THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **Sciences Economiques - Toulouse School of Economics**

Présentée et soutenue par

FUCHS Simon

le 12 juin 2018

Essays on Globalization:
Historical and Contemporary Evidence

Discipline : **Sciences Economiques**

Unité de recherche : **TSE-R**

Directeur de thèse : **Monsieur Thomas CHANEY, Professeur, Sciences Po**

JURY

Rapporteurs Monsieur Treb ALLEN, Professeur, Dartmouth College
Monsieur Thierry MAYER, Professeur, Sciences Po

Suffragants Monsieur Christian HELLWIG, Professeur, TSE

Essays on Globalization: Historical and Contemporary Evidence

Toulouse School of Economics

Simon Fuchs

23 May 2018

Acknowledgements

I am indebted to various people who supported me throughout my PhD. Most importantly, my advisors at TSE: Thomas Chaney and Christian Hellwig. Both Thomas' creative and rigorous approach to research and Christian's instincts and high standards for what constitutes valuable research have been a constant source of inspiration. For their guidance and their continuous support I am deeply thankful.

I also would like to extend my special gratitude to Treb Allen who invited me to visit Dartmouth College and whose comments have been incredibly important towards finishing the first chapter of this thesis. His guidance is a continuing source of inspiration.

I also want to thank the TSE community as a whole and in particular the Macroeconomics group. Rather than the classroom, I felt that it was during the interactions, workshops, reading groups and seminars of this group that my education took place. I want to thank current and past members of that group, in particular, Martial Dupaigne, Patrick Féve, Tim Lee, Marti Mestieri, Franck Portier, Vincent Rebeyrol, Robert Ulbricht and Nicolas Werquin.

During my PhD, I was also fortunate to visit the economics department at Sciences Po and had many fruitful discussions there. I particularly want to extend my gratitude towards Johannes Boehm, Philippe Martin, Thierry Mayer and Florian Oswald.

As mentioned before, I was fortunate to visit Dartmouth college and the time there was a special moment during my PhD and - I feel - has had a lasting impact on my ongoing education and development. For that and many - hopefully - fruitful discussion as well as the opportunity to present at a crucial moment I would like to thank the community at Dartmouth and in particuarl: Andy Bernard, Emily Blanchard, Jim Feyrer, Sharat Ganapati, Rob Johnson, Paul Novosad, Nina Pavcnik, Chris Snyder and Bob Staiger.

During my time at TSE I enjoyed many stimulating discussions with friends and colleagues. Too many names would need to be mentioned, but I would like to express my thanks to all of them.

I also want to thank my family and in particular my parents: For their support, their encouragement and for an environment that instilled in me the love for learning and study.

A special thanks goes to Begoña Alarcon. Words cannot do justice to the support that she has given to me throughout the years. This thesis is as much her accomplishment as it is mine.

Declarations

I acknowledge financial support via ERC grant No 337272-FiNet for both the first and second chapter of this thesis. I also acknowledges financial support via the Agence National de la Recherche (ANR-11-IDEX-0002-02) via which my final PhD year was financed. Finally, the work for the second chapter benefited from access to French confidential data and is thus supported by a public grant overseen by the French National Research Agency (ANR) as part of the « Investissements d'avenir » program (reference : ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD). All remaining errors are my own.

Abstract - English

This thesis contains three essays on globalization with a special focus on historical and contemporary empirical evidence and labor market effects.

In the first chapter, I study to what extent labor market frictions limit the gains from market integration. I use an external demand shock to the Spanish economy as a natural experiment to identify and quantify the effect of labor mobility costs on Spain's development. Using newly digitized trade and labor market data, I show that during WWI (1914-1918) a large, temporary and sectorally heterogeneous demand shock emanated from belligerent countries, as a result of which Spain expanded its manufacturing employment and exports, while income growth between the north and south in Spain diverged. To quantify and analyse the role of mobility costs I build and estimate a multi-sector economic geography model that allows for sectoral and spatial mobility costs. Spatial mobility costs dominated with an estimated 80% of reallocation of labor taking place *within* rather than *between* provinces. I use the estimated model to calculate counterfactuals to examine the effects of and interaction between output and input market integration: Comparing to the non-shock counterfactual I find that the WWI-shock increased manufacturing employment by 10%, and induced highly uneven spatial development with the north growing 27% faster. The shock constituted a 6% increase in market size and increased aggregate real incomes by 20%. Lowering mobility costs by 10% increases real income gains from the WWI-shock by an additional 3%, and exceeds gains in the non-shock scenario, suggesting that labor market integration and output market integration are complements.

In the second chapter, François de Soyres and I introduce a new framework to evaluate the effects of regional diversification. We observe that in the presence of mobility frictions workers are exposed to local shocks and that in a multi-sectoral framework this induces a trade-off: Regions can specialize in their comparative advantage industries, but at the same time such specialization increases labor market risk due to sector specific shocks. If mobility costs are high, then welfare effects from lack of diversification can be

substantial. We measure the segmentation of the French labor market and introduce a new spatial equilibrium model that incorporates labor market frictions, unemployment, and mobility cost into an otherwise standard multi-sector economic geography model. We employ the model to simulate unemployment responses to sector specific shock and demonstrate the interaction between mobility frictions and matching frictions.

In the third chapter, Konrad Adler and I ask the question to what extent is the set of products that are available to a country driven by the composition of international markets? We develop a quantitative framework to determine and map the similarity between countries from observed market shares of identical products across markets. We apply our framework to the global movies market where we can abstract from price competition and observe identical products and their market shares across countries. As an application we evaluate the hypothesis that the observed large increase in the revenue share of sequels has been due to shifts in the composition of global demand away from traditional western markets inducing demand risk and with sequels providing insurance. While we find substantial shifts in the profit space and lower risk associated with sequels, our simulations suggest that the risk due to taste heterogeneity in the movies market is quantitatively insufficient to explain the increase in the revenue of sequels, suggesting that other forces such as scale economies might be at play.

Abstract - French

Cette thèse contient trois essais sur la mondialisation avec un accent particulier sur les preuves empiriques historiques et contemporaines ainsi que seseffets sur le marché du travail.

Dans le premier chapitre, j'étudie dans quelle mesure les frictions du marché du travail limitent les gains de l'intégration du marché. J'utilise un choc de demande externe pour l'économie espagnole comme expérience naturelle pour identifier et quantifier l'effet des coûts de mobilité des travailleurssur le développement de l'Espagne. À l'aide des données du commerce et du marché du travail nouvellement numérisées, je montre que pendant la Première Guerre mondiale (1914-1918), un choc de demande important, hétérogène en temps et en secteur a émanédes pays belligérants ce qui a permis à l'Espagne d'accroître son emploi manufacturier et ses exportations. Dans le même temps, la croissance des revenus a divergé entre le nord et le sud de l'Espagne. Pour quantifier et analyser le rôle des coûts de mobilité, je construis et j'estime un modèle de géographie économique multisectoriel qui permet d'intégrer des coûts de mobilité sectoriels et spatiaux. Les coûts de la mobilité spatiale ont dominé avec environ 80 pc de la réallocation du travail ayant lieu plutôt au sein des provinces qu'entre elles. J'utilise le modèle estimé pour calculer les contrefactuels pour examiner les effets de l'interaction entre l'intégration du marché des intrants et des produits : comparé au contrefactuel non-choc, le choc de la Première Guerre mondiale a augmenté l'emploi manufacturier de 10 pc et induit un développement spatial très inégal avec le nord ayant une croissance de 27 % plus rapide que le sud. Le choc a constitué une augmentation de 6% de la taille du marché et une augmentation des revenus réels agrégés de 20%. L'abaissement de 10% des coûts de mobilité augmente les gains de revenu réel générés par le choc de la Première Guerre mondiale de 3 pc supplémentaires et dépasse les gains du scénario sans choc ce qui suggère que l'intégration du marché du travail et l'intégration de marché des biens sont complémentaires.

Dans le deuxième chapitre, François de Soyres et moi introduisons une nouvelle méthodolo-

gie pour évaluer les effets de diversification régionale. Nous observons qu'en présence de frictions de mobilité, les travailleurs sont exposés aux chocs locaux et que dans un cadre multisectoriel cela induit un compromis : les régions peuvent se spécialiser dans leurs industries d'avantage comparatif, mais en même temps cette spécialisation augmente le risque du marché du travail en raison de chocs propres à leur secteur. Si les coûts de la mobilité sont élevés, les effets du manque de diversification sur le bien-être peuvent être considérables. Nous mesurons la segmentation du marché du travail français et introduisons un nouveau modèle d'équilibre spatial qui incorpore les frictions du marché du travail, le chômage et le coût de la mobilité dans un modèle de géographie économique multisectoriel standard. Nous utilisons ce modèle pour simuler les réponses du chômage aux chocs sectoriels et démontrons l'interaction entre les frictions de mobilité et les frictions du chômage

Dans le troisième chapitre, Konrad Adler et moi posons la question dans quelle mesure l'ensemble des produits disponibles pour un pays dépend de la composition des marchés internationaux. Nous développons un cadre quantitatif pour déterminer et cartographier la similarité entre les pays à partir des parts de marché observées des produits identiques sur les marchés. Nous appliquons notre cadre au marché mondial des films où nous pouvons faire abstraction de la concurrence des prix et observer des produits identiques et leurs parts de marché dans les différents pays. En tant qu'application, nous évaluons l'hypothèse selon laquelle la forte augmentation observée de la part des revenus des séquelles est due à des changements dans la composition de la demande mondiale, plus précisément un éloignement des marchés occidentaux traditionnels, induisant un risque de demande et des suites fournissant des assurances. Alors que nous trouvons des changements substantiels dans l'espace de profit et un risque inférieur associé aux suites, nos simulations suggèrent que le risque dû à l'hétérogénéité du goût dans le marché du film est quantitativement insuffisant pour expliquer l'augmentation des revenus des suites, suggérant que d'autres forces comme les économies d'échelle pourraient être en jeu.

Contents

List of Figures

List of Tables

Chapter 1

The Spoils of War: Trade Shocks during WWI and Spain's Regional Development

Simon Fuchs

1.1 Abstract

This paper analyzes to what extent labor market frictions limit the gains from market integration. I use an external demand shock to the Spanish economy as a natural experiment to identify and quantify the effect of labor mobility costs on Spain's development. Using newly digitized trade and labor market data, I show that during WWI (1914-1918) a large, temporary and sectorally heterogeneous demand shock emanated from belligerent countries, as a result of which Spain expanded its manufacturing employment and exports, while income growth between the north and south in Spain diverged. To analyse the role of mobility costs I estimate a multi-sector economic geography model that allows for sectoral and spatial mobility costs. Spatial mobility costs dominated with an estimated 80% of reallocation of labor taking place *within* rather than *between* provinces. I use the estimated model to calculate counterfactuals to examine the effects of and interaction between output and input market integration: Comparing to the non-shock counterfactual I find that the WWI-shock increased manufacturing employment by 10%, and induced highly uneven spatial development with the north growing 27% faster. The shock constituted a 6% increase in market size and increased aggregate real incomes by 20%. Lowering mobility costs by 10% increases real income gains from the WWI-shock

by an additional 3%, and exceeds gains in the non-shock scenario, suggesting that labor market integration and output market integration are complements.

"Spain is today a bundle of small bodies tied together by a rope of sand."

Ford (1845), p.2

1.2 Introduction

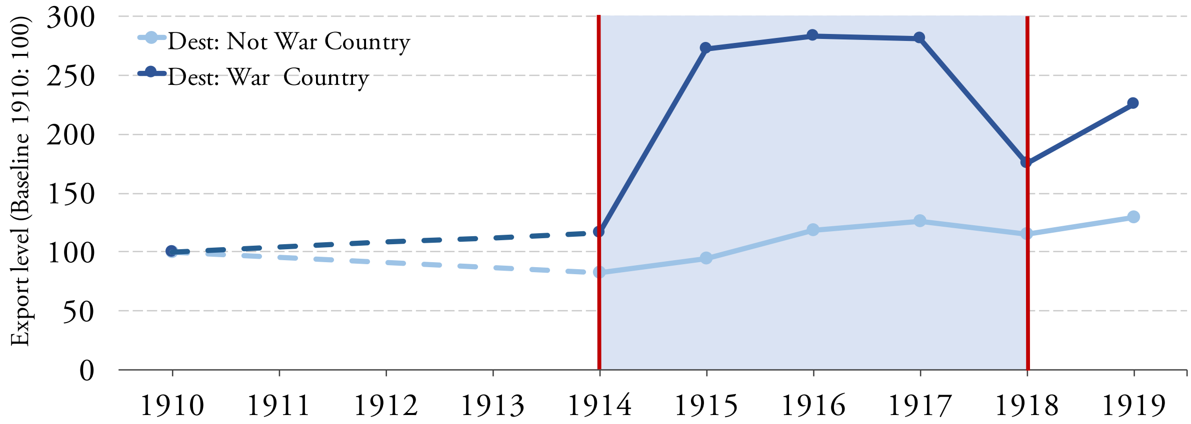
Why might an economy be trapped at a low level of economic development? Why is the adjstument to trade liberalization slow and often does not seem to effectively equilibrate local labor markets across space?[1] A common explanation to these questions is that high mobility costs and low initial gains to the worker from migration might prevent labor reallocation towards higher productivity sectors and regions. This can in turn limit the gains from market integration and undermine development, but to what extent this might be the case is difficult to determine. Understanding and quantifying these frictions is therefore of primary importance in understanding the obstacles to growth and structural change in developing countries as well as the welfare effects of trade liberalization episodes.

However, empirically verifying and quantifying these frictions is challenging, since neither labor market frictions nor the counterfactual gains from reallocation are directly observed. This paper overcomes this problem by using a natural experiment where a foreign demand shock reallocates labor across sectors and space. The reallocation patterns are informative about the sectoral and spatial mobility frictions that inhibit labor movements even in the absence of a shock. Using the shock in tandem with an economic geography model, I show how to estimate the gains from reallocation as well as the labor market frictions. The key point of this paper is to illustrate how and to what extent labor market frictions can limit gains from market integration and how this can be analyzed in a setting where a temporary foreign demand shock reallocates labor by creating temporary gains that offset adjustment costs. This analysis singles out mobility costs as a key factor in determining the size of welfare gains from market integration as well as the spatial distribution.

This study examines a trade shock to the Spanish economy that was caused by the participation of Spain's key trading partners - in particular France - in the first World War (1914-1918) while Spain remained neutral. Prior to the shock, Spain had experienced a

[1] The empirical finding that local labor markets only adjust slowly to shocks goes back to Blanchard and Katz (1992). Dix-Carneiro and Kovak (2017) show that labor market frictions and slow adjustment processes can permanently prevent spatial arbitrage.

Figure 1.1: Aggregate Trade levels



Notes: This figure compares aggregate export levels in constant pre-war prices between destination countries that participated in WWI and those that did not. To adjust for additional spatial disruptions of the frontline the belligerent countries are made up of France, Italy and the United Kingdom. The non-belligerent countries exclude the United States and other later participants of WWI. Data is not available for the years between 1910 and 1914 therefore a trend line is imputed. The blue shaded area indicates the period of WWI. The source data are the digitized product-destination level trade statistics.

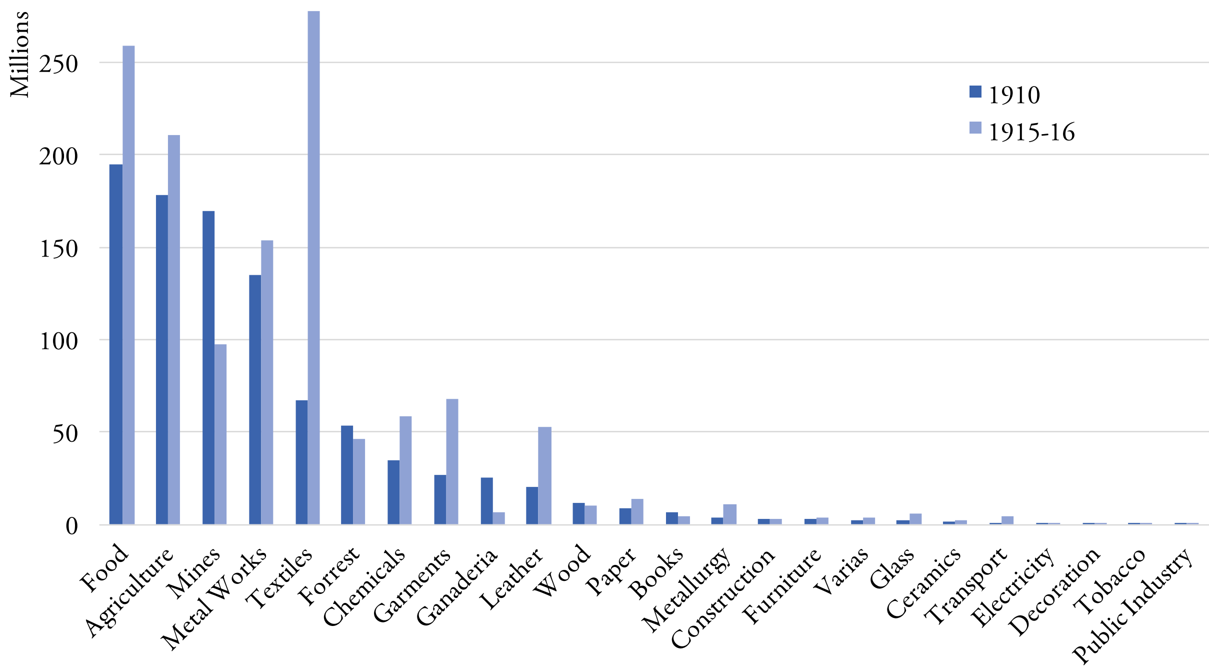
prolonged period of low GDP growth with little structural change (Prados de la Escosura; 2017). Using newly collected trade data on Spanish product level exports between 1910-1919, as well as labor market data on wages and employment across 48 different provinces and 24 different sectors before and after the war, I document five stylized facts about the shock and its impact. Firstly, the trade shock was large, increasing aggregate exports by 40% at constant prices, and additionally the shock was spatially biased with most of the aggregate increase being due to higher volumes of trade with belligerent countries with France being by far the most important destination. Secondly, the trade shock was asymmetric across sectors. Comparing the trade increase between belligerent and non-belligerent countries before and during the war, I find that exports to belligerent countries increased in particular for garments, textiles, paper and products from the heavy industry. Thirdly, sector-level income growth was spatially tilted towards the French border, with each additional 100km distance to the French border decreasing - on average - the growth rate by 4 percent. Fourthly, provinces with a higher specialization in industries favored by the shock experienced faster population growth compared to their pre-trend, with the opposite being true for provinces with less favorable industrial composition. Finally, regional industry dynamics depended on the tightness of the local labor market, indicating an important role for spatial frictions in segregating labor markets and thus preventing arbitrage between geographically segregated labor markets.

The general point that provinces with a prior specialization in sectors that benefited from the war shock had an opportunity to expand their production can be illustrated with an example: Already before the War the *Sociedad Minera y Metalúrgica de Peñarroya* operated a factory for fertilizer and other chemical goods in Cordoba. During the War the factory faced higher wages and input prices, but they also experienced a substantial increase in both domestic and foreign demand, allowing them to expand their output of superphosphate - a fertilization agent - by 30 percent while expanding their workforce by 20 percent between 1914 and 1917 (Instituto de Reformas Sociales; 1916). Companies like the *Sociedad Minera* make up the individual industries considered in this paper. With their industrial capacity in place they were well positioned to benefit from the shock, but had to attract labor from other provinces and sectors. In doing so, industries found themselves competing with each other to attract workers from the agricultural hinterland. The focus of this paper is to learn more about the labor market conditions and frictions that shaped the response to the shock.

I develop a quantitative economic geography model to understand the *aggregate* impact of that shock, accounting for the *disaggregated* geographical margins of adjustments. The models is consistent with the stylized facts and focuses on taking explicitly into account the spatial linkages in the labor market and the patterns of comparative advantage across provinces, as well as the sectoral switching costs within provinces. I build on the existing quantitative economic geography literature model - as recently surveyed by Redding and Rossi-Hansberg (2016) - and extend a baseline model into several directions. Labor demand is determined by a framework where multiple sectors conduct intra-national and international trade subject to geographical frictions. Differently to most of the commonly used models in the literature, I do not take a stance on the strength of industry level scale economies. Rather, the patterns of comparative advantage across space and sectors are partially endogenous, with higher labor densities translating into productivity gains, depending on the strength of a set of sector specific parameters that determine industry level scale economies. The adopted models - first introduced into the international trade literature by Kucheryavyy et al. (2016) - can be represented by a tractable log linear gravity formulation and is consistent with a Ricardian multi-sectoral trade model with external scale economies, but also nests multiple other canonical models currently used in the literature, depending on the interpretation and values of the parameters.

Labor supply is determined by a nested discrete choice framework where workers first make a decision about reallocating across space subject to incurring switching costs, and then upon arrival in the new province sort into sectors. A two-staged sequence of

Figure 1.2: Sectoral Export Composition (1910, 1915/1916)



Notes: This figure reports the aggregate export composition in sectoral terms. The product level trade has been aggregated to sector level trade data to match the level of aggregation of the labor market panel. The total value of exports for each section in 1910 as well as the mean exports for 1915/1916 is reported. The source data are the digitized product-destination level trade statistics.

preference shocks from a Fréchet distribution make the framework tractable. Two kinds of switching costs are introduced: Firstly, workers who leave a sector incur a switching cost that is specific to the sector and proportional to the expected utility of its destination, secondly, a worker who reallocates to a different province incurs a switching cost that scales with distance. This framework extends the commonly used economic geography models by allowing for stickiness in employment at the sectoral and provincial level - a key feature of the data. At the same time the number of parameters that is being introduced is limited.

I then show how the structure of the model and the exogenous variation due to the natural experiment can be combined to obtain credible estimates for structural parameters that pin down the gains from reallocation. The general intuition for my estimation strategy is that benchmark economic geography models can be inverted to obtain a unique set of province-sector specific market share shifters. This is related to inverting market shares to obtain mean utilities when estimating demand in differentiated product demand markets as recently applied in the trade literature in Adao et al. (2017). These market share shifters are structurally related to prices adjusted for the curvature of the demand function. In a large class of commonly used economic geography models the responsiveness of this price measure to wages is directly informative about how trade patterns respond to wage changes, and the responsiveness with regard to industry scale is informative about scale effects.

More specifically, my approach can be described as follows: Conditional on specifying the strength of geographical frictions in input markets, the structure of the model together with income data can be used to solve for the origin-specific prices. The strategy behind this is that economic geography models allow to decompose total sectoral income into two parts, a first determinant of income that is due to proximity to lucrative destination markets, and a second part that describes how given market acess lower marginal costs translates into a higher captured trade share across all locations, with this part being theoretically interpreted as an origin-specific price and is empirically related to the origin fixed effect in a gravity equation. These origin specific prices can be regressed on (log) wages and (log) employment sizes of sectors to obtain elasticities that describe how changes in wages and sectoral employment translate into higher trade shares and thus higher incomes. The elasticity with regard to wage changes is commonly referred to as *trade elasticity*, while the other elasticity determines scale effects. I will refer to it as *scale elasticity* in the remainder of the text.[2] An obvious problem in this estimation is the endogeneity of wages and labor densities. I utilize instruments that effectively exploit differential shock exposure across provinces interacted with differences in labor market tightness to estimate the parameters. A challenge is that wage and labor changes are correlated, thus *differential* variation is needed to distinguish the independent effect of each variable. Labor market tightness induces variation in the extent to which the shock is being absorbed into wages or employment levels, making it possible to identify the trade and scale elasticity.

The estimated parameters point to the presence of *decreasing* returns to scale in the medium run, effectively limiting the immediate gains from reallocating labor in the absence of the shock. Similar estimates for scale economies over a 10-year horizon indicate that decreasing returns vanish over the long(er) run. The estimation also gives insights into the performances of a broad class of economic geography models in capturing adjustment patterns. Specifically, the fit of the regression expresses how much of the observed variation in residual income shifters can be explained by the endogenous mechanism provided by the model. The model can explain half of that residual variation.

[2]Note that this elasticity does not correspond to an output scale elasticity, but rather combines how scale translates into productivity gains which translate into lower prices and thus into higher market shares across all trading partners where the responsiveness of the trade growth depends on the trade elasticity.

For the estimation of labor market frictions the structure of the model is combined with additional data. Usually such an estimation requires flows of workers across space and sectors. However, in a historical context this type of data is rarely available. I show how to estimate labor market frictions in the absence of such data. The structure of the model allows for a conveniently separable estimation of geographical and sectoral frictions. Geographical frictions are being estimated by fitting the model to additional data available in the censuses. The data decomposes the stock of residents along their place of birth and is available in 1920 and 1930. Following Silvestre (2005), comparing the stocks between 1920 and 1930, I can obtain net migration rates between provinces, thus providing implicitly geographical information to estimate the impact of distance on migration flows. The estimation itself is a minimum distance estimation that fits the geographical stage of the labor supply model to the data.

In order to estimate sectoral switching costs I fit the model to changes in labor market conditions at the province-sector level from before to after the war. A key concern is that migration decisions were made during the war based on wage dynamics that are not available. I overcome this data limitation by using the estimated labor demand model together with the trade shock to simulate unobserved wages during the war and estimate sectoral frictions consistent with those wages. As has been pointed out by Silvestre (2005), levels of internal migration during that period were markedly low, amounting to decennial net flows of less than 5 percent out of the population. Consistent with that, the estimated model indicates high frictions to labor mobility across sectors and in particular across space, implying similarly low levels of migration with less than 3 percent moving over the 6 year period that is being considered.

As a result of the estimation I obtain simulated reallocation patterns of labor that are consistent with the changes in labor market conditions due to the war. The implied reallocation patterns strongly suggest that spatial frictions dominate sectoral adjustment frictions, with 83 percent of the adjustment happening across sectors *within* provinces, rather than *between* provinces. Finally, I use the estimated model to obtain the counterfactual evolution of the Spanish economy in the absence of the World War I shock. This exercise shows that the War increased the overall size of the manufacturing sector by 13 percent, while shifting the national industry composition towards more advanced industrial sectors such as chemicals, metal works, textiles and garments. The model can also be used to calculate changes in nominal income in the counterfactual. As suggested above, during the War Spain experienced a differential growth pattern between northern provinces (defined as above Madrid in terms of proximity to France), and southern provinces. Northern provinces experienced around 30 percent larger (nominal) income growth than southern provinces. The counterfactual without the War indicates only a minimal spatial gradient of 4 percent, residual productivity trends can explain a further 15 percent with the reamining 11 percent being attributed directly to the War. Since the model only allows a parametrically limited channel this can be understood as a lower bound for the effect of the War shock on spatial inequality.

The current study implies a tentative explanation for the lack of development prior to the shock. The presence of decreasing returns in the short run suggest that even if labor reallocation took place it would not generate a fortuitous circle of productivity gains, higher wages and further reallocation. Rather sectoral productivity would be on average decreased as a result of employment growth and only recover in the medium run. Such dynamics in productivity gains would inhibit structural change, especially when combined with a high level of labor market frictions. If furthermore workers reside in low productivity sectors in provinces that are distant from the provinces that feature highly productive sectors then the high level of *spatial* labor market frictions is particularly prohibitive. In the Spanish case, the analysis seems to suggest that the pre-shock wage differential between the industrializing North and the agricultural South was insufficient to surmount spatial labor market frictions. An additional obstacle to reallocation might be present if workers do not respond effectively to individual sectors' wage dynamics but rather make migration decisions based on the general appeal of provinces as a whole - an approach that is implicit in the two stage labor supply model formulated in the current study and is consistent with the data.

This interaction between decreasing returns and labor market frictions can actually be beneficial in the presence of the shock. Labor market frictions effectively lock in labor in new sectors until the decreasing returns vanish, inducing a delayed industrial dynamism as a response to the shock that can be related to the economic take off observed in Spain in the 1920s, long after the shock of the War had faded away.

The current setting has three distinct advantages that make the analysis possible. Firstly, the shock is large as well as spatially and sectorally asymmetric and plausibly exogenous towards prior industrialization patterns in Spain. This provides a large amount of independent variation that allows to identify the parameters. Secondly, there is prior substantial variation in sectoral specialisation across cities, allowing for an uneven impact of the shock across space. Finally, the policy response was limited. During the War the central government in Madrid was dominated by the land-based oligarchy, who took little interest in the economic needs of the business community in Catalonia or the Basque country (Harrison; 1978). Policy remedies only came late and in a limited form.[3]

Related Literature This paper contributes to a growing literature that looks at how industries and regions within countries might respond to an external shock. What sets the current paper apart is that it looks at a natural experiment that affected the whole economy while also accounting for the sectoral dynamics. As such it is a convenient setting to shift the focus towards the effect of labor market frictions and scale economies as well as their interaction. In doing so, this paper brings together different aspects that have been looked at separately before. One of these aspects is the endogenous productivity response to trade changes and how that in turn affects industry dynamics as in Juhasz (2017). She studies how temporary import protection can induce import substitution and productivity improvements in the textile industry during the Napoleonic blockade. In the current setting the model allows for endogenous productivity responses in a more abstract way as a function of the observable scale of employment.

