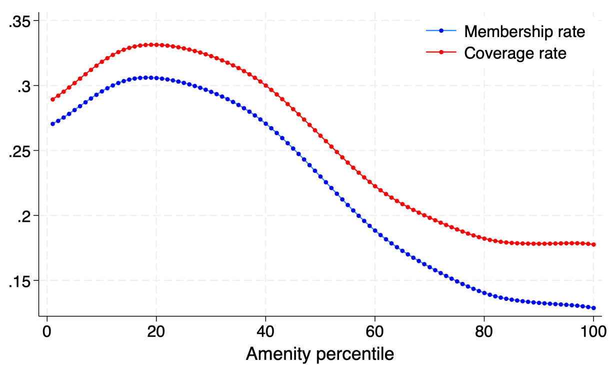
(b) Hourly wages



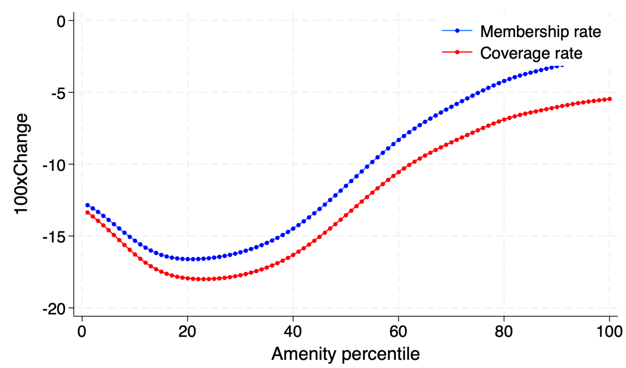
Note: Routine cognitive refers to the major occupation group *clerical/retail sales*; routine manual to blue-collar occupation groups *production/craft*, *transportation/construction/ mechanics/...*, and *machine operators/assemblers*; and abstract refers to *managers/professionals/ technicians/....*

Figure A.24: Union membership and coverage rates by amenity percentiles

(a) Rates in 1983



(b) Change rates between 1983 and 2015



Note: The aggregate data on union membership and coverage rates by occupation is from https://www.unionstats.com/ (data described in 115 and 139).

Table A.14: Union membership and coverage by occupation groups

Occupation group	Membership rate (1983)	Coverage rate (1983)	Change membership (1983-2015)	Change coverage (1983-2015)
Managers/professionals/ technicians/finance/public safety	0.14	0.18	-2.8	-5.3
Production/craft	0.31	0.34	-16.2	-17.6
Transportation/construction/ mechanics/mining/ farm	0.34	0.36	-18.5	-19.6
Machine operators/assemblers	0.36	0.38	-22.6	-24.1
Clerical/retail sales	0.19	0.23	-5.6	-8.2
Service occupations	0.14	0.16	-5.8	-6.8

A.6 Details growth accounting

A.6.1 Theoretical results

First, it is convenient to define the workers' problem in terms of a choice rule $x_i()$ as in (146). For worker i the function $x_i(j)$ is equal to one if worker i chooses occupation j and zero otherwise. Then, we define the worker i's problem as:

$$V_i = \max_{c_i, x_i()} \sum_j x_i(j) U_i(c_i, \{A_{jn}\}) \tag{A.4}$$
$$s.t. \qquad c_i - \sum_j x_i(j) w_j(\theta_i) = Qk_i.$$

With the choice rules, we can write the following equilibrium definition.

Definition 2. *An equilibrium consists of allocations $\{c_i\}_i$, $\{x_i()\}_i$, $\{h^j(\theta)\}_{\theta,j}$, K and prices $\{w_j(\theta)\}_{\theta,j}$, Q such that:*

1. *Given wages, the consumption c_i and the choice rule $x_i()$ solve the worker's problem (A.4) for all i.*

2. *Given the employment distributions $\{h_j(\theta)\}_{\theta,j}$, wages and capital satisfy the optimality conditions:*

$$MPL_j(\theta) = w_j(\theta) \tag{A.5}$$

$$z\mathcal{F}_K = Q \tag{A.6}$$

for all θ and j.

3. *Market clearing:*

   - *Labor market: for all j and θ, $h^j(\theta)$ is consistent with workers' optimal choice. Formally, $h^j(\theta) = \int x_i(j)\alpha_i(\theta)di$ where $\alpha_i(\theta) = 1$ if i has productivity type θ and $\alpha_i(\theta) = 0$ otherwise.*
   - *Capital market: $K = \int k_i di$.*
   - *Goods market: $Y = \int c_i di$.*

Proof of Proposition 1

The proof is similar to Proposition 2 in (146) but with no uncertainty and in a production economy instead of an endowment economy. Fix an equilibrium $(\{\tilde{c}_i\}_i, \{\tilde{x}_i()\}_i, \{\tilde{h}^j(\theta)\}_{\theta,j}, \tilde{K}, \{w_j(\theta)\}_{\theta,j}, Q)$. Notice first that if an alternative allocation $(c_i, x_i())$ is strictly preferred for some worker i, this allocation cannot be budget feasible:

$$c_i - \sum_j x_i(j) w_j(\theta_i) - Qk_i < 0. \tag{A.7}$$

As usual, this follows from utility maximization and local nonsatitation of preferences, which is satisfied here because I assumed that the utility functions are increasing in wages and amenities. Now consider an alternative allocation $(\{c_i\}_i, \{x_i()\}_i, \{h^j(\theta)\}_{\theta,j}, K)$ that Pareto dominates the initial equilibrium. The allocation must be feasible, which implies from the goods market clearing condition that:

$$\begin{aligned}\int c_i &= Y \\ &= \sum_j \int MPL_j(\theta)h^j(\theta)d\theta + \mathcal{F}_K K \\ &= \sum_j \int w_j(\theta)h^j(\theta) + QK,\end{aligned}$$

where the second equality uses the constant returns to scale assumption and the second the firms' optimality conditions (A.5)-(A.6). Then, using the labor and capital markets clearing conditions, we can write:

$$\begin{aligned}
\int c_i di &= \sum_j \int w_j(\theta) h^j(\theta) d\theta + Q \int k_i di \\
&= \sum_j \int w_j(\theta) \left(\int x_i(h) \alpha_i(\theta) di \right) d\theta + Q \int k_i di \\
&= \sum_j \int x_i(j) \left(\int w_j(\theta) \alpha_i(\theta) d\theta \right) di + Q \int k_i di \\
&= \int \left(\sum_j x_i(j) w_j(\theta_i) + Q k_i \right) di \\
&> \int c_i di,
\end{aligned}$$

where the last inequality uses the fact that the strictly preferred allocation cannot be budget feasible (Equation (A.7)), which gives us a contradiction.

Proof of Proposition 2

To show the proposition, we only need to express the term of the reallocation of workers in Equation (1.23) in terms of amenity prices so that it can be canceled out in (1.28). The steps are:

$$\begin{aligned}
\sum_{j \in \mathcal{J}(\theta)} \int \frac{w_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta &= \int \sum_{j \in \mathcal{J}(\theta)} \frac{(w_j(\theta) - w_{j_{min}(\theta)}(\theta)) h^j(\theta)}{Y} d \log h^j(\theta) d\theta \\
&= \int \sum_{j \in \mathcal{J}(\theta)} - \frac{\Delta_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) \theta,
\end{aligned}$$

where the first equality uses $dh^{\underline{j}(\theta)}(\theta) = -\sum_{j \in \mathcal{J}(\theta) \setminus \underline{j}(\theta)} dh^j(\theta)$.

A.6.2 Decompositions of total compensation growth by demographic groups

In Figure A.25, I repeat the analysis and compute the change in the average income and amenity value separately for the non-college and the college-educated, and in Figure A.26, I do so by gender. Although the non-college educated experienced a much lower increase in wages, I find that their increase in the amenity value was of a more similar magnitude—7984.57 dollars compared to 9312.96 for the college-educated. As a result, the growth in total compensation is 330% (37%) larger than the growth in wages for the non-college (college) educated. Regarding the gender comparison, while women experience larger increases in wage income, the bulk of the increase in the amenity value accrued to men—of 9647.36 compared to only 1540.12 for women. This is consistent with the fact that the blue-collar occupations, which tend to have the lowest amenities, largely employed non-college educated men.

Figure A.25: Changes in income and amenity value by education level

(a) Non-college educated (b) College educated

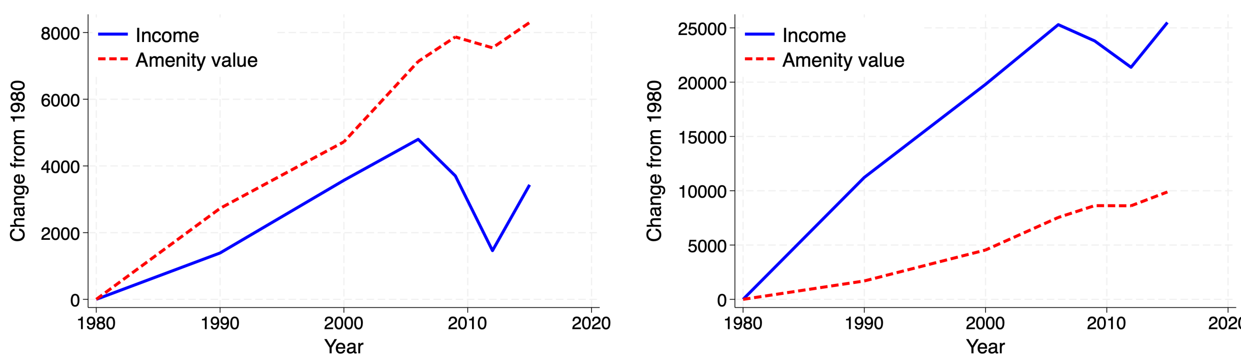
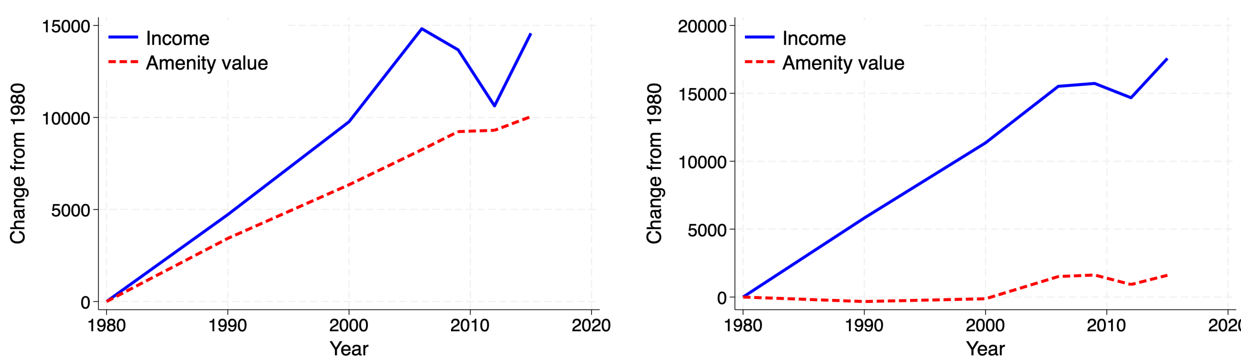


Figure A.26: Changes in income and amenity value by gender

(a) Men (b) Women



A.7 Extensions

A.7.1 Endogenous production of amenities by firms

I now extend the model by allowing firms to provide some amenities endogenously, show that it delivers a similar mismeasurement in output and that it can be corrected with amenity prices. Broadly, amenities can be classified into two types: those intrinsic to an occupation, such as its artistic nature, and those chosen by firms, like office perks. Throughout the paper, I have focused on the first, but similar ideas apply for the growth accounting in the latter.

Let A^I denote the amenity endogenously provided by the firm.[4] The worker's utility now writes $U_i(w, \{A_n\}, A^I)$. The cost of producing the endogenous amenity for firms in occupation j is $g_j(A^I; q_j^I)$, where q_j^I is a (Hicks-neutral) technology shifter to the cost of producing the amenity. I assume this cost is in terms of the final good so that these amenities will be treated as an intermediate.

Firms in every occupation can choose the amenity level for every one of their workers $A_{j,i}^I$. Since firms are competitive, the marginal product of every worker is equal to the cost of her compensation, which is equal to the wage plus the cost of producing the amenity, i.e.:

$$\mathcal{F}_{y_j} \frac{\partial m^j}{\partial h^j(\theta_i)} = w_{j,i} + g_j(A_{j,i}^I; q_j^I).$$

Then, the combination $(w_{i,j}, A_{j,i}^I)$ is chosen to maximize the worker's utility subject to the above condition.[5] Hence, we can think of firms as offering a menu of wages and amenities contingent on the worker's productivity given by the wage function:

$$w_j(A_{j,i}^I; \theta_i) = \mathcal{F}_{y_j} \frac{\partial m^j}{\partial h^j(\theta_{i,j})} - g_j(A_{j,i}^I; q_j^I).$$

Therefore, the local price of the amenity is given by

$$\delta_j^I(A_{j,i}^I) = -\frac{\partial w_j(A_{j,i}^I; \theta_i)}{\partial A_{j,i}^I} = \frac{\partial g_j(A_{j,i}^I; q_j^I)}{\partial A_{j,i}^I}. \tag{A.8}$$

[4] It is straightforward to extend the analysis to multiple endogenously provided amenities. With continuous amenity choices, we can define the price for each amenity as in the empirical part.

[5] This is similar to the setups in (120) and (160), where firms provide amenities as in (160), but the worker's total compensation is determined by its productivity.

Then, the worker's optimality condition pins down the chosen amenity level:

$$\frac{\frac{\partial U_i}{\partial A^I_{j,i}}}{\frac{\partial U_i}{\partial c_i}} = \delta^I_j(A^I_{j,i}).$$

Because the endogenous amenity is produced with the final good, the market clearing condition for the final good writes:

$$Y = \int c_i di + \int \sum_j x_i(j) g_j(A^I_{j,i}; q^I_j) di. \tag{A.9}$$

For exposition, I assume that all workers with the same type θ have the same preferences for wages and A^I, so that we can write

$$Y = \int c_i di + \int \sum_j h^j(\theta) g_j(A^I_j(\theta); q^I_j) d\theta, \tag{A.10}$$

where $A^I_j(\theta)$ is the optimal level of the endogenous amenity for workers of type θ in occupation j. Similarly, this assumption allows us to write a unique wage for every occupation and worker type $w_j(\theta)$.

Again, if income measures do not include the compensation or the value of the amenities is not counted in the firm sales, the measured output is equal to:

$$\begin{aligned}
\widetilde{Y} &\equiv \int c_i di \\
&= Y - \int \sum_j g_j(A^I_j(\theta); q^I_j) h^j(\theta) d\theta \\
&= \int \sum_j w_j(\theta) h^j(\theta) d\theta + QK.
\end{aligned}$$

As in the main model with exogenous amenities, the consumption of amenities would not be included in any of the three approaches to measuring output. An increase in the cost of amenities production—either through an increase in $A^I(\theta)$, a reallocation to occupations with a higher amenity production, or direct increases in the costs (dq^I_j)—will lead to a decrease in measured output and TFP. The intuition is that an increase in the cost of amenity production is measured as an increase in the firm's production costs, but the value of amenities is not accounted as part of either the firm's sales or the workers' compensation. Thus, this negative effect of higher amenity expenditures on TFP is akin to that of an increase in the cost of a factor input.

Defining the growth in measured TFP as in (1.22) and differentiating the market clearing condition, we obtain:

$$
\begin{aligned}
d\log TFP =& \frac{Y}{\widetilde{Y}} d\log z + \sum_j \frac{W_j}{\widetilde{Y}} d\log z_j + \int \sum_j \frac{MPL_j(\theta) h^j(\theta)}{\widetilde{Y}} d\log h^j(\theta) d\theta \\
& - \int \sum_j \frac{g_j(A_j^I(\theta); q_j^I) h^j(\theta)}{\widetilde{Y}} d\log h^j(\theta) d\theta - \int \sum_j \frac{\frac{\partial g_j(A_j^I(\theta); q_j^I)}{\partial A_j^I(\theta)} A_j^I(\theta) h^j(\theta)}{\widetilde{Y}} d\log A_j^I(\theta) d\theta \\
& + \int \sum_j \frac{g_j(A_j^I(\theta); q_j^I) h^j(\theta)}{\widetilde{Y}} (-d\log q_j^I) d\theta.
\end{aligned}
\tag{A.11}
$$

Again, this formula shows that Hulten's theorem does not hold under the conventional measures of output and TFP growth because changes in the endogenous allocations ($\{d\log h^j(\theta)\}$ and $\{d\log A_j^I(\theta)\}$) have first-order effects on measured TFP growth.

Augmented output and Hulten's theorem. I now show that we can also derive a Hulten's theorem by redefining output to include the value of the endogenous amenities. I define the total price of the endogenous amenity as

$$
\Delta_j^I(A_j^I) = \int_0^{A_j^I} \delta_j^I(\widetilde{A}_j^I) d\widetilde{A}_j^I. \tag{A.12}
$$

Then, the augmented output is

$$
\widetilde{Y}^a = \widetilde{Y} + \int \sum_{j \in \mathcal{J}(\theta)} \Delta_j(\theta) h^j(\theta) d\theta + \int \sum_{j \in \mathcal{J}(\theta)} \Delta_j^I(A_j^I(\theta)) h^j(\theta) d\theta.
$$

The growth rate of augmented TFP ($d\log ATFP$) is defined as in (1.27) and the growth $d\log \widetilde{Y}^a$ at constant prices implies that the functions $\{\delta_j^I(A_j)\}_j$ are also held constant. The following proposition derives a Hulten's theorem for the growth in augmented TFP.

Proposition 15. *The growth rate in augmented TFP with endogenous amenities satisfies*

$$d\log ATFP = \frac{Y}{\widetilde{Y}^a} d\log z + \sum_j \frac{W_j}{\widetilde{Y}^a} d\log z_j + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\delta_j(\theta; dA_j) A_j h^j(\theta)}{\widetilde{Y}^a} d\log A_j d\theta$$
$$+ \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j^I(A_j^I(\theta)) h^j(\theta)}{\widetilde{Y}^a} (-d\log q_j^I) d\theta.$$

Proof. From the definition, the growth in augmented TFP satisfies

$$d\log ATFP = \frac{\widetilde{Y}}{\widetilde{Y}^a} d\log TFP + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j(\theta) h^j(\theta)}{\widetilde{Y}^a} d\log h^j(\theta) d\theta + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\delta_j(\theta; dA_j) A_j h^j(\theta)}{\widetilde{Y}^a} d\log A_j d\theta$$
$$+ \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j^I(A_j^I(\theta)) h^j(\theta)}{\widetilde{Y}^a} d\log h^j(\theta) d\theta + \int \sum_{j\in\in\mathcal{J}(\theta)} \frac{\delta_j^I(A_j^I(\theta)) A_j^I(\theta) h^j(\theta)}{\widetilde{Y}^a} d\log A_j^I(\theta) d\theta,$$

where $d\log TFP$ is given by (A.11). Then, using the definitions of the prices $\Delta_j(\theta)$, $\Delta_j^I(\theta)$, and $\delta_j^I(A_j^I)$, we can cancel out all the terms that depend on the changes in the endogenous variables. □

Corrected TFP. Finally, as in the main model, we may also be interested in using amenity prices to measure a corrected TFP growth. I now define corrected TFP growth as:

$$d\log CTFP = \frac{Y}{\widetilde{Y}} d\log z + \sum_j \frac{W_j}{\widetilde{Y}} d\log z_j + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j^I(A_j^I(\theta)) h^j(\theta)}{\widetilde{Y}^a} (-d\log q_j^I) d\theta.$$

Note that since the endogenous amenity is an intermediate good that directly affects the measured output, I include the changes in its production cost ($d\log q_j^I$) in the corrected TFP. The corrected TFP growth can now be computed as:

$$d\log CTFP = d\log TFP + \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j(\theta) h^j(\theta)}{\widetilde{Y}} d\log h^j(\theta) d\theta$$
$$+ \int \sum_{j\in\mathcal{J}(\theta)} \frac{\Delta_j^I(A_j^I(\theta)) h^j(\theta)}{\widetilde{Y}} d\log h^j(\theta) d\theta + \int \sum_{j\in\in\mathcal{J}(\theta)} \frac{\delta_j^I(A_j^I(\theta)) A_j^I(\theta) h^j(\theta)}{\widetilde{Y}^a} d\log A_j^I(\theta) d\theta.$$

A.7.2 Inefficient economy with wage markdowns

In recent years, a large literature has emerged quantifying the aggregate effects of misallocation in inefficient economies (158 and 118). Frictions, such as markups or taxes, create wedges between prices and marginal products, which imply that resources are not allocated efficiently, thereby creating output, productivity, and welfare losses. For growth accounting, (22) provide an aggregation result for inefficient economies with rich input-output linkages that can be used to measure the sources of TFP growth. This idea has also proved important in the labor market. For instance, (33) quantify the aggregate welfare and output losses from labor market power, while (117) quantify the growth from an improved allocation of talent due to a reduction in discrimination, better access to education, and changes in social norms and preferences.