Another aspect is how trade shocks can fuel differential population dynamics between cities and provinces creating persistent differences. This has been explored before by Hanlon (2014) in the case of a negative supply shock caused by the U.S. Civil War (1861-1865) which dramatically reduced the cotton imports to the English cotton textile industry. This differentially affected cities that were more specialised in that industry compared to those that were not. The same effect is present in the current setting, but crucially the data in combination with the structure allows us to examine the labor market interactions across multiple sectors and provinces, granting valuable insights on how labor market frictions shape regional dynamics as a response to the shock. Furthermore, as suggested above, the combination of labor market frictions and productivity dynamics suggest an interesting perspective on delayed and to some degree persistent effects of external shocks.

Finally, some of the findings in this paper are reminiscient of a study conducted by Dix-Carneiro and Kovak (2017). They examined Brazil's regional dynamics as a result of trade reforms and trade liberalization and find slow adjustment and steadily increasing divergent trade effects driven by a mechanism where high labor market frictions and slow capital accumulation drive the adjustment pattern. Their empirical and theoretical setting is very different: While they focus on a permanent change in the trade envi-

[3] For example only in 1917 did the Spanish government introduce a *law for the protection of new industries and the extension of existing ones* earmarking 10 million pesetas for the use of industries falling far short of the demand of the industry lobby to establish a foreign exchange bank and a commission house to faciliate the financing of exports instead.

ronment of a country, I focus on how a temporary shock can reallocate inputs across provinces and sectors and the analysis is more focused on counterfactuals employing an extended quantitative spatial equilibrium model. However, some of their results are reflected in the current paper, such as the prolonged effect of the shock as well as the limited labor mobility across space.

Secondly, the paper adds to the quantitative economic geography literature as recently surveyed by Redding and Rossi-Hansberg (2016). I contribute by showing how to adapt a baseline economic geography paper to examine reallocation patterns of labor by accounting for several key aspects of the data. Firstly, the model is more flexible with regard to the presence of productivity returns to scale, which are important in determining the gains from reallocating labor. Secondly, the proposed model manages to match the observed persistence of employment at the province and sector level despite large wage differentials, by introducing sectoral and spatial labor market frictions, making it possible to compute unobserved patterns of reallocation and allowing to distinguish within and between provincial adjustments. A key underlying theme of the current work is how to combine sufficient structure to augment the paucity of the historical data. A constraint is that a model with rich spatial interactions usually requires flow data to infer the structural parameters and to disentangle different labor market frictions. I demonstrate how to leverage the structure of the model to estimate its parameters, relying on a separate treatment of labor demand and labor supply as well as a convenient separation of spatial and sectoral labor market frictions.

Finally, the paper adds to the literature on Spanish economic history by showing that the WWI shock had an important impact on the Spanish economy, not necessarily by creating large output and productivity gains directly, but by reallocating factors across space and sectors to provide the preconditions for an economic take-off in the 1920s. As such it is a middle ground between the two opposing views in the literatue. The established view, represented by Roldan and Delgado (1973), interprets the war as a large turning point for economic development. Using his own constructed GDP series, Prados de la Escosura (2016) emphasises that the World War shock actually decreased GDP per head and instead he points towards the 1920s as a much more important decade for Spain's development. My analysis implicitly connects the two events by pointing towards the reallocation of labor across sectors as a fertile ground for capital fuelled growth in the 1920s.

The remainder of the paper is structured as follows. Section 1.3 discusses the historical background, describing both the situation in Spain before the War and during the

War. Section 1.4 describes the various data sources as well as the construction of the labor market panel that underlies most of the analysis. Section 1.5 gives reduced form evidence on the trade shock and its effect on regional population dynamics. Section 1.6 describes the theoretical model that guides the estimation and analysis. Section 1.7 then proceeds with describing the estimation procedure. In Section 1.8 I then use the quantatitative model to simulate Spain in the absence of the War before discussing the results. Finally, in section 3.7 I conclude.

1.3 Historical Background

This section describes the historical circumstances. The first part gives an overview of the state of the Spanish economy towards the beginning of the war. The second section gives an overview of the historical circumstances of the World War itself and how Spain itself was connected to it.

1.3.1 Spanish Economy at the beginning of the 20th Century

After missing the first wave of the industrial revolution in the first half of the 19th century (Harrison; 1978), the Spanish economy underwent a period of rapid industrialization in the second half of the 19th century, fuelled by market integration due to the expansion of the railroad network which in turn resulted in the devolution of industrial capacity to the peripheral provinces with the cotton industry in Catalonia and Metallurgy in the Basque country developing especially rapidly (Nadal; 1975). However, industrialization soon came to an early halt with the census data showing little increase in industrial employment from 1887 onwards as can be seen in figure 1.6. This is also mirrored by very low GDP per head growth rates averaging 0.6 percent between 1883-1913 (Prados de la Escosura; 2017). Some authors attribute the low levels of growth to limited demand for manufacturing goods domestically as well as little capacity to compete with goods from countries such as Germany, France and the UK that are more advanced in terms of their industrialization (Harrison; 1978).

As a result, at the beginning of the 20th century, the industrial sector barely continued to expand and Spain remained at a low level of industrial development. According to census data, in 1900 roughly 70% of the working population worked in agriculture and only 12.5% worked in industrial/manufacturing sectors. Industrialization only proceeded slowly, with the industrial sector only growing marginally in total employment by 3%, adding a little bit less than 40,000 jobs nation-wide in the first decade of the century. At that time, the largest share of the industrial sector was made up by sectors associated with primary goods, such as the exploitation of mines or the production of construction material.

In terms of the spatial distribution of the population, most of the population was still concentrated in predominantly rural and agricultural areas such as Andalucia[4] or Castilla y Leon[5]. However, looking beyond the larger regional aggregation and looking at individual provinces, it is precisely such major urban centres such as Oviedo, Valencia, Bilbao, Madrid and Barcelona that increasingly attracted and concentrated the Spanish population. The provinces that contained these urban centres tended to concentrate most of the industrial activity as can be seen by the map in figure 1.7 indicating the spatial distribution of manufacturing employment. While internal migration was perenially low, with net migration amounting to less than 5% of the population before 1920, the two largest cities, Barcelona and Madrid, tended to nevertheless attract a large share of agricultural workers from other provinces, making them unique magnets for migrants around 1900 (Silvestre et al.; 2015).

The industrial structure of those urban centres was heterogenous. For example, Barcelona was highly specialised in the cotton textile industry, while Valencia specialized in garments. Because of natural endowments mining and associated downstream industries dominated in Oviedo and Jaen. The Basque country had an early advantage in the heavy metal industries, featuring numerous Martin-Siemens open hearth furnaces for steel production as well as other installations. This degree of agglomeration of specific industries even at this early stage of industrialisation suggests some degree of agglomeration externalities.

In terms of external markets, at the end of the 19th century, (former) colonies and other Latin American markets played a particularly important role, while after the loss of the colonies Spain's exports shifted more towards European countries with France and Great Britain taking up the biggest share of exports. Most of the exports were raw materials or agricultural products consistent with the low developmental status of Spain at the time as depicted in figure 1.2. In general Spain ran a trade deficit for most of the beginning of the 20th century except for the short period under consideration in this paper.

[4]Andalucia comprises eight provinces: Almería, Cádiz, Córdoba, Granada, Huelva, Jaén, Málaga and Seville, with major industrial activity located in Seville and Mining employment in Huelva

[5]Castilla y Leon comprises nine provinces: Avila, Burgos, Leon, Palencia, Salamanca, Segovia, Soria, Valladolid and Zamora with major industrial activity centred in Valladolid.

In summary, it can be stated that at the beginning of the 20th century Spain was a predominantly agricultural economy with a low level of industrial activity and while there was some rural urban migration, there was in general little dynamism towards further industrialisation.

1.3.2 The Spanish Economy and World War I

The assassination of the Austrian Archduke Franz Ferdinand on 28 June 1914 by Yugoslavist revolutionaries, triggered a series of declarations of Wars that set off the first World War on 28 July 1914, with the allied powers spearheaded by France, the British Empire, Russia, and later on the United States, fighting the central powers, composed of the German Empire, Austria-Hungary, the Ottoman Empire and other co-belligerents. The consensus is that a conflict limited in terms of duration and extent was expected, but instead it would become one of the largest wars in history, spreading across all major populated continents and lasting until 11 November 1918.

At the onset of the war Spanish society was divided into two opposing camps, with liberal fractions supporting the allied powers, and the remainder of the population supporting the central powers. However, a participation in the war itself was not considered feasible (Harrison; 1978), so Spain remained neutral throughout the war.

The effects of the first World War on the Spanish economy are well documented in the reports by the *Instituto de Reformas Sociales* (Instituto de Reformas Sociales; 1916). They can be broadly summarised into two categories. Firstly, the war brought about opportunities to provide war materials to the belligerent nations. This spawned increased demand for textiles, garments, and for the heavy metal industry. Secondly, a lack of British, French and German competition in the home market provided an opportunity for domestic producers to produce import substitutes. The report mentions new factories that produced goods as varied as supplies for cars, paper folders, perfumes, small machinery, lightbulbs and others. I will examine the effect in more detail in the reduced form section below.

1.4 Data

Labor Market Panel Data The main source for labor market data is an industry survey that covers the years 1914, 1920, 1925 (Ministerio de Trabajo; 1927). This industry survey was published by the Ministery for Labor and Industry and is based on surveys conducted at all public firms and large private enterprises in cities that are larger than 20,000 inhabitants (Casanovas 2004). It covers 23 different industries[6] and 48 different provinces.[7]

While the industry survey covers a large range of the manufacturing sector, it does not give further information on the remaining economy. As mentioned before, a crucial feature of the Spanish economy was a large agricultural sector. In order to account for that, I digitized the occupation-province specific section of the census for 1900, 1910, 1920 and 1930. I use the 1920 data on agricultural employment[8] to augment the 1920 data.[9] For the 1914 data, I use the 1910 province specific agricultural employment data and extrapolate by calculating province specific fertility trends until 1914. Finally, I use data contained in the official Spanish statistical yearbooks on province specific agricultural mean wages for 1915 and 1920.

Trade Data The trade data is taken from annual trade records released by the Spanish custom agency. Using crowdsourcing services, I digitized the trade statistics for the years 1910 and 1914-1919. For those years, the quantity of exports in 383 product categories across 77 different destination countries are available. Furthermore, the border agency uses a system of product level prices to obtain total export values. These prices do not vary throughout the period of consideration and can be interpreted to give the relative pre-war prices across goods.

[6]The industries included are called: Books, Ceramics, Chemicals, Construction, Decoration, Electricity, Food, Forrest, Furniture, Garments, Glass, Leather, Metal Works, Metallurgy, Mines, Paper, Public, Public Industry, Textiles, Tobacco, Transport, Varias, Wood.

[7]The census for 1910 lists 49 different provinces. They mostly correspond to the modern administrative units called *provincias* - provinces - which are in turn roughly the NUTS3 level administrative units of Spain. There are some minor differences, e.g. in how different off-continental administrative units are being treated. For my analysis I drop the Canary islands from the sample since their distance from the mainland makes it hard to argue that they are similarly integrated as other provinces.

[8]More specifically I add the Agriculture (Owner) section and the workers in fishery, forrest and agriculture together to obtain an aggregate size of the agricultural sector at the time in each province

[9]When merging the census data with the industry survey, I adjust for the fact that the survey does not cover the universe of workers, while the census does. In order to maintain the correct relative size of agriculture to manufacturing sector, I compare the total size of industry employment in the survey data with the census - with the census potentially accounting for informal employment as well as industries in smaller villages. On average, the manufacturing employment size of the survey data represents at least 44% of the manufacturing sector in the census data. I scale the agricultural employment accordingly when merging the census and survey data.

Correspondence In order to construct a correspondence between product-level trade data and industry-level labor market data, I used an additional publication that lists the official correspondence between industries and occupations (Instituto Nacional de Prevision Social; 1930), often explicitly stating the associated product as occupation name for an industry. From that I constructed a correspondence table that matches products to industries.[10] While some products can be uniquely associated to one industry, others can be at least matched with two industries. In matching exports to industry levels, I add the export values for those products to both relevant industries.

Migration Data In order to infer labor mobility costs, data on migration flows is necessary. I follow Silvestre (2005) and use the province level data on inhabitants that are *Born in Another Province* which is contained in the censuses. For 1920 and 1930 additional information is available listing not only the stock of migrants which were born in another province, but their origin province as well. The difference between 1930 and 1920 in the stock of migrants - adjusted for decennial survivability rates - is informative about net migration. In order to construct net migration, I follow Silvestre (2005) and use the decennial census survivability rate between 1921-1930, $S \equiv 0.86$. Net internal migration can be obtained by constructing the survivability adjusted change in stock of migrants, i.e.

$$\text{Internal migrations}_{1930,1920,i,j} = BAP_{i,j,1930} - S \times BAP_{i,j,1920}$$

where $BAP_{i,j,1920}$ refers to the stock of residents in i who where born in province j in 1920.

Distance Using GIS software, I georeferenced the Spanish railroad network in 1920. Then, using MATLAB's internal shortest path function, I obtain bilateral distances between provincial capitals along the shortest path of the railroad network. In order to obtain distances to Paris, I augmented the graph with the French railroad network and further added maritime linkages between important ports in France and Spain. Again using the shortest path functionality of MATLAB I can obtain the shortest distance along this transportation network between provincial capitals in Spain and Paris.

[10]The correspondence table is available upon request.

Housing The housing expenditure share as well as stock and rental rates can be imputed combining different data sources. The statistical yearbooks make available the number of buildings available in a province as well as the inhabitants and thus the effective occupancy rate, the inverse of which is the share of a building that is rented by an average resident. Additionally, average yearly rental expenditure is selectively available across provinces in the Boletins of the *Instituto de Reformas Sociales*. This yearly rate can be adjusted towards an hourly rate in a province, r_i. Total expenditure on housing can be imputed by firstly multiplying the rental rate and the inverse of the occupancy rate - call this the unit rental rate - with the stock of housing. Calculating total expenditure on housing as a share of total labor income across all provinces defines the expenditure share on housing, which I will refer to as δ.

1.5 Reduced Form Evidence

In this section I develop five stylized fact that characterize the nature of shock, as well as the impact it had on regional development within Spain. The stylized facts will guide the choice of the model and will inform the empirical estimation.

Stylized Fact 1: The Trade Shock was large & spatially biased The export shock was large from an aggregate point of view. In 1915 aggregate exports increased by 40% compared to 1914 and stayed at a high level for as long as the war lasted.[11] Most of the increase was due to differential increase of belligerent countries compared to non belligerent countries as shown in figure 1.1: The trade to belligerent countries tripled, while trade with non-belligerent countries remained at a relatively low level and only grew in the later war years above pre-war levels. Most of the increase in trade with belligerent country stems solely from export increases to France. Since the trade shock originated mostly from France, provinces close to the French border had a more favorable position since they facing lower transport costs when shipping towards France. If transport costs matter, then the fact that most of the increase was due to France implies a spatial bias in the trade shock.

[11]This increase is probably underestimated since official statistics kept the price for the calculation of values of exported goods at a constant level during the decade under consideration, while it is plausible that increased demand has further increased the price.

Stylized Fact 2: The Trade Shock was asymmetric across Sectors Most of the increased demand can be associated with war needs, such as Textiles, Garments, Metal Works and Leather goods which is evident in the shift in the sectoral composition of exports from Spain to France. However, it is not clear whether these changes in sectoral trade flows are driven by plausibly exogenous demand side effects or by potentially endogenous domestic supply side trends. In order to obtain a sector specific measure of the foreign demand shock, I construct a theoretically consistent measure by leveraging a standard gravity trade equation,

$$X_{od,s,t} = \tau_{od}^{-\epsilon_s} w_{o,s,t}^{-\epsilon_s} A_{o,s,t}^{\epsilon_s} P_{d,s,t}^{\epsilon_s} E_{d,s,t}$$

where $X_{od,s,t}$ denotes the export level from origin (o) to destination (d) in sector s which depends on bilateral resistance term, τ_{od}, as well as the marginal cost of production in the origin country, $w_{o,s,t}/A_{o,s,t}$, positively on the sectoral expenditure in the destination country, $E_{d,s,t}$ and the price index, $P_{d,s,t}$, measuring the competitiveness in the destination market, and where ϵ_s denotes the sector specific trade elasticity. Constructing the growth of exports, $\hat{X}_{od,s,t} \equiv \frac{X_{od,s,t}}{X_{od,s,t-1}}$, and comparing the growth rate across destination countries, one can obtain the following expression,

$$\Delta_{o,s,t} \equiv \frac{\hat{X}_{od,s,t}}{\hat{X}_{od',s,t}} = \left(\frac{\hat{P}_{d,s,t}}{\hat{P}_{d',s,t}}\right)^{\epsilon_s} \times \left(\frac{\hat{E}_{d,s,t}}{\hat{E}_{d',s,t}}\right)$$

where hat variables refer to changes. In words, this double difference states that export growth from origin o to destination d compared to export growth from o to some other destination d', $\hat{X}_{od,s,t}/\hat{X}_{od',s,t}$, is a function of relative changes in the price index in the two destination countries, $\hat{P}_{d,s,t}/\hat{P}_{d',s,t}$, as well as relative growth in their expenditure levels $\hat{E}_{d,s,t}/\hat{E}_{d',s,t}$. This double difference can be used to isolate destination specific effects, in particular, the relative changes in the expenditure and competitiveness of one desitnation market compared to some other, plausibly unaffected, comparison group.

When calculating this measure for the WWI shock, I compare sectoral export growth to belligerent countries to non-belligerent countries. However, some adjustments are necessary to account for secondary effects of the war. First of all, the war made trade across the frontline and maritime trade after 1917 difficult. Therefore the sample of belligerent countries that I focus on only includes France, Italy and the United Kingdom and I construct export growth by comparing the mean export levels for 1915/1916 with the baseline export in 1910, thus avoiding additional distortions after 1916 and the partial-year war effect of 1914. For the non-belligerent comparison group I exclude belligerent countries as well as the United States, to avoid any war preparations to pollute the measure. The sectoral results can be seen in the appendix in figure 1.10. The sectors that benefited from particularly high levels of demand during the war are Garments, Glass, Metal Works, Mines, Paper and Textiles. These sectors experienced between 5-20 times more growth from belligerent countries than they did from non-belligerent countries.[12]

Stylized Fact 3: Regional Dynamics exhibited a Spatial Gradient The shock induced a demand shock that had spatial and sectoral characteristics, but how did the shock affect regional dynamics? I use the labor market data introduced in the previous section to construct income growth at the sector-province level. In order to examine whether the spatially biased shock induced regional development that was spatially tilted, I run the following regression,

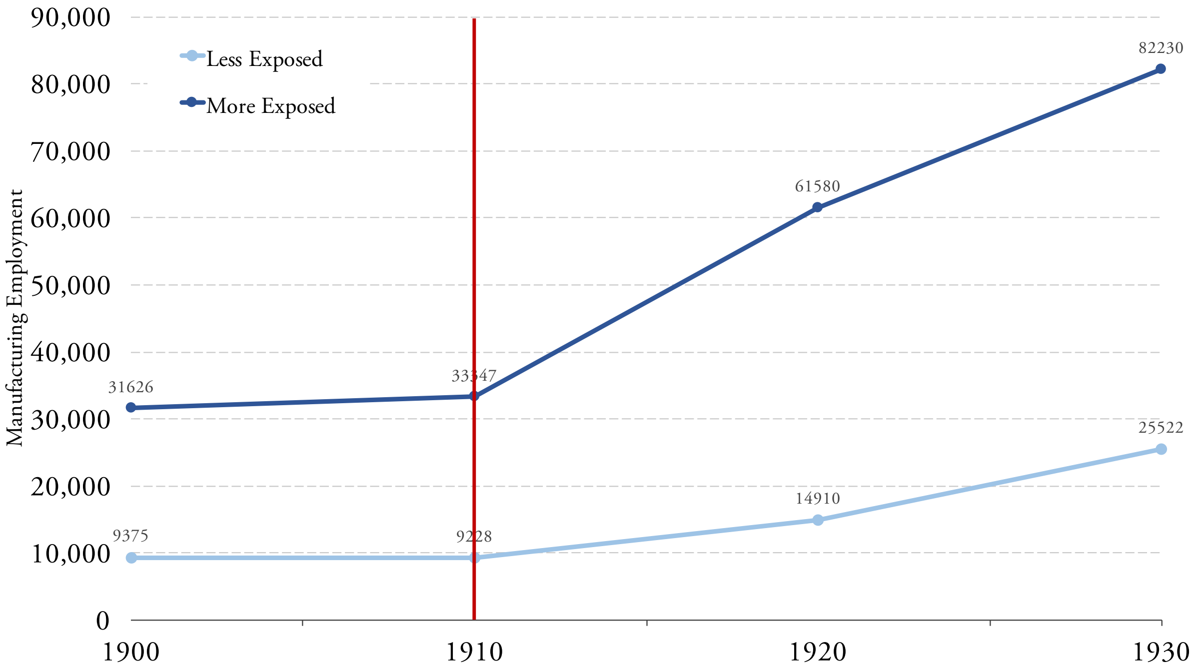
$$\frac{Y_{i,s,1920}}{Y_{i,s,1914}} = \alpha + \beta_1 distance_{i,Paris} + \epsilon_{i,s}$$

where $Y_{i,s,1920}$ is the total labor income of sector s in province i in 1920, that is $Y_{i,s,1920} = w_{i,s,1920} L_{i,s,1920}$ with $w_{i,s,1920}$ referring to the wage in that province-sector and $L_{i,s,1920}$ referring to the total number of employees, and finally $distance_{i,Paris}$ refers to the shortest distance along the railroad network or maritime linkages between the capital of province i and Paris. The fitted line is depicted in figure 2.3. I find that each additional 100km distance to Paris translates into 4 percent lower income growth. This stylized fact is also robust at the sectoral level and controlling for labor market tightness - as proxied by the own sector size relative to the province size - as well as initial differences in comparative advantage - as proxied by the sectoral employment share in the national industry - as can be seen in regression table 1.7.

Stylized Fact 4: Regional Dynamics & Industrial Capacity To understand the differential impact that the shock had at the province level, I use the sectoral shocks to construct an exposure measure to the shock, i.e.

[12] As can be seen in the table Mining exports to non-belligerent countries all but disappeared in the period under considerations. According to the historical reports, this is not due to demand factors, but capacity constraints in Spain, a feature that is not inherent in the standard gravity approach.

Figure 1.3: Manufacturing Employment Growth and Shock Exposure



Notes: This figure shows the evolution of average manufacturing employment of most and least exposed provinces. Most and least exposed provinces are defined as above or below the median value for the exposure index defined below. The red line indicates the observation after which the WWI shock (1914-1918) is taking place. The data is taken from the population censi 1900-1930 and export statistics.

$$E_i \equiv \sum_s \frac{L_{i,s,1914}}{\sum_j L_{j,s,1914}} \times (g_{Spain,Bel,s} - g_{Spain,Non-Bel,s}) \times X_{Spain,France,s,1914}$$

where $g_{Spain,Non-Bel,s}$ and $g_{Spain,Bel,s}$ refers to the growth rate of exports to non belligerent and belligerent countries respectively as calculated above. The difference is the excess growth in sector s associated with WWI. The exposure term summarizes therefore at the provincial level the expected incidence of the trade shock given the pre-existing industrial capacity within a sector proxied by the employment share in the *national* sectoral employment and given the estimated increase of French exports due to the WWI shock. In order to examine the impact of this exposure measure on regional dynamics and in order to illustrate pre trends I rely on additional data from the Spanish population censi on manufacturing employment. In order to analyse the responsiveness to the continuous exposure variable I examine a continuous treatment Diff-in-Diff specification in the spirit of Acemoglu et al. (2004), i.e.

$$\ln y_{it} = \delta_i + \delta_t + \gamma_1 \times d_{1920} + (\gamma_2 + \varphi \times d_{1920}) \times \ln dist_{i,Paris} + (\varphi_1 \times d_{1920} + \varphi_2 \times \delta_t) \times \ln E_i + \epsilon_{i,t}$$

The left hand side variable, y_{it} is manufacturing employment in province i at time t, where manufacturing employment is available in 1900, 1910, and 1920, δ_i refers to the full set of province specific fixed effects, d_{1920} is a dummy for the 1920 which is the first observation after the WWI shock,

Figure 1.3 illustrates the results and the regression results are reported in table ??. The coefficient of interest is φ_1 which is the responsiveness of manufacturing employment towards increases in the exposure measure - which measures a province's ability to exploit sector specific shocks given the scale of its prior industrial capacity in the affected industries. Comparing the 10th to the 90th percentile this gives an estimated effect of $(8.65 - 5.89) * \varphi = 0.19$ log points. The regression as well as the figure above point towards parallel trends prior to the shock, as can be seen by the coefficient φ_2 which is not significantly different from zero.

Stylized Fact 5: Local Labor Supply can inhibit Regional Dynamics In the presence of spatial labor market frictions, which would be consistent with the low level of decennial internal migration at the time in Spain (Silvestre; 2005), labor supply is partially localized and must be sourced from other sectors within the same province. This implies that the larger an industry's share in the local labor market the more limited the pool of workers it can source from. Regressing (nominal) income growth on the sectoral share of total provincial employment before the war which is defined as Employment Share of Sector in Province) $\equiv \frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$, I find that an increase by 1 log point translates into .1 log points lower nominal growth rates. The linear fitted line can be seen in figure 1.8. This finding is robust to controlling for comparative advantage as proxied by the size of the province-sector in the national industry, and level size affect as proxied for by (log) employment in 1914 of that industry as can be seen in table 1.7.

1.6 Theoretical Framework

The theoretical framework is informed by the stylized facts shown above. As indicated by the spatial gradient, spatial frictions in the output and input market will play a prominent role, thus shifting the attention towards economic geography models. Furthermore, the setting requires a multi-sectoral model to account for the sectoral heterogeneity of the shock. Finally, the last two stylized facts suggest that provinces compete for labor inputs and that labor supply can be - to some extent - localized. In order to accommodate that, I will extend the standard economic geography model to account for a fairly general set of labor market frictions, introducing switching costs that make labor sticky at the provincial and the sectoral level.

1.6.1 Setting

Consider an economy with a fixed number of I locations indexed by $i, j, k \in \mathcal{N}$. Locations are heterogeneous in their exogenously fixed housing supply, H_i, and their geographical location relative to one another. Each location produces goods in S sectors $r, s \in \mathcal{S}$. There are only two periods and the initial distribution at time 0 of the population across locations and sectors, $[L_{i,s,0}]_{\forall i,s}$, is given.

1.6.2 Labor Demand

Labor demand is being determined by a multi-sector Ricardian model with industry level economies of scale along the lines of Kucheryavyy et al. (2016), that allows for intranational trade between provinces within a country and international trade with foreign countries. The only factor of production is labor. Each country has a representative consumer with upper tier Cobb Douglas preferences across housing - with an expenditure share δ - and industry bundles, with industry specific expenditure shares given by $\beta_r \in (0, 1)$, such that $\sum_r \beta_r = 1 - \delta$. Trade costs are of the standard iceberg type implying that delivering a unit of any good in industry s from province i to province j requires shipping $\tau_{ij,s} \geq 1$ units of the good. Trade shares take on the following functional form,

$$\lambda_{ij,s,t}(\mathbf{w}_{s,t}, \mathbb{L}_{s,t}) = \frac{S_{i,s,t} L_{i,s,t}^{\alpha_s} (w_{i,s,t} \tau_{ij,s})^{-\epsilon_s}}{\sum_k S_{k,s,t} L_{k,s,t}^{\alpha_s} (w_{k,s,t} \tau_{kj,s})^{-\epsilon_s}}$$

where $\mathbf{w}_s$ and $\mathbb{L}_s$ refers to the vector of sectoral wages and employment levels across provinces respectively, $S_{i,s,t}$ is a province-sector specific productivity shifter, $w_{i,s,t}$ are the province-sector specific wages, and $L_{i,s,t}$ the quantity of labor employed, and $\tau_{ij,s}$

refers to the iceberg trade cost as defined above. Higher labor densities increase productivity via the parameter ψ_s, which in turn increases trade shares mitigated via the *trade elasticity*, ϵ_s, formally being defined as $\epsilon_s \equiv -\frac{\partial ln(\lambda_{ij,s}/\lambda_{ii,s})}{\partial ln \tau_{ij,s}}$. Together the effect can be summarised as $\alpha_s \equiv \psi_s \times \epsilon_s$ which is the elasticity of changes in trade flows as a response to changes in employment size of a sector, which I refer to as *scale elasticity*. Finally, the trade elasticity ϵ_s also governs the sensitivity of trade flows with regard to changes in the destination specific marginal cost pricing, in particular if they are driven by changes in the input cost, that is the local wage, $w_{i,s,t}$.

The current framework, which allows for industry level economies of scale, is consistent with a Ricardian model with external scale economies but is also sufficiently general to nest multiple other trade models including trade models that feature internal scale economies as pointed out by Kucheryavyy et al. (2016).[13] Within a given period the labor allocation is fixed. The static equilibrium can be defined as follows,

Definition 1 (Static Equilibrium). *The static equilibrium within a period t, given the labor distribution, is given by goods market clearing, balanced trade and housing market clearing.*

$$w_{is,t}L_{is,t} = \sum_j \lambda_{ij,s}\beta_s Y_{j,t} \qquad \forall(s,i) \in \mathcal{S} \times \mathcal{N} \tag{1.1}$$

$$E_{is,t} = \sum_j \lambda_{ji,s}\beta_s Y_{i,t} \qquad \forall(s,i) \in \mathcal{S} \times \mathcal{N} \tag{1.2}$$

$$r_{i,t} = \frac{\delta Y_{i,t}}{H_{i,t}} \qquad \forall(i) \in \mathcal{N} \tag{1.3}$$

1.6.3 Labor Supply

The initial allocation of households across across locations and sectors, $[L_{i,s,0}]_{\forall i,s}$, is given. Between the first period - $t = 0$ - and the second period - $t = 1$ - Households can make a decision to move across provinces and sectors. The moving decision is based on a nested discrete choice, where workers first decide which province to move to - and implicitly to leave their own sector - and then upon arrival in the province decide which

[13] As Kucheryavyy et al. (2016) show, the framework can map into multi-sector variants of Eaton and Kortum (2002), Krugman (1979) and Melitz-Pareto type trade models (Chaney; 2008)

sector to work in. Indirect utility is given by

$$V_{j,s,t} = \left(\frac{\rho_j w_{j,s,t}}{P_{j,t}^{1-\delta} r_{j,t}^{\delta}} \right) \times \kappa_{j,t} \times \iota_{s,t}$$

where ρ_j represent location specific amenities, $r_{j,t}$ the market clearing rental rate for housing, $P_{j,t}$ represents a local price index which aggregates sector level local price indices according to the Cobb Douglas preferences specified above, that is $P_{i,t} = \tilde{\beta}_n \prod_s P_{i,s,t}^{\beta_s}$ and where the sectoral price index $P_{i,s,t}$ is defined as follows,

$$P_{i,s,t} = \mu_{i,s} \left(\sum_{i \in \mathcal{N}} S_{i,s,t} L_{i,s,t}^{\alpha_s} (w_{i,s,t} \tau_{ij,s})^{-\epsilon_s} \right)^{-1/\epsilon_s}$$

where $\mu_{i,s}$ and $\tilde{\beta}_i$ are some constants, where $S_{i,s,t}$ is a province-sector specific productivity shifter, $w_{i,s,t}$ is the province-sector specific wage, and $L_{i,s,t}$ the quantity of labor employed, and $\tau_{ij,s}$ refers to the bilateral iceberg trade cost, α_s is the scale elasticity and ϵ_s the trade elasticity.[14] Finally, $\kappa_{j,t}$ and $\iota_{s,t}$ represent idiosyncratic preference shocks that capture preference heterogeneity at the micro-level. I adopt the assumption that they are Fréchet distributed.

Assumption 1. *The preference shocks are sequentially drawn and identically and independently distributed across provinces and sectors according to a Fréchet distribution with respective dispersion parameters ν and γ*

$$F(\kappa_{j,t}) = e^{-\kappa_{j,t}^{-\nu}}, \quad \nu > 1, \quad F(\iota_{s,t}) = e^{-\iota_{s,t}^{-\gamma}}, \quad \gamma > 1$$

Assumption 1 allows for convenient closed form solutions of the shares of workers across sectors and space. ν and γ are the respective dispersion parameters and which will be shown to pin down the responsiveness of migration flows to changes in indirect utility. A household which in period t is residing in province i and working in sector s faces the following problem,

[14]The constant $\mu_{n,k}$ depends on the specific model being adopted. $\tilde{\beta}_n$ is the standard Cobb Douglas term

$$\max \left[\underbrace{\mathbb{E}_t V_{i,s,t+1}}_{\text{Remain}}, \underbrace{\mathbb{E}_t \frac{V_{1,s,t+1}}{\mu_{i1}}, \ldots, \mathbb{E}_t \frac{V_{I,s,t+1}}{\mu_{iI}}}_{\text{Remain Sec/Change Prov}}, \underbrace{\mathbb{E}_t \frac{V_{i,1,t+1}}{\mu_s}, \ldots, \mathbb{E}_t \frac{V_{i,S,t+1}}{\mu_s}}_{\text{Change Sec/Remain Prov}}, \underbrace{\mathbb{E}_t \frac{V_{1,1,t+1}}{\mu_s \mu_{i1}}, \ldots, \mathbb{E}_t \frac{V_{I,S,t+1}}{\mu_s \mu_{iI}}}_{\text{Change Prov \& Sec}} \right]$$

that is she can decide to remain in the current sector and in the current province, change just the sector, change just the province or change both. Effectively, the worker compares the indirect utility of remaining in the current province with the expeceted indirect utility of reallocating to any other province subject to incurring a switching cost, where μ_{ij} and μ_s are the geographical and sector specific switching costs that capture the difficulty of switching sectors and provinces.[15] The population that remains in a province is pinned down by the geographical mobility cost μ_{ij} effectively discounting options that involve out migration. Similarly, the population that remains in a sector is pinned down by the sectoral mobility cost μ_s effectively discounting options that involve sectoral switching.

I assume that this decision problem is being done sequentially with the worker first observing the location specific preference shocks, $\boldsymbol{\kappa}_t$, but not yet knowing the vector of sector specific preference shocks, $\boldsymbol{\iota}_t$. In the first stage the worker forms expectation over the maximized outcome in the second stage. Given the Frechet distribution it can be shown that this implicit value has a closed form solution.

Proposition 1. *The expectation of the maximization problem over J alternatives, where the benefit accrued is $\delta_i \times \epsilon_i$ and where ϵ_i is Fréchet distributed with CDF $F(x) = e^{-x^{-a}}$, is given by the following expression,*

$$\sum_i E\left[\max_i (\delta_i \times \epsilon_i)\right] = \left(\sum_j \delta_j^a\right)^{\frac{1}{a}} \Gamma\left(1 - \frac{1}{a}\right)$$

where $\Gamma(\cdot)$ is the Gamma function.

Proof. See appendix. □

This mirrors the implicit value commonly used in nested discrete choice estimations us-

[15] Different interpretations are possible: For agriculture the sectoral switching cost might absorb some of the cost of moving to a major urban center, for other sectors they might simply signify the loss of sector specific human capital. For the geographical part the reallocation cost might absorb the lost utility due to disrupted social connections, a psychic cost or the actual economic moving cost.

ing Gumbel distributed additive preference shocks instead of Frechet distributed multiplicative preference shocks. Based on comparing these implicit values subject to geographical switching cost that are proportional to the expected utility in the destination and specific to bilateral pairs of provinces the Household chooses the optimal location to move towards. Therefore the upper level problem of a worker residing in sector s and in province i reduces to,

$$\max \left[\mathbb{E}_t \frac{\tilde{V}_{1,t+1|s}}{\mu_{i1}}, \ldots, \mathbb{E}_t \frac{\tilde{V}_{i,t+1|s}}{\mu_{ii}}, \ldots, \mathbb{E}_t \frac{\tilde{V}_{I,t+1|s}}{\mu_{iI}} \right]$$

where μ_{ij} is the bilateral spatial mobility cost, where μ_{ii} is normalized to 1, and where the implicit value $\tilde{V}_{i,t+1|s}$ indicates the expected utility obtained after observing the preference shocks in the second stage and making the utility optimizing decision. Due to sectoral switching costs the implict value depends on the initial sector the worker is currently working in. The closed form is given by,[16]

$$\tilde{V}_{i,t+1|s} \propto \mathbb{E}_t \left(\frac{\rho_j}{P_{j,t+1}^{1-\delta} r_{j,t+1}^{\delta}} \right) \left(w_{i,s,t}^{\gamma} + \mu_s^{-\gamma} \sum_{k \neq s} w_{j,k,t+1}^{\gamma} \right)^{\frac{1}{\gamma}}$$

An attractive property of this formulation of the labor reallocation problem is that bilateral flows between provinces is primarilily driven by a measure of aggregate attractiveness of the destination province rather than specifically tied to sector specific dynamics within that destination provinces. This is a more realistic choice in a setting where migrants in faraway provinces have little information about the specific conditions in specific sectors but might have some information about the general attractiveness of a destination. Crucially, the key determinant of the direction of migration flows is the relative size of spatial versus sectoral switching costs, pinning down to what extent labor adjusts between provinces rather than within provinces. I will return to this point during the quantitative analysis. Given the Fréchet distributed preference shocks, standard properties imply the following closed form for the shares of workers who move across provinces,

[16] In the current setting the Fréchet dispersion parameter γ is symmetric across locations, therefore we can abstract from additional multiplicative term that determines the scale of the expectation, $\Gamma(1 - \frac{1}{\gamma})$. However, one can easily extend the current setting to account for heterogeneity of local sectoral labor supply elasticities.

$$\sigma_{ij,s}\left(\tilde{\mathbb{V}}_{t+1|s}\right) = \frac{\left(\mathbb{E}_t \tilde{V}_{j,t+1|s} \times \frac{1}{\mu_{ij}}\right)^{\nu}}{\Omega_{i,s,t}}$$

where $\Omega_{i,s,t} \equiv \sum_j \left(\mathbb{E}_t \tilde{V}_{j,t+1|s} \times \frac{1}{\mu_{ij}}\right)^{\nu}$ summarises the option value of a person currently working in sector s and residing in province i, where $\tilde{\mathbb{V}}_{t+1|s}$ is the vector of implicit values $\tilde{V}_{j,t+1|s}$ as defined above, where μ_{ij} refers to the geographical switching cost, and where implicit values depend on expected wages, rental rates, price indices and the sectoral switching cost μ_s, finally ν defines the elasticity with regard to changes in the implicit values or alternatively the switching cost.

Conditional on reallocating and upon arrival in the province the worker uncovers her vector of sector specific preference shocks, $\boldsymbol{\iota}_t$ and makes a choice selecting a sector. Again assuming Fréchet distributed preference shocks with dispersion parameter γ, one can obtain the following closed form for the share of workers that flow into industry r in province i and where prior to that in industry s,

$$\sigma_{i,s,r}(\mathbf{w}_{i,t}) = \frac{\mu_s^{-\gamma} w_{i,r,t}^{\gamma}}{w_{i,s,t}^{\gamma} + \mu_s^{-\gamma} \sum_{k \neq s} w_{j,k,t+1}^{\gamma}} \qquad \text{for} \qquad s \neq r$$

$$\sigma_{i,s,s}(\mathbf{w}_{i,t}) = \frac{w_{i,s,t}^{\gamma}}{w_{i,s,t}^{\gamma} + \mu_s^{-\gamma} \sum_{k \neq s} w_{j,k,t+1}^{\gamma}} \qquad \text{for} \qquad s = r$$

where $\mathbf{w}_{i,t}$ is the vector of wages in province i and $w_{i,r,t}$ refers to the wage in sector r and in province i. Since the other determinants of indirect utility enter symmetrically across all options, they do not affect the sectoral shares. Finally, one can state the flows from province i and sector r to province j and sector s, as,

$$\sigma_{ij,sr}(\tilde{\mathbf{V}}, \mathbf{w}) = \sigma_{ij,s}\sigma_{j,s,r}$$

where $\tilde{\mathbf{V}}$ represents the vector of expected indirect utilities across provinces, and where province-sector specific flows are separable between, $\sigma_{ij,s}$, that is the bilateral flows between province i and province j, and the sorting into sector r within province i, $\sigma_{i,s,r}$. Total labor supply is then given by a market clearing condition, that is,

$$L_{i,s,t+1} = \sum_{j,r} \sigma_{ji,rs}(\tilde{\mathbf{V}}, \mathbf{w}) L_{j,r,t}$$

1.7 Estimation

In order to use the model described in the previous section for a quantitative analysis of the World War I shock, one needs to obtain estimates of the key parameters. On the labor demand side, we need to obtain trade elasticities, $\{\epsilon_s\}$, scale elasticities, $\{\alpha_s\}$, and productivity shifters, $\{A_{i,s}\}$. On the labor supply side we need to estimate switching costs and geographical and sectoral supply elasticities. The estimation of the parameters determining labor demand can be done separately, since changes in the spatial equilibrium are sufficiently informative to estimate them. Given those estimates we can then estimate the parameters associated with the labor supply model.

1.7.1 Labor demand

The estimation of the key parameters that determine labor demand relies mainly on the labor market data - that is wages and employment size for each province-sector. I demonstrate how to use that data in conjunction with the model structure in order to estimate the key parameters that determine labor demand. In the first step I use a structural approach to separate out origin specific marginal cost prices and market access. In a second step, I then regress the obtained prices on wages and labor densities to obtain the structural parameters.

Obtaining Origin-Prices

From the static (spatial) equilibrium, one can obtain the following two equations,

$$Y_{i,s} = \sum_j X_{ij,s} = \sum_j \tau_{ij}^{-\epsilon_s} p_{is}^{-\epsilon_s} P_{js}^{\epsilon_s} E_{js}$$
$$E_{i,s} = \sum_j X_{ji,s} = \sum_j \tau_{ji}^{-\epsilon_s} p_{js}^{-\epsilon_s} P_{is}^{\epsilon_s} E_{is}$$

where the first equation states that total income in province i and sector s, $Y_{i,s}$, must equal the cumulative export sales for that sector, that is the sum of all export flows from the origin province i to any province j, i.e. $\sum_j X_{ij,s}$. Since export flows follow the gravity structure the second equality follows. The second equation states that total expenditure in province i on goods from sector s, must equal total incoming export flows from all origin provinces j, that is $\sum_j X_{ji,s}$. Combining and rearranging, one can obtain a system of equations in terms of prices only,

$$p_{is}^{\epsilon_s} = \sum_j \tau_{ij}^{-\epsilon_s} \left(\sum_k \tau_{kj}^{-\epsilon_s} p_{ks}^{-\epsilon_s} \right)^{-1} \frac{E_{js}}{Y_{is}}$$

where $p_{i,s}^{\epsilon_s}$ refers to the origin prices introduced above. Standard results in economic geography imply that this equation can be solved to find the unique vector of provincial origin prices (up to normalization) for each sector, $\mathbf{p}_s^{\epsilon_s}$.

Using the labor market data before and after the war - that is for 1914 and 1920 - and using the housing market data to construct disposable income across provinces, $E_{is} \equiv \beta_s Y_i$, one can implement the inversion described in the previous paragraph. In the implementation, I first calculate the Cobb Douglas expenditure shares as the national income share of an industry out of aggregate labor income. This is theoretically consistent with one input economic geography model described above. The housing expenditure share δ is obtained as described in section 1.4. I use the shortest distance along the railroad graph between Spanish provincial capitals and furthermore add France as an additional location, where the distance to France is the shortest distance to Paris across railroad and maritime linkages. The iceberg transport cost is calibrated to be, $\tau_{ij} = distance_{ij}^{-1}$, calibrating the distance elasticity to the canonical value of -1 (Head and Mayer; 2013). Since I do not have coherent labor market data for France, I only include the total value of sectoral exports[17] as additional demand into the economic geography system.

Price Regression

In the second step, I can use marginal cost pricing, which implies that $p_{i,s} = \frac{w_{i,s}}{A_{i,s} L_{i,s}^{\alpha_s}}$, to obtain a log-linear expression of prices as a function of sector-province employment levels and wages. Taking the first difference, I obtain the following equation,

$$\epsilon_s \log \frac{\tilde{p}_{is,t+1}}{\tilde{p}_{is,t}} = \delta_i + \epsilon_s \log \frac{\tilde{w}_{is,t+1}}{\tilde{w}_{is,t}} - \alpha_s \log \frac{\tilde{L}_{is,t+1}}{\tilde{L}_{is,t}} - \log \frac{\tilde{A}_{is,t+1}}{\tilde{A}_{is,t}} \tag{1.4}$$

where relative changes in origin-prices of sector s in province i, $\frac{\tilde{p}_{is,t+1}}{\tilde{p}_{is,t}}$, are a function of relative changes in wages and employment levels in that sector-province and where $\tilde{x}$

[17]French exports are at the yearly level while the labor market data is in terms of the hourly wage and only covers a subset of the overall economy of Spain. When introducing the exports into the model I divide the total value by 54×50 to translate the value into hourly exports. Then I multiply it by the share of the industry that is represented in the sample, that is .44.

indicates that the variable x has been normalized relative to a sector-specific baseline province. The responsiveness of origin prices with regard wages and employment levels is pinned down by the trade elasticity, ϵ_s, and the scale elasticity α_s, respectively. The scale elasticity itself is combination of productivity externalities and how these productivity externalities in turn translate into income gains, that is $\alpha_s = \psi_s \times \epsilon_s$. We can define the structural residual as $\eta_{i,s,t} \equiv \log \frac{\tilde{A}_{is,t+1}}{\tilde{A}_{is,t}}$, which is the unobserved productivity evolution at the sector-province level. Additionally, I include the full set of province specific fixed effects δ_i to control for province specific confounding shocks.

Endogeneity A natural concern is the endogeneity of both wages, $w_{i,s}$, and employment, $L_{i,s}$. The model implies that as a result of increases in productivity, $\frac{\tilde{A}_{is,t+1}}{\tilde{A}_{is,t}} > 0$, labor demand will increase and move along the upward sloping labor supply curve, with increases in wages and employment levels as a result. This implies that the model structure indicates a positive correlation between the residual, $\eta_{i,s,t}$, and the wages and employment levels, which will in turn induce an upward bias for the estimation of α_s and a downward bias for the estimation of ϵ_s. The naive OLS results depicted in table 1.2 shows theoretically invalid negative trade elasticities and large estimates for the external scale parameter, consistent with the model implied bias. An instrument is therefore necessary to remedy the situation. The exclusion restriction for any instrument is that

$$E\left[\eta_{is,t}|\mathbf{z}_t\right] = E\left[\log \frac{\tilde{A}_{is,t+1}}{\tilde{A}_{is,t}}|\mathbf{z}_t\right] = 0$$

where $\mathbf{z}_t$ denotes the vector of instruments and $\eta_{is,t} = \log \frac{\tilde{A}_{is,t+1}}{\tilde{A}_{is,t}}$ denotes the structural error as discussed above. The setting is more challenging than a standard endogeneity problem because of the presence of two - potentially correlated - endogenous variables. An appropriate instrument needs to induce sufficient independent and differential variation in the endogenous variables to separately identify their impact on the dependent variable. The model suggests that labor supply shifters interacted with the incidence of the shock can serve as a source of such variation. Intuitively, while the foreign demand shock translates into a labor demand shock that stems from the industries desire to expand their production, the curvature of local labor supply will determine whether the additional demand is being absorbed mostly into higher wages or larger sectoral size as measured by employment. As illustrated in the figures below.

Historical evidence as well as the stylized facts suggest that spatial frictions are high and that labor supply is highly localized. I exploit this by using the (log) distance to Paris interacted with the (log) employment share of a sector within a province as a first instru-

Figure 1.4: Inelastic Labor Supply

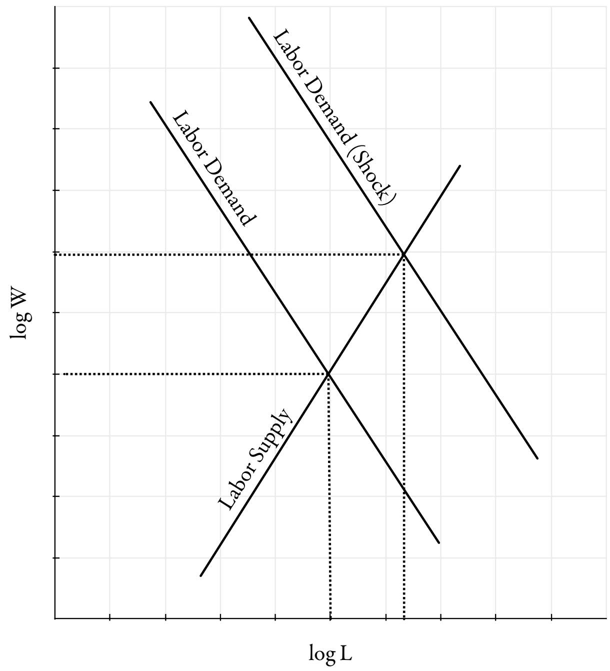
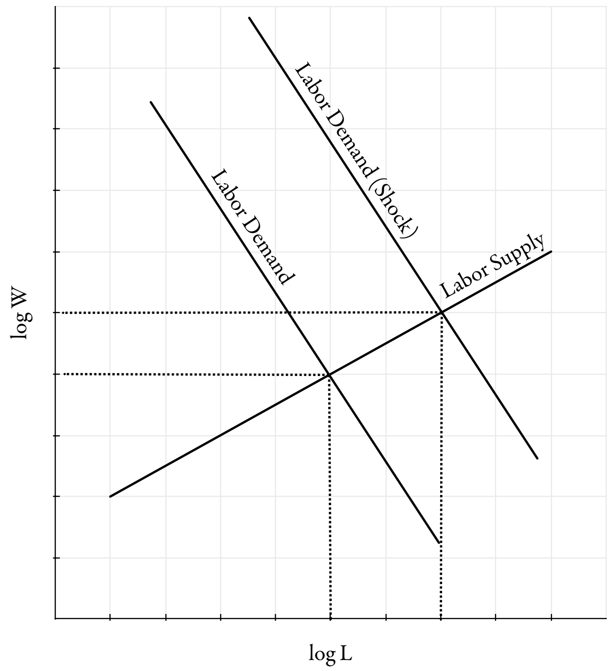


Figure 1.5: Elastic Labor Supply



Notes: The figure illustrates the underlying premise of the instrument. In the presence of two endogenous - potentially correlated - variables, independent variation is needed that differentially shifts the two variables. In the current setting a labor demand shock induces an outward shift of the labor demand curve from $LaborDemand$ to $LaborDemand(Shock)$, inducing an increase in both wages and employment levels. The extent to which the shock is being absorbed by prices or quantities depends on the curvature of labor supply. If labor supply is tight - due to a small size of the local labor market - the curve will be upward sloping and wages will increase rather than employment levels. The opposite is true if labor supply is highly elastic.

ment, where log(Employment Share of Sector in Province) $\equiv log\frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$ acts as a labor supply shifter and is interacted with distance to Paris as a reduced form proxy for differences in geographical advantages vis-a-vis the French destination market. A second instrument is given by a Harris Market Potential measure for the input market leaving own size out as a labor supply shifter, constructed as $LMA_{i,s} = \sum_{j\neq i,r\neq s}\frac{1}{distance_{i,j}}L_{j,r}$. The first stages are reported separately in table 1.6 and are sufficiently strong. Furthermore, there are no apparent pre trends neither along the distance margin nor along the sectoral shares as can be seen in the results for a regression that correlates wage growth prior to the war - that is between 1909 and 1914 - with distance to Paris and sectoral share, as can be seen in table **??**.

Results The results can be seen in the table 1.1. The trade elasticities are theoretically consistent, positive and of comparable magnitude to sectoral trade elasticities currently found in the literature, though due to different aggregation and different time periods not directly comparable. The scale elasticity, α_s, are mostly very imprecisely estimated with the result that for most sectors one cannot reject the presence of constant returns to scale, that is $\alpha_s = 0$. However, in some cases α_s is significantly different from zero and negative, indicating decreasing returns to scale in chemicals, mining, metallurgy, metal works and textiles. These industries tend to require fixed installations, and thus decreasing returns in those sectors in the short and medium run seem plausible. The R squared is a natural measure of the fit of the model. To understand why, recall that prices solve the spatial equilibrium conditions, thus effectively functioning as residual income shifters, once one controls for market access differences. The R squared then measures how much of the variation can be explained by the log linear regression. The fit indicates that the model can explain half of the variation in the residual income shifter. Additionally, the model including labor densities performs much better than the model that only accounts for wage effects - as would be the case in the absence of any scale effects. The model without labor densities can only account for a quarter of the observed variation. Finally, the same estimation strategy can be used for the changes in labor market conditions from 1914 to 1925. The estimated scale elasticity, α_s^{LR}, is reported in the table alongside the previous estimates. As can be seen, decreasing returns are no longer present in the industries in which they were present previously, suggesting decreasing returns to be a medium term phenomenon rather than a constant feature of these industries.

1.7.2 Labor supply

The estimation of the labor supply parameters proceeds in two steps and each step relies on different data sources.

Geographical Frictions

In the first step, I rely on data that shows the decennial change in the number of workers who live in a certain province but were born in another province, that is $BAP_{i,j,t}$ for a worker who was born in province i but now lives in province j. The difference in this stock of foreign born workers, $BAP_{i,j,t} - S \times BAP_{i,j,t-1}$ - adjusted for survivability rate S as explained in section 1.4 - is informative about the net inflow of foreign born workers, either directly from the province under consideration or indirectly from other provinces. The data is adjusted so that the 1920s data shows the same number of total inhabitants born in a given province as the 1930s data, adding the additional population in their origin provinces. Using the closed forms from the previous section I can construct the model equivalent of this moment. The (estimated) stock of workers born in province i and currently residing in province k is given by,

$$B\hat{A}P_{i,k,1930} = \sum_{j,s} \sigma_{jk,s}(\tilde{\mathbb{V}}_{1930|s}, \mathbf{w}) \times \pi_{i,s,1920} \times S \times BAP_{i,j,1920}$$

where $B\hat{A}P_{i,k,1930}$ refers to the simulated stock of workers born in province i and currently residing in province k, $\pi_{j,r,1920}$ refers to the industry share of industry r in province j in 1920 and where the closed form for the share of flows between province j and province k originating from sector s is given by $\sigma_{ij,s}\left(\tilde{\mathbb{V}}_{t+1|s}\right) = \frac{\left(\mathbb{E}_t \tilde{V}_{j,t+1|s} \times \frac{1}{\mu_{ij}}\right)^{\nu}}{\Omega_{i,s,t}}$. Implicitly, this is assuming that there is no sorting across industries of different groups of inhabitants, which in the absence on additional information is a necessary assumption. In the baseline estimation, I assume that wages and price indices follow a random walk. The geographical switching cost is calibrated as a function of distance that is

$$\mu_{ij} = \zeta_{cons} \times \zeta_i^1 \times distance_{ij}^{\zeta^2}$$

where $distance_{ij}$ is the shortest distance across railroad and maritime travelling routes from the province capital in i to the province capital in j in km. The structural estimation chooses the parameter vector $\boldsymbol{\beta} = (\zeta_1^1, \ldots, \zeta_1^I, \rho_1, \ldots, \rho^I, \zeta^2, \nu, \gamma, \mu_1, \ldots, \mu_S)$ to match the observed moments, that is minimizing the error between imputed and observed quantities of workers born in another province,

$$\eta_{i,j}(BAP_{i,j,1930}, \boldsymbol{\beta}) = BAP_{i,j,1930} - B\hat{A}P_{i,j,1930}$$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta \in B} \boldsymbol{\eta}\left(BAP_{1930}, \boldsymbol{\beta}\right)' \boldsymbol{\eta}\left(BAP_{1930}, \boldsymbol{\beta}\right)$$

where $\boldsymbol{\eta}$ is the stacked vector of structural errors, $\eta_{i,j}$.

Identification The origin varying scalar, ζ_i^1, determines the out-province migration share. Conditional on moving out of a province, the distance between the origin province and the destination province is informative about how geographical frictions affect migration flows and thus determines the distance elasticity, ζ^2. The incoming migration to specific provinces above and beyond what is justified by wage differences informs the province specific amenities, ρ_i. The responsiveness of in migration to dispersion in wages across sectors within a given province pins down the local supply elasticity, γ, while the response to dispersion of imputed indirect utilities across provinces informs the estimation of the spatial migration elasticity, ν.

Sectoral Switching Costs

In order to estimate sectoral switching costs, I fit the model to changes in labor market conditions at the province-sector level from before to after the war. A key concern is that migration decisions were made during the war based on wage dynamics that are not part of the available data. In order to overcome this limitation I propose to use the estimated labor demand model together with sectoral trade data from 1915 to simulate the market clearing wages in the presence of the World War shock. I proceed by first using the 1914 data to impute the residual productivities, $\{A_{i,s,1914}\}$, and then feed in the trade shock to back out the simulated market clearing sectoral wage vectors, $\hat{\mathbf{w}}_{s,1915}$. Using these sectoral wage vectors as expected wages, and calibrating the spatial friction to the estimated values from the previous section, I use the closed forms to match the observed changes in employment size between 1914 and 1920,

$$\hat{L}_{i,s,1920} = \sum_{j,r} \sigma_{ji,rs}(\hat{\mathbf{w}}) L_{j,r,1914}$$

where $\hat{L}_{i,s,1920}$ refers to the estimated stock of workers in province i and sector s in 1920, and $L_{j,r,1914}$ refers to the observed size of industry r and province j, and $\sigma_{ji,rs}(\hat{\mathbf{w}})$ is the closed form for migration flows between province j to province i and sector r to sector s. Recall that,

$$\sigma_{ij,rs}(\tilde{\mathbf{V}}, \hat{\mathbf{w}}) = \sigma_{ij,s}\sigma_{s,r,j}$$

that is the bilateral migration flows between sectors and provinces is a composite between outgoing migration between province i and province j in sector s and workers who upon arrival in province i sort into sector r. The structural error is given by,

$$\eta_{i,s}(\beta) = L_{i,s,1920} - \hat{L}_{i,s,1920}$$

$$\hat{\beta} = \arg\min_{\beta \in B} \boldsymbol{\eta}(\boldsymbol{\beta})' \boldsymbol{\eta}(\boldsymbol{\beta})$$

where $\boldsymbol{\eta}$ is the stacked vector of structural errors, $\eta_{i,j}$ and where the structural procedure chooses $\beta = (\mu_{agriculture,1}, \dots, \mu_{agriculture,I}, \mu_2, \dots, \mu_S, \gamma)$ to minimize the distance between the observed and the estimated employment size of each sector-province observation. Notice that the parameter vector includes province-specific switching costs for agriculture. The reason for this is twofold. Firstly, the switching costs associated with agriculture have a natural interpretation that is related to rural-urban migration, thus also involving some form of within province spatial friction. Since provinces differ in their size, this needs to be accounted for. Secondly, the changes of the agricultural sector are quantitatively important to match.