Throughout, I have studied a competitive economy where the first welfare theorem holds. First, because an efficient economy provides a benchmark to isolate the role of amenities and compensating differentials. But second, to highlight how, with compensating differentials, we can measure misallocation in an efficient economy. However, it is interesting and important to extend the analysis to an inefficient economy economy. Especially to understand the interaction between labor market power and compensating differentials.

Occupation-level wage markdowns. To study an inefficient economy in a parsimonious way, I introduce exogenous occupation-level wage markdowns $\{\mu_j\}$.[6] Hence, the wage paid to workers of type θ in occupation j satisfies

$$w_j(\theta) = \mu_j^{-1} MPL_j(\theta). \tag{A.13}$$

The firm's profits originating from the markdowns are rebated back to the workers. For ease of exposition, I assume no aggregate capital and no aggregate technology shifter, so the production of the final good is $Y = \mathcal{F}(\{y_j\}_j)$.

Because now there are no intermediates, the production of the final good is equal to the measured output, i.e., $\widetilde{Y} = Y$. Letting $dh^j(\theta_s) = \frac{\partial h^j(\theta_s)}{\partial \log z_j} d\log z_j + \frac{\partial h^j(\theta_s)}{\partial \log \mu_j} d\log \mu_j$ denote the changes in employment in response to changes changes in technology and markdowns,

[6] A more complex way of introducing inefficiencies would be to have market power or distortions at the firm level. In this case, we would need to introduce firm-specific preferences as in (33) or have a model with rationing.

we can write a decomposition for the growth in measured TFP growth similar to (1.23):

$$\begin{aligned} d \log TFP &= \sum_j \frac{p_j y_j}{Y} d \log z_j + \int \sum_j \frac{MPL_j(\theta)}{Y} dh^j(\theta) d\theta \\ &= \sum_j \frac{p_j y_j}{Y} d \log z_j + \int \sum_j \frac{\mu_j w_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta. \end{aligned}$$

Similar to the efficient case, changes in the allocation of production inputs $\{d \log h^j(\theta)\}$ have first-order effects on TFP. However, this is now caused by both the compensating differentials and the markdowns. Hence, as is well known, even without amenities Hulten's theorem will not hold in this economy.

In what follows, I first derive an aggregation result similar to that of (22) to characterize the changes in the measured TFP in this economy with markdowns and compensating differentials. Then, I show how the corrected TFP measures the change in allocative efficiency net of the effect of compensating differentials.

Aggregation result. The following proposition characterizes the change in TFP in response to changes in technology and markdowns.

Proposition 16. *The change in TFP in response to changes in technology ($\{d \log z_j\}_j$) and markdowns ($\{d \log \mu_j\}_j$) satisfies:*

$$d \log TFP = \sum_j \lambda_j [d \log z_j - d \log \mu_j - d \log W_j] + \int \sum_{j \in \mathcal{J}(\theta)} \mu_j \frac{w_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta,$$

where $\lambda_j = \frac{p_j y_j}{\widetilde{Y}}$ is the Domar weight of occupation j and $W_j = \int w_j(\theta) h^j(\theta) d\theta$ is the wage bill in occupation j.

The proposition is proven in Subsection A.7.2. The proposition is simpler than of (22) in that there is no input-output structure. However, the conceptual distinction is that in this economy, the price of each production factor (here θ) is different across the producers (here occupations). This implies that the last two terms are different. In particular, we need to keep track of the reallocation of the production factors (here $\{d \log h^j(\theta)\}$) across producers. In (22), it is sufficient to keep track of the changes in the income shares of each factor (here $W(\theta) = \sum_j w_j(\theta) h^j(\theta)$).

To understand the last term, recall that the terms $\mu_j w_j(\theta)$ equal the marginal product of labor (MPL) of θ in occupation j. Hence, measured TFP increases if workers reallocate to high MPL occupations. However, the differences in the MPLs are due to the amenities and the markdowns. For exposition, assume there are two occupations, $\{H, L\}$ with $A^H > A^L$, and no skill heterogeneity. Then, the last term can be written as:

$$\frac{1}{Y}[\mu_H w_H - \mu_L w_L] dh^H,$$

which implies that a reallocation to the high-amenity occupation decreases the measured TFP if:

$$\Delta > (\frac{\mu_L}{\mu_H} - 1) w^L,$$

where the total amenity price is defined as usual: $\Delta = -(w^H - w^L)$. That is, if the markdown in the low-amenity occupations is not too high compared to the markdown in the high amenity one and the amenity prices. If the markdowns are the same, a reallocation of workers to the high-amenity occupation always decreases the measured TFP.

Corrected TFP. I now show how the corrected TFP, computed as in Equation (1.31), measures a change in allocative efficiency net of the differences in wages from amenity prices. First, notice that we can define the total amenity prices $\Delta_j(\theta)$ exactly as in Section 1.6.3. They are the prices that workers face and are constructed from the differences in observed wages. Hence, they can be mapped to the amenity price estimates in the same way. Moreover, these amenity prices still measure the valuation of the amenity for the workers that reallocated. However, the wedges imply that they are not equal to the differences in the MPL across occupations, and so to the costs of producing the amenities. The following proposition characterizes the corrected TFP growth.

Proposition 17. *The response of corrected TFP to a change in technology ($\{d\log z_j\}_j$) and markdowns ($\{d\log \mu_j\}_j$) satisfies:*

$$d\log CTFP = \sum_j \lambda_j [d\log z_j - d\log \mu_j - d\log W_j] + \int \sum_{j\in\mathcal{J}(\theta)} (\mu_j - 1)\frac{w_j(\theta)h^j(\theta)}{Y} d\log h^j(\theta) d\theta,$$

where

$$\int \sum_{j\in\mathcal{J}(\theta)} (\mu_j - 1)\frac{w_j(\theta)h^j(\theta)}{Y} d\log h^j(\theta) d\theta =$$

$$\int \sum_{j\in\mathcal{J}(\theta)} \left[(\mu_j - \mu_{\underline{j}(\theta)}) h^j(\theta) \frac{w_{\underline{j}(\theta)}(\theta)}{Y} - (\mu_j - 1)\frac{\Delta_j(\theta) h^j(\theta)}{Y} \right] d\log h^j(\theta) d\theta$$

(Proof in subsection A.7.2). The corrected TFP measures changes in allocative efficiency net of the differences in wages due to amenity prices. Since amenity prices are measured from the prices paid by workers, they account for the valuation of the workers that switch. The proposition shows that this net change in allocative efficiency can be decomposed into two terms. The first term, $(\mu_j - \mu_{\underline{j}(\theta)})h^j(\theta)\frac{w_{\underline{j}(\theta)}(\theta)}{Y}$, measures the change in allocative efficiency due to the differences in markdowns across occupations. Without amenities, the term would be the same but with a unique wage for every θ, $w(\theta)$, instead of the wage in the lowest amenity occupation $w_{\underline{j}(\theta)}(\theta)$. This term implies that allocative efficiency improves if workers reallocate to high-markdown occupations, as these occupations have too few workers to begin with. The second term, $(\mu_j - 1)\frac{\Delta_j(\theta)h^j(\theta)}{Y}$, can be interpreted as a distortion in the sales of amenities. A markdown μ_j^{-1} in occupation j, can be seen as a markup μ_j in the sales of the amenity A_j. In this sense, this term is negative if workers purchase more amenities with larger (gross) markups.

Proof of Proposition 16

The first step is to derive the change in the occupations outputs' prices ($d\log p_j$) in response to changes in productivities and markups (the steps are similar to 22 or 21). Constant returns to scale imply:

$$p_j z_j m^j(\{h^j(\theta)\}_\theta) = \mu_j \int w_j(\theta)h^j(\theta)d\theta.$$

Log-differentiating and collecting terms:

$$\begin{aligned} d\log p_j &= -d\log z_j + \int \frac{\mu_j w_j(\theta)h^j(\theta)}{p_j y_j} d\log \mu_j d\theta + \int \frac{\mu_j w_j(\theta)h^j(\theta)}{p_j y_j} d\log w_j(\theta)d\theta \\ &= -d\log z_j + d\log \mu_j + \int \Omega_j(\theta) d\log w_j(\theta)d\theta, \end{aligned}$$

where $\Omega_j(\theta) = \frac{\mu_j w_j(\theta)h^j(\theta)}{p_j y_j}$ and the second equality uses $\int \mu_j w_j(\theta)h^j(\theta)d\theta = p_j y_j$.

The next step is to write the change in the final good as a function of the prices of the occupations outputs $\{p_j\}$. We can write the production of the final good as

$$Y(\{p_j\}, R) = \max_{\{y_j\}} \mathcal{F}(\{y_j\}) \ s.t \ \sum_j p_j y_j = R,$$

for some R. The envelope conditions are:

$$\frac{\partial Y}{\partial p_j} = -\mu y_j,$$

where μ is the multiplier. And the FOCs: $\mathcal{F}_{y_j} = \mu p_j$, which combined with the constant return assumption imply $\mu = 1$. Hence, if we consider a general perturbation of prices $\{dp_j\}$, using the envelope conditions we obtain:

$$dY_j = -y_j dp_j,$$

or in log differences:

$$d\log Y = -\sum_j \frac{p_j y_j}{Y} d\log p_j.$$

Substituting the changes in prices using the equations above:

$$d\log \widetilde{Y} = \sum_j \lambda_j d\log z_j - \sum_j \lambda_j \mu_j d\log \mu_j - \sum_j \lambda_j \int \mu_j \Omega_j(\theta) d\log w_j(\theta) d\theta. \tag{A.14}$$

Letting $W_j(\theta) = w_j(\theta) h^j(\theta)$, we can decompose the third term as

$$\sum_j \lambda_j \int \mu_j \Omega_j(\theta) d\log w_j(\theta) d\theta = \sum_j \lambda_j \int \mu_j \Omega_j(\theta) d\log W_j(\theta) d\theta - \sum_j \lambda_j \int \mu_j \Omega_j(\theta) d\log h^j(\theta) d\theta. \tag{A.15}$$

Letting $W_j = \int W_j(\theta) d\theta$, we rewrite the first term as:

$$\begin{aligned}
\sum_j \lambda_j \int \mu_j \Lambda_j(\theta) d\log W_j(\theta) d\theta &= \sum_j \lambda_j \frac{\mu_j}{p_j y_j} \int dW_j(\theta) d\theta \\
&= \sum_j \lambda_j \frac{\mu_j W_j}{p_j y_j} \frac{dW_j}{W_j} \\
&= \sum_j \lambda_j d\log W_j.
\end{aligned}$$

Proof of Proposition 17

Adding and subtracting $(\mu_j - 1)w_{\underline{j}(\theta)}(\theta)$, we have:

$$\int \sum_{j \in \mathcal{J}(\theta)} (\mu_j - 1) \frac{w_j(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta =$$

$$\int \sum_{j \in \mathcal{J}(\theta)} [(\mu_j - 1) \frac{(-\Delta_j(\theta)) h^j(\theta)}{Y} + (\mu_j - 1) \frac{w_{\underline{j}(\theta)}(\theta) h^j(\theta)}{Y}] d \log h^j(\theta) d\theta.$$

Then, using $\sum_{j \in \mathcal{J}(\theta)} h^j(\theta) d \log h^j(\theta) = 0$, the second term can be rewritten as:

$$\int \sum_{j \in \mathcal{J}(\theta)} (\mu_j - \mu_{\underline{j}(\theta)}) \frac{w_{\underline{j}(\theta)}(\theta) h^j(\theta)}{Y} d \log h^j(\theta) d\theta. \tag{A.16}$$

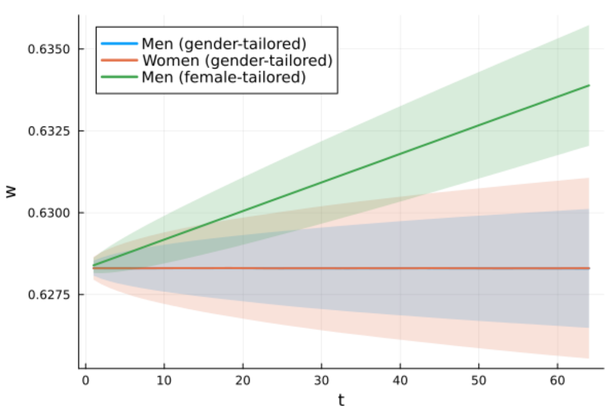
Appendix B

Appendix of Chapter 2: The Hidden Demand for Flexibility - A Theory of Gendered Employment Dynamics

B.1 Extra Tables and Figures

Figure B.1: Men under women-tailored contract

(a) Average wage dynamics



(b) Wage distributions (at $t = 64$)

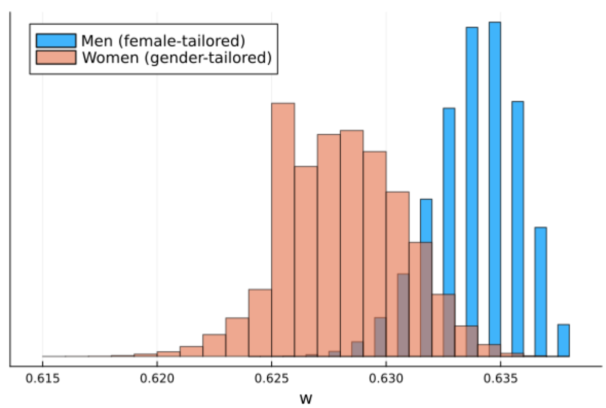
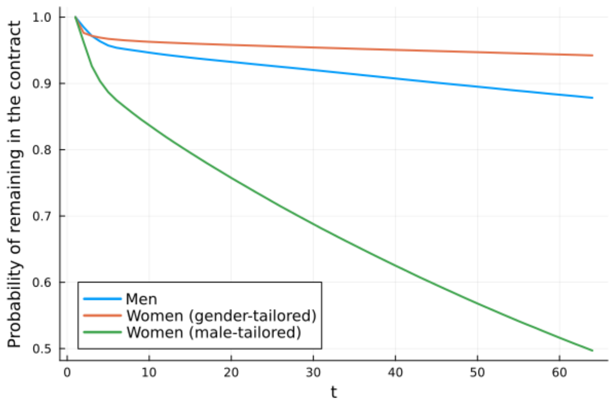
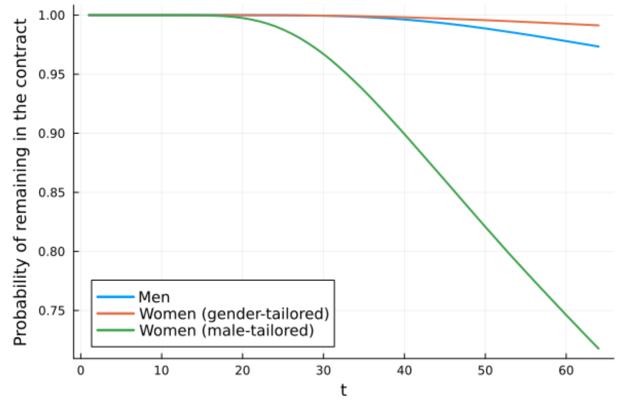


Figure B.2: Termination dynamics with low and high v_0



(a) Low v_0



(b) High v_0

Figure B.3: Comparative statics on α for women's wage dynamics under male-tailored contracts

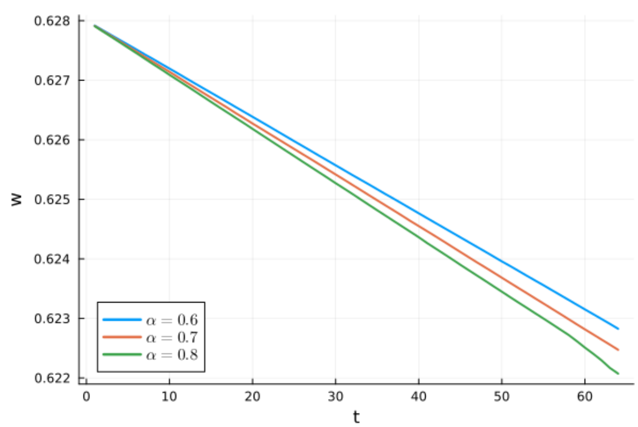


Table B.1: Comparative statics with respect to the targeted reduction in hours

	p_{men}	p_{women}	Avg. growth rate of womens' wages	Avg. wage gap after 16 years
Baseline (25% reduction)	0.06	0.15	-0.014%	0.88%
20% reduction in hours	0.08	0.21	-0.018%	1.12%
30% reduction in hours	0.04	0.1	-0.01%	0.64%

B.2 Proofs

Throughout, we use the following notation for the continuation utilities:

$$\omega^H = U(w^H, h^H; f^H) + \beta v^H$$
$$\omega^L = U(w^L, h^L; f^L) + \beta v^L.$$

Proof of Proposition 3

In the simple case when the employer can observe the employee's time availability f and there is full commitment, the Lagrangian of the principal's problem writes as

$$\mathcal{L} = (1-p)\left[g(h^H) - w^H + \beta\Pi(v^H)\right] + p\left[g(h^L) - w^L + \beta\Pi(v^L)\right]$$
$$+ \lambda\left[(1-p)\left(u(w^H) - (1-f^H)\psi(h^H) + \beta v^H\right) + p\left(u(w^L) - (1-f^L)\psi(h^L) + \beta v^H\right) - v\right].$$

Then, for any $j \in \{H, L\}$, the first order conditions are

$$w^j: \quad \frac{1}{u'(w^j)} = \lambda \tag{B.1}$$

$$h: \quad \frac{g'(h^j)}{(1-f)\psi'(h^j)} = \lambda \tag{B.2}$$

$$v^j: \quad \Pi(v^j) = -\lambda \tag{B.3}$$

and the envelope condition

$$\widehat{\Pi}'(v) = \lambda. \tag{B.4}$$

Combining (B.1) and (B.2) we get the first result of the proposition:

$$g'(h^j) = \frac{(1-f^j)\psi'(h^j)}{u'(w^j)}. \tag{B.5}$$

From (B.1) and (B.3), it also follows that contracts feature full insurance, i.e. $w^H = w^L$ and $v^H = v^L$. From (B.3) and (B.4), it also follows that $v = v^j$ so there is perfect intertemporal smoothing. Finally, because v is constant over time, the principal's value is also constant. Hence, it is never optimal to terminate the contract if it is optimal to enter it.

Proof of Proposition 4

We start establishing some preliminary results on the relevant constraints. We first show that (IC) constraint of the high type must bind in the optimal allocation. We assume throughout that $h^H \geq h^L$ and verify ex-post that this condition is indeed satisfied. In this case, incentive compatibility requires that either $u(w^H) \geq u(w^L)$, $v^H \geq v^L$, or both. By contradiction, assume that the high type's (IC) does not bind. Then the principal could lower the utility of the high type while satisfying (IC), (SUST) and (LC), and redistribute to the low type while satisfying the (PK) constraint. Because the cost of increasing the utility of the high type is smaller, the principal can obtain a direct resource gain from the perturbation.