Identification With spatial frictions being calibrated, the size of the sectoral switching cost, μ_s, is informed by the persistence of sectoral employment size in the presence of local wage disparities between sectors. An important caveat is that sectoral switching costs can only be identified in a scenario where workers do not reallocate despite a positive wage differential.

Results The results of the migration cost estimation are reported in table 1.8 in the appendix. Spatial frictions are prohibtively high implying low levels of internal migration with 2.7 percent of the population reallocating during the fitted period which is a gross measure. This is consistent with reported decennial *net* internal migration of 2.8 percent between 1911 and 1920 (Silvestre; 2005). Conditional on migrating distance is an important determinant with the composite distance elasticity, $\zeta^2 \times \nu$, giving a value of 2.38. Finally, labor is highly sticky, with a high degree of heterogeneity across sectors. Agriculture as a sector tends to be especially sticky across all provinces with a high degree of heterogeneity, nevertheless absolutely speaking agriculture releases most of the labor. This is to say that wage differentials are so large that high switching costs are necessary to justify the lack of mobility.

1.8 Quantitative Analysis: Spain without WWI

Implementation Having estimated the parameters that determine both labor supply and demand, one can now use the model to determine the counterfactual evolution of the Spanish economy in the absence of the WW1 shock. Since labor flows depend on the expectations of utilities across province-sectors, and since those utilities themselves depend on the migration choices - via the scale economies - there is a potential for multiple equilibria in this class of model, and a necessity for equilibrium selection when conducting the counterfactual.[18] The baseline results presented here assume that wages and price indices follow a random walk and therefore are expected to remain at the level of the initial equilibrium observed in 1914. That is, workers coordinate using the current wages and price indices. Alternatives to that baseline can be explored. Conditional on implied reallocation patterns market clearing wages can be calculated and conclusions about impacts on income evolution can be drawn. In the following I compare the counterfactual 1920 wages and labor distribution with the observed state of the economy in 1920.

Sectoral Employment Growth One informative aspect of the counterfactual is to compare the aggregate industry sizes between the two scenarios. The results of such a comparison are presented in figure 1.13. The results indicate two important aspects: Firstly, there is high degree of reallocation from the agricultural sector towards the manufacturing sector, with the manufacturing sector as a whole growing by 1 percent as a result. A second important pattern is the heterogeneous response within the manufacturing sector with sectors that were particularly affected by the shock gaining substantially in size. Amongst those food, garments, textiles and metal works stand out, with the largest changes taking place in the textile sector.

Regional Employment Growth The same analysis can be conducted looking at province sizes rather than sectoral sizes. The results are presented in figure 1.16. There are very small difference in regional growth between the two scenarios, consistent with the finding that most of the adjustment is due to within provincial reallocation rather than between provincial allocation. Incidentally this is also consistent with a key characteristic of the migration choice framework highlighted above, that is that mgiration decisions do not respond effectively to individual industry dynamics but rather respond to the aggregate appeal of a destination, as captured in the estimation by the amenity values. Those amenities do not change in the counterfactual thus driving the patterns of the limited migration flows in either scenario.

Spatial Inequality The model can also be used to calculate changes in nominal income aggregated at the sector-province level in the counterfactual. In the data the spatial gradient described in the reduced form section 1.5 led to a differential growth

[18] An alternative approach is to bound the possible outcomes by setting up the counterfactual problem as an MPEC Reguant (2016). This approach is currently being examined but the results are not yet available.

pattern between northern provinces (defined as above Madrid in terms of proximity to France), and southern provinces. Northern provinces experienced around 30 percent larger (nominal) income growth than southern provinces. The counterfactual without the War indicates only a minimal spatial gradient of 4 percent, residual productivity trends can explain a further 15 percent with the reamining 11 percent being attributed directly to the War. Since the model only allows a parametrically limited channel this can be understood as a lower bound for the effect of the War shock on spatial inequality. The exact patterns of incidence can be seen in map 1.16, indicating the differences in nominal income between the two scenarios which is indicative about the extent to which individual provinces managed to capture and monetize the demand shock effectively. The spatial gradient is visible, but is mitigate by provincial heterogeneity in sectoral specialization.

Mobility costs and Market Integration Finally, the model can be used to conduct counterfactuals on the real income levels if one allows for lower mobility costs. In order to simulate a reduction in the spatial mobility cost, I lower the bilateral travel distance between province capitals by 10%. Lowering mobility costs by 10% increases real income gains from the WWI-shock by an additional 3%, increasing the aggregate gains in real income from 20% to 23.59%. This is larger than the welfare gains from lower migration costs in the counterfactual non shock scenario where welfare would have only increased by 2.4%. This suggests that labor market integration and output market integration are complements. The reason why labor market integration and output market integration are complements is due to the fact that a more fluid labor market increases labor supply to the most productive industries and weakens localized competition for labor supply that in the presence of mobility cost can limit the extent to which the shock can be effectively exploited.

1.9 Conclusion

My primary interest was to examine to what extent labor market frictions can inhibit economic development of a country. I used a newly collected historical dataset that combines trade and labor market data, to examine a unique historical episode: A temporary trade shock to a developing economy that prior to the shock only underwent slow structural transformation. I demonstrated the key features of the shock and its impact on regional development within Spain: The shock was temporary, sectorally heterogeneous, large and spatially biased. It induced spatially tilded regional development and affected provinces heterogeneously depending on their initial industrial specialization. I built a quantitative economic geography model that can account for the dynamic response to the temporary shock. A baseline economic geography model is extended to be better suited to match the regional dynamics of a temporary shock, by introducing and estimating labor market frictions that make employment sticky at the sectoral and provincial level as well as allowing for endogenous productivity feedbacks to determine the immediate productivity gains from reallocation.

An interesting aspect of the current work is that limited historical data can be complemented with structural models to improve both the estimation of objects of interest and in order to get further insights into phenomena that are not directly observed - as was done in this paper by obtaining unobserved sector-province labor reallocation patterns consistent with estimated migration costs and observed sectoral employment sizes.

The analysis suggests that high levels of labor market frictions and low immediate returns to reallocation due to the absence of scale economies and even the presence of decreasing returns in some industries prevented the Spanish economy from developing before the War. The shock induced reallocation across space and particularly between sectors within provinces, thus creating the fruitful preconditions for an economic take-off in the following decade. Finally, the analysis suggests that welfare gains from (output) market integration depend on the extent to which input markets are integrated.

This suggests four important conclusions: Firstly, labor market frictions are of primary importance for analysing (spatial) development of a country or the lack thereof. Secondly, the relative size of different labor market frictions determines the pattern of development as well as the extent to which spatial arbitrage is possible between space and between sectors, making a quantitative understanding of these frictions important, in particular when analysing patterns of spatial inequality. If spatial mobility costs are a reasonable concern in a developing country then policy makers need to take into account the distribution of labor as well as the spatial unevenness of the development process. Finally, labor market integration and output market integration ought to be considered in tandem to benefit from the complementary effects of both forms of market integration.

1.10 Tables

Table 1.1: Estimation Results - Labor demand parameters - Long Run

Industry	α_s	Std Err	α_s^{LR}	Std Err	ϵ_s	Std Err
Agriculture	1.20	(2.60)	−0.93	(1.55)	4.71***	(1.74)
Books	−0.19	(1.08)	0.10	(1.54)	5.17***	(1.84)
Ceramics	−0.04	(1.32)	2.04**	(0.91)	5.34**	(2.12)
Chemicals	−0.76	(1.39)	−0.97	(1.09)	5.18***	(1.85)
Construction	0.92	(1.54)	0.11	(0.80)	4.22**	(1.99)
Decoration	0.84	(1.03)	−0.49	(1.51)	5.53***	(2.03)
Electricity	0.00	(1.24)	−0.17	(1.01)	5.47***	(1.89)
Food	0.05	(1.13)	0.72	(1.25)	4.58**	(1.80)
Forest	1.16	(4.71)	−6.10	(10.16)	4.84	(3.30)
Furniture	0.38	(0.87)	0.21	(1.19)	5.38**	(1.92)
Garments	0.28	(1.06)	0.41	(0.97)	4.44***	(1.79)
Glass	1.51	(2.92)	0.78	(1.42)	5.96***	(2.22)
Leather	1.79*	(1.07)	1.58	(1.52)	5.92***	(1.88)
Metal Works	−0.66	(0.88)	0.39	(1.43)	4.43***	(1.80)
Metallurgy	−0.98	(1.74)	−0.85	(1.75)	5.54***	(1.73)
Mines	−2.34*	(1.40)	2.80	(4.95)	5.99***	(1.83)
Paper	−3.06*	(1.89)	−6.86	(4.07)	3.21*	(1.93)
Public	1.34	(6.33)	1.61	(1.61)	5.37	(3.71)
Public Industry	22.21	(23.41)	1.18	(1.34)	11.99	(10.54)
Textiles	−0.88	(0.95)	0.48	(1.39)	4.04**	(1.75)
Tobacco	10.36*	(6.10)	2.78	(5.55)	1.43	(2.68)
Transport	1.51	(1.53)	0.71	(1.37)	4.78***	(1.85)
Varias	−0.95	(1.64)	−3.36	(3.77)	4.68**	(1.97)
Wood	0.53	(1.90)	1.00	(1.26)	4.84**	(1.99)
Observations			625			
R2			0.5892			
Province FE			✓			

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports both the short run and long run results from the structural estimation of the labor demand parameters. The parameter ϵ_s refers to the trade elasticity, α_s for the composite external economies of scale parameter as discussed in the theory section. Additionally, α_s^{LR} is reported, which is the corresponding scale elasticity if estimated for the 1914/1925 time frame instead of the 1914/1920 timeframe. The estimates are obtained via 2SLS instrumenting for employment size of sector s in province i, $L_{i,s}$ and wages $w_{i,s}$, using $logdistance_{i,Paris} \times log$(Employment Share of Sector in Province) as a first instrument, where log(Employment Share of Sector in Province) $\equiv \frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$ works as a labor supply shifter and is interacted with distance to Paris as a reduced form proxy for differences in geographical advantages vis-a-vis the French destination market. A second instrument is given by a Harris Market Potential measure for the input market leaving own size out as a labor supply shifter, constructed as $LMA_{i,s} = \sum_{j \neq i, r \neq s} \frac{1}{distance_{i,j}} L_{j,r}$. The first stages for the 1914/1920 estimates are reported separately in table 1.6. The estimation is obtained on the sample that drops the 1% smallest industries, thus avoiding large leverage of outliers on estimates due to small measurement error in employment sizes and wages. Standard errors are obtained via bootstrap.

Table 1.2: Structural Estimation: Naive OLS

Industry	α_s	T-Stats	ϵ_s	T-Stats
Agriculture	0.91***	(6.97)	−1.20***	(35.61)
Books	1.00***	(8.39)	−0.99***	(13.16)
Ceramics	1.07***	(9.08)	−0.59***	(6.95)
Chemicals	0.85***	(9.79)	−0.37***	(5.35)
Construction	1.18***	(11.89)	−1.34***	(23.71)
Decoration	1.10***	(12.71)	−1.07***	(16.33)
Electricity	0.90***	(8.12)	−0.61***	(6.99)
Food	1.15***	(10.28)	−1.29***	(24.86)
Forest	0.87***	(9.83)	−3.30***	(22.60)
Furniture	1.04***	(9.86)	−0.98***	(16.02)
Garments	1.09***	(8.09)	−1.20***	(16.36)
Glass	0.99***	(5.80)	−1.05***	(12.06)
Leather	1.03***	(13.11)	−0.93***	(13.77)
Metal works	1.04***	(10.69)	−1.08***	(16.20)
Metallurgy	0.90***	(14.32)	−0.03	(0.46)
Mines	1.13***	(23.13)	0.77***	(13.56)
Paper	1.22***	(4.56)	−1.24***	(14.34)
Public	0.99***	(5.31)	−1.81***	(11.19)
Public Industry	0.61***	(6.26)	−2.80***	(25.83)
Textiles	1.06***	(9.65)	−1.24***	(17.07)
Tobacco	1.02***	(8.02)	−1.29***	(14.60)
Transport	1.06***	(10.10)	−1.06***	(15.39)
variants	0.98***	(3.61)	−0.89***	(6.91)
Wood	0.97***	(10.07)	−1.29***	(22.04)
Observations		625		
R2		0.8819		
Province FE		✓		

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table reports the results of estimating the structural equation 1.4 without correcting for the endogeneity of wages and employment size of a sector. The estimation procedure is OLS. The dependent variable lprice refers to log $\frac{p_{is,1920}}{p_{is,1914}}$ and the explanatory variables lwage and llabor refer to $\frac{w_{is,1920}}{w_{is,1914}}$ and $\frac{L_{is,1920}}{L_{is,1914}}$ respectively.

Table 1.3: Diff in Diff

	(1)	
	lworkers	
ldist	9.964***	(8.61)
treated	-0.0397	(-0.04)
treated=1 × ldist	0.00262	(0.02)
treated=1 × lexp	0.0686*	(2.41)
lexp × year_count	-0.00111	(-0.22)
Constant	-59.88***	(-7.45)
Observations	144	
Province FE	✓	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table reports the results of the diff-in-diff regression described in section

Table 1.4: Pre Trends

	(1)	
	lwage_growth	
ldist	-0.187	(-1.76)
lshare	-0.0140	(-0.83)
Constant	1.336	(1.77)
Observations	144	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table reports the results of a regression Source: Labor inspections (1909-1914)

Table 1.5: Spatial Gradient: Sectoral Estimates

	$\frac{Y_{i,s,1920}}{Y_{i,s,1910}}$	
Agriculture × log(DistanceParis)	-0.619***	(-3.49)
Books × log(DistanceParis)	-0.754***	(-4.01)
Ceramics × log(DistanceParis)	-0.671***	(-3.55)
Chemicals × log(DistanceParis)	-0.647***	(-3.44)
Construction × log(DistanceParis)	-0.650***	(-3.59)
Decoration × log(DistanceParis)	-0.717***	(-3.76)
Electricity × log(DistanceParis)	-0.695***	(-3.66)
Food × log(DistanceParis)	-0.671***	(-3.70)
Forest × log(DistanceParis)	-0.793***	(-4.20)
Furniture × log(DistanceParis)	-0.737***	(-3.90)
Garments × log(DistanceParis)	-0.669***	(-3.70)
Glass × log(DistanceParis)	-0.723***	(-3.70)
Leather × log(DistanceParis)	-0.697***	(-3.70)
Metal Works × log(DistanceParis)	-0.675***	(-3.71)
Metallurgy × log(DistanceParis)	-0.668***	(-3.57)
Mines × log(DistanceParis)	-0.653***	(-3.60)
Paper × log(DistanceParis)	-0.700***	(-3.58)
Public × log(DistanceParis)	-0.735***	(-3.66)
Public Industry × log(DistanceParis)	-0.713***	(-3.64)
Textiles × log(DistanceParis)	-0.678***	(-3.72)
Tobacco × log(DistanceParis)	-0.802***	(-4.19)
Transport × log(DistanceParis)	-0.664***	(-3.66)
Varias × log(DistanceParis)	-0.737***	(-3.98)
Wood × log(DistanceParis)	-0.684***	(-3.73)
log(ShareInSector)	0.0624	(1.02)
log(ShareInProvince)	-0.218**	(-2.79)
Constant	6.741***	(5.18)
Observations	685	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table reports the results of a regression correlating nominal income growth between 1920 and 1910 at the sector province level with the (log) distance from the provincial capital to Paris. The distance measure is the shortest path along the railroad network and maritime linkages in kilometers. The regression allows for different intercepts for each sector. Additionally, the regression controls for the (log) employment share of sector s in province i in the national industry as a proxy for comparative advantage, as well as the (log) employment share of the sector within the province as a proxy for local labor market tightness.

Table 1.6: Labor Demand estimation: First stage

	(1) $\log \frac{w_{is,1920}}{w_{is,1914}}$		(2) $\log \frac{L_{is,1920}}{L_{is,1914}}$	
log(DistancetoParis) x log(ShareinProvince)	0.00569***	(4.81)	-0.0133***	(-8.86)
Log(LMA)	-0.0159	(-1.71)	0.0488***	(4.13)
Constant	0.827***	(11.55)	-0.386***	(-4.24)
Observations	657		657	
F Stat	15.53		46.08	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table reports the results of the first stage for estimating the structural equation 1.4. The first stage predicts the endogenous variables $\log \frac{w_{is,1920}}{w_{is,1914}}$, denoting (log) wage changes between 1920 and 1914 at the province-sector level, and $\log \frac{L_{is,1920}}{L_{is,1914}}$, denoting employment changes for the same time period at the province sector level. The first instrument is $logdistance_{i,Paris} \times log$(Employment Share of Sector in Province), where log(Employment Share of Sector in Province) $\equiv \frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$ works as a labor supply shifter and is interacted with distance to Paris as a reduced form proxy for differences in geographical advantages vis-a-vis the French destination market. A second instrument is given by a Harris Market Potential measure for the input market leaving own size out as a labor supply shifter, constructed as $LMA_{i,s} = \sum_{j \neq i, r \neq s} \frac{1}{distance_{i,j}} L_{j,r}$.

Table 1.7: Local Labor Supply and Income Dynamics

	$\frac{Y_{i,s,1920}}{Y_{i,s,1910}}$	
log(ShareInSector)	-0.0629	(-1.30)
log(ShareInProvince)	-0.173*	(-2.25)
log(EmploymentSize)	0.0933	(1.10)
Constant	0.860	(0.88)
Observations	637	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table reports the results from a regression of nominal income growth between 1920 and 1910 at the sector-province level on three different variables. log(Employment Share of Sector in Province) $\equiv \frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$

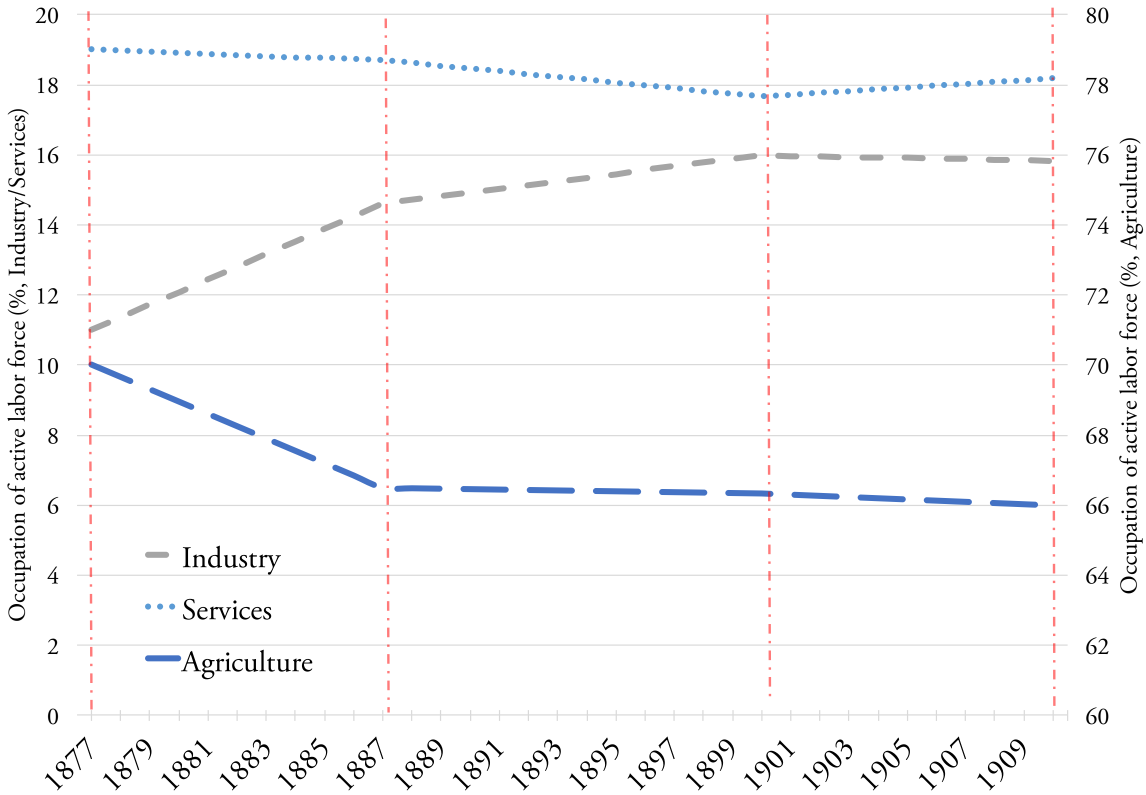
Table 1.8: Results: Migration Cost Estimation

Province	ρ_i	$log\zeta_i^1$	$\frac{1}{\mu_{agriculture,i}}$	Industries	$\frac{1}{\mu_s}$
Alava	1.09	0.92	0.03	Agriculture	0.2576
Albacete	0.20	3.30	0.00	Books	0.0452
Alicante	0.64	1.62	0.00	Ceramics	0.0601
Almeria	0.88	0.89	0.01	Chemicals	0.3394
Avila	0.65	1.95	0.01	Construction	0.0770
Badajoz	0.71	2.07	0.00	Decoration	0.3176
Baleares	0.00	11.07	0.06	Electricity	0.0793
Barcelona	10.01	0.18	0.04	Food	0.2611
Burgos	1.25	1.36	0.01	Forrest	0.0852
Caceres	1.11	2.18	0.00	Furniture	0.3219
Cadiz	0.50	1.91	0.02	Garments	0.0642
Castellon	1.00	1.34	0.01	Glass	0.3366
Ciudad Real	1.18	2.31	0.00	Leather	0.6854
Cordoba	1.07	1.80	0.00	Metal Works	0.0001
Coruna	1.60	1.21	0.00	Metallurgy	0.1664
Cuenca	0.69	2.06	0.00	Mines	0.1801
Gerona	1.80	1.21	0.00	Paper	0.3458
Granada	0.33	2.88	0.00	Public	0.4204
Guadalajara	1.05	2.49	0.00	Public Industry	0.3710
Guipuzcoa	1.52	0.63	0.01	Textiles	0.0714
Huelva	0.88	1.51	0.00	Tobacco	0.0922
Huesca	1.56	1.01	0.00	Transport	0.1703
Jaen	1.09	1.41	0.00	Varias	0.5858
Leon	0.95	1.79	0.02	Wood	0.0000
Lerida	1.09	1.45	0.00		
Logrono	1.03	1.02	0.01		
Lugo	1.13	1.60	0.01	Elasticities and Constants	
Madrid	5.57	0.36	0.04	ζ^2	1.49
Malaga	0.86	1.77	0.00	ν	1.59
Murcia	0.93	0.95	0.00	γ	1.35
Navarra	1.22	1.45	0.02	$log(\zeta_{cons})$	1.40
Orense	0.70	2.42	0.03		
Oviedo	0.79	2.06	0.02		
Palencia	0.59	1.79	0.03		
Pontevedra	1.63	1.21	0.02		
Salamanca	0.82	1.96	0.01		
Santander	0.83	0.75	0.02		
Segovia	0.90	2.02	0.00		
Sevilla	2.00	0.92	0.00		
Soria	1.07	1.31	0.01		
Tarragona	1.35	1.56	0.00		
Teruel	1.04	1.31	0.01		
Toledo	0.80	2.42	0.00		
Valencia	0.95	1.75	0.00		
Valladolid	0.98	1.31	0.03		
Vizcaya	1.09	0.67	0.04		
Zamora	0.64	2.18	0.01		
Zaragoza	0.39	2.93	0.01		

Notes: This table reports the results of the migration cost estimation. In the left column the amenity shifters associated with the different provinces are reported. Barcelona is normalized to 1, with the other provinces being expressed relatively to Barcelona. In the right column the sectoral switching cost parameter μ_s is reported as well as the key elasticities pinning down spatial migration cost $\mu_{ij} = \zeta_{cons} \times \zeta_i^1 \times distance_{ij}^{\zeta^2}$. The parameters are obtained via minimum distance estimation and the procedure is described in detail in section **??**.

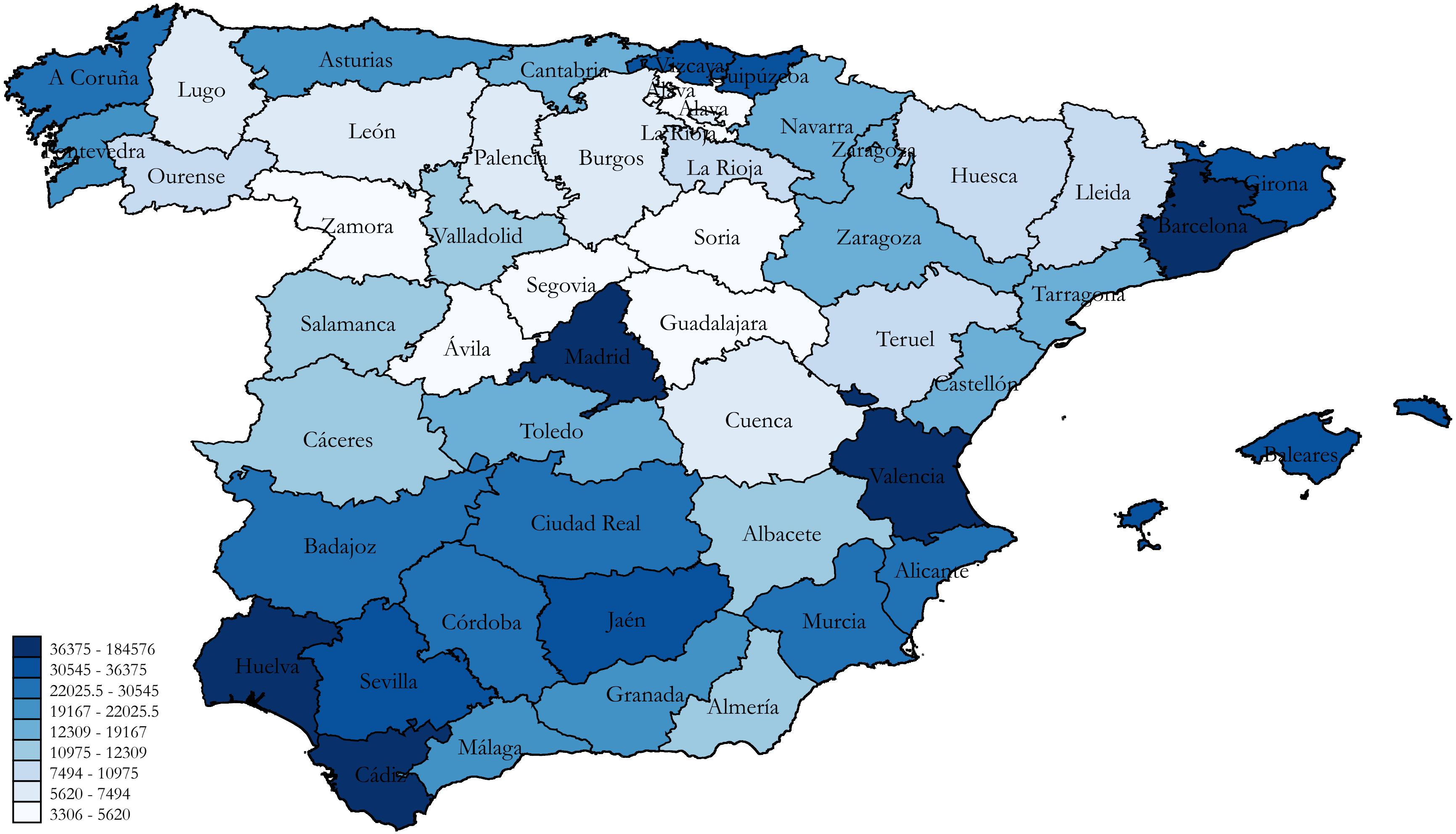
1.11 Figures

Figure 1.6: Structural Change in the 19th Century



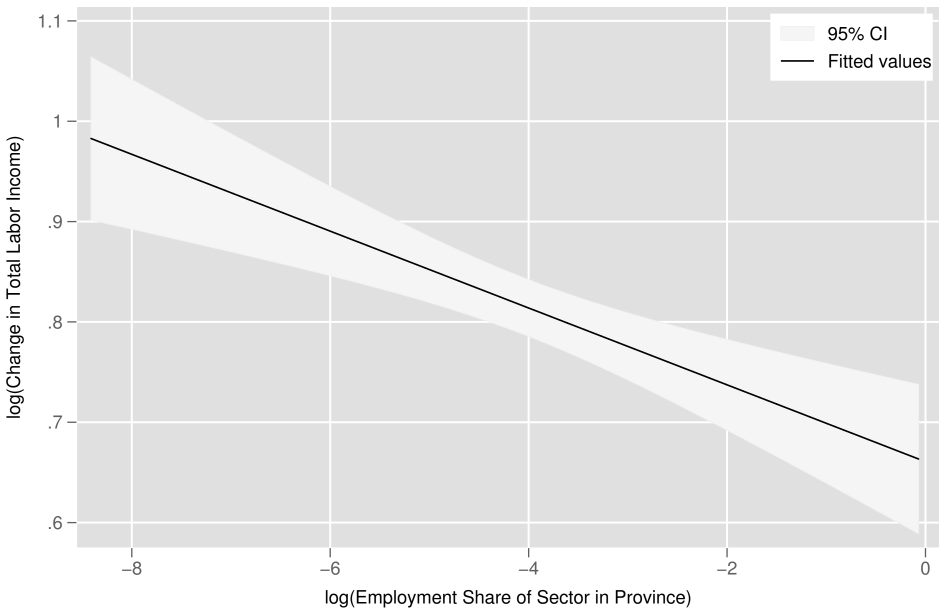
Notes: The figure depicts sectoral employment shares across the manufacturing sector/industry, agriculture and services. The shares are observed in the census data in 1877, 1887, 1900 and 1910 where census years are indicated by the red dotted line and the intervening years are imputed trend lines. Notice that while service and industry employment is plotted against the left y-axis, agricultural employment is plotted against right y-axis. The original computation of the aggregate employment share is due to Harrison (1978).

Figure 1.7: Spatial Distribution of Manufacturing Employment



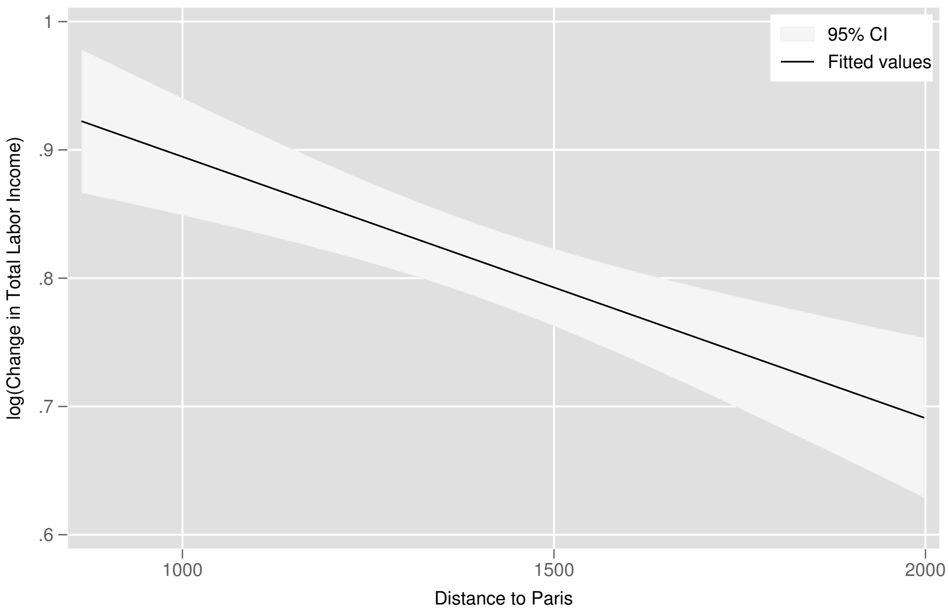
Notes: The map depicts total Manufacturing and Mining employment in the provinces in 1910 across Spain (without Canary Islands and North African possessions). The map is a chloropleth with darker shaded colors depicting higher absolute numbers. The data is obtained from the Population census from 1910.

Figure 1.8: Local Labor Supply and Income Growth



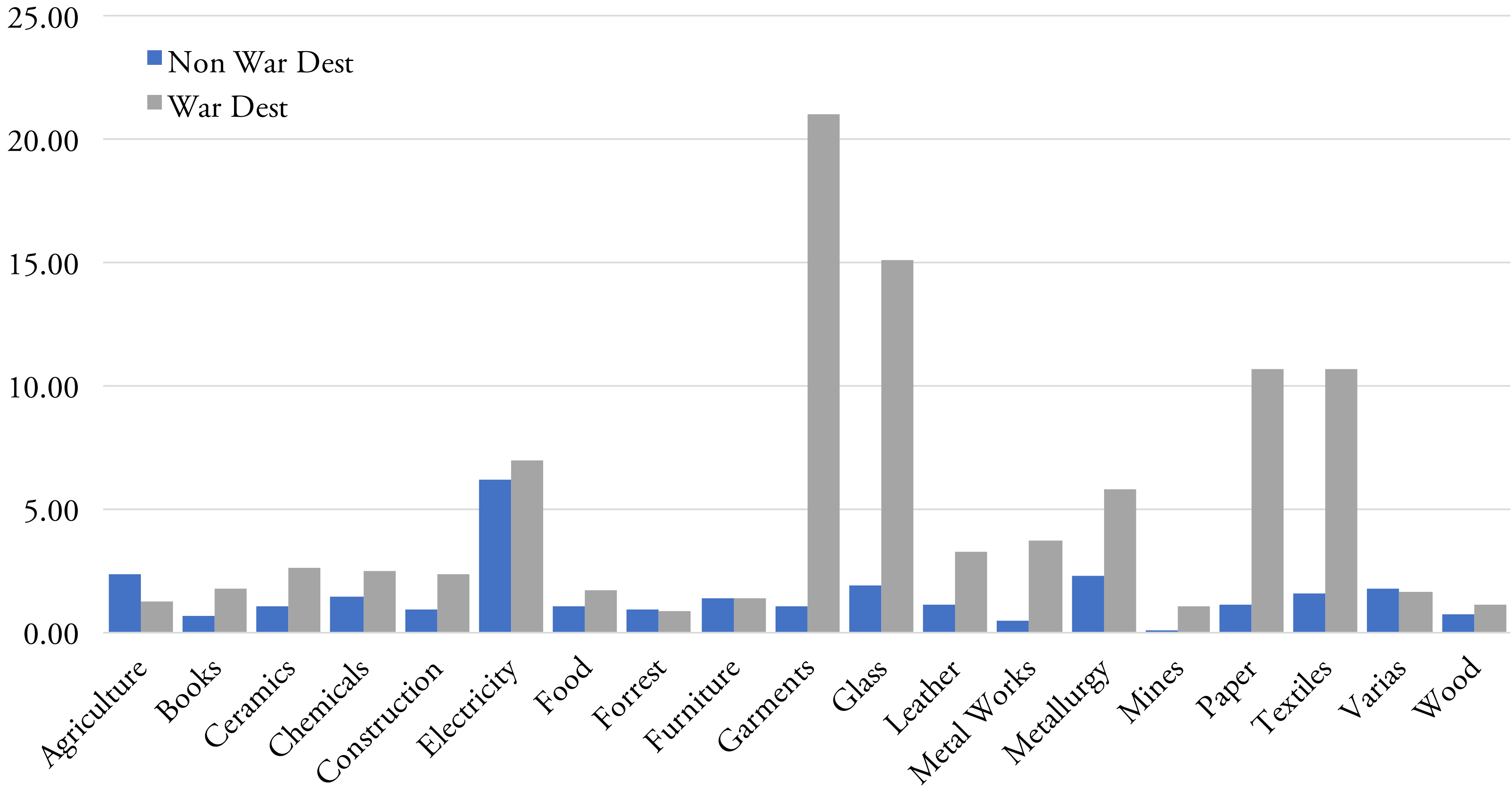
Notes: The graph shows the fitted line of a regression correlating (nominal) income growth at the sector province level between 1920 and 1914 with the log of the share of that sector in the total employed population in that province in 1914. Specifically, the variable on the x-axis is defined as log(Employment Share of Sector in Province) $\equiv \frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$. The data being used is the labor market panel introduced in the data section.

Figure 1.9: Spatial Gradient in Income Growth



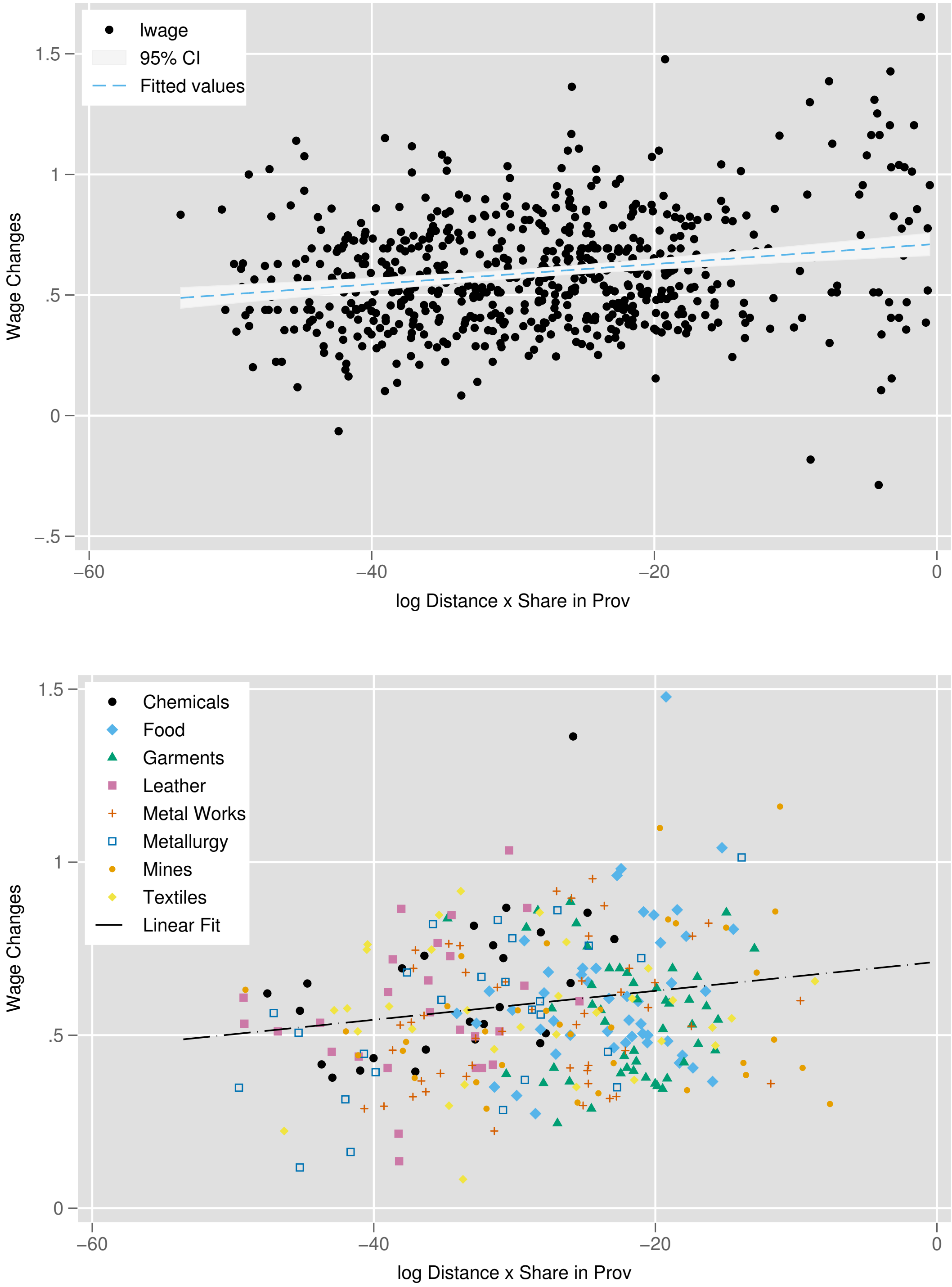
Notes: This figure reports the results of a regression correlating nominal income growth between 1920 and 1910 at the sector province level with the (log) distance from the provincial capital to Paris. The distance measure is the shortest path along the railroad network and maritime linkages in kilometers.

Figure 1.10: Sectoral Trade Growth: Belligerent vs Non Belligerent



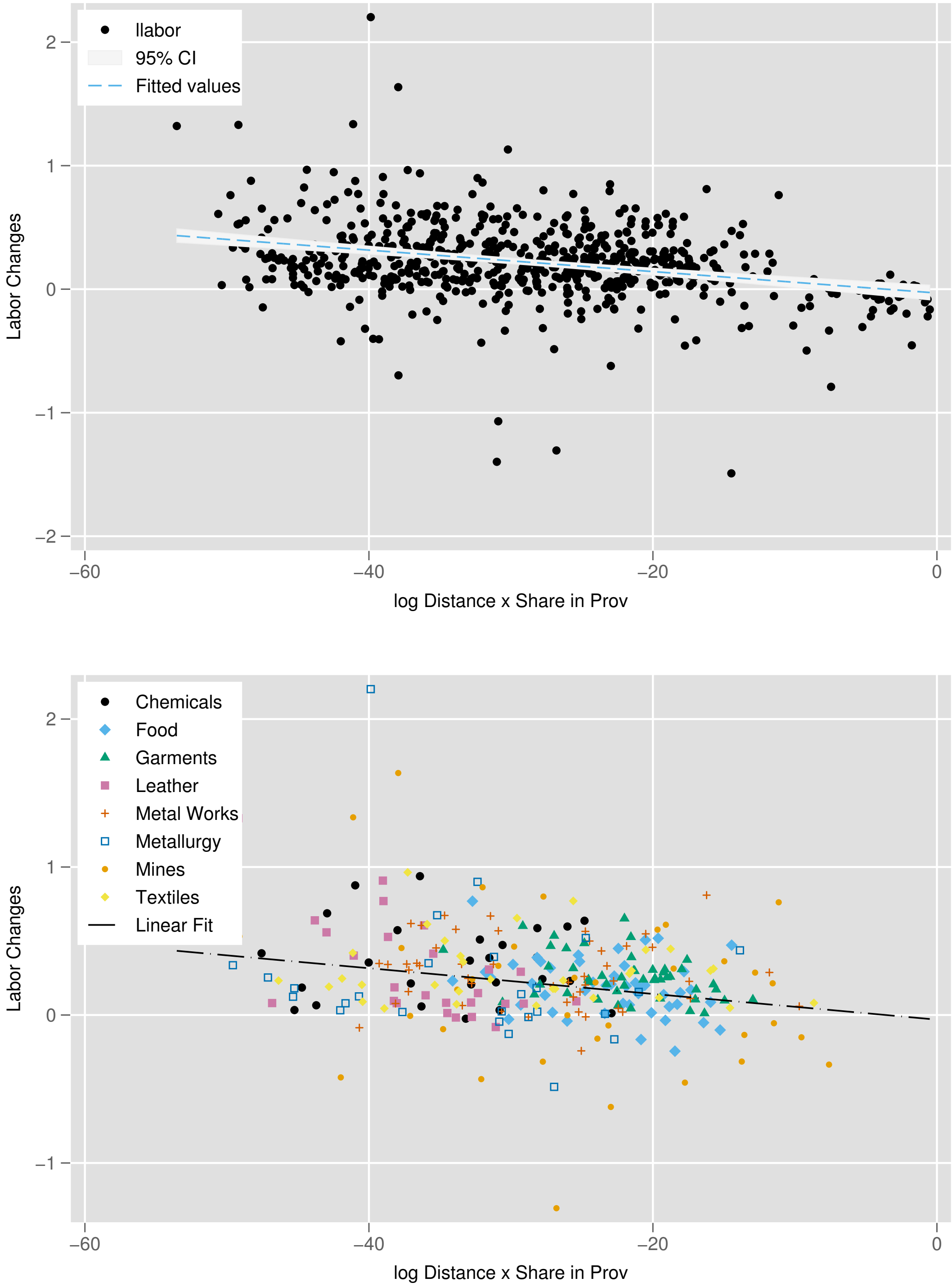
Notes: This figure reports the sectoral export growth for non belligerent destination countries - in blue - and belligerent destination countries in grey. The product level trade has been aggregated to sector level trade data to match the level of aggregation of the labor market panel. Growth rates are constructed by comparing the 1910 benchmark with average export values in 1915 and 1916, that is $g_X^{War} \equiv \frac{1/2 X_{Spain,War,1915} + X_{Spain,War,1916}}{X_{Spain,War,1910}}$ and correspondingly for non belligerent destinations. As discussed in section 1.5 I abstract from later years to avoid additional spatial frictions that perturbed international trade, in particular increased maritime warfare. To adjust for additional spatial disruptions of the frontline the belligerent countries are made up of France, Italy and the United Kingdom. The Non-belligerent countries exclude the United States and other later participants of WWI. The shock is being calculated using the official annual trade data in constant prices.

Figure 1.11: First stage for structural estimation (all industries)



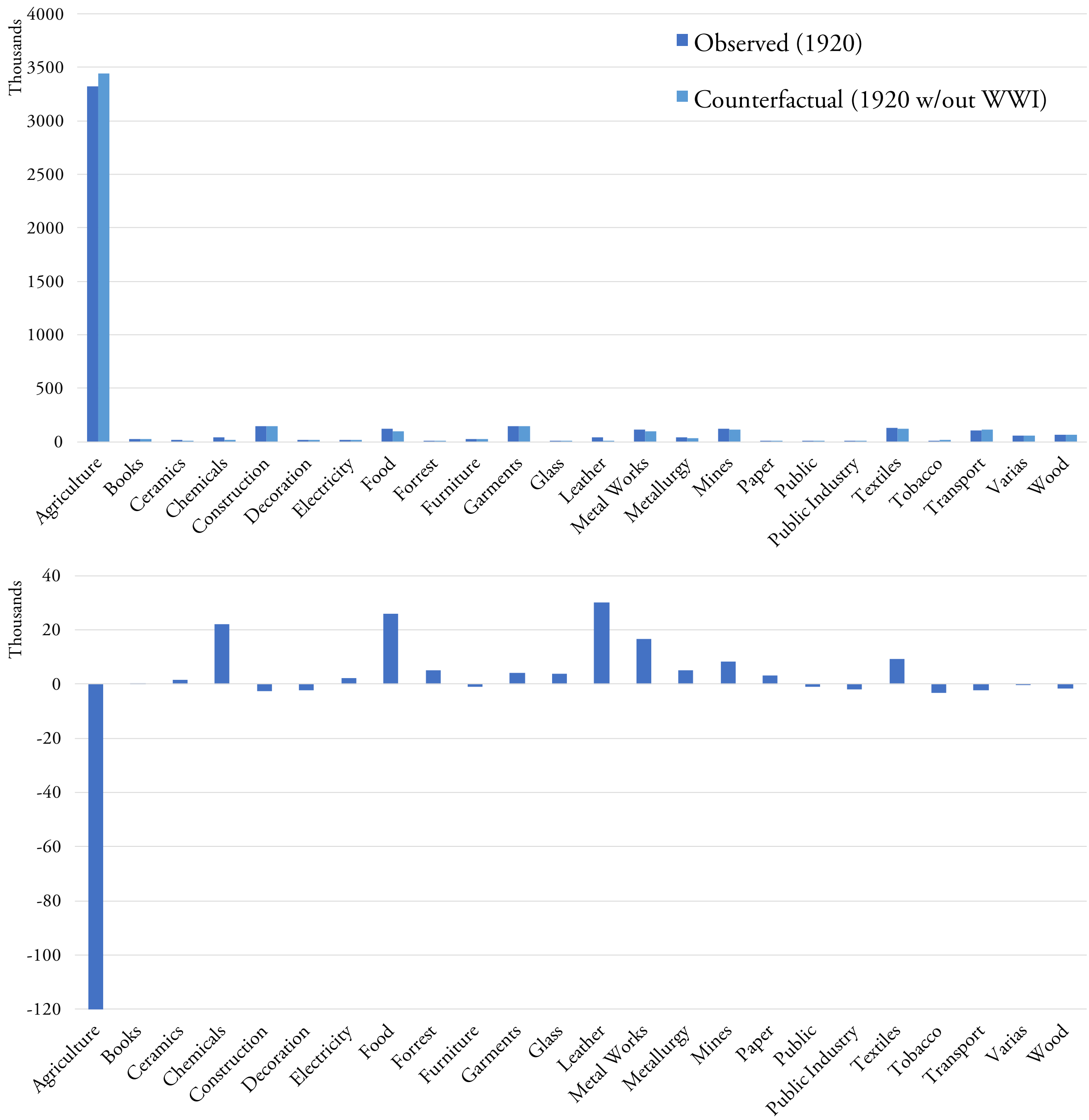
Notes: This figure shows the first stage regression predicting wage changes, $\frac{w_{i,s,1920}}{w_{i,s,1914}}$ at the province-sector level, using $logdistance_{i,Paris} \times log$(Employment Share of Sector in Province) as an instrument, where log(Employment Share of Sector in Province) $\equiv \frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$ works as a labor supply shifter and is interacted with distance to Paris as a reduced form proxy for differences in geographical advantages vis-a-vis the French destination market. Distance is calculated using the shortest path along a network of railroads and maritime linkages between province capitals in Spain and Paris in France. The figure depicts all industries. The data being used is the labor market panel introduced in section 1.4.

Figure 1.12: First stage for structural estimation (selected industries)



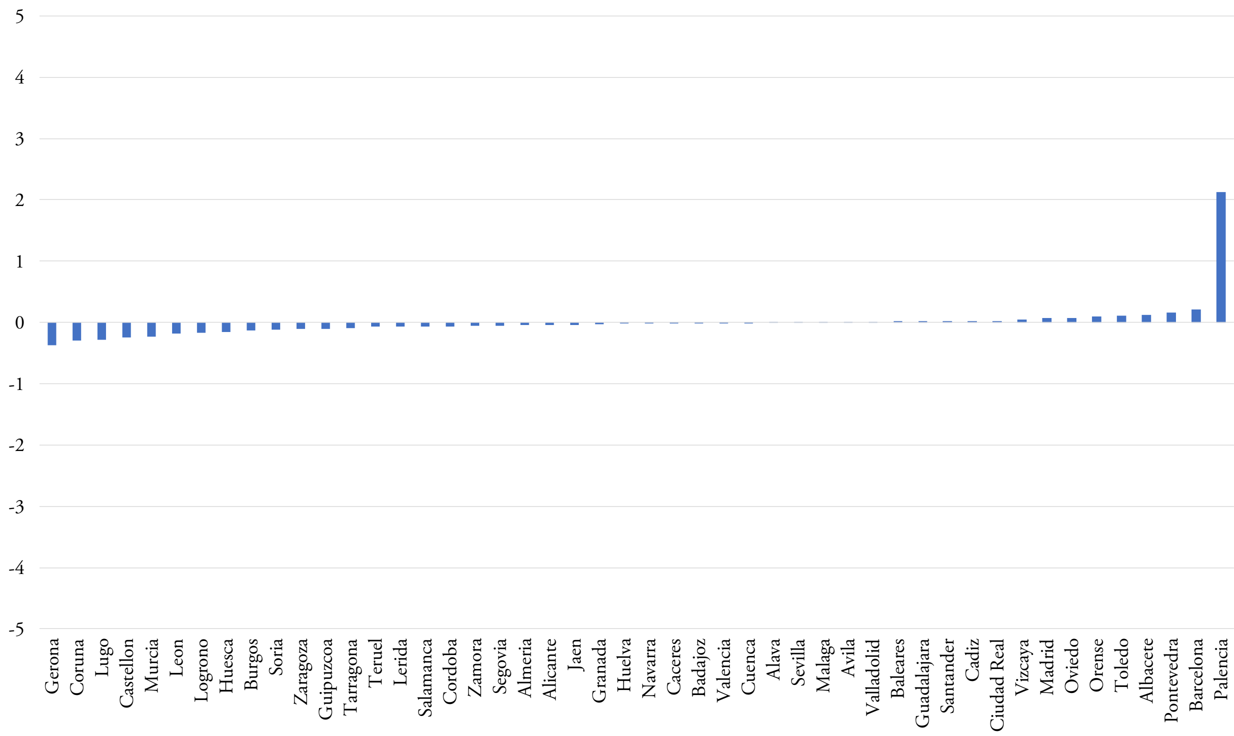
*Notes:*This figure shows the first stage regression predicting changes in employment size, $\frac{L_{i,s,1920}}{L_{i,s,1914}}$ at the province-sector level, using $logdistance_{i,Paris} \times log$(Employment Share of Sector in Province) as an instrument, where log(Employment Share of Sector in Province) $\equiv \frac{L_{i,s,1914}}{\sum_r L_{i,r,1914}}$ works as a labor supply shifter and is interacted with distance to Paris as a reduced form proxy for differences in geographical advantages vis-a-vis the French destination market. Distance is calculated using the shortest path along a network of railroads and maritime linkages between province capitals in Spain and Paris in France. The figure depicts all industries. The data being used is the labor market panel introduced in section 1.4.

Figure 1.13: Counterfactual: National industry size (Employment levels and differences, 1920)



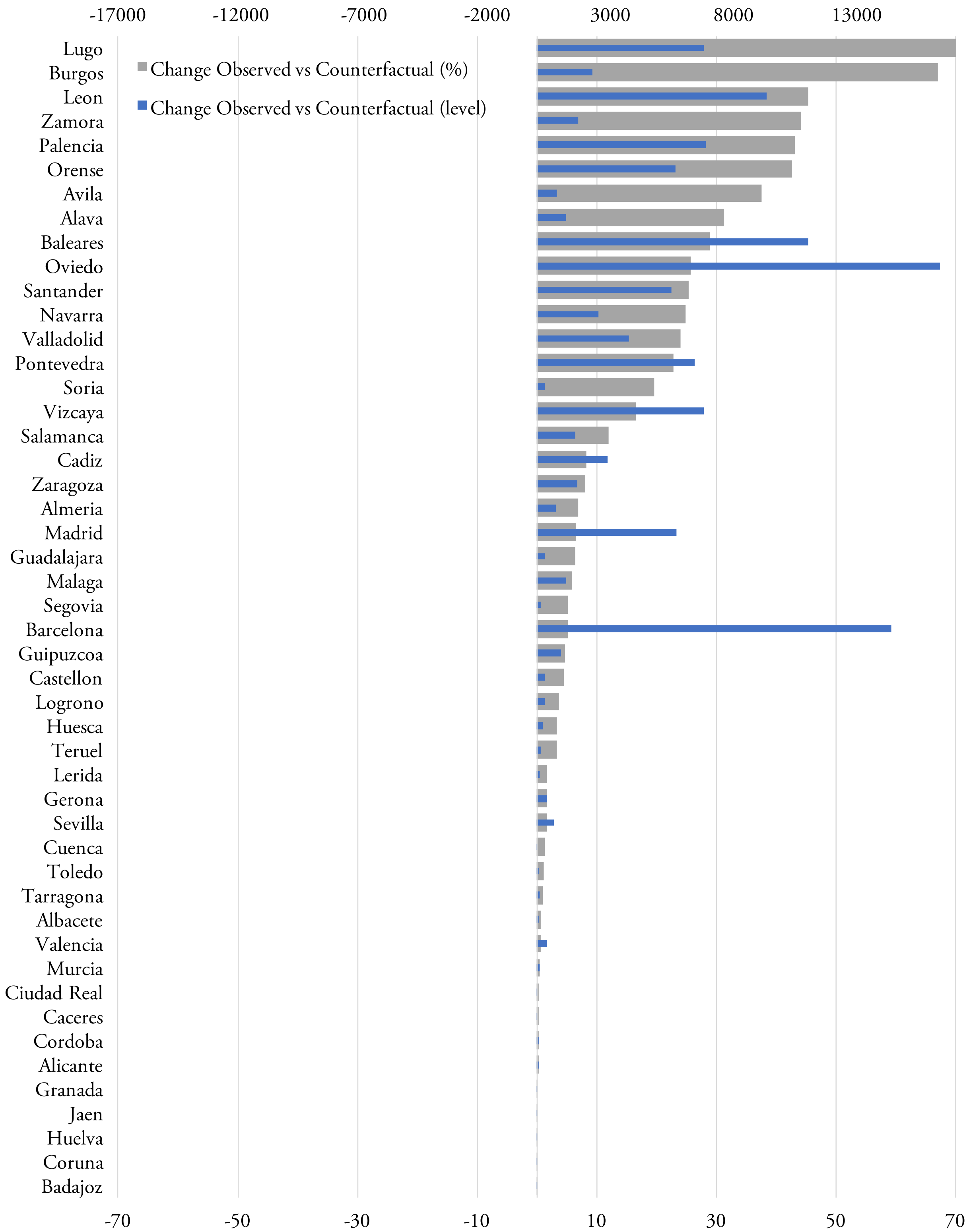
Notes: The upper graph depicts the aggregate sectoral employment for the observed data and the counterfactual simulation of Spain in the absence of WWI. The values are constructed in the following way, $L_s \equiv \sum_i L_{i,s,1920}$ where $L_{i,s,1920}$ refers to the observed employment size in province i and sector s. Similarly for the counterfactual, $L_s^{CF} \equiv \sum_i L_{i,s,1920}^{CF}$ where $L_{i,s,1920}^{CF}$ is the simulated counterfactual sectoral employment size using the estimated model as described in section 1.8. The lower graph shows the same figure in terms of difference between the counterfactual and observed data.