Then, note that the (IC) constraints can be written as

$$\omega^H = \omega^L + (f^H - f^L)\psi(h^L) \tag{B.6}$$

$$\omega^L \geq \omega^H + (f^L - f^H)\psi(h^H). \tag{B.7}$$

Using $h^H \geq h^L$, it is easy that the high type's (IC) binding implies that the (IC) of the low type will not bind. For the sustainability constraint, because $h^L \geq 0$, we must have $\omega^H \geq \omega^L$. Hence, if the allocation is incentive-compatible, the (SUST) constraint of the high type ($\omega^H \leq \overline{v}$) is implied from the (SUST) constraint of the low type, and so it can be ignored. Finally, we ignore the (LC) of the high type and verify ex-post that it does not bind.

The Lagrangian of the principal's problem is

$$\begin{aligned}
\mathcal{L} = &(1-p)\left[g(h^H) - w^H + \beta\Pi(v^H)\right] + p\left[g(h^L) - w^L + \beta\Pi(v^L)\right] \\
&+ \lambda\left[(1-p)\left(u(w^H) - (1-f^H)\psi(h^H) + \beta v^H\right) + p\left(u(w^L) - (1-f^L)\psi(h^L) + \beta v^L\right) - v\right] \\
&+ \mu\left[u(w^H) - (1-f^H)\psi(h^H) + \beta \hat{v}^H - u(w^L) + (1-f^H)\psi(h^L) + \beta v^L\right] \\
&+ p\gamma\left[u(w^L) - (1-f^L)\psi(h^L) + \beta v^L - \overline{v}\right] + p\xi\beta\left[v^L - \overline{v}\right]
\end{aligned}$$

The first order conditions are:

$$h^H: \frac{g'(h^H)}{(1-f^H)\psi'(h^H)} = \lambda + \frac{\mu}{1-p} \tag{B.8}$$

$$h^L: \frac{g'(h^L)}{(1-f^L)\psi'(h^L)} = \lambda + \gamma - \frac{\mu}{p}\frac{(1-f^H)}{(1-f^L)} \tag{B.9}$$

$$w^H: \frac{1}{u'(w^H)} = \lambda + \frac{\mu}{1-p} \tag{B.10}$$

$$w^L: \frac{1}{u'(w^L)} = \lambda + \gamma - \frac{\mu}{p} \tag{B.11}$$

$$v^H: \ -\Pi'(v^H) = \lambda + \frac{\mu}{1-p} \tag{B.12}$$

$$v^L: \ -\Pi'(v^L) = \lambda + \gamma + \xi - \frac{\mu}{p} \tag{B.13}$$

And the envelope condition:

$$\widehat{\Pi}'(v) = -\lambda. \tag{B.14}$$

Moreover, combining the envelope condition and the first order condition for v in problem (2.6), it is easy to see that $\widehat{\Pi}'(v_c) = \Pi'(v)$.

Part (i): Combining the first order conditions (B.8) and (B.10):

$$g'(h^H) = \frac{(1-f^H)\psi'(h^H)}{u'(w^H)}.$$

Combining (B.9) and (B.11):

$$g'(h^L) = \frac{(1-f^L)\psi'(h^L)}{u'(w^L)}\left(1 + u'(w^L)\frac{\mu}{p}\frac{f^H - f^L}{(1-f^L)}\right) > \frac{(1-f^L)\psi'(h^L)}{u'(w^L)}.$$

Part (ii): First note that because $v > \overline{v}$ we have $h^L > 0$ (see Lemma 2), which implies $\omega^H > \omega^L$. Moreover, we also know that $h^H > h^L$. This implies that we must have either $u^H > u^L$, $v^H > v^L$ or both, we now show it must be both. Consider the case where the (LC) constraint does not bind, i.e. $\xi = 0$. Then because $\mu > 0$, the FOCs (B.10)-(B.13) directly imply $u^H > u^L$ and $v^H > v^L$. Now assume the (LC) binds, so $v^L = \overline{v}$. The constraint for the high type implies that $v^H \geq v^L$, which combined with the FOC requires $\lambda + \frac{\mu}{1-p} \geq \lambda + \gamma + \xi - \frac{\mu}{p}$. Because $\xi > 0$ from the FOC for v^H and v^L we deduce that $u^H > u^L$. The last step is to show that $v^H > v^L$. By contradiction, assume $v^H = v^L$ is optimal. Optimality implies that the allocations following f^H and f^L must be the same. Now consider a perturbation where we decrease u^H by $\varepsilon > 0$ and increase v^H by $\frac{1}{\beta}\varepsilon$ so that ω^H

is kept constant. We increase v^H by increasing u^H_{t+1} by $\frac{1}{\beta(1-p)}\varepsilon$. The resource gain from this perturbation is

$$\frac{\Delta\widehat{\Pi}}{\varepsilon} \approx \frac{1}{u'(w^H_t)} - \frac{1}{u'(w^H_{t+1})}, \tag{B.15}$$

which is positive if $w^H_t > w^H_{t+1}$. The (SUST) and (IC) constraints imply that $\omega^H > \overline{v} = v^H$, which, because there is no distortion in hours for the high type, implies $w^H_t > w^H_{t+1}$.

Next, we show $v^H > v > v^L$ in the region without exp-post inefficiencies where constraints (SUST) and (*LC*) do not bind. Substituting the envelope condition (B.14) into the FOCs (B.12) and (B.13), we have

$$-\Pi'(v^H) = -\Pi'(v) + \frac{\mu}{1-p},$$

$$-\Pi'(v^L) = -\Pi'(v) - \frac{\mu}{p},$$

where we use $\Pi = \widehat{\Pi}$ because in this region there is no exit (see Proposition 5 and Corollary 1) and $\gamma = \xi = 0$ because the (SUST) and (*LC*) do not bind. Then, because Π is decreasing in this region (Proposition 5) and $\mu > 0$, it follows that $v^H > v$ and $v < v^L$.

Part (iii): Finally, when the (SUST) and (*LC*) do not bind, combining the FOC (B.10) and (B.11), and using sequential notation

$$\lambda_t = (1-p)\frac{1}{u'(w^H_t)} + p\frac{1}{u'(w^L_t)}.$$

Combining the FOC for the promised utility (equations (B.12) and (B.13)) and for wages, we observe that $-\Pi'(v^H) = \frac{1}{u'(w^H)}$ and $-\Pi'(v^L) = \frac{1}{u'(w^L)}$. Then, using the envelope condition implies $\lambda_t = \frac{1}{u'(w_{t-1})}$, which gives us the Inverse Euler equation in the proposition.

Proof of Lemma 2

First, notice that as $v \to \overline{v}$, the (PK) constraint converges to

$$(1-p)\omega^H + p\omega^L = \overline{v}. \tag{B.16}$$

Then the (SUST) constraint ($\omega^L \geq \overline{v}$) implies that we must have $\omega^H = \omega^L = \overline{v}$. Substituting into the (IC) constraint

$$\overline{v} = \overline{v} + (f^H - f^L)\psi(h^L), \tag{B.17}$$

which implies $h^L \to 0$.

Proof of Proposition 5

We first show that the (constrained) profit function Π^c is increasing in a neighborhood around $\overline{v}$. Starting from $v = \overline{v}$, consider perturbation where we increase promised utility by ε, i.e. $v = \overline{v} + \varepsilon$ and the hours of the low type by ε^h, i.e. $h^L = \varepsilon^h$ by Lemma (2). We now show that this perturbation can satisfy the (IC), (PK), (LC) and (SUST) constraints and deliver higher profits for the principal, i.e. $\Pi^c(\overline{v}) < \Pi^c(\overline{v} + \varepsilon)$. We keep v^L and v^H fixed at the optimal given $v = \overline{v}$, so the (LC) constraint holds. To satisfy the (SUST) constraint, we increase w^L to keep the low type's continuation utility, ω^L, constant:

$$\Delta w^L = \frac{(1 - f^L)\psi'(\varepsilon^h)\varepsilon^h}{u'(w^L)}. \tag{B.18}$$

As h^L increases, the information rent that must be given to the high type to preserve incentive compatibility also increases. To this end, we increase the high's type wage w^H by:

$$\Delta w^H = \frac{(f^H - f^L)\psi'(\varepsilon^h)\varepsilon^h}{u'(w^H)}. \tag{B.19}$$

Finally, the (PK) constraint must also be satisfied. Since ω^L is kept fixed, we only need to make sure that Δw^H increases the high's type utility enough. Hence,

$$(1 - p)(f^H - f^L)\psi'(\varepsilon^h)\varepsilon^h = \varepsilon, \tag{B.20}$$

which gives us the link between ε and ε^h. The next step is to show that for a small enough ε^h, this perturbation increases the principal's profits. The change in the principal's objective function is

$$\Delta\Pi^c \approx -(1 - p)\Delta w^H - p\Delta w^L + pg'(\varepsilon^h)\varepsilon^h. \tag{B.21}$$

Substituting for the wage changes

$$\frac{\Delta\Pi^c}{\varepsilon^h} \approx -(1 - p)\frac{(f^H - f^L)\psi'(\varepsilon^h)}{u'(w^H)} - p\frac{(1 - f^L)\psi'(\varepsilon^h)}{u'(w^L)} + pg'(\varepsilon^h) \tag{B.22}$$

$$= -\left[(1 - p)\frac{(f^H - f^L)}{u'(w^H)} + p\frac{(1 - f^L)}{u'(w^L)}\right]\psi'(\varepsilon^h) + pg'(\varepsilon^h). \tag{B.23}$$

The first term inside the squared brackets is a bounded constant. Moreover, we have the Inada conditions $\lim_{\varepsilon^h \to 0} \psi'(\varepsilon^h) = 0$ and $\lim_{\varepsilon^h \to 0} g'(\varepsilon^h) = \infty$, so for a small enough ε^h,

$\frac{\Delta\Pi}{\varepsilon^h} > 0$. Therefore $\Pi^c(\overline{v}+\varepsilon) \geq \Pi^c(\overline{v}) + \Delta\Pi^c\varepsilon^h > \Pi^c(\overline{v})$ which implies that Π^c must be increasing in a neighborhood around $\overline{v}$.

Finally, we show that Π^c must be decreasing at $v > \widetilde{v}$ for some $\widetilde{v} > \overline{v}$. Notice that for high enough $\widetilde{v}$, the (SUST) and (LC) constraints are not binding so $\gamma = \xi = 0$. Then combining the envelope condition (B.14), and adding up the FOCs (B.12) and (B.13) we obtain

$$\frac{\partial\Pi^c(v)}{\partial v} = -\left((1-p)\frac{1}{u'(w^H)} + p\frac{1}{u'(w^L)}\right) < 0,$$

which shows that Π^c is decreasing.

Proof of Proposition 6

We show the result first for wages. Then, we show it is straightforward to extend it for the promised utilities. Starting from the optimal allocation, we consider a variation where we decrease h^L, w^L and w^H while satisfying the (IC) and (PK) constraints and show that the resource gain is decreasing in p. For $\varepsilon > 0$ small, we lower the hours of the low type by $\Delta h^L = -\varepsilon$. We move along the indifference the low type's indifferent curve, i.e. we keep ω^L fixed, so u^L needs to be adjusted by

$$\Delta u^{L,\omega} = (1-f^L)\psi'(h^L)(-\varepsilon) < 0 \tag{B.24}$$

This relaxes the RHS of the incentive constraint (B.6) by $(f^H - f^L)\psi'(h^L)(-\varepsilon) < 0$, which allows as to decrease u^H by

$$\Delta u^{H,IC} = (f^H - f^L)\psi'(h^L)(-\varepsilon) < 0. \tag{B.25}$$

Because ω^L is fixed, the ex-ante utility decreases by $(1-p)\Delta u^{H,IC} < 0$. To satisfy the (PK), we increase both types' wage utilities uniformly by $\Delta u^{PK} = (1-p)\Delta u^{H,IC}$, which preserves incentive compatibility. The resulting total changes in the wage utility of each type are

$$\Delta u^{H,TOT} = \Delta u^{H,IC} + \Delta u^{PK} = p(f^H - f^L)\psi'(h^L)(-\varepsilon) < 0 \tag{B.26}$$

$$\Delta u^{L,TOT} = \Delta u^{L,\omega} + \Delta u^{PK} = [(1-f^L) - (1-p)(f^H - f^L)]\psi'(h^L)(-\varepsilon) < 0, \tag{B.27}$$

where the first inequality follows from $f^H > f^L$, and the second from $1 - f^L > 0, 1 - p < 1$ and $f^H < 1$, which verifies that $\Delta w^L < 0$ and $\Delta w^H < 0$. The principal's gain from this perturbation is

$$\Delta\Pi \approx (1-p)\left(-\frac{1}{u'(w^H)}\Delta u^{H,TOT}\right) + p\left(g'(h^L)(-\varepsilon) - \frac{1}{u'(w^L)}\Delta u^{L,TOT}\right). \tag{B.28}$$

Substituting for the changes in the wage utility and rearranging

$$\frac{\Delta\Pi}{\varepsilon} \approx p(1-p)\left(\frac{1}{u'(w^H)} - \frac{1}{u'(w^L)}\right)(f^H - f^L)\psi'(h^L) \tag{B.29}$$

$$+ p\left(\frac{(1-f^L)\psi'(h^L)}{u'(w^L)} - g'(h^L)\right). \tag{B.30}$$

Differentiating with respect to p

$$\frac{\partial\frac{\Delta\Pi}{\varepsilon}}{\partial p} \approx (1-2p)\left(\frac{1}{u'(w^H)} - \frac{1}{u'(w^L)}\right)(f^H - f^L)\psi'(h^L) \tag{B.31}$$

$$+ \left(\frac{(1-f^L)\psi'(h^L)}{u'(w^L)} - g'(h^L)\right). \tag{B.32}$$

We need to show that $\frac{\partial\frac{\Delta\Pi}{\varepsilon}}{\partial p} < 0$. For $p < 1/2$, the result is not direct because in the optimal allocation $\frac{1}{u'(w^H)} - \frac{1}{u'(w^L)} > 0$. Using the FOCs (B.10) and (B.11),

$$\frac{1}{u'(w^H)} - \frac{1}{u'(w^L)} = \lambda + \frac{\mu}{1-p} - (\lambda - \frac{\mu}{p}) = \frac{\mu}{p(1-p)}. \tag{B.33}$$

Moreover, from the optimality condition (2.7),

$$g'(h^L) - \frac{(1-f^L)\psi'(h^L)}{u'(w^L)} > \frac{\mu}{p}(f^H - f^L)\psi'(h^L). \tag{B.34}$$

Hence, $\frac{\partial\frac{\Delta\Pi}{\varepsilon}}{\partial p} < 0$ if

$$\frac{\mu}{p}(f^H - f^L)\psi'(h^L) > (1-2p)\frac{\mu}{p(1-p)}(f^H - f^L)\psi'(h^L), \tag{B.35}$$

which is equivalent to $p > 0$ and completes the proof.

To show v^H and v^L are also decreasing, notice we can follow the same variation as above but with $\Delta v^{H,TOT} = \frac{1}{\beta}\Delta u^{H,TOT}$ and $\Delta v^{L,TOT} = \frac{1}{\beta}\Delta u^{L,TOT}$.[1] Then, the resource gain

[1] In fact any variation where a fraction $\alpha \in [0,1]$ of the change in compensation is delivered through the flow utilities, u^H and u^L, and a fraction $1-\alpha$ trough promised utilities, v^H and v^L, would work.

for the principal is

$$\Delta\Pi \approx (1-p)\Pi'(v^H)\Delta u^{H,TOT} + p\left(g'(h^L)(-\varepsilon) + \Pi'(v^L)\Delta u^{L,TOT}\right). \tag{B.36}$$

Then using the FOC we can substitute $\Pi'(v^H) = -\frac{1}{u'(w^H)}$ and $\Pi'(v^L) = -\frac{1}{u'(w^L)}$ and follow the same steps as above.

Proof of Proposition 7

As discussed, with log-utility wages are a martingale under men's contract:

$$w_{t-1}^{\text{men}}(f^{t-1}) = p_{\text{men}} w_t^{\text{men}}(f^{t-1}, f^L) + (1-p_{\text{men}}) w_t^{\text{men}}(f^{t-1}, f^H).$$

Then, rearranging terms we can write:

$$\frac{w_t^{\text{men}}(f^{t-1}, f^H)}{w_{t-1}^{\text{men}}(f^{t-1})} - 1 = -\left(\frac{p_{\text{men}}}{1-p_{\text{men}}}\right)\left(\frac{w_t^{\text{men}}(f^{t-1}, f^L)}{w_{t-1}^{\text{men}}(f^{t-1})} - 1\right). \tag{B.37}$$

The expected wages for women under the *male-tailored* contract are

$$\mathbb{E}_{p_{\text{women}}}(w_t^{\text{men}}(f^t)) = p_{\text{women}} w_t^{\text{men}}(f^{t-1}, f^L) + (1-p_{\text{women}}) w_t^{\text{men}}(f^{t-1}, f^H).$$

So, the expected growth rate is:

$$\begin{aligned}
\mathbb{E}_{p_{\text{women}}}\left(\frac{w_t^{\text{men}}(f^t)}{w_{t-1}^{\text{men}}(f^{t-1})}\right) - 1 &= p_{\text{women}}\left(\frac{w_t^{\text{men}}(f^{t-1}, f^L)}{w_{t-1}^{\text{men}}(f^{t-1})} - 1\right) + (1-p_{\text{women}})\left(\frac{w_t^{\text{men}}(f^{t-1}, f^H)}{w_{t-1}^{\text{men}}(f^{t-1})} - 1\right) \\
&= \left(p_{\text{women}} - (1-p_{\text{women}})\frac{p_{\text{men}}}{1-p_{\text{men}}}\right)\left(\frac{w_t^{\text{men}}(f^{t-1}, f^L)}{w_{t-1}^{\text{men}}(f^{t-1})} - 1\right) \\
&= \left(\frac{p_{\text{women}} - p_{\text{men}}}{1-p_{\text{men}}}\right)\left(\frac{w_t^{\text{men}}(f^{t-1}, f^L)}{w_{t-1}^{\text{men}}(f^{t-1})} - 1\right),
\end{aligned}$$

where the second equality uses equation (B.37).