Figure 1.14: Counterfactual: Province size (Manufacturing and Agricultural Employment, 1920)



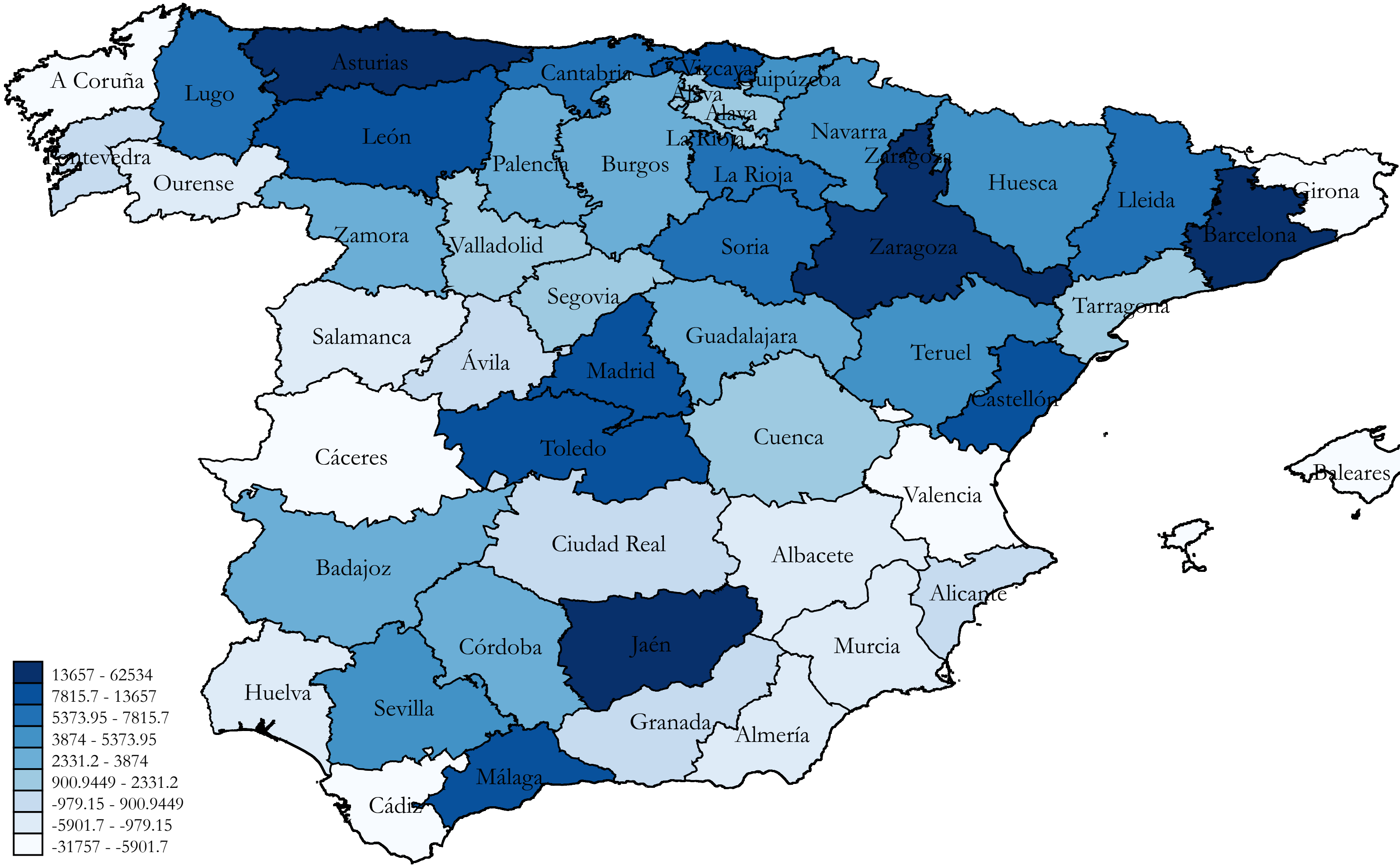
Notes: The figure depicts the change in aggregate provincial employment for the observed data and the counterfactual simulation of Spain in the absence of WWI. The values are constructed in the following way, $L_s \equiv \sum_s L_{i,s,1920}$ where $L_{i,s,1920}$ refers to the observed employment size in province i and sector s. Similarly for the counterfactual, $L_s^{CF} \equiv \sum_s L_{i,s,1920}^{CF}$ where $L_{i,s,1920}^{CF}$ is the simulated counterfactual sectoral employment size using the estimated model as described in section 1.8. Relative changes are indicated and calculated using those variables

Figure 1.15: Counterfactual: Manufacturing Employment (1920)



Notes: The figure depicts the change in manufacturing employment aggregated at the provincial level between the observed data and the counterfactual simulation of Spain in the absence of WWI. The values are constructed in the following way, $L_s \equiv \sum_s L_{i,s,1920}$ where $L_{i,s,1920}$ refers to the observed employment size in province i and sector s. Similarly for the counterfactual, $L_s^{CF} \equiv \sum_s L_{i,s,1920}^{CF}$ where $L_{i,s,1920}^{CF}$ is the simulated counterfactual sectoral employment size using the estimated model as described in section 1.8. Absolute and relative changes are indicated and calculated using those variables. The upper axis gives the relevant scale for absolute changes, while the lower axis gives the relevant scale for relative changes.

Figure 1.16: Counterfactual: Nominal Income Gains (1920)



Notes: The map depicts the provincial differences in nominal income in the observed data compared to the counterfactual without the war shock. The values for each province is calculated by first computing the nominal wages at the province level, that is $Y_{i,1920} = \sum_r w_{i,r,1920} L_{i,r,1920}$ and sbustracting the income in the counterfactual scenario, $Y^{CF}_{i,1920} = \sum_r w^{CF}_{i,r,1920} L^{CF}_{i,r,1920}$.

1.12 Data Sources

- *Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1910* (Instituto Geográfico; 1912)
  - This publication contains population data disaggregated by profession for each province of Spain in 1910.
- *Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1920* (Instituto Geográfico; 1922)
  - This publication contains population data disaggregated by profession for each province of Spain in 1920.
  - Furthermore, it also contains data on the origin of residents in each province that were born in another province.
- *Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1930* (Instituto Geográfico; 1932)
  - This publication contains population data disaggregated by profession for each province of Spain in 1930.
  - Furthermore, it also contains data on the origin of residents in each province that were born in another province.
- *Estadística general del comercio exterior de España con sus posesiones de ultramar y potencias extranjeras* (de Aduanas; 1910-1930)
  - This publication contains trade records decomposed along destination countries and product type.
- *Estadistica de salarios y jornadas de trabajo referida al periodo 1914-1925* (Ministerio de Trabajo; 1927)
  - This publication contains wage and quantity data by profession between for 1914, 1920 and 1925
- *Clasificación general de industrias, oficios y comercios 1931* (Instituto Nacional de Prevision Social; 1930)
  - This publication contains the official correspondence between industries and occupations.

1.13 Derivations

1.13.1 Proof of Proposition 1

Let J be the number of alternatives. Depending on the values of the vector $\boldsymbol{\epsilon} = \{\epsilon_1, \ldots, \epsilon_J\}$ the function $\max_i(\delta_i \times \epsilon_i)$ takes on different values. First, examine the case where $\max_i(\delta_i \times \epsilon_i) = \delta_1 \times \epsilon_1$. That is, we will integrate $\delta_1 \times \epsilon_1$ over the set $M_1 \equiv \{\boldsymbol{\epsilon} : \delta_1 \times \epsilon_1 > \delta_j \times \epsilon_j, j \neq i\}$:

$$
\begin{aligned}
&E_{\boldsymbol{\epsilon}\in M_1}[\max_i(\delta_i \times \epsilon_i)] = \\
&\int_{-\infty}^{\infty} (\delta_1 \times \epsilon_1) f(\epsilon_1) \left[\int_{-\infty}^{\frac{\delta_1\times\epsilon_1}{\delta_2}} \ldots \int_{-\infty}^{\frac{\delta_1\times\epsilon_1}{\delta_J}} f(\epsilon_2)\ldots f(\epsilon_J) d\epsilon_2 \ldots d\epsilon_J \right] d\epsilon_1 = \\
&\quad \int_{-\infty}^{\infty} (\delta_1 \times \epsilon_1) f(\epsilon_1) \left(\int_{-\infty}^{\frac{\delta_1\times\epsilon_1}{\delta_2}} f(\epsilon_2) d\epsilon_2 \right) \ldots \left(\int_{-\infty}^{\frac{\delta_1\times\epsilon_1}{\delta_J}} f(\epsilon_J) d\epsilon_J \right) d\epsilon_1 = \\
&\qquad\qquad \int_{-\infty}^{\infty} (\delta_1 \times \epsilon_1) f(\epsilon_1) F\left(\frac{\delta_1 \times \epsilon_1}{\delta_2}\right) \ldots F\left(\frac{\delta_1 \times \epsilon_1}{\delta_J}\right) d\epsilon_1
\end{aligned}
\tag{1.5}
$$

The final term in the last equation is the first of J such terms in $E[\max_i\,(\delta_i \times \epsilon_i)]$. Specifically,

$$
E\left[\max_i\,(\delta_i \times \epsilon_i)\right] = \sum_i E_{\boldsymbol{\epsilon}\in M_i}\left[\max_i\,(\delta_i \times \epsilon_i)\right]. \tag{1.6}
$$

Now we apply the functional form of the Fréchet distribution, where the CDF is given by $F(x) = e^{-x^{-x}}$, and the PDF is given by $f(x) = ax^{-1-a}e^{-x^{-a}}$, where a is the dispersion parameter. This gives,

$$
\begin{aligned}
&E_{\boldsymbol{\epsilon}\in M_i}\left[\max_i\,(\delta_i \times \epsilon_i)\right] \\
&= \int_{-\infty}^{\infty} (\delta_i \times \epsilon_i) a\epsilon_i^{-a-1} e^{-\epsilon_i^{-a}} e^{-\left(\frac{\delta_i\epsilon_i}{\delta_2}\right)^{-a}} \ldots e^{-\left(\frac{\delta_i\epsilon_i}{\delta_J}\right)^{-a}} d\epsilon_i \\
&= \int_{-\infty}^{\infty} (\delta_i \times \epsilon_i) a\epsilon_i^{-(a+1)} \prod_j e^{-\left(\frac{\delta_i\epsilon_i}{\delta_j}\right)^{-a}} d\epsilon_i \\
&= \int_{-\infty}^{\infty} (\delta_i \times \epsilon_i) a\epsilon_i^{-(a+1)} \exp\left(\sum_j -\left(\frac{\delta_i\epsilon_i}{\delta_j}\right)^{-a}\right) d\epsilon_i \\
&= \int_{-\infty}^{\infty} (\delta_i \times \epsilon_i) a\epsilon_i^{-(a+1)} \exp\left(\epsilon_i^{-a} \times \sum_j -\left(\frac{\delta_i}{\delta_j}\right)^{-a}\right) d\epsilon_i
\end{aligned}
\tag{1.7}
$$

where the second step comes from collecting one of the exponentiated terms into the product, along with the fact that $\delta_j/\delta_i = 1$ if $i = j$. Now we define $D_i \equiv \sum_j \left(\frac{\delta_i}{\delta_j}\right)^{-a}$ and make the substitution $x = D_i\epsilon_i^{-a}$ so that $dx = -a\epsilon_i^{-a-1}D_i d\epsilon_i \Rightarrow -\frac{dx}{D_i} = a\epsilon_i^{-(a+1)}d\epsilon_i$ and $\epsilon_i = \left(\frac{x}{D_i}\right)^{-\frac{1}{a}}$. Note that as ϵ_i approaches infinite, x approaches 0, and as ϵ_i approaches negative infinity, x approaches infinity.

$$
\begin{aligned}
&E_{\epsilon\in M_i}\left[\max_i\left(\delta_i \times \epsilon_i\right)\right] = \\
&\int_{\infty}^{0}\left(\delta_i\left(\frac{x}{D_i}\right)^{-\frac{1}{a}}\right)\left(-\frac{1}{D_i}\right)\exp\left\{-x\right\}dx \\
=&\frac{1}{D_i}\int_{0}^{\infty}\left(\delta_i\left(\frac{x}{D_i}\right)^{-\frac{1}{a}}\right)e^{-x}dx
\end{aligned}
\tag{1.8}
$$

Recall that $D_i \equiv \sum_j \left(\frac{\delta_i}{\delta_j}\right)^{-a} = \frac{\sum_j \delta_j^a}{\delta_i^a}$. Notice that the familiar frechet choice probabilities $P_i = \frac{\delta_i^a}{\sum_j \delta_j^a}$ are inverses of the D_i's or in other words $P_i = 1/D_i$. Also note that $\sum_i P_i = 1$.

$$= P_i \int_0^{\infty}\left(\delta_i\left(\frac{x}{D_i}\right)^{-\frac{1}{a}}\right)e^{-x}dx$$

$$= P_i\ \delta_i D_i^{\frac{1}{a}}\int_0^{\infty} x^{\frac{1}{a}}e^{-x}dx$$

The Gamma function is defined as $\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x}dx$. This implies that the integral term is equal to $\Gamma\left(1-\frac{1}{a}\right)$. Furthermore, since $D_i^{1/a} = \frac{1}{\delta_i}\left(\sum_j \delta_j^a\right)^{\frac{1}{a}}$, we obtain,

$$= \left(\sum_j \delta_j^a\right)^{\frac{1}{a}}\Gamma\left(1-\frac{1}{a}\right) \times P_i$$

Finally summing over all alternatives,

$$\sum_i E_{\boldsymbol{\epsilon} \in M_i}\left[\max_i \left(\delta_i \times \epsilon_i\right)\right] = \left(\sum_j \delta_j^a\right)^{\frac{1}{a}} \Gamma\left(1 - \frac{1}{a}\right) \times \sum_i P_i = \left(\sum_j \delta_j^a\right)^{\frac{1}{a}} \Gamma\left(1 - \frac{1}{a}\right) \quad \square$$

Chapter 2

Regional Diversity and the Geography of Unemployment

François de Soyres and Simon Fuchs

2.1 Abstract

We introduce a new framework to evaluate the effects of regional diversification. We observe that in the presence of mobility frictions workers are exposed to local shocks and that in a multi-sectoral framework this induces a trade-off: Regions can specialize in their comparative advantage industries, but at the same time such specialization increases labor market risk due to sector specific shocks. If mobility costs are high, then welfare effects from lack of diversification can be substantial. We measure the segmentation of the French labor market and introduce a new spatial equilibrium model that incorporates labor market frictions, unemployment, and mobility cost into an otherwise standard multi-sector economic geography model. We employ the model to simulate unemployment responses to sector specific shock and demonstrate the interaction between mobility frictions and matching frictions.

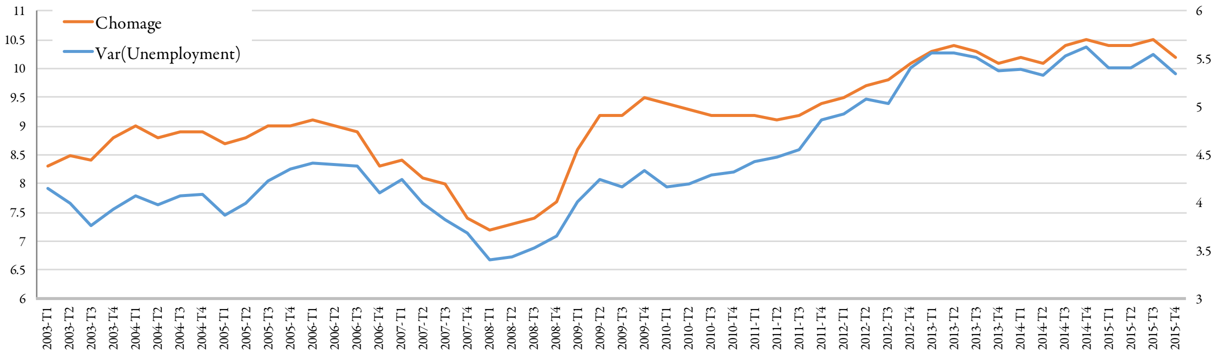
2.2 Introduction

With the increasing importance of China and other emerging economies, developed countries are facing more competition in international markets. High degrees of regional specialization which prior might have been optimal, now might subject workers and whole regions to the risk of individual sectors losing out in international markets, inducing unemployment risk, and protracted local adjustments. At a regional level these new development might shift the focus to an important trade-off: Either a region specializes to exploit comparative advantage, scale economies and foreign market opportunities or a region diversifies its sectoral base to provide insurance against sector specific shocks in the form of a more diversified labor market. This begs the question: How to characterize, evaluate and quantify this trade-off in the data?

Our argument in this paper is that in order to evaluate this question labor mobility and geography should be central. Consider the following hypothetical example: An engineer is working in Toulouse for Airbus on engines for the narrowbody twin engine aircraft A320 neo. Due to the surprisingly successful launch of the Comac C919 - a Chinese aerospace manufacturer - Airbus begins to anticipate less orders and downsizes its engineering team - the worker is being let go. If he is mobile he can easily move away from Toulouse and find a job elsewhere. What matters here is the geography of his occupation-specific job market. For example he might be able to find another job at the Rolls-Royce plant in Dahlewitz, Germany to produce airplane engines there, but this would require a costly - not just in monetary terms - distant reallocation. If he is not mobile then he might have to look for other work locally and if furthermore Airbus or Airbus suppliers are the only or largest employers for engineers in Toulouse then the likelihood of finding employment as an engineer in Toulouse again might be low and the worker might have to incur a wage cut by transitioning to a different occupation where his human capital level is effectively lower. To make things worse, his likelihood of finding employment might be worse if he is at the same time competing with other - freshly unemployed - engineers from his team at Airbus. What matters for the welfare of the worker is therefore the worker's mobility, the geography of his alternative job markets, the transition cost to other occupations and regions and the local alternatives. What is needed is therefore a tractable model that takes these forces into account and helps us to quantify the different magnitudes (unemployment probabilities across geographically and occupationally distinct labor markets and mobility costs).

A second motivation for such a model is the observed persistent increase in spatial variance of unemployment during the great recession both in France (see figure 2.1) and

Figure 2.1: Spatial Variance during Great Recession in France



Notes: The graph indicates the national unemployment rate as well as the (spatial) variance of unemployment across departments between 2003-2015. The source of the data is INSEE.

in the US (Yagan; 2017): This is difficult to reconcile with economic geography models that abstract from impediments to labor mobility, but is an important phenomenon with potentially large welfare effects.

In this paper we introduce a new theoretical framework that builds on standard multi-sectoral economic geography model, but adds an enriched descriptions of the labor market to capture the forces described above. We introduce three particular features: Firstly, we break the traditional symmetry where firms produce across sectors and labor markets are segmented across sectors as well. Instead, firms produce across sectors and hire workers across different occupationally segregated labor markets. This allows us to characterize several things: (1) The heterogeneous workforce across sectors, (2) the degree to which particular occupation-region specific labor markets depend on (undiversified) sectoral demand, (3) a meaningful mobility choice on the worker side between occupation-region specific labor markets with heterogeneous demand conditions and (4) mobility cost between different occupational labor markets (which potentially represent human capital losses). This feature also allows us to incorporate in our strucutral model mobility costs that we can estimate from rich micro-level French panel data.

Secondly, we introduce matching frictions. Sectors hire across occupation specific local labor markets by posting vacancies and job seeking workers are being matched up subject to matching frictions as represented by a matching function. This induces search cost on the firm site and therefore a wedge between the market clearing wage in the absence of matching frictions and the employment cost of the worker, resulting in involuntary unemployment. The model is kept tractable by introducing non-persistent in the matches, i.e. matches will be separated at the end of each period. While this limits the extent to which the model can be used to analyze the persistence of temporary shocks (i.e. persistence of frictional unemployment) it increases the tractability substantially and still allows us to employ the model to analyse transitions between steady states as well as the interaction between labor mobility and frictional unemployment. Matching frictions and the associated congestion effects in the local labor market can then capture to what extent the size of the shock (and therefore the lack of diversification) might matter.

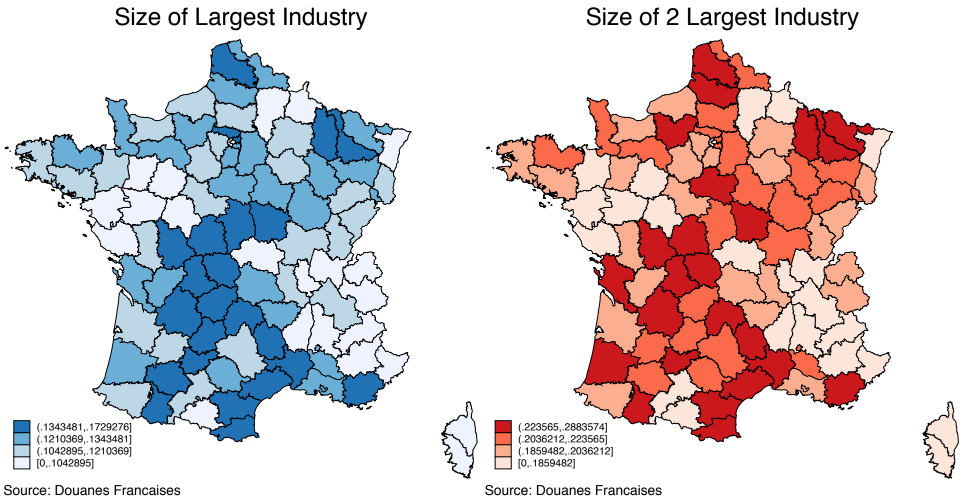
Thirdly, we introduce imperfect labor mobility. Workers can choose to relocate between occupation-region specific labor markets but relocation is subject to occupation specific and geographical switching costs. As mentioned before, occupation specific switching costs might be related to the loss of human capital, while geographical switching cost might capture psychic, informational and monetary switching cost associated to relocation between regions. The framework can be connected to switching costs estimated using dynamic estimation techniques on rich panel data and can therefore provide a connecting point between a macroeconomic structural framework and detailed micro-level administrative data. We show how to extent the current approach of estimating occupational switching costs to incorporate geographical switching costs as well. We work towards using the estimated parameters in our simulations.

Finally it is worth noting that mobility costs interact with the matching friction: Higher mobility costs means more locally unemployed workers given the size of a separation shock and the congestion effect in the local labor market increases the unemployment response and welfare cost, therefore the total welfare effects of a shock are determined by *the geography of its unemployment response*.

The paper connects to three separate strands in the literature: Firstly, ever since the fundamental contributions of Henderson (1974) and Fujita and Ogawa (1982) the economic geography literature has had at its center what has been dubbed the 'fundamental trade-off in spatial economics' that is the trade off between agglomeration benefits from increased city size compared to higher congestion costs which in tandem determine the optimal city size. The current analysis introduces an additional dimension to be considered that is the degree of sectoral diversification at the city or region level and the implied unemployment risk of higher specialization.

Secondly, there is also an ongoing discussion on how to define and quantify economic resilience (cp. for example Martin and Sunley (2015)). We provide a framework to evaluate economic resilience using the latest generation of quantative economic geography models and utilizing a large array of micro-level data.

Figure 2.2: Specialization Patterns in France



Notes: The two maps indicate the specialization patterns across French departments. On the left hand side the chloropleth for the size of the largest industry is presented and on the right hand side the chloropleth for the two largest industries. The measure is calculated based on an imputed measure of human capital or effective labor units introduced later in this text. The source data is DADS.

Finally, the literature on international trade is increasingly moving away from making ad-hoc assumptions on labor mobility and instead moves towards measuring mobility costs more precisely. Our work is aiming towards closing the gap between improved estimation of labor market segmentation and quantitative models that can be used to understand the aggregate effects of local or external shocks. This type of work was recently survey by McLaren (2017). In comparison to the literature we offer the first endeavour to measure joint occupational, spatial and sectoral mobility costs and incorporate them into a rich quantative economic geography framework. Additionally, we are amongst the first to explicitly incorporate multiple sectors, unemployment and segregated labor markets into a unified framework.

The paper proceeds as follows: Section 2 gives a simple illustration of the argument for sectoral diversification. Section 3 measures introduces our technique for measuring occupational and geographical switching costs on French administrative data, effectively measuring the labor market segmentation in France. Section 4 introduces the theoretical framework as well as the calibration and simulation. Section 5 concludes.

Figure 2.3: Optimal Sector Size (Constant MP)

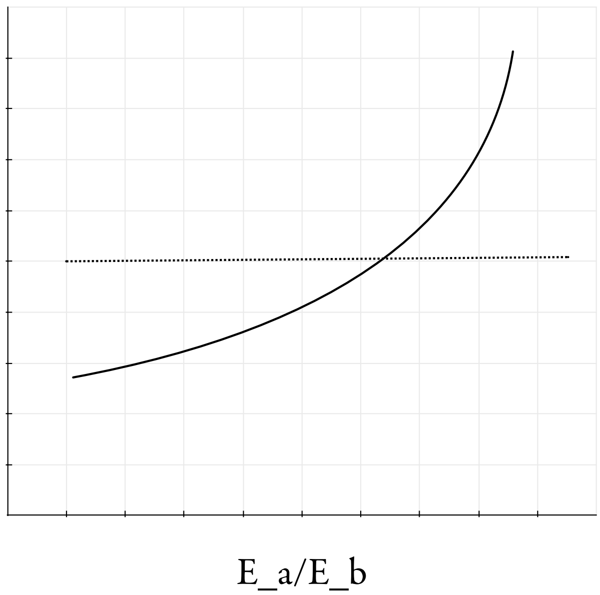
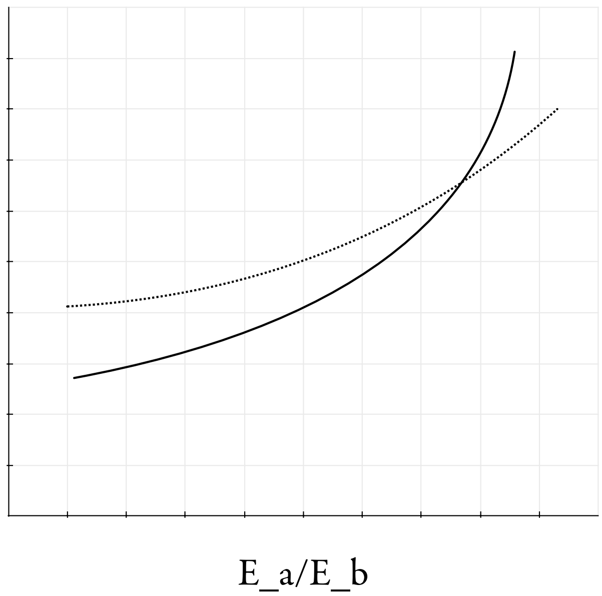


Figure 2.4: Optimal Sector Size (Inc MP)



Notes: The two graphs depict the Social Planner's trade-off: Increasing the relative size of sector (via relocating workers from E_b to E_a) imposes an increasing marginal cost (due to the congestion externality in the labor market) while providing constant marginal benefits in the Ricardian example (left graph) and increasing marginal benefits in an example where sector-level scale economies are present (right graph).

2.3 Insurance from Diversification: Simple Illustration

Consider a simple economy with two sectors a and b. Each sector uses currently employed labor, e_a, to produce a final consumption good using the production function, $f_a(\cdot)$, which is assumed to be increasing and weakly convex. Workers are either initially employed by a sector or unemployed in a sector specific labor market. The stock of unemployed workers in sector a and b are given by u_a or u_b respectively and they ultimately become job seekers denoted as x_a and x_b respectively. The economy is subject to sector specific shocks. Each sector has a probability of $\frac{1}{2}$ to be hit by a shock and if so loses a proportion δ of their employed workforce. Workers who due to the shock lose their employment join the stock of unemployed to become job seekers in their attached sector. We denote the stock of job seekers in the shocked scenario as $\hat{x}_a = u_a + \delta e_a$ and $\hat{x}_b = u_b + \delta e_b$. A matching function is taking the stock of job seekers as an input and generates new matches as an output, i.e. $n_i = g(x_i)$ where $g(\cdot)$ is assumed to be concave and increasing in x_i. This results in a new stock of employment $e_i' = e_i + n_i$ which is used to produce the consumption good. The timing is as follows: First, the separation shock hits the economy, then the matches are being generates and finally based on e_i'

there will be sector specific production. There is no inter-sectoral mobility.

Now consider a social planner who given the initial employment and unemployment stock in sector a and b, aims to reallocate workers from sectoral employment in a to sectoral employment in b before the separation shock is being realized in order to maximize expected welfare. That is he chooses l to change adjust $e_a = \bar{e}_a - l$ and $e_b = \bar{e}_b + l$. In such a setting the following can be show:

Proposition 2. *The Social Planner faces a trade-off between allocative efficiency and welfare losses due to higher (expected) unemployment.*

The intuition of the result is simple: The Social Planner balances the allocative efficiency with the insurance provided from higher diversification of the sectoral base. The crucial assumption is that there is some concavity in the matching function such that a shock to the larger sector perturbs the local labor market to a larger extent thus creating higher welfare costs.

Finally, while the current example features the extreme case of immobile workers, the degree to which a worker in a given sector will be affected by an adverse shock depends crucially on his mobility: If the worker can easily move towards sectors and areas of the economy that are unaffected then she can easily escape unemployment and as a side effect of her move the local labor market will be less congested. This crucially depends on the empirical segmentation of labor markets, that is to what extent workers can move between potentially differently affected local sectoral labor markets. The next section seeks to quantify the mobility cost across occupations and space in France.

2.4 French Labor Market Segmentation

As a first step we examine the mobility cost and thus segmentation between occupation-department specific labor markets in France. This section introduces the French data as well as our estimation strategy to estimate mobility costs between geographically and occupationally separated labor markets.

2.4.1 Data

To analyse and quantify the segmentation of the French labor market, we use data from the *Declarations Annuelles des Donnees Sociales* (DADS). In this data the French state makes available mandatory employer reports of the earnings for each employee in the private sector. The universe of workers subject to the payroll tax are available in the cross-section. However, in order to estimate mobility costs a panel is necessary. The panel data version includes since 2002 all individuals born in October, i.e. 8.3 pc subsample of the working population. For each employee the duration of a job along with the wage is recorded. Furthermore some demographics are available such as sex and age as well as a geographical notion of residence and work (up to community level that is 36.000 units in France, however we are currently using departements). The job is furthermore described in terms of an occupation codes (PCS). Finally, firm identifiers allow us to match the worker level observations with firm level datasets on balance sheets and recorded exports.

For our current estimation we focus on only five occupation classes defined at the PCS1 level which correspond to white collar workers (PCS1 category 3), intermediate professions/technicians (PCS1 category 4), employees (PCS1 category 5), blue collar workers (PCS1 category 6), and other artisanal workers and self-employed (PCS1 category 2). Furthermore, we focus on continental departments (95 overall). Firms are identified by sectors according to the APE which corresponds roughly to the NAF where we characterize firms at the 2 digit level which gives roughly 100 different sectors.

2.4.2 Model

The model is based on recent contributions by Traiberman (2017) and Scott (2014) and formulates a worker's dynamic problem of chosing a sequence of labor markets where mobility is costly and therefore imperfect. We follow Traiberman (2017) in adopting techniques along the lines of Hotz and Miller (1993) to remedy concerns about unobservable continuation values. While our approach extends his by incorporating geographical mobility cost[1] the exposition here follows his closely and the reader is referred to his paper for more detail.

Workers, indexed by i, are characterized by a vector of observable demographics and unobservable occupation specific comparative advantages, denoted by ω. They are initially located in a region-occupation specific labor market (n, o) and can choose to reallocate subject to switching costs $\tau_{n_{t-1},n_t}^{o_{t-1},o_t}$. Let $v_t(o, n, \omega, \epsilon)$ by the value function of a worker with demographics ω, previous occupation o, previous location n, vector of switching

[1]Though our approach is at the same time more limited since we do not yet present results adapting the EM algorithm to control for unobserved heterogeneity.

cost shock ϵ and define the integrated value function as $V_t(o,n,\omega) = \int v_t(o,n,\omega,\epsilon)dG(\epsilon)$. The worker's problem can the be written in recursive form:

$$v_t(o,n,\omega_{it},\epsilon_{it}) = \max_{o',n'} \Big(w_{o',n',t} EH_{o',n'}(\omega_{it}) + \eta_{o',n'} + \rho\epsilon_{o',n',i,t} - \tau^{o,o'}_{n,n'} \\ + \beta E_t V_{t+1}(o',n',T(\omega_{it},o',n')) \Big) \tag{2.1}$$

where o, o' denotes the current and future occupation of the worker, n, n' denotes the current and future region, $w_{o,n}$ is the occupation-region specific wage that is being paid per unit of human capital which is measured by $H_{o,n}(\omega_{it})$ and depends on the worker's type at the time, ω_{it}, $\epsilon_{o',n',i,t}$ is the preference shock and $\tau^{o_{t-1},o_t}_{n_{t-1},n_t}$ the mobility cost. The last term denotes the continuation value which in turn depends on the chosen region-occupation labor market and the transition of the worker's characteristics. We assume a log linear human capital function along the lines of the classical mincer regression literature, that is,

$$\log(H_o(\omega_{it}\vartheta_{it})) = \beta^o_1 \times age_{it} + \beta^o_2 \times age^2_{it} + \beta^o_3 \times ten_{it} + \theta_{oi} + \sigma_o\varsigma_{iot}$$

Switching costs are a function of occupational bilateral switching cost and distance,

$$\tau^{o_{t-1},o_t}_{n_{t-1},n_t} = \delta_{o_t,o_{t-1}} + \delta_{geo} + \varrho\log(dist_{n_t,n_{t-1}})$$

To emphasise the connection to the full class of multinomial choice models we can define the auxiliary function, $\tilde{v}_t(o,n,o',n',\omega)$, such that,

$$v_t(o_{i,t-1},n_{i,t-1},\omega_{it},\epsilon_{it}) = \max_{o',n' \in \mathcal{O}\times\mathcal{N}} \tilde{v}_t(o,n,o',n',\omega) + \rho\epsilon_{o',n',i,t} \tag{2.2}$$

Assuming GEV distributed cost shocks we obtain closed form transition probabilities,

$$\pi(o',n'|o,n,\omega) = \frac{\exp(\tilde{v}_t(o,n,o',n',\omega)/\rho)}{\sum_{o'',n''}\exp(\tilde{v}_t(o,n,o'',n'',\omega)/\rho)}$$

2.4.3 Estimation Strategy

The challenge in estimating mobility costs is twofold. Firstly, dynamic problems can be computationally burdensome if the estimation strategy requires solving the dynamic program. Secondly, unobservable heterogeneity might induce bias: In the case of imperfect labor mobility and switching costs between labor markets, there is an option value of being attached to a specific labor market creating unobserved continuation values that can induce omitted variable bias if not controlled for. To remedy this problem Hotz and Miller (1993) and Arcidiacono and Miller (2011) introduced techniques that are based on the properties of the GEV distribution to remove the unobservable continuation or option values and allow us to obtain non-linear estimating equations that do not require solving the complete DP. In the following we will give a short description.

We first collect moving costs and incomes into a flow payoff denoted by $u_t(o, n, o', n', \omega)$ and define the inclusive value as

$$D_t(\omega, o, n) = \sum_{o,n \in \mathcal{O} \times \mathcal{N}} \exp[u_t(o, n, o', n', \omega) + \beta E_t V_{t+1}(T(\omega, o, n, o', n'), o', n')$$

The probability of observing a career path from time t to τ can then be written as a discounted sum of flow payoffs, discounted sum of worker's expectation error and the cinlusive value and an unobserved future continuation value (Hotz and Miller; 1993)

$$\sum_{s=t}^{\tau} \beta^{s-t} \log \pi_s(\omega_s, o_{s-1}, n_{s-1}, o_s, n_s) = \sum_{s=t}^{\tau} \beta^{s-t} u_s(\omega_s, o_s, n_s, o_{s-1}, n_{s-1}) + \sum_{s=t}^{\tau} \beta^{s-t} \zeta_s \\ + \beta^{\tau+1} E_\tau V_{\tau+1}(\omega_{\tau+1}, o_\tau, n_\tau) - \log D_t(\omega_s, o_{s-1}, n_{s-1}) \tag{2.3}$$

where ζ_s refers to expectation errors. What is important to note is that if we assume renewal action (a special case of finite dependence) that is that past occupational choices do not affect your flow payoff in future occupational choices except through changes in your worker level demographics, then we can compare worker career trajectories with identical starting and end positions and the unobservable terms cancel out. Specifically, comparing two different career trajectories for workers, i and j

$$\begin{aligned} &\sum_{s=t}^{\tau} \beta^{s-t} \log \frac{\pi_s(\omega_{is}, o_{i,s-1}, n_{i,s-1}, o_{i,s}, n_{i,s})}{\pi_s(\omega_{j,s}, o_{j,s-1}, n_{j,s-1}, o_{j,s}, n_{j,s})} \\ &= \sum_{s=t}^{\tau} \beta^{s-t}[u_s(\omega_{i,s}, o_{i,s}, n_{i,s}, o_{i,s-1}, n_{i,s-1}) - u_s(\omega_{j,s}, o_{j,s}, n_{j,s}, o_{j,s-1}, n_{j,s-1})] \\ &\quad + \sum_{s=t}^{\tau} \beta^{s-t}[\zeta_{i,s} - \zeta_{j,s}] + \beta^{\tau+1} E_\tau(V_{\tau+1}(\omega_{i,\tau+1}, o_{i,\tau}, n_{i,\tau}) - V_{\tau+1}(\omega_{j,\tau+1}, o_{j,\tau}, n_{j,\tau})) \\ &\quad - [\log D_t(\omega_{i,s}, o_{i,s-1}, n_{i,s-1}) - D_t(\omega_{j,s}, o_{j,s-1}, n_{j,s-1})] \end{aligned} \tag{2.4}$$

If workers have same initial inclusive values and same terminal continuation values then the unobservable terms disappear. The left hand side can be estimated directly from data, and the right hand side is a nonlinear function of observables with an additive error term.

2.4.4 Estimation

For the estimation we follow Traiberman (2017) and focus on one shot deviations, i.e. career trajectories that diverge for one period both along the occupation and geography margin, but then converge again to the same occupation-region labor market. Mathematically, we estimate the following equation:

$$\log \frac{\pi_t(\omega, o, n, o', n')}{\pi_t(\omega, o, n, o, n)} + \beta \log \frac{\pi_t(\omega', o', n', o'', n'')}{\pi_t(\omega'', o, n, o'', n'')}$$
$$= -\tau_{n,n'}^{o,o'}(\omega) - \tau_{n',n''}^{o',o''}(\omega') + \tau_{n,n''}^{o,o''}(\omega'') + \frac{1}{\rho}(w_{o',n'} - w_{o,n}) + \frac{1}{\rho}(\eta_{n'o'} - \eta_{no}) + \zeta_{o,n,t} + m_{o,o',o'',n,n',n'',t} \tag{2.5}$$

Since there are too many combinations that can be used and computed[2] we instead focus on geographical trajectories that end in departements that form the Parisian region (Île-de-France). This gives us more than six million observations.

In order to construct the left hand side, we obtain the conditional choice probabilities by forming bins by age groups, distance to destination and sex. We exploit the fact that once we introduce a occupation-region(-time) specific fixed effect that controls for different choice sets across labor markets, the probability of choosing one location rather than another is merely a function of the bilateral distance and the labor market's attractiveness, i.e. real wage, of a location.

2.4.5 Results

Preliminary results with regard to the occupational mobility costs are presented in figure 2.6. Occupational mobility costs can be as high as 60 percent of the yearly income of a worker and severely limit occupational mobility. Spatial mobility cost is still work in progress.

[2]Note, that compared to Traiberman (2017) the number of possible combinations is even larger since we introduce the additional geographic dimension.

2.5 Quantitative Analysis: Optimal Diversification

In this section we first introduce the theoretical framework before we illustrate its calibration and finally its workings via simulation.

2.5.1 Theoretical Framework

Consider an economy with multiple locations indexed by $n, i \in \mathcal{N}$. Locations are heterogeneous in their exogenously fixed housing supply, $\overline{H}_n$, and their geographical location relative to one another. Each location produces goods in multiple sectors $s \in \mathcal{S}$. Total population at time t is exogenously fixed at L_t. Firms hire workers from different occupations, $o \in \mathcal{O}$, and there is an initial distribution at time 0 of the population across locations and occupations, $[L_{i,o,0}]_{\forall i,o}$. For the exposition of the static equilibrium we drop the time subscripts. We will reintroduce them later on when we introduce the dynamic adjustment via worker mobility.

Preferences and Endowments

Workers belong to a labor market (n, o) which is location-occupation specific. In a given labor market (n, o), workers can be either unemployed or employed with a process described below. Each worker is endowed with preferences for consumption of goods from different sectors and residential land use. The flow utility of worker i in labor market (n, o) is defined by:

$$U_{n,o}(i) = H_n(i)^{1-\sum_j \alpha_j} \prod_{j=1}^{\mathcal{S}} C_j^{n,o}(i)^{\alpha_{n,j}}$$

with $C_j^{n,o}(i)$ the consumption of industry j-good by a worker i in labor market (n, o). This consumption does not take the same value if the worker is unemployed of employed. The sectoral goods consumption index is defined over consumption of a fixed continuum of goods $k \in [0, 1]$:

$$C_j^{n,o} = \left(\int_0^1 \left(c_{n,o}^j(k) \right)^\rho dk \right)^{1/\rho}$$

where the CES parameter ρ captures the elasticity of substitution between goods defined as $\sigma = 1/(1-\rho)$. It follows that the expenditure of any consumer on a given variety k is given by

$$q_{n,o,s}(k) = \left(\frac{p_{n,s}(k)}{P_{n,s}}\right)^{1-\sigma} \alpha_s I_{n,o} \tag{2.6}$$

where $I_{n,o}$ is the revenue of the representative consumer in labor market (n, o) and the corresponding dual price index at the sectoral level is:

$$P_{n,s} = \left(\int_0^1 p_{n,s}(k)^{1-\sigma} dk\right)^{1/1-\sigma}$$

The indirect flow utility that worker i obtains at time t from living in labor market (n, s) can be written in the following way:

$$U_{n,o}(i) = \frac{v_{n,o}(i)}{\mathbb{P}_{n,t}} \tag{2.7}$$

where $\mathbb{P}_n$ is the price of the consumption bundle in location n defined by

$$\mathbb{P}_n = \mu_{n,s} r_n^{1-\sum_s \alpha_{n,s}} \prod_{s=1}^{\mathcal{S}} P_{n,s}^{\alpha_s} \tag{2.8}$$

where $\mu_{n,s}$ represents the Cobb Douglas multiplier [3] and $v_{n,o}(i)$ is the income of worker i in labor market (n, o). The income of the worker depends on its status, that is if she is employed or unemployed,

$$\text{If i is employed} \quad v^e_{n,o} = w_{n,o} \tag{2.9}$$

$$\text{If i is unemployed} \quad v^u_{n,o} = b_{n,o} \tag{2.10}$$

With $w_{n,o}$ the wage, $b_{n,o}$ is the unemployment benefit which is locally financed via proportional taxation and $v_{n,o}$ the average income of workers in labor market (n, s). Noting $E_{n,o}$ the number of *employed* workers in labor market (n, o), $v_{n,o}$ is simply defined by:

$$v_{n,o} = \frac{E_{n,o}}{L_{n,o}} v^e_{n,o} + \frac{L_{n,o} - E_{n,o}}{L_{n,o}} v^u_{n,o} \tag{2.11}$$

[3] This is given by

$$\mu_{n,s} \equiv \left(\prod_{s=1}^{\mathcal{S}} \alpha_s^{-\alpha_s}\right) \times (1 - \sum_s \alpha_{n,s})^{-(1-\sum_s \alpha_{n,s})}$$

Production

Within each location n, each sector s is endowed with an exogenous baseline productivity level $A_{n,s,t}$. Firms employ a constant returns to scale production function,

$$Q_{n,s}(k) = z_{n,s}(k)A_{n,s} \times \tilde{E}_{n,s}(k) \tag{2.12}$$

where $A_{n,s,t}$ is the implied sector wide productivity and where $\tilde{E}_{n,s}$ is a Cobb-Douglas composite of labor inputs across different occupations,

$$\tilde{E}_{n,s}(k) = \prod_{o=1}^{O} E_{n,o,s}(k)^{\beta_{o,s}} \tag{2.13}$$

Each variety k is also associated with an idiosyncratic productivity, $z_{n,s}(k)$, which is drawn from a Frechet distribution as in Eaton and Kortum (2002),

$$Prob(z \leq Z) = e^{-Z^{-\theta_s}}$$

where θ_s determines the variance of productivities being drawn and therefore pins down the trade elasticity. We will see that only a subset $X_{n,s} \subset [0,1]$ is actually produced in location n in every industry s.

Labor Market

Labor markets are segmented across occupations and space and can be denoted by the tuple (n,o). Each labor market (n,o) is characterized by matching frictions. An infinitely elastic supply of potential firms may enter the labor market by opening vacancies. At the beginning of any period, all workers are unmatched and are looking for a job. Matches last one period only. Firms can post vacancies $V_{n,o}$ at a cost c_o. The number of matches is governed by:

$$E_{n,o} = \Gamma_{n,o} L_{n,o}^{\chi} V_{n,o}^{1-\chi} \tag{2.14}$$

where $\Gamma_{n,o}$ is a measure of the aggregate efficiency of the matching technology in labor market (n,o). Furthermore, several sectors might post vacancies in the same occupation specific labor market implying that aggregate posted vacancies in market labor market (n,o) is the sum of vacancies posted by firms in individual sectors searching in that labor market, i.e. $V_{n,o} = \sum_s V_{n,o,s}$. The number of matches specific to sector s is proportional to its share of vacancies posted.

$$E_{n,o,s} = \frac{V_{n,o,s}}{V_{n,o}} \Gamma_{n,o} L_{n,o}^{\chi} V_{n,o}^{1-\chi} = V_{n,o,s} \Gamma_{n,o} \lambda_{n,o}^{-\chi} \tag{2.15}$$

where we define the labor market tightness in labor market (n, o) as $\lambda_{n,o} = V_{n,o}/L_{n,o}$. As in Helpman and Itskhoki (2010, 2014) matching frictions lead to a non trivial hiring cost for firms as well as matching rents over which firms and workers bargained, as explained below.

Proposition 3. *The cost of hiring one worker for a firm in sector s and in labor market (n, o) is:*

$$\kappa_{n,o,s} = \omega_{n,o,s}^{-1} a_{n,o} \lambda_{n,o}^{\chi}, \quad \textit{with} \quad a_{n,o} = \frac{c_o}{\Gamma_{n,o}} \quad \textit{and} \quad \omega_{n,o,s} = \frac{V_{n,o,s}}{V_{n,o}} \tag{2.16}$$

Proof. The vacancy filling rate for any vacancy in labor market (n, o) is

$$\psi_{n,o} = \Gamma_{n,o} \lambda_{n,o}^{-\chi}$$

The expected cost of filling a vacancy posted by sector s in labor market (n, o) is the cost of posting the vacancy multiplied by the vacancy filling rate and the probability that the filled vacancy was posted by sector s, that is,

$$\kappa_{n,o,s} = \frac{c_o}{\omega_{n,o,s} \psi_{n,o}} = \frac{c_o}{\Gamma_{n,o}} \omega_{n,o,s}^{-1} \lambda_{n,o}^{\chi}$$

where $\omega_{n,o,s} = \frac{V_{n,o,s}}{V_{n,o}}$ is the probability that the filled vacancy belongs to sector s. Noting $a_{n,o} = \frac{c_s}{\Gamma_{n,o}}$, one has the result. □

Similarly, we can calculate the cost of finding a unit of the composite worker, $\tilde{E}_{n,s,t}$.

Proposition 4. *The cost of hiring one composite worker for a firm in labor market (n, s) is given by,*

$$\tilde{\kappa}_{n,s} \propto \sum_{o=1}^{O} \beta_{o,s} \left(\frac{\kappa_{n,o,s}}{\kappa_{n,o,s} + w_{n,o}} \right) \tag{2.17}$$

Proof. Define as $x_o(\mathbf{w}, \mathbf{k}, 1)$ the factor demand to obtain one composite unit of labor given the vector of wages and hiring costs as defined above. The cost of hiring one composite unit can be written as the sum of the conditional demand multiplied by the cost of hiring, that is,

$$\tilde{\kappa}_{n,s} \equiv \sum_{o=1}^{O} x_o(\mathbf{w}, \mathbf{k}) \times \kappa_{n,o,s} = C_{n,s}(\mathbf{w}, \mathbf{k}) \times \sum_{o=1}^{O} \beta_{o,s} \left(\frac{\kappa_{o,n,s}}{\kappa_{n,o,s} + w_{n,o}} \right)$$

where the equality stems from the fact that the conditional demand is given by $x_o(\mathbf{w},\mathbf{k},1) = \left(\frac{\beta_{o,s}}{\kappa_{n,o,s}+w_{n,o}}\right) \times C_{n,s}(\mathbf{w},\mathbf{k})$, where $C_{n,s}(\mathbf{w},\mathbf{k})$[4] refers to the unit cost function and is common to the conditional demand functions for all factors and can thus be factored out obtaining the result in the proposition. □

Finally, we can obtain a sector-location specific matching function noting that,

$$\tilde{E}_{n,s} = \underbrace{\left(\prod_{o=1}^{O} (\Gamma_{n,o})^{\beta_{o,s}}\right)}_{\equiv \tilde{\Gamma}_{n,s}} \times \left(\underbrace{\prod_{o=1}^{O} (L_{n,o})^{\beta_{o,s}}}_{\equiv \tilde{L}_{n,s}}\right)^{\chi} \times \left(\underbrace{\prod_{o=1}^{O} \left(\left(\frac{V_{n,o,s}}{V_{n,o}}\right)^{\frac{1}{1-\chi}} V_{n,o}\right)^{\beta_{o,s}}}_{\equiv \tilde{V}_{n,s}}\right)^{1-\chi} = \tilde{\Gamma}_{n,s}\tilde{L}_{n,s}^{\chi}\tilde{V}_{n,s}^{1-\chi}$$

Note that the cost of vacancy filling is increasing in vacancies posted. This comes from the concavity of the matching function with regard to vacancies posted and mirrors a localized congestion effect in the labor market.

Prices and Wages

Since firms need to participate in the labor market by posting vacancies in order to hire the marginal cost of a worker is equal to the market wage plus the vacancy posting cost. This introduces a wedge into the local labor market which in turn induces rationing as mitigated via the matching function. The firm takes that into account when making its decisions. The production function is linear in the composite of occupational labor and there is perfect competition in the good market. This implies that the size of a firm is irrelevant: the firm bargains with each worker independently.[5] The revenue generated by each unit of a composite worker, $\tilde{E}_{n,s}(k)$, when producing variety $k \in X_{n,s}$ is given by

$$r_{n,s}(k) = p_{n,s}(k) z_{n,s}(k) A_{n,s}$$

[4]The unitcost function of the above Cobb-Douglas aggregator across occupations is given by

$$C_{n,s}(\mathbf{w},\mathbf{k}) = \left(\frac{\prod_{o=1}^{n}(\kappa_{n,o,s} + w_{n,o})^{\beta_{o,s}}}{\prod_{o=1}^{n} \beta_{o,s}^{\beta_{o,s}}}\right)$$

[5]This is the same as in Carrere, Grujovic and Robert-Nicoud (2015). If firms were monopolists within their variety, they would take into account the demand curve and revenues would be concave with respect to labor, as in Helpman and Itskhoki (2010, 2014). In such a case, one would have to use the Stole-Zwiebel bargaining protocol.

Assuming that firms and workers have equal bargaining shares, the composite wage that producer of variety k is willing to pay is equal to:

$$\tilde{w}_{n,s}(k) - b_n = \frac{r_{n,s}(k)}{2} \tag{2.18}$$

Firms must pay a cost $\tilde{\kappa}_{n,s}$ to hire an additional unit of the composite worker, this cost is exogenous to any individual firm k. Firms optimally choose to post vacancies until the marginal revenue they derive from the worker (net of wage) is equal to the cost of hiring her, $\tilde{\kappa}_{n,s}$. Since firms and workers equally share the matching rent as stated above, this then implies that the wage rate satisfies $\tilde{w}_{n,s}(k) = 2b_n + \tilde{\kappa}_{n,s}$. A key consequence is that all firms in industry s pay the same unit wage regardless of their productivity. Moreover, with perfect competition firms price at their marginal cost so that the market price for variety k, in industry s, from origin n and sold in the destination market i is:

$$p_{n,i,s}(k) = \zeta_{ni}\frac{\tilde{w}_{n,s} + \tilde{\kappa}_{n,s}}{z_{n,s}(k)A_{n,s,t}} = \zeta_{ni}\frac{2\tilde{\kappa}_{n,s} + 2b_n}{z_{n,s}(k)A_{n,s}} \tag{2.19}$$

where ζ_{ni} the iceberg cost when shipping from n to i and $(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})$ is the total cost incurred by a firm when hiring a unit of the composite worker, accounting for the wage and hiring cost. Finally we can also define the wage bill for hiring one unit of the composite worker.

Proposition 5. *The cost of hiring one unit of the composite worker for a firm in location-sector* (n, s) *is:*

$$\tilde{w}_{n,s} \propto \sum_{o=1}^{O} \beta_{os}\left(\frac{w_{n,o}}{\kappa_{n,o,s} + w_{n,o}}\right) \tag{2.20}$$

Proof. The proof follows the proof for proposition 3. Define as $x_o(\mathbf{w}, \mathbf{k}, 1)$ the factor demand to obtain one composite unit of labor given the vector of wages and hiring costs as defined above. The wage cost of employing one composite unit can be written as the sum of the conditional demand multiplied by employment cost, that is,

$$\tilde{w}_{n,s} \equiv \sum_{o=1}^{O} x_o(\mathbf{w}, \mathbf{k}) \times w_{n,o} = C_{n,s}(\mathbf{w}, \mathbf{k}) \times \sum_{o=1}^{O} \beta_{o,s}\left(\frac{w_{n,o}}{\kappa_{n,o,s} + w_{n,o}}\right)$$

where the equality stems from the fact that the conditional demand is given by $x_o(\mathbf{w}, \mathbf{k}, 1) =$

$\left(\frac{\beta_{o,s}}{\kappa_{n,o,s}+w_{n,o}}\right) \times C(\mathbf{w},\mathbf{k})$, where $C(\mathbf{w},\mathbf{k})$[6] refers to the unit cost function and is common to the conditional demand functions for all factors and can thus be factored out obtaining the result in the proposition. $\square$

Housing Market

We assume that there is a fixed expenditure share as housing as introduced above, $\delta \equiv (1 - \sum_j \alpha_j) > 0$ [7]. Furthermore, we follow Monte et al. (2017) and assume that the housing stock is owned locally by landlords who consume traded goods only. This implies that total demand in a location is equal to total factor earnings.

Trade shares

Consumers buy any variety k from the location charging the cheapest price. Standard calculations imply the following trade shares,

$$\pi_{n,i,s} = \frac{A_{n,s}^{\theta_s}((\tilde{\kappa}_{n,s} + b_n)\zeta_{ni})^{-\theta_s}}{\Phi_{i,s}} \tag{2.21}$$

with

$$\Phi_{i,s} = \sum_k A_{k,s}^{\theta_s}(\zeta_{ki}(\tilde{\kappa}_{k,s} + b_k))^{-\theta_s} \tag{2.22}$$

representing the multilateral resistance term in location i and sector s, characterizing how competitive the destination market is. The price index in location i is related to the multilateral resistance term in the following way,

$$P_{i,s}^{-\theta_s} = \gamma\left[\Phi_{i,s}\right] \tag{2.23}$$

where γ is the gamma function,

$$\gamma = \left[\Gamma\left(\frac{\theta_s - (\sigma - 1)}{\theta_s}\right)\right]$$

[6]The unitcost function of the above Cobb-Douglas aggregator across occupations is given by

$$C_{n,s}(\mathbf{w},\mathbf{k}) = \left(\frac{\prod_{o=1}^{n}(\kappa_{n,o,s} + w_{n,o})^{\beta_{o,s}}}{\prod_{o=1}^{n}\beta_{o,s}^{\beta_{o,s}}}\right)$$

[7]In later empirical applications this will be calibrated to a value of .23 as in Combes et al. (2018)

Note that location-sector-specific labor market conditions (described by $\kappa_{n,s}$) and location-sector-specific technologies (described by $A_{n,s,t}$) have observationally identical effects on trade volumes. Note that equation (2.23) can be written as

$$P_{i,s} = Q \left(\sum_k A_{k,s}^{\theta_s} ((2\tilde{\kappa}_{n,s} + 2b_n)\zeta_{ki})^{-\theta_s} \right)^{-1/\theta_s} \tag{2.24}$$

where Q is a constant equal to the gamma function evaluated at $1 + (1 - \frac{\sigma}{\theta_s})$. Before introducing the static equilibrium, we make one last adjustment to the standard model and introduce a (foreign) sector specific demand shifter, i.e. there exists a fixed exogenous sector specific demand, $\Xi_{F,s,t}$, where exports to that destination e.g. from region i in sector s are modelled as follows,

$$X_{i,s,t}^F = \pi_{i,F,s,t} \times \Xi_{F,s,t}$$

where $\Xi_{F,s,t}$ is a stochastically evolving shock variable. Given the description of the model we are now in a position to define the static equilibrium conditional on a worker allocation.

To enhance clarity and elucidate the structure of the model we introduce the static equilibrium in two steps: In a first step we introduce the conditions that define the trade equilibrium conditional on sector-location (n, s) specific demand and composite wages. In a second step we introduce the conditions that determine occupation specific variables. The first part of the static equilibrium is defined by the following equations,

Definition 2 (Static Equilibrium (Location- and Sector-level outcomes)). *Given data on labor allocations aggregated to the location-sector specific composite, $[\tilde{L}_{n,s}]_{\forall n\times s}$, stocks of location specific unemployed workers, $[U_n]_{\forall n}$, the housing stock, $[\bar{H}_n]_{\forall n}$, baseline productivities, $[A_{n,s}]_{\forall n\times s}$, values for the parameters $[\beta_{o,s}]_{\forall o\times s}, [\theta_s]_{\forall s}, [\alpha_s]_{\forall s}, [\Xi_{F,s}]_{\forall s}, [\tilde{\Gamma}_{n,s}]_{\forall n\times s}, [\tilde{c}_s]_{\forall s}, [b_n]_{\forall n}$, the static trade equilibrium is defined by a set of bilateral expenditure shares between regions $[\pi_{i,n,s}]_{\forall i\times n\times s}$, wages $[\tilde{\omega}_{n,s}]_{\forall n\times s}$, hiring costs $[\tilde{\kappa}_{n,s}]_{\forall n\times s}$, vacancies posted $[\tilde{V}_{n,s}]_{\forall n\times s}$, and employment levels $[\tilde{E}_{n,s}]_{\forall n\times s}$ such that the following conditions are satisfied.*

- *For each industry-location (n, s), total income equals total factor expenditure,*

$$(\tilde{\omega}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s} = \left(\sum_i \pi_{n,i,s}\alpha_s Y_i \right) + X_{nF,s} \qquad \forall (n,s) \in \mathcal{N} \times \mathcal{S} \tag{2.25}$$

- *where the share of total spending devoted to products produced in i in location n and sector s is given by,*

$$\pi_{n,i,s} = \frac{A_{n,s}^{\theta_s}((\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\zeta_{ni})^{-\theta_s}}{\Phi_{i,s}} \qquad \forall (i,n,s) \in \mathcal{N} \times \mathcal{N} \times \mathcal{S} \tag{2.26}$$

with $\Phi_{n,s}$ defined by (2.22).