Proof of Proposition 8

A recursive formulation of the problem is challenging as the continuation utilities for men and women under a common contract do not coincide. Hence, it is more convenient to work with the sequential problem. Without sustainability and limited commitment constraints, the employer's problem consists of maximizing

$$\Pi(v_0) = \sum_{t=0}^{\infty}\sum_{f^t}\beta^t \left[sP^{\text{men}}(f^t)\pi(w(f^t),h(f^t)) + (1-s)P^{\text{women}}(f^t)\pi(w(f^t),h(f^t))\right] \quad \text{(B.38)}$$

subject to the incentive constraints for men and women:

$$\sum_{t=0}^{\infty}\sum_{f^t}\beta^t P^{\text{men}}(f^t)U(w(f^t),h(f^t);f^t) \geq \sum_{t=0}^{\infty}\sum_{f^t}\beta^t P^{\text{men}}(f^t)U(w(\hat{f}_t(f^t)),h(\hat{f}_t(f^t));f^t) \quad \text{(B.39)}$$

$$\sum_{t=0}^{\infty}\sum_{f^t}\beta^t P^{\text{women}}(f^t)U(w(f^t),h(f^t);f^t) \geq \sum_{t=0}^{\infty}\sum_{f^t}\beta^t P^{\text{women}}(f^t)U(w(\hat{f}_t(f^t)),h(\hat{f}_t(f^t));f^t) \quad \text{(B.40)}$$

for all type histories $f^\infty \in \{f^L, f^H\}^\infty$ and reporting strategies $\hat{f} : \{f^L, f^H\}^\infty \to \{\hat{f}^L, \hat{f}^H\}^\infty$, and the time-0 participation constraints:

$$\sum_{t=0}^{\infty}\sum_{f^t}\beta^t P^{\text{men}}(f^t)U(w(f^t),h(f^t);f^t) \geq v_0 \quad \text{(B.41)}$$

$$\sum_{t=0}^{\infty}\sum_{f^t}\beta^t P^{\text{men}}(f^t)U(w(f^t),h(f^t);f^t) \geq v_0. \quad \text{(B.42)}$$

Then, assuming that there exists a contract that satisfies constraints (B.39)-(B.42), we can derive the Inverse Euler equation in the proposition with the usual perturbation argument. Fix a history f^t, and consider a perturbation where we decrease the wage utility by $\delta u(f^t) = -\varepsilon$ for $\varepsilon > 0$ small, and we increase the wage utility in the following period of both types by $\delta u(f^t, f^L) = u(f^t, f^H) = \frac{\varepsilon}{\beta}$. Due to the uniform change in utilities, this perturbation preserves all the incentive and participation constraints. Then, using $P^{\text{avg}}(f^t) \equiv sP^{\text{men}}(f^t) + (1-s)P^{\text{women}}(f^t)$ the change in the employer's value is

$$\delta\Pi = \beta^t P^{\text{avg}}(f^t)\frac{1}{u'(f^t)}(-\varepsilon) + \beta^{t+1}\left(P^{\text{avg}}(f^t,f^L)\frac{1}{u'(f^t,f^L)}\frac{\varepsilon}{\beta} + P^{\text{avg}}(f^t,f^H)\frac{1}{u'(f^t,f^H)}\frac{\varepsilon}{\beta}\right).$$

In an optimal contract, the gains from this perturbation must be zero, setting $\delta\Pi = 0$ and collecting terms

$$\frac{1}{u'(f^t)} = \frac{P^{\text{avg}}(f^t,f^L)}{P^{\text{avg}}(f^t)}\frac{1}{u'(f^t,f^L)} + \frac{P^{\text{avg}}(f^t,f^H)}{P^{\text{avg}}(f^t)}\frac{1}{u'(f^t,f^H)},$$

where

$$\begin{aligned}\frac{P^{\text{avg}}(f^t, f^L)}{P^{\text{avg}}(f^t)} &= \frac{sP^{\text{men}}(f^t, f^L) + (1-s)P^{\text{women}}(f^t, f^H)}{sP^{\text{men}}(f^t) + (1-s)P^{\text{women}}(f^t)} \\ &= \frac{sP^{\text{men}}(f^t)p_{\text{men}} + (1-s)P^{\text{women}}(f^t)p_{\text{women}}}{sP^{\text{men}}(f^t) + (1-s)P^{\text{women}}(f^t)} \\ &= p_{\text{avg}}(f^t)\end{aligned}$$

and

$$\frac{P^{\text{avg}}(f^t, f^L)}{P^{\text{avg}}(f^t)} = \frac{sP^{\text{men}}(f^t)(1-p_{\text{men}}) + (1-s)P^{\text{women}}(f^t)(1-p_{\text{women}})}{sP^{\text{men}}(f^t) + (1-s)P^{\text{women}}(f^t)} = 1 - p_{\text{avg}}(f^t).$$

Finally, using $P^{\text{men}}(f^t), P^{\text{women}}(f^t) \in (0,1)$ for all $f^t \in \{f^L, f^H\}^t$, $s \in (0,1)$ and $p_{\text{men}} < p_{\text{women}}$ it follows that $p_{\text{men}} < p_{\text{avg}}(f^t) < p_{\text{women}}$ for all $f^t \in \{f^L, f^H\}^t$.

B.3 Details Numerical Solution

There are two challenges to solving the optimal contract problem numerically. The first one, which is common to all dynamic contracting problems, is that the constraints in the dynamic programming problem are forward-looking, and as a result, the set of feasible promised utilities is not known ex-ante. This prevents using standard dynamic programming techniques. One solution is to follow (142), which consists of solving a recursive Lagrangian. The second challenge is that this approach is known to fail when the Pareto frontier is not strictly concave (57), which is the case in the termination region.

Our approach consists of first solving the principal's value function Π with the recursive Lagrangian method in a region where the (SUST) constraint does not bind: $\{v_{min}^{MM}, ..., v_{max}^{MM}\}$ with $v_{min}^{MM} >> \overline{v}$. With this solution, we can then solve the problem with a direct promised utility approach (i.e. standard dynamic programming optimizing over promised utilities) in the region $\{\overline{v}, ..., v_{min}^{MM}\}$. Because the value function has already been computed with high promised utilities, we know that we lie in the range of feasible promised utilities.

B.3.1 Recursive Lagrangian

As discussed, for high values of promised utility where the (SUST) is far from binding, we solve the model following (142). Let $\mathbf{x} = \{h^H, h^L, w^H, w^L\}$ and denote by λ and μ

the multipliers on the (PK) and (IC) constraints, respectively. We define the recursive Lagrangian as

$$\begin{aligned}\mathcal{L}(\lambda,\mu,\mathbf{x}) = (1-p)&\left[g(h^H) - w^H + (\lambda + \frac{\mu}{1-p})\left(u(w^H) - (1-f^H)\psi(h^H)\right) + \beta W(\lambda + \frac{\mu}{1-p})\right] \\ &+ p\left[g(h^L) - w^L + (\lambda - \frac{\mu}{p})u(w^L) - \left(\lambda(1-f^L) - (1-f^H)\frac{\mu}{p}\right)\psi(h^L) + \beta W(\lambda - \frac{\mu}{p})\right],\end{aligned}$$

where W solves the saddle-point problem:

$$W(\lambda) = \min_{\mu \geq 0} \max_{\mathbf{x}} \mathcal{L}(\lambda,\mu,\mathbf{x}). \tag{B.43}$$

With standard value function iteration, we can solve W and compute the policy functions $\mathbf{x}$ on a grid $\{\lambda_{min},...,\lambda_{max}\}$. For every (λ,μ) we compute the policy functions using the FOCs (B.8)-(B.11). With the policy functions, we can then also compute Π and v by VFI on the grid for λ. Combining the two we can also compute Π on a grid $\{v^{MM}_{min},...,v^{MM}_{max}\}$. Finally, we verify that $v^{MM}_{min} = v(\lambda_{min})$ is high enough such that the (SUST) is far from binding. Otherwise, we increase λ_{min} and solve again.

B.3.2 Promised Utility Approach

We now have a solution for the principal's value function Π on a grid $\{v^{MM}_{min},...,v^{MM}_{max}\}$. The next step is to solve the problem in the region $(\overline{v},...,v^{MM}_{min})$ with a promised utility approach. We solve separately for both Π^c, i.e. when the principal is not allowed to terminate, and Π. The algorithm to solve for Π^c is the following:

Algorithm Π^c:

1. Guess Π^c on a grid $\{\overline{v},...,v^{MM}_{min}\}$.
2. For every $v \in \{\overline{v},...,v^{MM}_{min}\}$, optimize over (h^H,h^L,v^L).
    - The remaining policy variables are obtained from the constraints. For every h^H, use the FOC (B.10) to solve:

$$w^H = \left(\frac{1-f^H}{z\alpha}(h^H)^{(\eta+1-\alpha)}\right)^{-\frac{1}{\sigma}}. \tag{B.44}$$

Then, at every point, combining the (IC) and (PK) constraints we can compute:[2]

$$w^L = u^{-1}\left(v - \beta v^L + \left[(1-f^L) + (1-p)(f^L - f^H)\right]\psi(h^L)\right) \tag{B.45}$$
$$v^H = \frac{1}{\beta}\left(u^L - u^H + (1-f^H)(\psi(h^H) - \psi(h^L))\right) + v^L. \tag{B.46}$$

- Check if the optimal policy x^* satisfies the (SUST) constraint. If it's not satisfied, add a large penalty in the objective.
- Compute the objective and find the point $x^* = (h^{H*}, h^{L*}, v^{L*})$ that maximizes the principal's value.

3. Update the value function, if it doesn't satisfy the tolerance go back to step 2.

Algorithm Π and $\widehat{\Pi}$: First, we check that the outside is such that $\overline{\Pi} > \Pi^c(\overline{v})$. Otherwise, there is no exit, and $\Pi^c = \Pi$. Then the algorithm proceeds as follows:

1. Start with a guess of the continuation value function $\widehat{\Pi}^{guess}$. We can use the solution of the constrained frontier Π^c from the previous part.[3]

2. Compute the guess for the value function Π^{guess}. For this, first find the grid point v^* such that the line from $(\overline{v}, \overline{\Pi})$ to $(v^*, \widehat{\Pi}^{guess}(v^*))$ is weakly above $\widehat{\Pi}^{guess}$. Compute the slope of this line as $b = \frac{\overline{\Pi} - \widehat{\Pi}^{guess}(v^*)}{v^* - \overline{v}}$. Then for $v \in \{\overline{v}, v^*\}$, compute

$$\Pi^{guess}(v) = \overline{\Pi} - b(v - \overline{v}) \tag{B.47}$$

and for $v \in \{v^*, v^{MM}_{min}\}$ set $\Pi^{guess}(v) = \widehat{\Pi}^{guess}(v)$.

3. Solve

$$\widehat{\Pi}^{pol}(v) = \max(1-p)\left[g(h^H) - w^H + \beta\Pi^{guess}(v^H)\right] + p\left[g(h^L) - w^L + \beta\Pi^{guess}(v^L)\right] \tag{B.48}$$

subject to (IC), (PK), (SUST) and (LC).[4] To solve this, repeat the step 2. of the algorithm for Π^c.

4. With $\widehat{\Pi}^{pol}$ follow the same procedure as step 2. to compute Π^{pol}. Check the distance (can do for both Π and $\widehat{\Pi}$), update guess and go back to 3. until convergence.

[2] If need to extrapolate, use the solution computed with the other method.
[3] As Π^c and Π are very similar in the region without termination. Π^c gives a good initial guess and the value function converges fast.
[4] As before, if need to extrapolate, use the solution computed with the other method

B.4 Data

The American Time Use Survey (ATUS) is a nationally representative U.S. time diary survey with detailed information on how many minutes a certain time of the day respondents spent on various activities, including work, care and leisure. Our sample comprises information from 2003 to 2022. We focus on working population aged 20 to 65, excluding self-employed. We restrict information to time diaries of typical working days, Monday to Friday. In accordance with out framework, in the baseline calibration, we focus on parents who have a least one child below the age of 12. Further to not rely on time-diary entries on working time we keep observations with answers on usual working hours.[5] To capture full-time workers, we limit the usual working hours to be at least 35 hours per week and cap maximum 60 working hours per week. This leaves with a small sample of 1072 observations. Given this being a particularly selected sample of full time workers, the sample is comprised of 627 men and 445 women.

To construct our data set we rely on tools provided by IPUMS. In particular, IPUMS allows to easily construct customized time-use variables. We create variables for both care and work activities that indicate on an hourly basis how many minutes of that hour were dedicated to the corresponding task. In particular, we look at care activities between 9 am and 5 pm, which we assume to be typical working hours.[6] Table B.2 shows summary statics for three key variables: age, number of children and the age of the youngest child. By construction the age of the youngest child is below 12 and respondents have on average 2 children.

Table B.2: Summary Statistics for Key Variables

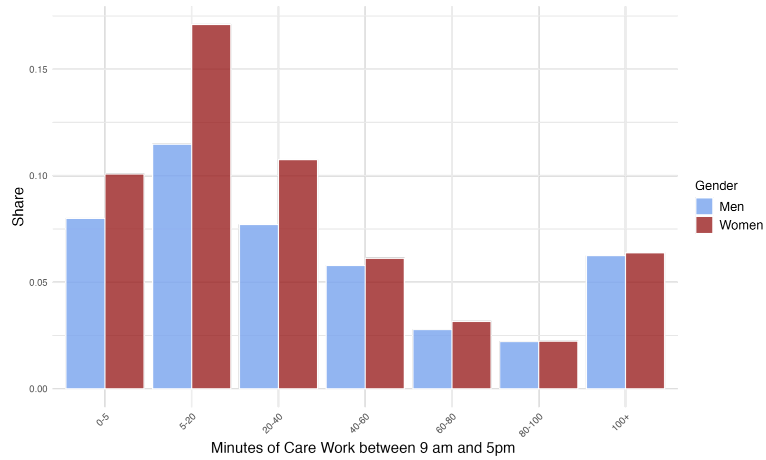
	Mean	Min	Max
Age	39	21	64
Number of Children	1.95	1	9
Age of Youngest Child	6.02	0	12

Figure B.4 shows the distribution of minutes of care work during 9 am and 5pm by gender. We see that women have a particularly high frequency of 5-20 minute interruptions, but are always more likely to do care work for any size of interruption.

[5] We do so to not inflate the care-work ratio with lower working hours resulting from the care activities on that particular day.

[6] By doing so we most likely underestimate the results for our care-work ratio because many observations report more than 40 usual working hours per week.

Figure B.4: Share of men and women with different minutes of care activities.



The care-work ratio and probabilities p_{men} and p_{women} are constructed as explained in Section 2.5. Standard errors for the latter are computed using the bootstrap method with 1000 resamples.

B.5 Incentive Compatibility in Male-tailored Contracts

In this section, we provide conditions for incentive compatibility in male-tailored contracts, i.e. when women take men's contract, and then verify numerically that they are satisfied. In a static model, this would be straightforward as the probability p would not show up in the incentive constraints. So, if a contract was incentive-compatible for men, it would also be incentive-compatible for women. However, in a dynamic setting, the continuation values will be different when women take men's contract because they depend on p. Hence, incentive compatibility does not follow directly.

We start with the IC constraint for type f^H. Let $v_w(v^H)$ and $v_w(v^L)$ denote the continuation utilities of a women taking a men's contract after f^H and f^L, respectively. Given a fixed v, the IC constraint of a woman under men's contract writes:

$$u(w^H) - (1 - f^H)\psi(h^H) + \beta v_w(v^H) \geq u(w^L) - (1 - f^H)\psi(h^L) + \beta v_w(v^L). \tag{B.49}$$

At the same time, the IC constraint for men (which binds) implies

$$v^H - v^L = \frac{1}{\beta}\left[u(w^L) - u(w^H) - (1 - f^H)(\psi(h^L) - \psi(h^H))\right]. \tag{B.50}$$

Combining the two incentive constraints, we get that incentive compatibility for women with high time availability is satisfied if:

$$v_w(v^H) - v_w(v^L) \geq v^H - v^L. \tag{B.51}$$

Second, we also need to verify that the constraint that prevents a woman with type f^L from reporting f^H, i.e.:

$$u(w^L) - (1 - f^L)\psi(h^L) + \beta v_w(v^L) \geq u(w^H) - (1 - f^L)\psi(h^H) + \beta v_w(v^H). \tag{B.52}$$

Using the incentive constraint for f^H of men (Equation B.50) and collecting terms we get that the previous constraint is equivalent to:

$$v^H - v^L + \frac{1}{\beta}(f^H - f^L)(\psi(h^H) - \psi(h^L)) \geq v_w(v^H) - v_w(v^L) \tag{B.53}$$

Intuitively, both conditions will be satisfied ((B.51) and (B.53)) if the difference $f^H - f^L$ is sufficiently larger than the difference between p_{men} and p_{women} so that $v_w(v^H) - v_w(v^L)$ is close to $v^H - v^L$.

To check these conditions numerically, we approximate v_w on a grid for v with Montecarlo simulations over a sufficiently long time horizon. That is, for every v, we first get $v^H(v)$ and $v^L(v)$ from the men's policy functions. Then, we compute $v_w(v^H(v))$ and $v_w(v^L(v))$ with a Montecarlo simulation using the policies of a man with promised utilities $v^H(v)$ and $v^L(v)$, respectively, and the probability p_{women}. To make a more accurate comparison, we also approximate the men's values $v^H(v)$ and $v^L(v)$ using the same Montecarlo simulations. Figure B.5 plots the differences in continuation utilities $v_w(v^H(v)) - v_w(v^L(v))$ and $v^H(v) - v^L(v)$ over a grid for the continuation utility v. The difference for women under men's contracts is always higher (blue line), which verifies that – in our calibration – the condition of equation (B.51) holds, and so incentive compatibility is preserved for type f^H. Figure B.6 verifies that the IC constraint for f^L also holds. The green line is much higher (in our calibration $f^H - f^L = 13.5$), so the condition (B.53) should generally be very slack.

An equivalent condition is required for men under women's contract. That is, letting $v_m(v^H)$ and $v_m(v^L)$ denote the men's continuation utilities under a women's contract, incentive compatibility for the type with high time availability is preserved if: $v_m(v^H) - v_m(v^L) \geq v^H - v^L$. Figure B.7 shows that this condition is not satisfied, so incentive compatibility is not preserved when men take women's contracts.

Figure B.5: Verification IC for women under men's contract

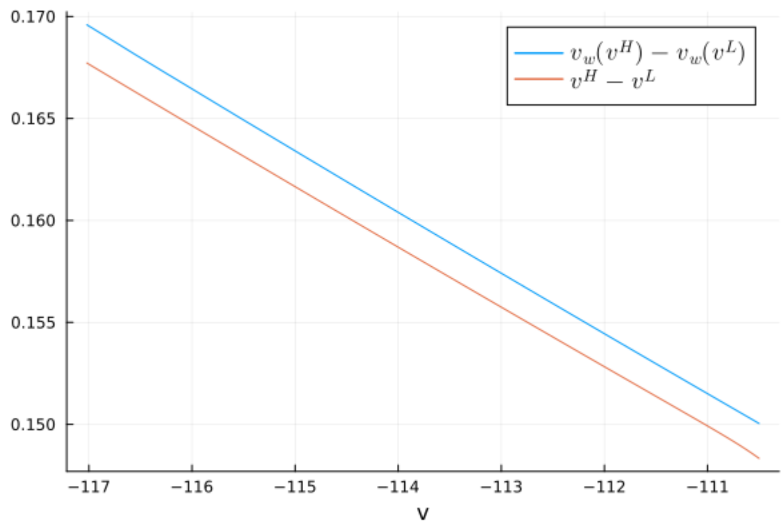


Figure B.6: Verification IC for women under men's contract (lower and upper bounds)

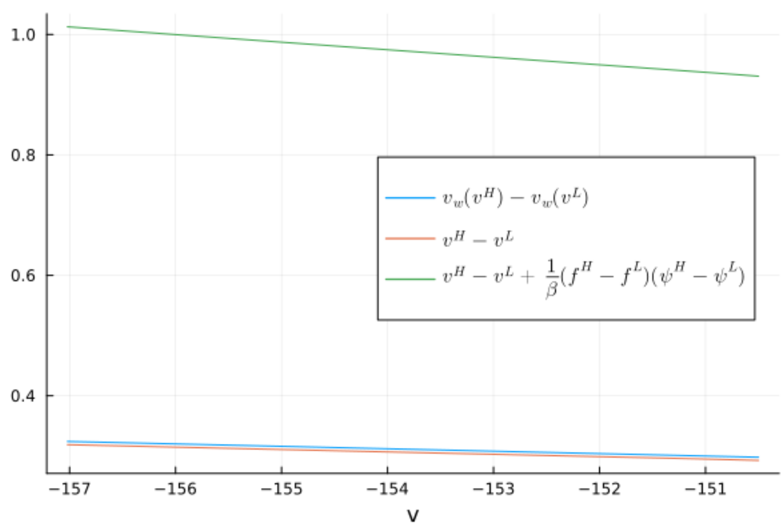
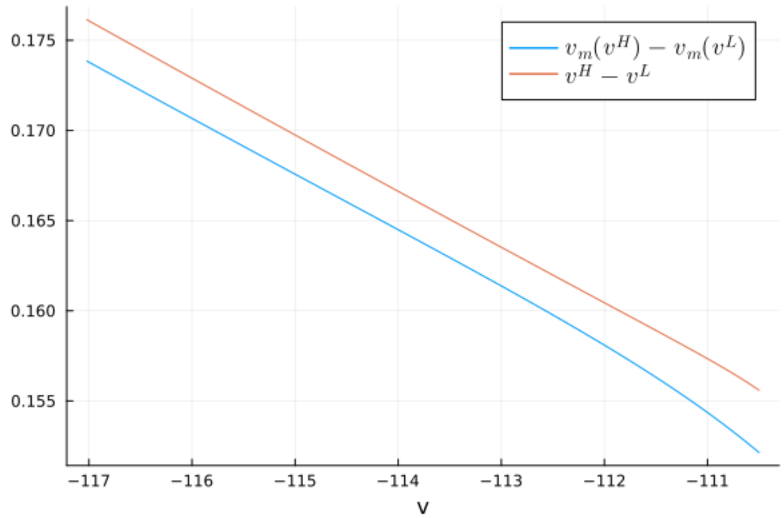


Figure B.7: Verification IC for men under women's contract



B.6 Extensions

B.6.1 Stochastic Outside Option

As discussed in the main text, we extend the model by allowing the outside option to be stochastic in order to generate termination. We assume $\overline{v}$ can take values on a grid $\{\overline{v}_1,..,\overline{v}_i,...,\overline{v}_I\}$ with corresponding probabilities $\{\overline{p}_i\}_{i=1}^I$. We also assume that this outside option is observable by the employer and realized at the end of the period but before the termination decision. Hence, the recursive problem of the employer following no termination and with realized outside option $\overline{v}_i$ writes

$$\widehat{\Pi}(v,\overline{v}_i) = \max_{\substack{w^H,h^H,v^H \\ w^L,h^L,v^L}} (1-p)\left[\pi(w^H,h^H) + \beta\sum_{j=1}^{I}\overline{p}_j\Pi(v^H,\overline{v}_j)\right] + p\left[\pi(w^L,h^L) + \beta\sum_{j=1}^{I}\overline{p}_j\Pi(v^L,\overline{v}_j)\right]$$

subject to the usual (PK) and (IC) constraints, the (SUST) constraints based on the current outside option

$$U(w^H,h^H;f^H) + \beta v^H \geq \overline{v}_i$$
$$U(w^L,h^L;f^L) + \beta v^L \geq \overline{v}_i,$$

and with the (LC) constraints based on the highest outside option

$$v^H, v^L \geq \overline{v}_I.$$

If the (LC) constraints were not based on the highest outside option, we could get that the promised utility is smaller than the outside option in some states. But then the allocation cannot satisfy simultaneously the (PK) and (SUST) constraints. Finally, the employer's problem before the termination decision and with realized outside option $\overline{v}_i$ is

$$\Pi(v, \overline{v}_i) = \max_{q \in [0,1], v_c} (1-q)\overline{\Pi} + q\widehat{\Pi}(v_c, \overline{v}_i)$$
$$\text{subject to } (1-q)\overline{v}_{i'} + qv_c = v.$$

B.6.2 Unpredictable Hours for the Employer

Our model focuses on the flexibility in working hours on the worker side. However, regular and predictable schedules have also been found to be equally as important for women (53). To capture this, we now study an extension where the employer needs flexibility in hours in the sense that, with some probability, she would like the employee to work more hours than usual. Then, we assume that the cost of working these extra hours for the employee is stochastic –e.g. some days the employee can stay longer in the office because the other parent can pick up the child from school– and unverifiable. The private information of the cost of extra hours implies similar dynamics and that (qualitatively) all results go through.

Consider a version of the model where, with probability p^N, the employer asks the employee to work "regular" hours, but with probability $(1-p^N)$, the employer needs the employee to work "extra" hours. When the employer needs regular hours, the production function is given by $z^N g(h)$, but when the employer needs extra hours by $z^E g(h)$ with $z^E > z^N$.

To capture extra costs of working overtime, we assume that if hours are below a threshold h^*, the employee disutility is $(1-f^H)\varphi(h)$. However, if hours are higher than h^*, the disutility increases to $(1-f^L)\varphi(h)$ with probability p. We assume the following GHH utility function

$$U(w,h;f) = u(w - (1-f)\varphi(h)), \tag{B.54}$$

so that hours are always invariant to compensation. When $z = z^N$, the hours of work are always given by

$$z^N g'(h^N) = (1 - f^H)v'(h^N) \tag{B.55}$$

if $h^N \leq h^*$ because there is no private information and no income effects. We assume that $h^N = h^*$, so the hours demanded will be higher than h^* whenever $z = z^E$.

For simplicity and to focus on the wage induced by the flexibility needs of the employer, we assume full commitment of the employee. Hence, we can drop the (SUST) and (LC) constraints and abstract from the termination decision. It is easy to show that the results on the optimal termination extend to this model. The employer's problem writes

$$\begin{aligned}\widehat{\Pi}(v) = \max p^N[z^N g(h^N) - w^N + \beta\Pi(v^N)]\\ + (1 - p^N)[(1-p)(z^E g(h^{E,H}) - w^{E,H} + \beta\Pi(v^{E,H})) + p(z^E g(h^{E,L}) - w^{E,L}) + \beta\Pi(v^{E,L})]\end{aligned}$$

subject to

$$\begin{aligned}&p^N[U(w^N, h^N, f^H) + \beta v^N]+\\ &(1-p^N)[(1-p)(U(w^{E,H}, h^{E,H}, f^H) + \beta v^{E,H}) + p(U(w^{E,L}, h^{E,L}, f^L) + \beta v^{E,L})] = v\end{aligned}$$

and

$$U(w^{E,H}, h^{E,H}, f^H) + \beta v^{E,H} \geq U(w^{E,L}, h^{E,L}, f^H) + \beta v^{E,L}.$$

Notice that we only need to consider the incentive constraint for the case where $z = z^E$ because if $z = z^N$, the disutility is not private information. We let λ be the multiplier on the promise-keeping constraint and $(1 - p^N)\mu$ the multiplier on the incentive constraint. Combining the FOC for w^N, v^N, and the envelope condition

$$u'_t(N) = u'_{t+1}(N). \tag{B.56}$$

So, the compensation is constant over time when the employer needs regular hours. However, when $z = z^E$, we get similar results as before. Combining the FOC $w^{E,H}$ and $h^{E,H}$ we get

$$z^E g'(h^{E,H}) = (1 - f^H)v'(h^{E,H}) \tag{B.57}$$

and

$$z^E g'(h^{E,L}) = (1 - f^L)\psi'(h^{E,L}) \left(1 + u'(w^{E,L})\frac{\mu}{p}\frac{f^H - f^L}{1 - f^L})\right) > (1 - f^L)\psi'(h^{E,L}). \tag{B.58}$$

Finally, the FOC for $v^{E,H}$ and $v^{E,L}$ give

$$-\hat{\Pi}'(v^{E,H}) = \lambda + \frac{\mu}{1-p} \tag{B.59}$$

$$-\hat{\Pi}'(v^{E,L}) = \lambda - \frac{\mu}{p}. \tag{B.60}$$

Then, it is easy to see that we obtain similar dynamics with $v^{E,H} > v^{E,L}$. Therefore, when the employer needs extra hours, the employee is penalized in case of low time availability for the extra hours. It is also easy to verify that all the results and comparisons of *gender-tailored*, *male-tailored* and *team-tailored* contracts also go through in this model.

B.6.3 Stochastic and time-varying p and the non-convergence of wages

In this section, we extend the model to allow for a stochastic and time-varying process for the probability of a low time availability p. We assume $p_t \in \{p^1, ..., p^I\} \equiv \mathcal{P}$ follows a time-dependent Markov process with transition probabilities $Q_t(p_t|p_{t-1})$. Notice that this formulation nests a deterministic process where $p_{\text{women}} < p_{\text{men}}$ for T periods and $p_{\text{women}} = p_{\text{men}}$ afterwards (e.g. when the children grow up). However, we maintain the assumption that p is observable for the employer.

For this section, we assume full commitment of the agent so that we can drop the (SUST) and (LC) constraints and abstract from the termination decision. Before the realization of p_t, the principal's state variables are (v_t, p_{t-1}, t) and its objective is

$$\Pi_t(v_t, p_{t-1}) = \max \sum_{p_t \in \mathcal{P}} Q_t(p_t|p_{t-1})$$
$$\times \{(1-p_t)\left[\pi_{t+1}(w^H(p_t), h^H(p_t)) + \beta\Pi_t(v^H_{t+1}(p_t), p_t)\right] + p_t\left[\pi(w^L(p_t), h^L(p_t)) + \beta\Pi_{t+1}(v^L_{t+1}(p_t), p_t)\right]\}$$

It will be convenient to split the promise-keeping constraints and denote by $\widetilde{v}(p_t)$ the ex-post utility after the realization of p_t. That is, we have the following constraints

$$\sum_{p_t \in \mathcal{P}} Q_t(p_t|p_{t-1})\widetilde{v}_t(p_t) = v_t \tag{B.61}$$

and for all p_t

$$(1-p_t)\left(U(w^H(p_t),h^H(p_t);f^H)+\beta v^H_{t+1}(p_t)\right)+p_t\left(U(w^L(p_t),h^L(p_t);f^L)+\beta v^L_{t+1}(p_t)\right)=\widetilde{v}_t(p_t). \tag{B.62}$$

We place multiplier λ_t on constraint (B.61) and multipliers $Q_t(p_t|p_{t-1})\widetilde{\lambda}_t(p_t)$ on constraints (B.62). The incentive constraints are as before, but we have one constraint for each p_t, and we place multipliers $Q_t(p_t|p_{t-1})\mu(p_t)$. The FOCs are:

$\widetilde{v}_t(p_t)$:

$$\lambda_t=\widetilde{\lambda}_t(p_t) \tag{B.63}$$

$w^H(p_t)$:

$$\widetilde{\lambda}_t(p_t)+\frac{\mu(p_t)}{1-p_t}=\frac{1}{u'(w^H(p_t))} \tag{B.64}$$

$w^L(p_t)$:

$$\widetilde{\lambda}_t(p_t)-\frac{\mu(p_t)}{p_t}=\frac{1}{u'(w^L(p_t))} \tag{B.65}$$

Adding up the two FOCs and using (B.63), we get that for all p_t:

$$\lambda_t=(1-p_t)\frac{1}{u'(w^H(p_t))}+p_t\frac{1}{u'(w^L(p_t))}. \tag{B.66}$$

Hence, the employee is insured against changes in p_t because the expected inverse marginal utilities are equalized across all realizations. With log-utility ($u(c)=\log(c)$), this implies that expected wages are the same for all p_t,

$$\lambda_t=(1-p_t)w^H(p_t)+p_t w^L(p_t). \tag{B.67}$$

In fact, this result holds for all future periods. The FOCs for $v^H_{t+1}(p_t)$ and $v^L_{t+1}(p_t)$ give

$$\left(-\frac{\partial\Pi_{t+1}(v^H_{t+1}(p_t),p_t)}{\partial v^H_{t+1}(p_t)}\right)=\widetilde{\lambda}_t(p_t)+\frac{\mu(p_t)}{1-p_t} \tag{B.68}$$

$$\left(-\frac{\partial\Pi_{t+1}(v^L_{t+1}(p_t),p_t)}{\partial v^L_{t+1}(p_t)}\right)=\widetilde{\lambda}_t(p_t)-\frac{\mu(p_t)}{p_t}. \tag{B.69}$$

Adding up the two FOCs and using (B.63), for all p_t we have

$$\lambda_t = (1-p_t)\left(-\frac{\partial \Pi_{t+1}(v^H_{t+1}(p_t), p_t)}{\partial v^H_{t+1}(p_t)}\right) + p_t\left(-\frac{\partial \Pi_{t+1}(v^L_{t+1}(p_t), p_t)}{\partial v^L_{t+1}(p_t)}\right)$$
$$= (1-p_t)\lambda^H_{t+1} + p_t\lambda^L_{t+1}.$$

where the second line substitutes the envelope conditions. Iterating forwards and using history notation, we have, for all $\tau \geq 1$

$$\lambda_t = \mathbb{E}\left[\lambda_{t+\tau}(f^{t+\tau})|f^t\right] \tag{B.70}$$

$$= \mathbb{E}\left[\frac{1}{u'(w_{t+\tau}(f^{t+\tau}))}|f^t\right] \tag{B.71}$$

and assuming log utility

$$\lambda_t = \mathbb{E}_{t-1}\left[w_{t+\tau}(f^{t+\tau})|f^t\right], \tag{B.72}$$

which holds for all p_t.

Appendix C

Appendix of Chapter 3: Firm Size and Compensation Dynamics with Risk Aversion and Persistent Private Information

C.1 Appendix: Derivations and proofs

Optimality conditions problem (3.15). The optimality condition for $b_t(\theta^t)$ is:

$$\xi_t(\theta^t) = \frac{1}{u'(\theta^t)} \left[1 + \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} \iota f_\theta(\theta^t) u''(\theta^t) \right] \tag{C.1}$$

The envelope conditions are

$$\frac{\partial K_{t+1}}{\partial v_t(\theta^t)} = \lambda_{t+1}(\theta^t) \tag{C.2}$$

$$\frac{\partial K_{t+1}}{\partial \Delta_t(\theta^t)} = \gamma_t(\theta^t) \tag{C.3}$$

$$\frac{\partial K_{t+1}}{\partial k_{t+1}(\theta^t)} = \mathbb{E}\left[-\xi_{t+1}(\theta^{t+1}) u'(\theta^{t+1}) f_k(\theta^{t+1}) | \theta_t \right] + \mathbb{E}\left[\frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_t)} \left(u''(\theta^{t+1}) \iota f_\theta(\theta^{t+1}) f_k(\theta^{t+1}) + u'(\theta^{t+1}) \iota f_{\theta k}(\theta^{t+1}) \right) | \theta_t \right]. \tag{C.4}$$

Using the envelope conditions (C.2) and (C.3), the optimality conditions for $v_t(\theta^t)$ and $\Delta_t(\theta^t)$ write:

$$\lambda_{t+1}(\theta^t) = \frac{\beta}{q}\xi_t(\theta^t) \tag{C.5}$$

$$\gamma_{t+1}(\theta^t) = -\frac{\beta}{q}\frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta_{t-1})}. \tag{C.6}$$

Substituting (C.1) and (C.4) into the FOC for $k_{t+1}(\theta^t)$ we get

$$\frac{1}{q} = \mathbb{E}\left[f_k(\theta^{t+1}) - \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_t)}u'(\theta^{t+1})\iota f_{\theta k}(\theta^{t+1})|\theta_t\right]. \tag{C.7}$$

Finally, the law of motion for the co-state is

$$\dot{\mu}_t(\theta^t) = -\left[\xi_t(\theta^t) - \lambda_t - \gamma_t\mathcal{E}(\theta_t,\theta_{t-1})\right]\varphi_t(\theta_t|\theta_{t-1}) \tag{C.8}$$

Proof of Proposition 9. Set $\mu_t(\theta^t) = 0$ for all θ^t, then from Equation (C.7) we obtain point 3. For point 2, note that with $\mu_t(\theta^t) = 0$, Equation (C.8) becomes $\xi_t(\theta^t) = \lambda_t$. From Equation (C.1), $\frac{1}{u'(\theta^t)} = \xi_t(\theta^t)$ and using (C.5) gives point 2. Point 1 holds in the first best and second best allocations.

Proof of Proposition 10. From the FOC for $k_{t+1}(\theta^t)$ (Equation (C.7)), multiplying the second term inside the expectation by $\frac{f_k(\theta^{t+1})}{f_k(\theta^{t+1})}$ and letting

$$\tau^k(\theta^{t+1}) = \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_t)}u'(\theta^{t+1})\iota\frac{f_{\theta k}(\theta^{t+1})}{f_k(\theta^{t+1})}, \tag{C.9}$$

we have $\frac{1}{q} = \mathbb{E}[f_k(\theta^{t+1})(1-\tau^k(\theta^{t+1}))|\theta_1]$. Combining with the definition of $\tilde{\tau}^k(\theta^t)$,

$$(1-\tilde{\tau}^k(\theta^t))\mathbb{E}\left[f_k(k_{t+1}(\theta^t),\theta_{t+1})|\theta^t\right] = \mathbb{E}\left[f_k(k_{t+1}(\theta^t),\theta_{t+1})\left(1-\tau^k(\theta^{t+1})\right)|\theta^t\right],$$

or

$$\tilde{\tau}^k(\theta^t) = \frac{\mathbb{E}\left[f_k(k_{t+1}(\theta^t),\theta_{t+1})\tau^k(\theta^{t+1})|\theta^t\right]}{\mathbb{E}\left[f_k(k_{t+1}(\theta^t),\theta_{t+1})|\theta^t\right]}$$

Finally, to write the return-dependent wedge as in the proposition $\tau^k(\theta^{t+1})$ multiplying by $\frac{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)}{1-\Phi_{t+1}(\theta_{t+1}|\theta_t)}$ and rearrange terms.

Proof of Proposition 11. The proof follows similar steps as Proposition 1 in Hellwig (114). Substitute $\xi_t(\theta^t)$ in the LOM of the co-state (C.8)

$$\dot{\mu}_t(\theta^t) + \mu_t(\theta^t)\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} = -\left[\frac{1}{u'(\theta^t)} - \lambda_t - \gamma_t\mathcal{E}(\theta_t,\theta_{t-1})\right]\varphi_t(\theta_t|\theta_{t-1}),$$

substitute $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}$, using the boundary conditions $\mu_t(\underline{\theta}) = 0$ and $\mu_t(\bar{\theta}) = 0$ and integrating upwards

$$-\mu_t(\theta^t)m(\theta^t) = \int_{\theta_t}^{\bar{\theta}}\left[\lambda_t + \gamma_t\mathcal{E}(\theta',\theta_{t-1}) - \frac{1}{u'(\theta',\theta^{t-1})}\right]\varphi_t(\theta'|\theta_{t-1})m(\theta^t)d\theta'.$$

Using the definition of the incentive-adjusted measure

$$\begin{aligned}\mu_t(\theta^t)m(\theta^t) =&(1-\Phi(\theta_t|\theta_{t-1}))\mathbb{E}\left[m(\theta^{t-1},\theta')|\theta' > \theta_{t+1},\theta_t\right] \\ &\times\left\{\hat{\mathbb{E}}\left[\frac{1}{u'\left(\theta',\theta^{t-1}\right)} \mid \theta' \geq \theta_t,\theta^{t-1}\right] - \gamma_t\hat{\mathbb{E}}\left[\mathcal{E}(\theta',\theta_{t-1}) \mid \theta' \geq \theta_t,\theta^{t-1}\right] - \lambda_t\right\}.\end{aligned} \tag{C.10}$$

To get λ_t, note that using the boundary conditions we have

$$0 = \int_{\underline{\theta}}^{\bar{\theta}}\left[\lambda_t + \gamma_t\mathcal{E}(\theta',\theta_{t-1}) - \frac{1}{u'(\theta',\theta_{t-1})}\right]\varphi_t(\theta_t|\theta^{t-1})m(\theta^t)d\theta'$$

or

$$\lambda_t = \hat{\mathbb{E}}\left[\frac{1}{u'\left(\theta^t\right)} \mid \theta^{t-1}\right] - \gamma_t\hat{\mathbb{E}}\left[\mathcal{E}(\theta_t,\theta_{t-1}) \mid \theta^{t-1}\right].$$

Substituting back λ_t into Equation (C.10), using the definition of $\hat{\rho}(\theta^t)$ (Equation (3.21)), multiply both sides by $u'(\theta^t)$, collecting terms and using $\widetilde{\mu}_t(\theta^t) = \frac{\mu_t(\theta^t)}{1-\Phi_t(\theta_t|\theta_{t-1})}u'(\theta^t)$ we get the equation in the propositon. Finally, the inequalities $\hat{MB}(\theta^t) \geq 0$ and $\hat{\rho}_t(\theta^t) \geq 0$ follow by using that $\frac{1}{u'(\theta')}$ and $\mathcal{E}(\theta',\theta_{t-1})$ are non-decreasing due to incentive compatibility and assumption A.1, respectively.

Proof of Proposition 12. Equation (C.10) also holds for period $t = 1$, hence we only need to characterize $\gamma_1(\theta_0)(\equiv -MB_0(\theta_0))$ from the time 0 problem. I start with the case $\kappa = 1$. Combining the optimality condition for $\Delta_0(\theta_0)$ and the Envelope condition (C.3) gives:

$$\gamma_1(\theta_0) = -\frac{1}{q}\frac{\mu(\theta_0)}{h(\theta_0)} \tag{C.11}$$

Using $\dot{w}_0(\theta_0) = \beta \dot{v}_0(\theta_0)$ and the Envelope condition (C.2), the LOM of the co-state satisfies

$$\dot{\mu}(\theta_0) = -\left[q\lambda_1(\theta_0) - \beta\lambda_0\right] h(\theta_0).$$

Integrating, we get $q\mathbb{E}_h\left[\lambda_1(\theta_0)\right] = \beta\lambda_0$, and

$$\frac{1}{q}\frac{\mu(\theta_0)}{h(\theta_0)} = \frac{1-H(\theta_0)}{h(\theta_0)}\left\{\mathbb{E}_h\left[\lambda_1(\theta_0')|\theta_0' > \theta_0\right] - \mathbb{E}_h\left[\lambda_1(\theta_0)\right]\right\} = MB_0(\theta_0)$$

where the last equality uses (C.11). The inequality $MB_0(\theta_0) \geq 0$ holds because $v_0(\theta_0)$ is increasing in θ_0 (by the IC constraint) and the multiplier $\lambda_1(\theta_0)$ is increasing in $v_0(\theta_0)$.

For the case $\kappa = 0$, notice we can set $\lambda_0 = 0$ and because $w_0(\underline{\theta}_0)$ is not a free variable we only have the boundary condition $\mu(\overline{\theta}_0) = 0$. Integrating the LOM:

$$\mu(\theta_0) = \mu(\underline{\theta}_0) - qH(\theta_0)\mathbb{E}\left[\lambda_1(\theta_0')|\theta_0' < \theta_0\right].$$

So we have $\mu(\underline{\theta}_0) = \mathbb{E}\left[\lambda_1(\theta_0)\right]$ and $\frac{1}{q}\frac{\mu(\theta_0)}{h(\theta_0)} = \frac{1-H(\theta_0)}{h(\theta_0)}\mathbb{E}\left[\lambda_1(\theta_0')|\theta_0' > \theta_0\right]$. Finally, if θ_0 is fixed we can drop the incentive constraint so $\mu(\theta_0) = 0$ and $MB_0(\theta_0) = 0$.

Proof of Proposition 13. This proof also follows similar steps to Theorem 1 in Hellwig (114). Using the characterization of λ_t in Proposition 11 and substitute the multipliers $\lambda_{t+1}(\theta^t)$ and $\gamma_{t+1}(\theta^t)$ from the optimality conditions (C.5) and (C.6), and using equation (C.1) to substitute for ξ_t:

$$\frac{1}{u'(\theta^t)} + \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta_{t-1})}\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} = \frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] + \frac{\mu(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})}\hat{\mathbb{E}}\left(\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right), \tag{C.12}$$

Then, we can show that

$$\hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right] = \hat{\mathbb{E}}\left[\rho(\theta^{t+1})\frac{u''(\theta^{t+1})\iota f_\theta(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right].$$

First, notice that

$$\hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right] = \frac{1}{\mathbb{E}\left[m(\theta^{t+1})|\theta_t\right]}\int_{\underline{\theta}}^{\overline{\theta}}\left(-\int_{\theta_{t+1}}^{\overline{\theta}}\mathcal{E}(\theta',\theta_t)\varphi(\theta'|\theta_t)d\theta'\right)' m(\theta^{t+1})d\theta_{t+1}.$$

Integrate by parts and use $\mathbb{E}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta_t\right] = \int_{\underline{\theta}}^{\overline{\theta}}\frac{\partial\varphi(\theta_{t+1}|\theta_t)}{\partial\theta_t}d\theta_{t+1} = 0$. Then using the definition of $\rho(\theta^{t+1})$ and $\frac{m'(\theta^t)}{m(\theta^t)} = \frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}$,

$$\hat{\mathbb{E}}\left[\mathcal{E}(\theta_{t+1},\theta_t)|\theta^t\right] = \int_{\underline{\theta}}^{\overline{\theta}}\int_{\theta_{t+1}}^{\overline{\theta}} \mathcal{E}(\theta_{t+1},\theta_t)\varphi_{t+1}(\theta'|\theta^t)d\theta' \frac{m'(\theta^{t+1})}{\mathbb{E}\left[m(\theta^{t+1})|\theta_t\right]}d\theta_{t+1}$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{\varphi_{t+1}(\theta_{t+1}|\theta^t)}\int_{\theta_{t+1}}^{\overline{\theta}} \mathcal{E}(\theta',\theta^t)\varphi_{t+1}(\theta'|\theta^t)d\theta' \frac{m'(\theta^{t+1})}{m(\theta^{t+1})}\frac{m(\theta^{t+1})}{\mathbb{E}\left[m(\theta^{t+1})|\theta_t\right]}\varphi_{t+1}(\theta_{t+1}|\theta_t)d\theta_{t+1}$$

$$= \int_{\underline{\theta}}^{\overline{\theta}} \rho(\theta^{t+1})\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)}\hat{\varphi}_{t+1}(\theta_{t+1}|\theta^t)d\theta_{t+1}.$$

Substitute back and use Equation (C.9) to substitute $\frac{\mu_t(\theta^t)}{\varphi(\theta_t|\theta^{t-1})}$:

$$\frac{1}{u'(\theta^t)} + \frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)}\frac{\tau^k(\theta^t)}{u'(\theta^t)}\frac{u''(\theta^t)\iota f_\theta(\theta^t)}{u'(\theta^t)} =$$
$$\frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] + \frac{f_k(\theta^t)}{f_{\theta k}(\theta^t)}\frac{\tau^k(\theta^t)}{u'(\theta^t)}\hat{\mathbb{E}}\left[\rho(\theta^{t+1})\frac{u''(\theta^{t+1})\iota f_\theta(\theta^{t+1})}{u'(\theta^{t+1})}|\theta^t\right].$$

Finally, collecting terms and using the definition of the savings wedge $s(\theta^t)$ we get the equation in the proposition.

Proof of Proposition 14. If $s(\theta^t) \geq 0$, we have

$$\frac{1}{u'(\theta^t)} \leq \hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \frac{\mathbb{E}\left[M(\theta^{t+1})|\theta_t\right]}{\mathbb{E}\left[u'(\theta^{t+1})M(\theta^{t+1})|\theta_t\right]},$$

where $M(\theta^{t+1}) = \frac{m(\theta^{t+1})}{u'(\theta^{t+1})}$, rearranging $\frac{\mathbb{E}\left[u'(\theta^{t+1})M(\theta^{t+1})|\theta_t\right]}{\mathbb{E}\left[M(\theta^{t+1})|\theta_t\right]} \leq u'(\theta^t)$. Because $u'(\theta^{t+1})$ is decreasing, to show $\mathbb{E}\left[u'(\theta^{t+1})|\theta^t\right] \leq u'(\theta^t)$ we only need to show that $M(\theta^{t+1})$ is weakly decreasing. Differentiating

$$\frac{d}{d\theta_{t+1}}\left(M(\theta^{t+1})\right) = M(\theta^{t+1})\frac{u''(\theta^{t+1})}{u'(\theta^{t+1})}\left(\iota f_\theta(\theta^{t+1}) - c'(\theta^{t+1})\right)$$

from the the local IC constraint $\frac{dw(\theta^{t+1})}{d\theta_{t+1}} = \frac{\partial w(\theta^{t+1})}{\partial\theta_{t+1}}$ we have

$$c'(\theta^{t+1}) + \beta\frac{\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right)}{u'(\theta^{t+1})} = \iota f_\theta(\theta^{t+1}).$$

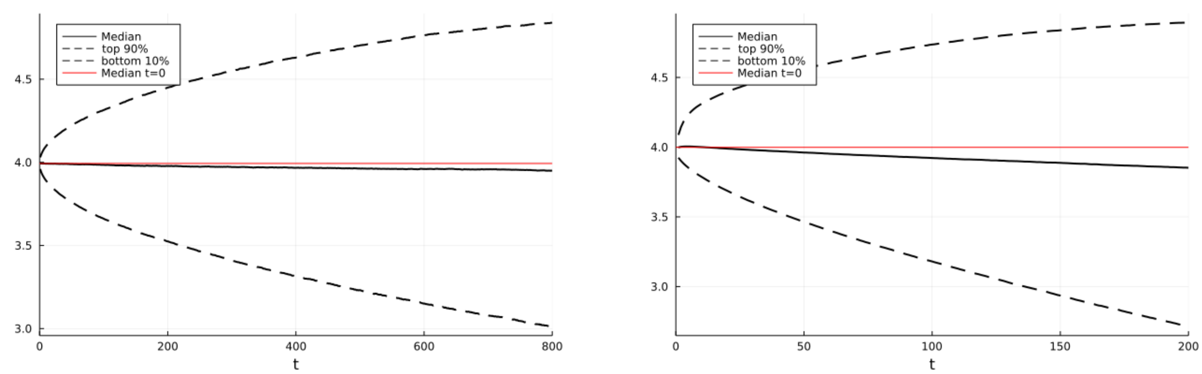
Then, the assumption $\mathbb{E}\left(\frac{dw(\theta^{t+2})}{d\theta_{t+1}}|\theta_{t+1}\right) \geq 0$ implies $\iota f_\theta(\theta^{t+1}) - c'(\theta^{t+1}) \geq 0$, and so $M(\theta^{t+1})$ is weakly decreasing. For the second part, assume $s(\theta^t) \geq 0$ for all θ^t, then u' follows a non-negative super-martingale. By Doob's super-martingale convergence theorem, u' converges almost surely to a finite limit. By contradiction, assume u' converges to a positive limit $\overline{u'} > 0$. Then, almost sure convergence implies that for some τ we have $u'(\theta^\tau) = u'(\theta^\tau, \theta_{\tau+1}) = ... = \overline{u'}$, which would violate incentive compatibility.[1] Hence, we must have $u' \to 0$ almost surely.

C.2 Additional tables and figures

Figure C.1: Immiseration in the long run

(a) Consumption i.i.d ($\rho = 0$)

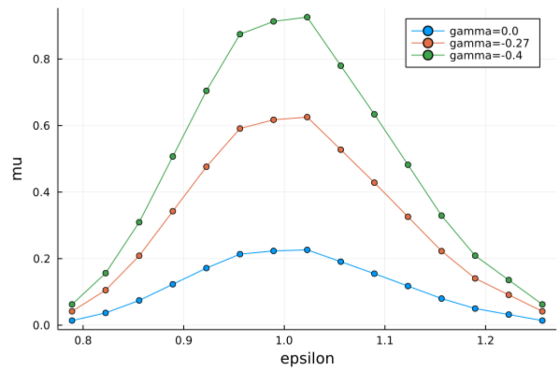
(b) Consumption with persistence ($\rho = 0.7$)



Note: The figures show the median, 10%, and 90% quantiles of the distribution of consumption at every period. For reference, the red line displays the median at period $t = 0$. The median monotonically decreases and the growth of the 90% quantile decreases over time. This implies that consumption will converge to its lower bound. However, we also observe that this convergence is very slow.

[1] Formally, this is statement is true because the proposition assumes fixed capital. Otherwise, this would be incentive compatible by setting $k_{t+1}(\theta^t) = 0$, which is generally not optimal because of the Inada condition $\lim_{k \to 0} f_k(k, \theta) = \infty$.

Figure C.2: Shadow cost information rents μ at different γ



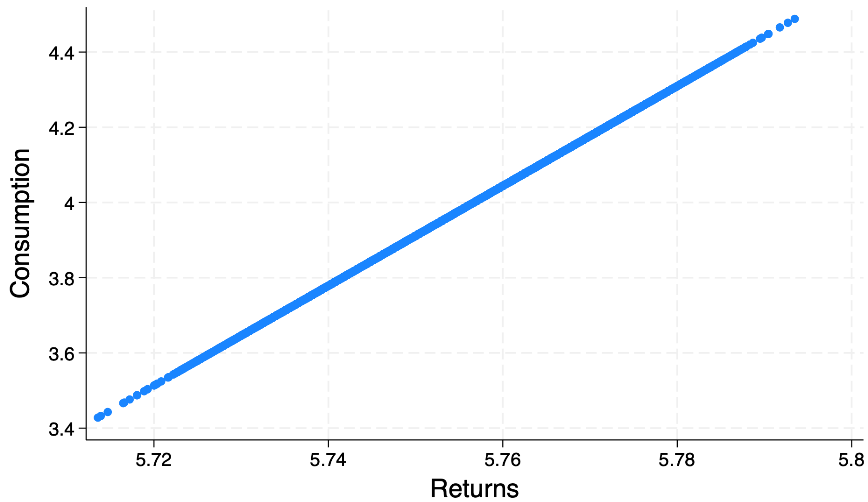
Note: For a fixed (λ_-, k, θ_-), the figure shows the shadow cost μ as a function of the shock ε for different γ_-. The dynamic information rent Δ_- is increasing in γ_-. So when the agent is promised lower information rents (low Δ_- and γ_-), the shadow costs μ are higher. The increase is more pronounced for the types in the middle.

Table C.1: Regressions with i.i.d type process

	(1) c_t	(2) c_t	(3) c_t	(4) c_t	(5) c_t
$returns_t$	0.0489	0.0511	0.0508	0.0704	
	(10558.17)	(399.15)	(108.58)	(926.10)	
v_{t-1}	0.792		0.790	0.792	
	(14968.42)		(1459.73)	(15010.68)	
$returns_{t-5}$		0.0490			
		(382.92)			
$returns_t * v_{t-1}$			0.000386		
			(4.13)		
$returns_t^2$				-0.00185	
				(-283.17)	
c_{t-1}					0.998
					(5300.13)
N	2400000	1900000	2400000	2400000	2300000
R^2	0.999	0.139	0.999	0.999	0.924

t statistics in parentheses

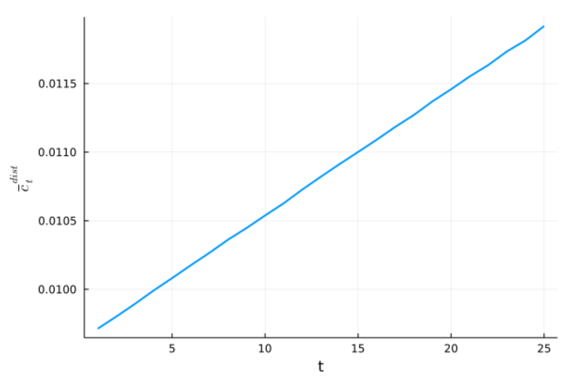
Figure C.3: Consumption and returns residualized (i.i.d)



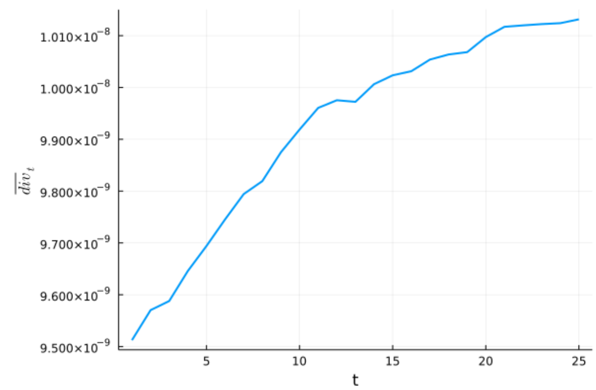
Note: Using simulated data, I residualize both consumption and returns on v_{t-1}.

Figure C.4: Simulations implementation i.i.d

(a) Average distance consumption SB and implementation



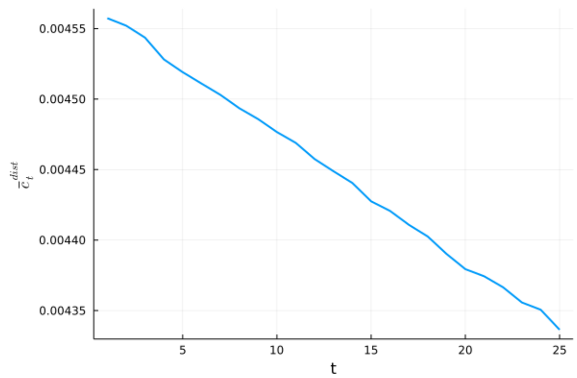
(b) Average fraction of diverted funds



Note: The left figure shows, for every period, the average distance between consumption in the optimal contract (c^{SB}) and the implementation (c^{I}), i.e. $\bar{c}_t^{dist} = \frac{1}{N}\sum_i \sqrt{(c_t^{SB}(\{\varepsilon_{i,\tau}\}_{\tau=1}^{t}) - c_t^{I}(\{\varepsilon_{i,\tau}\}_{\tau=1}^{t}))^2}$. The right figure shows the average of the diverted funds as a fraction of total returns, i.e. $\overline{div}_t = \frac{1}{N}\sum_i \frac{f(k_{SB},\theta_i^t) - f(k_{SB},\widetilde{\theta}_t(\theta_i^t))}{f(k_{SB},\theta_i^t)}$

Figure C.5: Simulations implementation i.i.d with log utility ($\sigma = 1$)

(a) Average distance consumption SB and implementation



(b) Average fraction of diverted funds

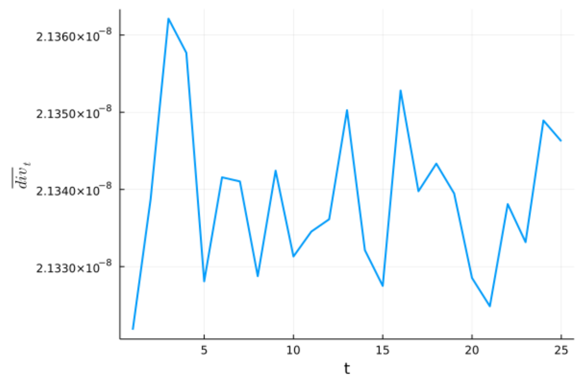
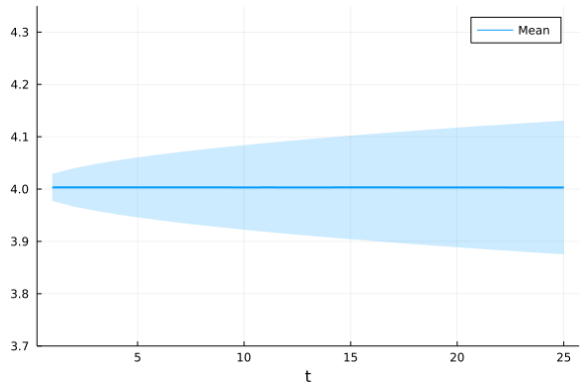
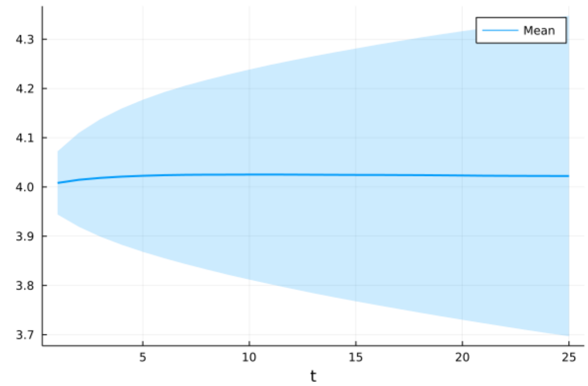


Figure C.6: Log utility ($\sigma = 1$)

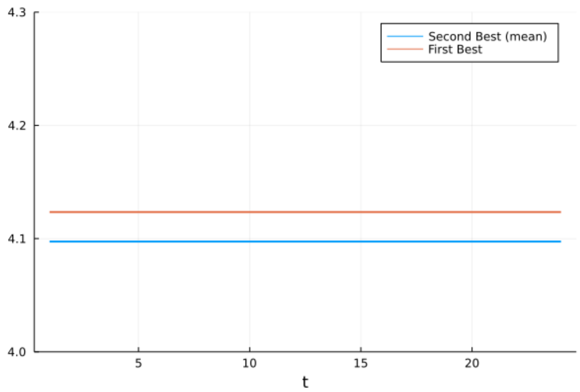
(a) Consumption i.i.d ($\rho = 0$)



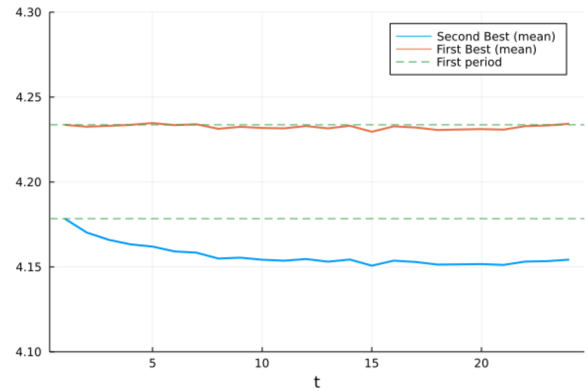
(b) Consumption with persistence ($\rho = 0.7$)



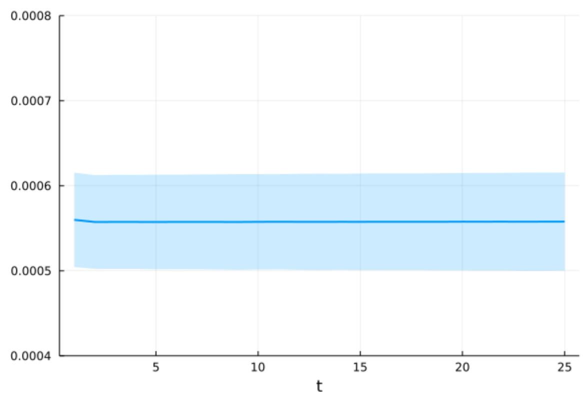
(c) Firm size i.i.d ($\rho = 0$)



(d) Firm size with persistence ($\rho = 0.7$)



(e) Investment wedge i.i.d ($\rho = 0$)



(f) Investment wedge with persistence ($\rho = 0.7$)

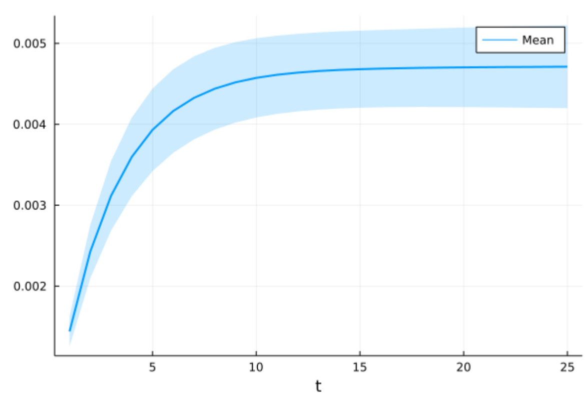
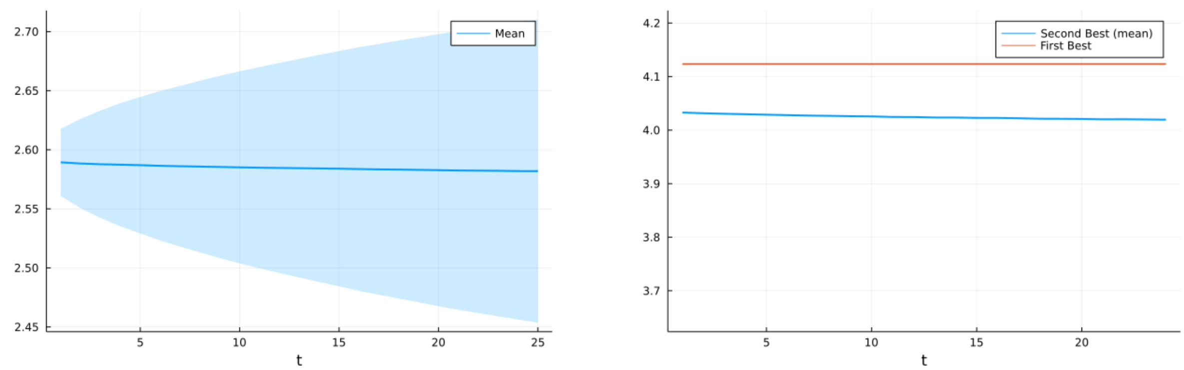


Figure C.7: CARA utility

(a) Consumption i.i.d ($\rho = 0$)

(b) Firm size i.i.d ($\rho = 0$)



C.3 Extensions

C.3.1 Limited commitment

In this section, I relax the assumption of full commitment of the entrepreneur. Limited commitment leads to firm size and compensation dynamics that are very different from those with the private information friction. The limited commitment works as follows. At every period, before knowing the realization of his productivity, the entrepreneur can divert and consume all the funds advanced by the lender and terminate the project. In this case, I assume the entrepreneur would obtain utility $h(k_{t+1}(\theta^t))$, where h is increasing and concave. Therefore, the agent will not terminate the project at period $t+1$ if $h(k_{t+1}(\theta^t)) \leq v_t(\theta^t)$.[2] This limited commitment constraint can be added directly to the planning problem (3.15). Because the limited commitment constraint does not affect the within-period insurance and incentives trade-off, the characterization of the shadow cost of information rents (Proposition 11) is not affected by the limited commitment assumption.

However, the limited commitment constraint does modify the consumption and firm size dynamics. Let $\eta_t(\theta^t)$ be the multiplier on the limited commitment constraint. Then,

[2]A natural specification of the function h is $\frac{u((1-\iota)(1-q)k_{t+1})}{1-\beta}$, this is the value that the agent would obtain if he could keep a fraction $(1-\iota)$ of the capital and then save outside the contract at rate $\frac{1}{q}$.

the GIEE is given by

$$\frac{q}{\beta}\hat{\mathbb{E}}\left[\frac{1}{u'(\theta^{t+1})}|\theta^t\right] = \frac{1}{u'(\theta^t)}(1+s(\theta^t)) + \frac{\eta_t(\theta^t)}{\beta}.$$

Because $\eta_t(\theta^t) \geq 0$, the limited commitment gives a force to have a downward drift in marginal utilities. As is well known, in models with only limited commitment, the agent's consumption is backloaded, and consumption follows a sub-martingale. Therefore, the private information and limited commitment frictions will generally have opposite effects on consumption dynamics.

The investment wedge is now given by

$$\tau^{k,LC}(\theta^{t+1}) = \tau^k(\theta^t) + \eta_t(\theta^t)\frac{h'(k_{t+1}(\theta^t))}{f_k(k_{t+1}(\theta^t),\theta_t)} \geq 0,$$

where $\tau^k(\theta^t)$ is the wedge from the private information friction in Proposition 10. Because $\eta_t(\theta^t)h'(k_{t+1}(\theta^t)) \geq 0$, the limited commitment friction also lowers firm size relative to the first. However, if the promised utility increases over time, the limited commitment constraint will eventually not bind ($\eta_t(\theta^t) = 0$). Therefore, this friction still gives an incentive to have firm size increasing over time.

C.3.2 Endogenous termination

In this section, I show how the model can be extended to allow for endogenous termination of the contract. As is well known, in regions of the state space where the contract becomes very inefficient, the principal may be better off terminating the project or randomizing between terminating and continuing the contract at an efficient point. I assume that after termination, the lender receives a scrap value S. At period t, based on θ^t, the lender can choose a probability $\alpha_{t+1}(\theta^t)$ of termination at $t+1$. In that event, the principal can also give the entrepreneur a compensation of $Q_{t+1}(\theta^t)$. In case of no termination at period t the objective of the principal is

$$\int \left[-b(\theta^t) + \alpha_{t+1}(\theta^t)q\left(S - Q_{t+1}(\theta^t)\right) + \left(1-\alpha_{t+1}(\theta^t)\right)\left(k_{t+1}(\theta^t) + qK_{t+1}(v_t(\theta^t),\Delta_t(\theta^t),\theta^t,k_{t+1}(\theta^t))\right)\right]$$
$$\times \varphi_t(\theta_t|\theta_{t-1})d\theta_t.$$

I assume that after terminating the contract, the entrepreneur can freely save $Q_{t+1}(\theta^t)$ and obtains a per period gross return $\frac{1}{q}$. Then, his value after terminating the contract is $\frac{u\left((1-q)Q_{t+1}(\theta^t)\right)}{(1-q)}$. The continuation utility now becomes

$$w_t(\theta^t) = u(c(\theta^t)) + \beta \left[\alpha_{t+1}(\theta^t) \frac{u\left((1-q)Q_{t+1}(\theta^t)\right)}{(1-q)} + (1 - \alpha_{t+1}(\theta^t)) v_t(\theta^t) \right],$$

and the local IC

$$\dot{w}_t(\theta^t) = u'(c(\theta^t)) \iota f_\theta(k_t, \theta_t) + \beta(1 - \alpha_{t+1}(\theta^t)) \Delta_t(\theta^t).$$

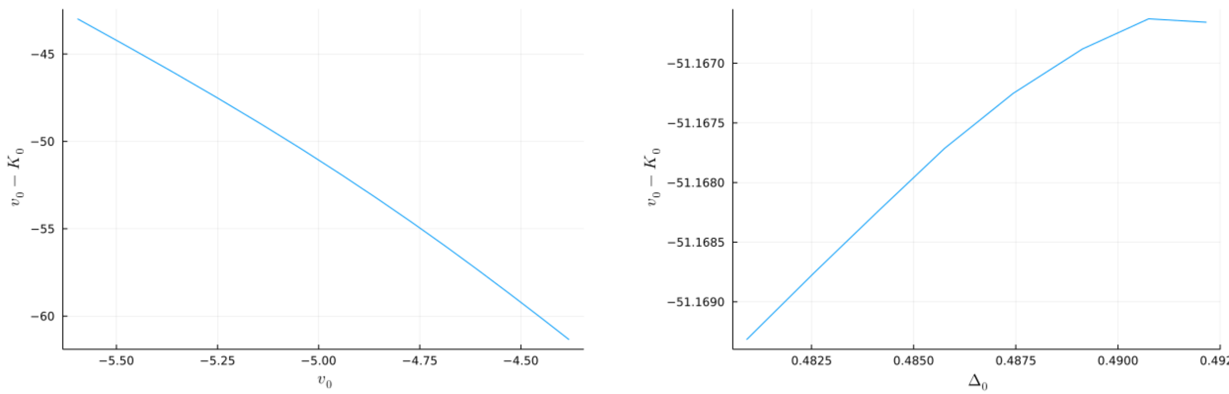
It is then easy to see that the optimality conditions for $b(\theta^t)$, $k_{t+1}(\theta^t)$,$v_t(\theta^t)$, $\Delta_t(\theta^t)$ and $w_t(\theta^t)$ are the same as in the main model. Therefore, although it may be optimal to terminate the contract, the characterizations of the optimal allocation presented in the paper do not rely on the assumption of no termination.

It is interesting to understand when and where (in the state space) termination may occur in this model and how it compares with the risk neutral case. With risk neutrality (54) the Pareto frontier is increasing in the promised utility. Consequently, termination occurs when the promised utility is low. Moreover, because utilities drift upward, the termination probability tends to decrease over time. With risk aversion, the Pareto frontier is decreasing in promised utility (see Figure C.8a). So, the motives for termination differ from the risk neutral case.

Two other sources of inefficiencies could motivate the termination of the firm. First, low promised information rents (i.e. low Δ_0) is inefficient. Panel C.8b shows that the Pareto frontier is increasing in Δ_0. Over time, if the lender has promised too low information rents to the entrepreneur, termination could potentially become optimal. Second, introducing a limited commitment constraint as in Section C.3.1 could also generate endogenous termination. If $v_t(\theta^t)$ decreases sufficiently, the limited commitment constraint may require that $k_{t+1}(\theta^t) \to 0$. Then, as shown in Dovis (78), the Inada condition $\lim_{k\to 0} f_k(k,\theta) = \infty$ implies that in this region, the Pareto frontier is increasing in v. When the frontier is increasing, there may be a range of scrap values S where it is optimal for the lender to randomize between termination and continuing at a higher v. Because the shadow costs of information rents and the variance of promised utility tend to increase, both inefficiencies should imply that the termination probabilities tend to increase over time. Again, these are the opposite dynamics of what is found with risk neutrality.

Figure C.8: Pareto frontier

(a) As a function of promised utility v_0 (b) As a function of promised information rent Δ_0



Note: For Panel (a), the promised information rent is set at the optimal level at $t = 0$, i.e. I set $\gamma_0 = 0$. For Panel (b), λ_0 is adjusted for every value of Δ_0 so that v_0 is kept fixed.

Persistent shocks and optimal termination probabilities. Another interesting observation is that whenever termination is optimal, the lender may have more incentives to increase the termination probabilities when the persistence of the shocks is higher. The reason is that a higher termination probability decreases the dynamic information rents. The intuition is similar to the equity purchases and the distortions in firm size. Imagine that, at history θ^{t-1}, the lender increases the termination probability of type θ_t and compensates him by increasing $Q(\theta^t)$ such that his ex-ante continuation utility is kept constant. Types $\theta' > \theta_t$ know they are expected to obtain higher returns at $t+1$, so they have a relatively higher preference for continuing to operate the firm. Therefore, the increase in $\alpha(\theta^t)$ makes deviations less attractive for $\theta' > \theta_t$, and so it lowers the cost of screening types.

I show this intuition more formally in a simplified two-period and two-type version of the model. Assume the entrepreneur's productivity can take values $\{\theta^H, \theta^L\}$ with $\theta^H > \theta^L$. In the first period, $P(\theta_1 = \theta^H) = p^1$, and in the second one, $P(\theta_2 = \theta^H | \theta_1 = \theta^H) = p^H$ and $P(\theta_2 = \theta^H | \theta_1 = \theta^L) = p^L$. Let $\rho = p^H - p^L \geq 0$, if $\rho > 0$ we say types are persistent. We assume the production function is of the form $f(\theta, k) = \theta$, so we can abstract away from the choice of firm size. In the second (and last) period, we assume that the entrepreneur consumes all its endowment, so there is no repayment. The principal's objective is

$$K(v) = p^1 \left[-b^H + q\alpha^H (S - Q^H) \right] + (1 - p^1) \left[-b^L + q\alpha^L (S - Q^L) \right].$$

The values of the high and low types are

$$
\begin{aligned}
w^H &= u(\theta^H - b^H) + \beta\left[\alpha^H u(Q^H) + (1-\alpha^H)\mathbb{E}^H\left(u(\theta_2)\right)\right] \\
w^L &= u(\theta^L - b^L) + \beta\left[\alpha^L u(Q^L) + (1-\alpha^L)\mathbb{E}^L\left(u(\theta_2)\right)\right],
\end{aligned}
$$

where for $j \in \{H, L\}$, $\mathbb{E}^j\left(u(\theta_2)\right) = p^j u(\theta^H) + (1-p^j)u(\theta^L)$. The participation constraint is

$$
p^1 w^H + (1-p^1)w^L = v
$$

and the IC constraint can be written as

$$
w^H = w^L + \underbrace{u(\theta^H - b^L) - u(\theta^L - b^L)}_{\text{static info rent}} + \underbrace{(1-\alpha^L)\beta\rho(u(\theta^H) - u(\theta^L))}_{\text{Dynamic info rent}}.
$$

Notice that the dynamic information rent is increasing in ρ and decreasing in α^L. I directly assume that the parameters are such that $\alpha^L \in (0,1)$ is optimal and show that the principal increases the termination probability when the persistence increases.

Proposition 18. *If $\alpha^L \in (0,1)$ is optimal, the optimal contract is such that $\frac{\partial \alpha^L}{\partial \rho} > 0$.*

Proof. The proof is as follows. Starting from the optimal contract, we consider a perturbation where we increase α^L while preserving the IC and PK constraints, and show that the resource gains are increasing in ρ . To this end, let $\Delta\alpha^L = \varepsilon > 0$, for ε small. We perturb the allocation along the low type's indifference curve, so to keep w^L constant, we increase Q^L by

$$
Q^L = \frac{\left[u(Q^L) - \mathbb{E}^L\left(u(\theta_2)\right)\right]}{\alpha^L u'(Q^L)}\varepsilon^L.
$$

The perturbation lowers the dynamic information rents, and so it relaxes the IC constraint. This allows us to lower the high type's period one utility by

$$
\Delta u^H = -\beta\rho(u(\theta^H) - u(\theta^L))\varepsilon.
$$

Because w^L is kept fixed, this changes the ex-ante utility by $p^1\Delta u^H$. Then, to satisfy the PK constraint, we increase the period one utility of both types in an incentive-compatible manner. Because information rents depend on consumption, increasing utilities uniformly would not be incentive compatible. If we increase the low type's utility by Δu^L, the IC constraint requires increasing the utility of the high type by

$$\Delta u^{H,IC} = \frac{u'(\theta^H - b^L)}{u'(\theta^L - b^L)}\Delta u^L.$$

The ex-ante utility is kept fixed if

$$p^1 \frac{u'(\theta^H - b^L)}{u'(\theta^L - b^L)}\Delta u^L + (1-p^1)\Delta u^L = -p^1 \Delta u^H,$$

which implies

$$\Delta u^L = -\frac{p^1 u'(\theta^L - b^L)}{p^1 u'(\theta^H - b^L) + (1-p^1)u'(\theta^L - b^L)}\Delta u^H.$$

Therefore, the total change in the high type utility is

$$\begin{aligned}\Delta u^{H,TOT} &= \Delta u^H + \Delta u^{H,IC} \\ &= (1-p^1)\frac{u'(\theta^L - b^L)}{p^1 u'(\theta^H - b^L) + (1-p^1)u'(\theta^L - b^L)}\Delta u^H.\end{aligned}$$

Finally, the resource gain from this perturbation is

$$\begin{aligned}\frac{\Delta K}{\varepsilon} &\approx p^1\left[\frac{1}{u'(\theta^H - b^H)}\Delta u^{H,TOT}\right] + (1-p^1)\left[\frac{1}{u'(\theta^L - b^L)}\Delta u^L\right] + \Omega \\ &\approx \left[\frac{1}{u'(\theta^L - b^L)} - \frac{1}{u'(\theta^H - b^H)}\right]\frac{u'(\theta^L - b^L)}{p^1 u'(\theta^H - b^L) + (1-p^1)u'(\theta^L - b^L)}(1-p^1)p^1\beta\rho(u(\theta^H) - u(\theta^L)) + \Omega\end{aligned}$$

where Ω collects all the terms that do not depend on ρ. Because the initial allocation is optimal, $\left[\frac{1}{u'(\theta^L - b^L)} - \frac{1}{u'(\theta^H - b^H)}\right] < 0$ and $u(\theta^H) - u(\theta^L) > 0$. Therefore, the principal's resource gain from this perturbation is increasing in ρ, i.e. $\frac{\partial \frac{\Delta K}{\varepsilon}}{\partial \rho} < 0$, which implies that $\frac{\partial \alpha^L}{\partial \rho} > 0$ is optimal. □

C.3.3 Screening model: divert funds before investing

In this section, I study a screening version of the model where the entrepreneur can choose what fraction of the funds available he invests in the project. The remaining funds are secretly diverted for consumption. Now, the lender can observe the entrepreneur's returns but not the entrepreneur's productivity nor invested and diverted funds. In this sense, the investment decision is similar to the labor/leisure choice in the Mirrlees taxation problem. This model yields the same characterization of the shadow costs μ_t, the GIEE, and the firm size dynamics. Moreover, we can directly define the investment wedge $\tau^k(\theta^t)$ as the wedge between invested and diverted funds relative to the first best.

Denote by B_t the funds advanced by the lender. The entrepreneur can use these funds to invest in the project k_t, but he can also divert a portion a_t of the funds for his consumption. Therefore, invested and diverted funds are subject to the constraint

$$k_t + a_t \leq B_t. \tag{C.13}$$

The lender now observes returns $f(k_t, \theta_t)$ but not productivity θ_t and how funds are used, i.e. k_t and a_t. Diverted funds are converted into consumption units according to the function $g(a_t)$, with $g'' \leq 0 < g'$, so the entrepreneur's consumption is

$$c_t = f(k_t, \theta_t) - b_t + g(a_t). \tag{C.14}$$

The principal's within period objective now is $B_t - b_t$. The envelope condition of the agent's problem is

$$\frac{\partial}{\partial \theta_t} w_t(\theta^t) = u'(c_t(\theta^t)) f_\theta(k_t(\theta^t), \theta_t) + \beta \Delta_t(\theta^t).$$

Now the investment wedge can be defined explicitly as the distortion in invested and diverted funds relative to the first best (where we would have $f_k(k_t(\theta^t), \theta_t) = g'(a_t(\theta^t))$. Define

$$\tau^k(\theta^t) \equiv 1 - \frac{g'(a(\theta^t))}{f_k(k(\theta^t), \theta_t)}.$$

The rest of the planning problem is the same but with the extra flow of funds constraint (C.13). The optimality condition for diverted funds is

$$\zeta_t(\theta^t) = g'(a_t(\theta^t)),$$

where $\zeta_t(\theta^t)$ is the multiplier on constraint (C.13). The FOC for investment is

$$\zeta_t(\theta^t) = f_k(k_t(\theta^t), \theta_t) - \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} u'(\theta^t) f_{\theta k}(k_t(\theta^t), \theta_t)$$

Then, combining the two optimality conditions, we get

$$\tau^k(\theta^t) = \frac{\mu_t(\theta^t)}{\varphi_t(\theta_t|\theta^{t-1})} \frac{f_{\theta k}(\theta^t)}{f_k(\theta^t)} u'(\theta^t) > 0,$$

which is the same as in Proposition 10. Because $\tau^k(\theta^t) > 0$, there is more cash diversion than in the first best. This is the standard screening result; the principal distorts effort (here investment k_t) downwards to screen types at a lower cost. When shadow costs $(\mu_t(\theta^t))$ are high, the principal increases distortions to reduce the costs of screening types. Moreover, this wedge also captures the distortions to firm size as in the cash flow diversion model. Combining the FOC for $B_{t+1}(\theta^t)$ and the envelope condition, we get

$$\begin{aligned}\frac{1}{q} &= \mathbb{E}\left[f_k(k_{t+1}(\theta^t),\theta_{t+1}) - \frac{\mu_{t+1}(\theta^{t+1})}{\varphi_{t+1}(\theta_{t+1}|\theta_{t+1})}u'(\theta^{t+1})f_{\theta k}(k_{t+1}(\theta^t),\theta_{t+1})|\theta_t\right] \\ &= \mathbb{E}\left[f_k(k_{t+1}(\theta^t),\theta_{t+1})\left(1-\tau^k(\theta^{t+1})\right)|\theta_t\right],\end{aligned}$$

which is exactly how the wedges were defined for the cash flow diversion model in equation (3.16). Then, it is also easy to verify that this model yields the same characterization for the shadow costs $\mu_t(\theta^t)$ and the GIEE.

C.4 Details numerical simulations

I follow a similar procedure as Farhi and Werning (87), Stantcheva (166) and Ndiaye (149). In these papers (and in Kapička (122) and Golosov *et al.* (101)), the model is solved with a geometric random walk process. This allows to normalize the principal's optimization problem and drop θ_{t-1} as a state variable. Here, the problem can also be normalized if the production function is assumed to be of the form $f(k,\theta) = z\theta^{1-\alpha}k^{\alpha}$. However, I am interested in performing comparative statics with respect to the persistence of the process (ρ). Therefore, I solve the full problem without renormalizing.

It is convenient to transform the problem to write the Hamiltonian as a function of the current shock ε_t instead of the current productivity θ_t. Denote the density function of the shock by $g_\varepsilon(\varepsilon_t)$, then it follows that

$$\varphi\left(\theta_t \mid \theta_{t-1}\right) = \frac{g_\varepsilon(\varepsilon_t)}{\theta_{t-1}^{\rho}}.$$

Moreover, we also have that

$$\frac{\partial\varphi\left(\theta_t \mid \theta_{t-1}\right)}{\partial\theta_{t-1}} = \frac{\rho}{\varepsilon_t\theta_{t-1}^{1+\rho}}\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}\frac{(\log\theta_t - \rho\log\theta_{t-1}-\mu)}{\sigma_\varepsilon^2}\exp\left\{-\frac{(\log\theta_t - \rho\log\theta_{t-1}-\mu)^2}{2\sigma_\varepsilon^2}\right\}$$

and

$$\frac{\partial g_\varepsilon(\varepsilon_t)}{\partial \varepsilon_t} = -\frac{1}{\varepsilon_t^2 \sigma_\varepsilon \sqrt{2\pi}} \left[\frac{(\log \varepsilon_t - \mu)}{\sigma_\varepsilon^2} + 1 \right] \exp\left\{ -\frac{(\log \varepsilon_t - \mu)^2}{2\sigma_\varepsilon^2} \right\}.$$

Therefore,

$$\tilde{g}_\varepsilon(\varepsilon_t) \equiv g_\varepsilon(\varepsilon_t) + \varepsilon \frac{\partial g_\varepsilon(\varepsilon_t)}{\partial \varepsilon_t} = -\frac{\theta_{t-1}^{1+\rho}}{\rho} \frac{\partial \varphi(\theta_t \mid \theta_{t-1})}{\partial \theta_{t-1}}.$$

Then note that $d\theta_t = \theta_{t-1}^{\rho} d\varepsilon_t$ implies

$$\varphi(\theta_t \mid \theta_{t-1})\, d\theta_t = g_\varepsilon(\varepsilon_t) d\varepsilon_t,$$

and

$$\frac{\partial \varphi(\theta_t \mid \theta_{t-1})}{\partial \theta_{t-1}} d\theta_t = -\rho \frac{\tilde{g}_\varepsilon(\varepsilon_t)}{\theta_{t-1}} d\varepsilon_t.$$

The planning problem over the shock ε_t is

$$\begin{aligned}
K(v_{t-1}, \Delta_{t-1}, k_t, \theta_{t-1}) = \min \int & \left(k_{t+1}(\varepsilon_t) - b_t(\varepsilon_t) + qK(v_t(\varepsilon_t), \Delta_t(\varepsilon_t), k_{t+1}(\varepsilon_t), \theta_{t-1}^{\rho}\varepsilon_t) \right) g_\varepsilon(\varepsilon_t) d\varepsilon_t \\
s.t \quad (PK) \quad & w_t(\varepsilon_t) = u(c_t(\varepsilon_t)) + \beta v_t(\varepsilon_t) \qquad [g_\varepsilon(\varepsilon_t)\xi_t(\varepsilon_t)] \\
& v_{t-1} = \int w_t(\varepsilon_t) g_\varepsilon(\varepsilon_t) d\varepsilon_t \qquad [g_\varepsilon(\varepsilon_t)\lambda_{t-1}] \\
(IC) \quad & \dot{w}_t(\varepsilon_t) = \theta_{t-1}^{\rho} \left(u'(c(\varepsilon_t)) \iota f_\theta(k_t, \theta_{t-1}^{\rho}\varepsilon_t) + \beta \Delta_t(\varepsilon_t) \right) \qquad [\mu_t(\varepsilon_t)] \\
& \Delta_{t-1} = \int w_t(\varepsilon_t)(-\frac{\rho}{\theta_{t-1}} \tilde{g}_\varepsilon(\varepsilon_t)) d\varepsilon_t \qquad [g_\varepsilon(\varepsilon_t)\gamma_{t-1}] \\
& (Feasibility) \quad c_t(\varepsilon_t) = f(k_t, \theta_{t-1}^{\rho}\varepsilon_t) - b_t(\varepsilon_t)
\end{aligned}$$

The optimality conditions are

$$\frac{q}{\beta}\lambda_t(\varepsilon_t) = \frac{1}{u'(c_t(\varepsilon_t))} \left[1 + \frac{\mu(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)} \theta_{t-1}^{\rho} \iota f_\theta(k_t, \theta_{t-1}^{\rho}\varepsilon_t) u''(c(\varepsilon_t)) \right] \tag{C.15}$$

$$\gamma_t(\varepsilon_t) = -\frac{\beta}{q} \theta_{t-1}^{\rho} \frac{\mu(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)} \tag{C.16}$$

and the two LOM

$$\dot{\mu}(\varepsilon_t) = -\left[\frac{q}{\beta}\lambda_t(\varepsilon_t) - \lambda_{t-1} + \gamma_{t-1} \frac{\rho}{\theta_{t-1}} \frac{\tilde{g}_\varepsilon(\varepsilon_t)}{g_\varepsilon(\varepsilon_t)} \right] g_\varepsilon(\varepsilon_t) \tag{C.17}$$

$$\dot{w}_t(\varepsilon_t) = \theta_{t-1}^{\rho}\left(u'(c(\varepsilon_t))\iota f_\theta(k_t, \theta_{t-1}^{\rho}\varepsilon_t) + \beta\Delta_t(\varepsilon_t)\right). \tag{C.18}$$

I truncate the distribution of ε_t at the 0.01 and 0.99 percentiles, the boundary conditions then need to be adjusted to $\mu(\bar{\varepsilon}) = -\gamma_{t-1}\frac{\rho}{\theta_{t-1}}\bar{\varepsilon}g_\varepsilon(\bar{\varepsilon})$ and $\mu(\underline{\varepsilon}) = -\gamma_{t-1}\frac{\rho}{\theta_{t-1}}\underline{\varepsilon}g_\varepsilon(\underline{\varepsilon})$.

To solve the model, the state space is modified to $(\lambda_-, \gamma_-, k, \theta_-)$, so the multipliers λ_- and γ_- are used instead of v_- and Δ_-, respectively. I use 20 grid points for λ_-, 14 for γ_-, 25 for k and 15 for θ_-. I interpolate on K, v and Δ with cubic splines and allow to extrapolate. To solve the model with an i.i.d type process, the algorithm is the same but with $\Delta = 0$ and without the state variables γ_- and θ_-.

Algorithm

Step 0: Guess the value function K', promised utility v' and promised marginal utility Δ' on the grid $(\lambda_-, \gamma_-, k, \theta_-)$

Step 1: Compute the policy functions for k_+ on a grid $(\lambda_{pol}, \gamma_{pol}, \theta)$ by minimizing

$$k_+ + qK'(\lambda_{pol}(i), \gamma_{pol}(i), k_+, \theta(i))$$

(Note: k_+ needs to be computed multiple times at every step while solving the ODE. But to improve speed, we can solve before the policies on a dense grid and then interpolate when solving the ode).

Step 2: For each point in $(\lambda_-, \gamma_-, k, \theta_-)$, solve the optimal control problem with a shooting method.

- a) Guess continuation utility of lowest type $w(\underline{\varepsilon}) = \underline{w}$.
- b) For each ε, solve $\lambda(\varepsilon)$ in equation (C.15) and $\gamma(\varepsilon)$ in equation (C.16). To compute $c(\varepsilon)$, first compute $k_+(\varepsilon)$ by interpolationg the array of policies on $(\lambda(\varepsilon), \gamma(\varepsilon), \theta_-^{\rho}\varepsilon)$. Then obtain $v(\varepsilon)$ by interpolation of v' on $(\lambda(\varepsilon), \gamma(\varepsilon), k_+(\varepsilon), \theta_-^{\rho}\varepsilon)$ and solve

  $$c(\varepsilon) = u^{-1}\left(w(\varepsilon) - \beta v(\varepsilon)\right).$$

  With these solutions solve the differential equations (C.17) and (C.18). Note when solving (C.17) also need to interpolate Δ' on $(\lambda(\varepsilon), \gamma(\varepsilon), k_+(\varepsilon), \theta_-^{\rho}\varepsilon)$.
- c) Check the boundary condition $\mu(\bar{\varepsilon}) = -\gamma_-\frac{\rho}{\theta_-}\bar{\varepsilon}g_\varepsilon(\bar{\varepsilon})$. If it does not satisfy the tolerance, go back to step a).

Step 3: Given the solution $(\mu(\varepsilon), w(\varepsilon))$, repeat step b) to obtain all policy functions on a grid $(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$, also compute $b(\varepsilon) = f(k, \theta_-^\rho \varepsilon) - c(\varepsilon)$.

Step 4: Compute the lender's value function, promised utility and expected marginal utility at every grid point

$$v(\lambda_-, \gamma_-, k, \theta_-) = \int w(\lambda_-, \gamma_-, k, \theta_-, \varepsilon) g_\varepsilon(\varepsilon_t) d\varepsilon_t$$

$$\Delta(\lambda_-, \gamma_-, k, \theta_-) = \int w(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)(-\frac{\rho}{\theta_-} \tilde{g}_\varepsilon(\varepsilon_t)) d\varepsilon_t$$

$$K(\lambda_-, \gamma_-, k, \theta_-) = \int \left(k_+(\lambda_-, \gamma_-, k, \theta_-, \varepsilon) - b(\lambda_-, \gamma_-, k, \theta_-, \varepsilon) + qK'(\lambda(\varepsilon), \gamma(\varepsilon), k_+(\varepsilon), \theta_-^\rho \varepsilon)\right) g_\varepsilon(\varepsilon_t) d\varepsilon_t.$$

Calculate the distance with previous guess of K', v' and Δ' , and repeat from **Step 1** until the convergence criteria is satisfied.

When solving the model without time-0 screening, for the Montecarlo simulation, the initial λ_0 and θ_0 can be fixed at arbitrary values. Because Δ_0 is a free variable, we must set $\gamma_0 = 0$. Then k_1 is chosen optimally given (λ_0, θ) and $\gamma_0 = 0$.

C.4.1 Time-0 problem

I also solve the time-0 problem with a shooting method by iterating on $v_0(\underline{\theta})$. To solve the problem, we use the value functions $v(\lambda, \gamma, \theta, k)$ and $\Delta(\lambda, \gamma, \theta, k)$ and the policy function $k(\lambda, \gamma, \theta)$ from the solution of the recursive problem.

Algorithm.

1. Guess the lowest type continuation utility $v_0(\underline{\theta})$
2. Solve the ODEs $\dot{v}(\theta_0) = \Delta_0(\theta_0)$ and $\dot{\mu}(\theta_0) = -[\lambda_1(\theta_0) - \lambda_0] g(\theta_0)$
   (a) At every step for every $(v(\theta_0), \mu(\theta_0))$
      i. Get $\gamma_1(\theta_0)$ from: $-\gamma_1(\theta_0) = \frac{\mu(\theta_0)}{g(\theta_0)}$
      ii. With $\gamma_1(\theta_0)$, θ_0 and $v_0(\theta_0)$ can back out the corresponding $\lambda_1(\theta_0)$. Note, v will be computed on a grid $(\lambda_1(\theta_0), \gamma_1(\theta_0), \theta_0, k_1(\theta_0))$, so for every $\lambda_1(\theta_0)$ need to get the optimal $k_1(\theta_0)$ (from the policy functions in the recursive problem) and then use this to get v until solve the root.

iii. With $(\lambda_1(\theta_0), \gamma_1(\theta_0), \theta_0)$ can in turn also back out $\Delta_0(\theta_0)$

iv. Update the LOM

3. Check $\mu(\overline{\theta}) = 0$, if it doesn't satisfy tolerance go back to 1. with a new guess

4. Repeat (a) to recover all the policy functions on a grid for θ_0

C.4.2 Check global IC constraints

The first-order approach consists of solving a relaxed problem where only the local incentive constraints are considered. A priori global incentive constraints may bind, in which case the solutions of the relaxed program (3.15) and the full program (3.9) would not coincide. I follow the approach outlined in Kapička (122) and Farhi and Werning (87) to verify ex-post that only the local incentive constraints bind.

The procedure is the following. First, after solving numerically the relaxed problem, we have obtained the policy functions $\lambda(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$, $\gamma(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$, $k_+(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$ and $b(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$ and the value function $v(\lambda_-, \gamma_-, k, \theta_-)$. Let $\widetilde{\varepsilon}$ denote the agent's report about the innovation to the productivity. Then we consider a problem where the entrepreneur takes as given the policy functions and can report any $\widetilde{\varepsilon} \in [\underline{\varepsilon}, \varepsilon]$ and verify that for every $(\lambda_-, \gamma_-, k, \theta_-, \varepsilon)$

$$
\begin{aligned}
\varepsilon = \arg\max_{\widetilde{\varepsilon} \in [\underline{\varepsilon}, \varepsilon]} & u(\widetilde{c}(\lambda_-, \gamma_-, k, \theta_-, \varepsilon, \widetilde{\varepsilon})) \\
& + \beta v(\lambda(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}), \gamma(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}), k_+(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}), \theta_-^{\rho} \widetilde{\varepsilon}) \\
s.t \quad & \widetilde{c}(\lambda_-, \gamma_-, k, \theta_-, \varepsilon, \widetilde{\varepsilon}) = \iota f(k, \varepsilon) + (1-\iota) f(k, \widetilde{\varepsilon}) - b(\lambda_-, \gamma_-, k, \theta_-, \widetilde{\varepsilon}).
\end{aligned}
$$

C.4.3 Solution implementation

With persistent shocks and constant equity, the entrepreneur's problem in the quasi-implementation is

$$
\begin{aligned}
\mathcal{W}(W_t, \theta_{t-1}, \widetilde{\theta}_{t-1}, \varepsilon_t) = & \max_{\widetilde{\theta}_t \leq \theta_{t-1}^{\rho} \varepsilon_t} u(\widetilde{c}_t) + \beta \mathcal{V}(W_{t+1}, \theta_{t-1}^{\rho} \varepsilon_t, \widetilde{\theta}_t) \\
s.t \qquad & W_{t+1} = qC(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) \\
& c_t = (1-q) C(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) \\
& \widetilde{c}_t = c_t + \iota(f(k_{SB}, \theta_{t-1}^{\rho} \varepsilon_t) - f(k_{SB}, \widetilde{\theta}_t))
\end{aligned}
$$

where

$$C(W_t, \widetilde{\theta}_t, \widetilde{\theta}_{t-1}) = \frac{1}{q} W_t + \chi(f(k_{SB}, \widetilde{\theta}_t) + q\overline{f}(k_{SB}, \widetilde{\theta}_t) - \overline{f}(k_{SB}, \widetilde{\theta}_{t-1}))$$

$$\overline{f}(k_{SB}, \theta_t) = \mathbb{E}\left[\sum_{\tau=1}^{\infty} q^{\tau-1} f(k_{SB}, \theta_{t+\tau}) | \theta_t\right]$$

and

$$\mathcal{V}(W_{t+1}, \theta_t, \widetilde{\theta}_t) = \int \mathcal{W}(W_{t+1}, \theta_t, \widetilde{\theta}_t, \varepsilon_{t+1}) g_\varepsilon(\varepsilon_{t+1}) d\varepsilon_{t+1}.$$

With i.i.d shocks, the problem is the same but without θ_{t-1} and $\widetilde{\theta}_{t-1}$ as state variables and with $\overline{f}(k_{SB})$ independent of θ_t. The problem is solved with standard value function iteration, and to have the closest comparison with the solutions of the optimal allocation, $\mathcal{V}(W_{t+1}, \theta_t, \widetilde{\theta}_t)$ is computed with numerical integration.