- *Total consumer spending Y_n in any location n is defined by*

$$Y_n = \left(\sum_{s \in \mathcal{S}} \tilde{w}_{n,s} \tilde{E}_{n,s} \right) \qquad \forall (n,s) \in \mathcal{N} \times \mathcal{S} \tag{2.27}$$

- *Fixed expenditure share on foreign products such that trade is balanced.*
- *Recruitment cost and wages are related according to*

$$\tilde{w}_{n,s} = 2b_{n,s} + \tilde{\kappa}_{n,s} \qquad \forall (n,s) \in \mathcal{N} \times \mathcal{S} \tag{2.28}$$

- *Total recruitment cost satisfies,*

$$\tilde{\kappa}_{n,s}\tilde{E}_{n,s} = \tilde{c}_{n,s}\tilde{V}_{n,s} \qquad \forall (n,s) \in \mathcal{N} \times \mathcal{S} \tag{2.29}$$

- *Sector location specific matching function*

$$\tilde{E}_{n,s} = \tilde{\Gamma}_{n,s}\tilde{L}_{n,s}^{\chi}\tilde{V}_{n,s}^{1-\chi} \qquad \forall (n,s) \in \mathcal{N} \times \mathcal{S} \tag{2.30}$$

- *Housing market clears,*

$$r_n = \frac{\delta Y_n}{\overline{H}_n} \qquad \forall n \in \mathcal{N} \tag{2.31}$$

Definition 3 (Static equilibrium (Occupation-level outcomes))**.** *If the static trade equilibrium is satisfied (that is data for $[\pi_{i,n,s}]_{\forall i \times n \times s}$, wages $[\tilde{w}_{n,s}]_{\forall n \times s}$, hiring costs $[\tilde{\kappa}_{n,s}]_{\forall n \times s}$, vacancies posted $[\tilde{V}_{n,s}]_{\forall n \times s}$, and employment levels $[\tilde{E}_{n,s}]_{\forall n \times s}$ are available) and the sector specific variables are linked to the occupation specific variables as follows,*

- *In each sector-location, the recruitment cost for one composite unit of labor satisfies,*

$$\tilde{\kappa}_{n,s} \propto \sum_{o=1}^{O} \left(\beta_{os} \frac{\kappa_{n,o,s}}{\kappa_{n,o,s} + w_{n,o}} \right) \tag{2.32}$$

- *The composite sector-location specific wage function is defined as follows,*

$$\tilde{w}_{n,s} \propto \sum_{o=1}^{O} \beta_{os} \left(\frac{w_{n,o}}{\kappa_{n,o,s} + w_{n,o}} \right) \tag{2.33}$$

- *Sector location specific employment is given by,*

$$\tilde{E}_{n,s} = \prod_{o=1}^{O} \left(E_{n,o,s}\right)^{\beta_{o,s}} \tag{2.34}$$

- *The arguments of the sector-location specific matching function are defined as follows,*

$$\tilde{\Gamma}_{n,s} \equiv \prod_{o=1}^{O} \left(\Gamma_{no}\right)^{\beta_{o,s}} \tag{2.35}$$

$$\tilde{L}_{n,s} \equiv \prod_{o=1}^{O} \left(L_{n,o}\right)^{\beta_{o,s}} \tag{2.36}$$

$$\tilde{V}_{n,s} \equiv \prod_{o=1}^{O} \left(\left(\frac{V_{n,o,s}}{V_{n,o}} \right)^{\frac{1}{1-\chi}} V_{n,o} \right)^{\beta_{o,s}} \tag{2.37}$$

and if furthermore data on location-occupation allocations, $[L_{n,o}]_{\forall n,o}$ and vacancy postings costs, $[c_o]_{\forall o}$, are available, then the static equilibrium is defined by occupation specific wages $[w_{n,o}]_{\forall n \times o}$, hiring costs $[\kappa_{n,o}]_{\forall n \times o}$, vacancies posted $[V_{n,o,s}]_{\forall n \times o \times s}$, and employment levels $[E_{n,o,s}]_{\forall n \times o \times s}$ such that the following conditions are satisfied.

- *In each labor market (n, s), the recruiting cost satisfies*

$$\kappa_{n,o,s} = \omega_{n,o,s}^{-1} a_{n,o} \lambda_{n,o}^{\chi} \qquad \forall (n,o,s) \in \mathcal{O} \times \mathcal{N} \times \mathcal{S} \tag{2.38}$$

where $a_{n,o} = c_o / \Gamma_{n,o}$, $\omega_{n,o,s} \equiv V_{n,o,s} / \sum_s V_{n,o,s}$

- *In each labor market (n, s), labor force and employed workers are related according to:*

$$E_{n,o} = \Gamma_{n,o} L_{n,o}^{\chi} V_{n,o}^{1-\chi} \qquad \forall (n,o) \in \mathcal{N} \times \mathcal{O} \tag{2.39}$$

- *The level of sectoral employment for a specific occupation is given by*

$$E_{n,o,s} = V_{n,o,s} \Gamma_{n,o} \lambda_{n,o}^{-\chi} \qquad \forall (n,o) \in \mathcal{N} \times \mathcal{O} \tag{2.40}$$

- *Each industry-location (n,s) is paying a fixed share of its income to each factor*

$$(w_{n,o} + \kappa_{n,o,s})E_{n,o,s} = \beta_{o,s}(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s} \qquad \forall (n,o,s) \in \mathcal{N} \times \mathcal{O} \times \mathcal{S} \tag{2.41}$$

- *Each occupation-location (n,o) specific labor market clears*

$$w_{n,o} \sum_s E_{n,o,s} + \sum_s \kappa_{n,o,s} E_{n,o,s} = \sum_s \beta_{o,s}(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s}$$

- *Total recruitment cost is equal to vacancy posting cost*

$$\kappa_{n,o,s} E_{n,o,s} = c_o V_{n,o,s}$$

Dynamics

The previous section describes wages, unemployment, prices and trade conditional on labor allocation. Between periods we allow labor to move between (o,n) labor markets to arbitrage away indirect utility differences subject to mobility costs. This introduces spatial dynamics where a negative demand shock that affects one region can spill over into another region in the form of out migration. In order for the quantitative model to generate realistic unemployment fluctuations we follow Chodorow-Reich and Wieland (2017) and introduce ad-hoc downward nominal wage rigidity.

Labor Mobility The previous subsection builds the solution to the equilibrium *given* the cross-sectional labor allocation. However, before the realization of external demand in period $t+1$, workers can choose to re-locate to a different labor market. We therefore follow Kennan and Walker (2011) in characterising the worker's decision as a dynamic discrete choice between different occupation-location specific labor markets subject to incurring a mobility cost. The worker is assumed to discount future periods, thus her lifetime utility takes the following form,

$$V_{0,n_0,o_0}(i) = \sum_{t \geq 0} \beta^t \left(U_{n_t,o_t}(i) - \tau^{o_{t-1},o_t}_{n_{t-1},n_t} \right)$$

where $V_{0,n_0,o_0}(i)$ denotes the value function of individual i located in an initial labor market (n,o) at time period 0 and who chooses a sequence of labor markets $[n_t, o_t]_{\forall t}$ over her lifetime, where $\tau^{o_{t-1},o_t}_{n_{t-1},n_t}$ indicates the mobility cost which takes a positive value if either $n_t \neq n_{t-1}$ or $o_t \neq o_{t-1}$.

The worker's dynamic discrete choice problem is to choose next periods location optimally by comparing the expected value of each location net the mobility cost of moving there. Since workers are risk neutral and the current employment status has no impact on the mobility decisions[8], we can write the problem as if there is perfect risk sharing between all workers in each (n, o) market. Noting the current cross-sectional labor allocation as $\mathcal{L}_t = \{L_{n,o,t}\}_{n\in\mathcal{N},o\in\mathcal{O}}$ and the current realization of external demand shocks as $\Xi_t = \{\Xi_{s,t}\}_{s\in\mathcal{S}}$, the state of the world is defined by $\Omega_t = \{\mathcal{L}_t, \Xi_t\}$. The problem writes:

$$V_{n,o}(\Omega_t) = U(v_{n,o}(\Omega_t)) + \max_{n',o'} \left[\beta \mathbb{E}_t V_{n',o'}(\Omega_{t+1}) - \tau_{n,n'}^{o,o'} + \nu \epsilon_t^{n',o'}\right]$$

where $v_{n,o,t}$ is the average income in labor market (n, o) defined by the static equilibrium for a state Ω_t. As is common in the dynamic discrete choice literature as well as in the labor mobility literature, we can follow McFadden (1977) and assume that the preference shocks are extreme value distributed and iid distributed, such that

Assumption 2 (McFadden 1977)**.** *Idiosyncratic location-occupation specific preference shifter, $\epsilon_t^{n',o'}$ is iid EV 1 distributed with dispersion parameter ν.*

This gives us the law of motion for labor,

$$L_{n',o',t+1} = \sum_{n,o} \mu_{n,n'}^{o,o'} L_{n,o,t} \tag{2.42}$$

where $\mu^{(i,j)\to(n,s)}$ is the implied transition rate, that is the share of workers from (i, j) to choose to migrate to (n, s). Standard computations imply that bilateral migration shares are given by,

$$\mu_{n,n'}^{o,o'} = \frac{\exp\left(\frac{1}{\nu}\left(\beta \mathbb{E}_t V_{n',o'}(\Omega_{t+1}) - \tau_{n,n'}^{o,o'}\right)\right)}{\sum_{n'',o''} \exp\left(\frac{1}{\nu}\left(\beta \mathbb{E}_t V_{a,b}(\Omega_{t+1}) - \tau_{n,n''}^{o,o''}\right)\right)} \tag{2.43}$$

where the value functions can be written exploiting McFadden's aggregation result,

$$V_{n,o}(\Omega_t) = U(v_{n,o}(\Omega_t)) + \nu \log\left[\sum_{n',o'} \exp\left(\frac{1}{\nu}\left(\beta \mathbb{E}_t V_{n',o'}(\Omega_{t+1}) - \tau_{n,n'}^{o,o'}\right)\right)\right] \tag{2.44}$$

[8] Recall that matches last only one period.

Nominal Downward Wage ridigity Wages are being determined competitively in the labor market but are dynamically subject to nominal wage ridigities in the spirit of Chodorow-Reich and Wieland (2017):

$$w_{n,o,t} = \max\left\{w^*_{n,o,t}, (1-\chi)w_{n,o,t-1}\right\} \tag{2.45}$$

where $w^*_{n,o,t}$ is the wage determined by the static equilibrium.

Steady State Finally we are in a position to define the steady state of the economy.

Definition 4 (Steady State). *A steady state solves the static equilibrium, imposes that the spatial labor allocation and frictional unemployment has converged. The steady state is defined as a vector of labor allocations* $[\bar{L}_{n,o}]_{\forall n\in\mathcal{N},o\in\mathcal{O}}$ *such that the following conditions are satisfied,*

- *Spatial labor allocation has converged such that,*

$$\bar{L}_{n',o'} = \sum_{n,o} \mu^{o,o'}_{n,n'} \bar{L}_{n,o} \tag{2.46}$$

- *where the bilateral mobility flows are defined by*

$$\mu^{o,o'}_{n,n'} = \frac{\exp\left(\frac{1}{\nu}\left(\beta\mathbb{E}_t V_{n',o'}(\bar{\Omega}) - \tau^{o,o'}_{n,n'}\right)\right)}{\sum\limits_{n'',o''} \exp\left(\frac{1}{\nu}\left(\beta\mathbb{E}_t V_{a,b}(\bar{\Omega}) - \tau^{o,o''}_{n,n''}\right)\right)} \tag{2.47}$$

- *where Value functions solve the following equation*

$$\bar{V}_{n,o}(\bar{\Omega}) = U(v_{n,o}(\Omega_t)) + \nu \log\left[\sum_{n',o'} \exp\left(\frac{1}{\nu}\left(\beta\mathbb{E}_t \bar{V}_{n',o'}(\bar{\Omega}) - \tau^{o,o'}_{n,n'}\right)\right)\right] \tag{2.48}$$

- *Downward wage rigidity is no longer binding, i.e.* $w_{n,o,t} = w^*_{n,s,t}$

2.5.2 Calibration

This section states how to use the model structure and sufficient data to back out the relevant parameters of the model. We first state the inversion result in general before giving an outline of the procedure of the calibration.

Proposition 6 (Calibration Static Equilibrium)**.** *Given data on* $[E_{n,o,s}]_{\forall n,o,s}$[9]*,* $[w_{n,o,s}]_{\forall n,o,s}$*,* $[X^F_{n,s}]_{\forall n,s}$*,* $[r_n]_{\forall n}$*,* $[L_{n,o}]_{\forall n,o}$*,* $[\zeta_{ni}]_{\forall ni}$*, and calibrating the value of* χ*,* $[\theta_s]_{\forall s}$ *the model can be inverted to obtain the parameters of the model,* $[\beta_{o,s}]_{\forall o,s}$*,*$[\alpha_s]_{\forall s}$*,* $[A_{n,s}]_{\forall n,s}$*,* $[\Gamma_{n,o}]_{\forall n,o}$*.*

The calibration proceeds in multiple steps which are detailed in the appendix. The procedure can be summarized by highlighting the key steps: (1) Back out $\beta_{o,s}$ as sector-occupation specific shares in the payroll of an industry.[10] (2) We solve the labor market clearing conditions for search costs and sector specific employment levels and wages. (3) From observed employment levels and unmeployment levels we back out the matching parameters either exploiting cross-sectional or time-series variation and assuming that matching efficiency is constant across time or across a occupations or space. (4) We finally use the static equilibrium and sector level employment, wages, search costs, income and calibrations for the trade elasticity to back out the underlying productivity of each sector-location.

2.5.3 Simulation Results

In this section we demonstrate simulation results to illustrate the mechanism of the model, and in particular the interaction between labor market matching frictions and mobility frictions. In order to do so we simulated a baseline economy with ten locations, three occupations and two sectors. The locations are randomly drawn from the geographical extent of the France, i.e. bilateral distances are limited to bilateral distance within France. We draw randomly sector-region-occupation employment numbers as well as wages and rental rates. We furthermore calibrate the trade elasticity, the matching function, the discount factor and housing expenditure share. Foreign demand is calibrated to 80 percent of the sector-location specific payroll.

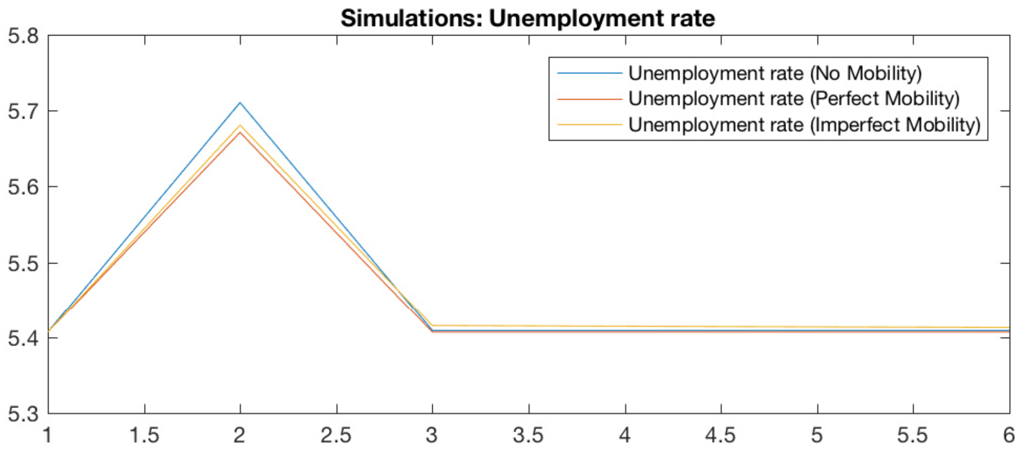
The simulation shocks a sector-location by decreasing the foreign demand in period 2 only and solves then for the dynamic counterfactual. The figure 2.5 demonstrates the global/national unemployment rate under three scenarios: No labor mobility across occupations and geography, perfect mobility and imperfect labor mobility.[11]

[9]The stock of employment can be computed by calculating the efficiency unit of labor equivalent for each worker as a multiple of a baseline wage the Mincer regression

[10]Notice that this is an appropriate estimator if and only if hiring costs are orthogonal to wages across locations.

[11]Imperfect labor mobility allows for some geographical mobility, but no occupational mobility.

Figure 2.5: Simulation: Unemployment Perfect vs Imperfect Mobility



Notes: Simulation results for three different scenarios: No mobility, perfect mobility and imperfect mobility. There is a temporary shock in the second period (a negative shock to a random location-sector specific foreign demand variable). Adjustment follows in the same period and thereafter. The first period is the steady state unemployment rate.

Under no labor mobility and perfect labor mobility the unemployment rate increases for 1 period and then reverts back to the steady state. In the current model, because of the limited duration of filled vacancies, there is no persistence of temporary demand shocks in those extreme cases. As demand returns either the workers return immediately (perfect mobility) or they will never have left their occupation-region specific labor market (no mobility). In either case the initial equilibrium can be directly reobtained. Notice however, that the peak unemployment due to the shock is higher in the no mobility case. This is because in the perfect mobility scenario workers can seek employment elsewhere and the impact of the shock and local congestion of the labor market is alleviated. In the no mobility scenario this cannot be achieved, thus higher national unemployment is the result. Intuitively, the mobility cost determines to what extent the factors can be re-allocated in an efficient manner as a response to the shock.

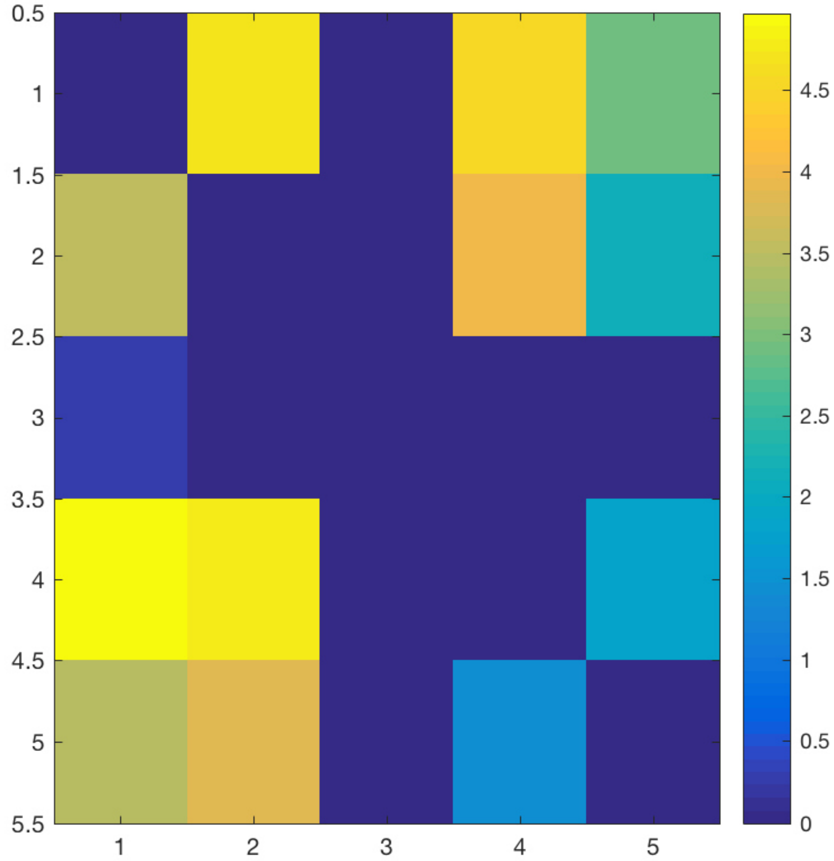
The imperfect mobility case indicates (slightly) persistent unemployment responses: The initial shock disperses workers into other labor markets. The initial unemployment response is below no mobility scenario. However, the return to the initial steady state is also being slowed down thus creating persistence in the national and spatial variance of unemployment.

2.6 Conclusion

In this paper we introduce a tractable framework that extends standard economic geography models to include a quantitative and realistic description of the labor market that can - amongst other things - be easily connected to micro- and macro-level data. We point towards a pathway to analyze both theoretically and empirically the costs and benefits from regional diversification and thereby extend the fundamental trade-off in spatial economics between scale economies and congestion by an additional dimension - that of diversification. The framework can also be used to examine related questions such as the optimal distribution of sectors between rather than within regions. Future application will include the extensive evaluation of regional unemployment risk in France as well as a comparison with the allocative efficiency motif presented in a stylized fashion in section 2.

2.7 Figures

Figure 2.6: Estimation: Occupational Mobility cost



Notes: The graph depicts the occupation mobility costs between occupation 1 to 5 (PCS1 code) to occupations 1 to 5. Darker/colder shaded colors indicate lower (additive) switching costs while lighter/warmer shaded colors indicate higher switching costs. Diagonal switching costs are normalized to 0.

2.8 Theoretical Illustration: Details

- Two sectors a and b
- For each sector two states of the world - shocked states (outcomes denoted by $\hat{x}$) and non-shocked states (outcomes denoted by x) - shock separates a fixed share, δ, of workers

- Initial employment allocation given, e_a, e_b, as well as stock of unemployed workers, u_a, u_b
- In each period unemployed (and freshly separated workers) search for a job, they make up the stock of job searchers

$$x_a = u_a$$

$$x_b = u_b$$

$$\hat{x}_a = u_a + \delta e_a$$

$$\hat{x}_b = u_b + \delta e_b$$

- Matching function takes job searchers as an input and produces employment (abstracting from vacancies posted)

$$n_a = g(x_a)$$

$$\hat{n}_a = g(\hat{x}_a)$$

$$n_b = g(x_b)$$

$$\hat{n}_b = g(\hat{x}_b)$$

- Variables, e'_a, e'_b, summarize the stocks of workers at the end of the period

$$e'_a = e_a + n_a$$

$$e'_b = e_b + n_b$$

$$\hat{e}'_a = (1 - \delta)e_a + \hat{n}_a$$

$$\hat{e}'_b = (1 - \delta)e_b + \hat{n}_b$$

- Employed workers produce and consume their consumption, unemployed workers receive fixed unemployment benefits z

$$C_a = f(e_a)$$

$$C_b = f(e_b)$$

$$\hat{C}_a = f(\hat{e}_a)$$

$$\hat{C}_b = f(\hat{e}_b)$$

- Social Planner maximizes expected welfare by reallocating employed workers from sector a to sector b, that is he chooses variable l such that,

$$e_a = \bar{e}_a - l$$

$$e_b = \bar{e}_b + l$$

$$\max_{l} \frac{1}{2}\left[\hat{e}_a'\hat{C}_a + \hat{u}_a z + e_a' C_a' + u_a z + \hat{e}_b'\hat{C}_b + \hat{u}_b z + e_b' C_b' + u_b z\right]$$

- First order conditions imply that l must satisfy,

$$\left(C_b' + e_b \frac{dC_b'}{dl} + \hat{e}_b' \frac{d\hat{C}_b'}{dl}\right) + \left(-C_a' + e_a \frac{dC_a'}{dl} + \hat{e}_a' \frac{d\hat{C}_a'}{dl}\right) = (-1)\left(\frac{d\hat{e}_a'}{dl}\hat{C}_a' + \frac{d\hat{e}_b}{dl}\hat{C}_b' + \frac{d\hat{u}_a'}{dl}z + \frac{d\hat{u}_b'}{dl}z\right)$$

- Where the LHS signifies the allocative efficiency loss due to reallocating workers from potentially higher productivity sectors to lower productivity sectors
- Where the RHS signifies gains associated with decreases in expected unemployment

2.9 Algorithm: Calibration

To describe the calibration in more detail we explain the individual steps of the code as well as data and calibration requirements.

Code 1 (Calibration Static Equilibrium)**.** *The calibration obtains the parameters of the model,* $[\beta_{o,s}]_{\forall o,s}$, $[\alpha_s]_{\forall s}$, $[A_{n,s}]_{\forall n,s}$, $[\Gamma_{n,o}]_{\forall n,o}$, *using the following data as input,* $[E_{n,o,s}]_{\forall n,o,s}$, $[w_{n,o,s}]_{\forall n,o,s}$, $[X^F_{n,s}]_{\forall n,s}$, $[r_n]_{\forall n}$, $[L_{n,o}]_{\forall n,o}$, $[\zeta_{ni}]_{\forall ni}$, *and calibrating the value of* χ, $[\theta_s]_{\forall s}$. *The calibration proceeds as follows,*

1. *We back out* $\beta_{o,s}$ *as sector-occupation specific shares in the payroll of an industry:*

$$\beta_{o,s} = \frac{\sum_n w_{n,o,s} E_{n,o,s}}{\sum_{n'}\sum_{o'} w_{n',o',s} E_{n',o',s}}$$

This is an appropriate estimator if and only if hiring costs are orthogonal to wages across locations.

- *Data Required:* $[E_{n,o,s}]_{\forall n,o,s}$, $[w_{n,o,s}]_{\forall n,o,s}$
- *Calibration Required: None.*
- *Data obtained:* $[\beta_{o,s}]_{\forall o,s}$

2. *Solve occupation-province labor market clearing for $\kappa_{n,o,s}$ ($N \times O \times S$ equations)*

$$(w_{n,o} + \kappa_{n,o,s})E_{n,o,s} = \beta_{o,s}(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s}$$

$$\tilde{w}_{n,s} = C_{n,s}(\mathbf{w}, \mathbf{k}) \times \sum_{o=1}^{O} \beta_{o,s} \left(\frac{w_{n,o}}{\kappa_{n,o,s} + w_{n,o}} \right)$$

$$\tilde{\kappa}_{n,s} = C_{n,s}(\mathbf{w}, \mathbf{k}) \times \sum_{o=1}^{O} \left(\beta_{o,s} \frac{\kappa_{n,o,s}}{\kappa_{n,o,s} + w_{n,o}} \right)$$

$$C_{n,s}(\mathbf{w}, \mathbf{k}) = \left(\frac{\prod_{o=1}^{n} (\kappa_{n,o,s} + w_{n,o})^{\beta_{o,s}}}{\prod_{o=1}^{n} \beta_{o,s}^{\beta_{o,s}}} \right)$$

$$\tilde{E}_{n,s} = \prod_{o=1}^{O} (E_{n,o,s})^{\beta_{o,s}}$$

- *Data Required:* $[E_{n,o,s}]_{\forall n,o,s}$, $[w_{n,o}]_{\forall n,o}$[12], $[\beta_{o,s}]_{\forall o,s}$
- *Calibration Required: None.*
- *Data obtained:* $[\kappa_{n,o,s}]_{\forall n,o,s}$, $[\tilde{E}_{n,s}]_{\forall n,s}$, $[\tilde{\kappa}_{n,s}]_{\forall n,s}$, $[\tilde{w}_{n,s}]_{\forall n,s}$

3. *We back out α_s as nation-wide share of industry s in total revenue,*

$$\alpha_s = \frac{\sum_n (\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s} - X_{n,s}}{\sum_{n,s} (\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s} - X_{n,s}}$$

- *Data Required:* $[\tilde{E}_{n,s}]_{\forall n,s}$, $[\tilde{\kappa}_{n,s}]_{\forall n,s}$, $[\tilde{w}_{n,s}]_{\forall n,s}$
- *Calibration Required: None.*
- *Data obtained:* $[\alpha_s]_{\forall s}$

4. *Back out model implied housing stock from*

$$r_n = \frac{\delta Y_n}{\overline{H}_n}$$

[12] Notice that while in the data $w_{n,o,s}$ are available the model only allows for wages that are heterogeneous across occupation-province segmented labor markets that is $w_{n,o}$

$$Y_n = \left(\sum_s \tilde{\kappa}_{n,s} \tilde{E}_{n,s} + \sum_{s \in \mathcal{S}} \sum_o w_{n,o} E_{n,o,s} - \frac{Y_n}{\sum_n Y_n} \sum_{n,s} X^F_{n,s} \right)$$

$$Y_n = \left(1 - \frac{\bar{X}}{\bar{Y}} \right) \left(\sum_s (\tilde{w}_{n,s} + \tilde{\kappa}_{n,s}) \tilde{E}_{n,s} \right)$$

where $\bar{X} \equiv \sum_{n,s} X^F_{n,s}$ *and* $\bar{Y} \equiv \sum_n Y_n$

- *Data Required:* $w_{n,o}, E_{n,o,s}, \beta_{o,s}, r_n$
- *Calibration Required:* $\delta = 0.23$ *following Combes et al. (2018) which in turn relies on the French ministry that oversees housing (CGDD; 2011)*
- *Data obtained:* $[\overline{H}_n]_{\forall n}$

5. *Back out matching parameters and model implied posted vacancies*

$$E_{n,o} = \Gamma_n L^{\chi}_{n,o} V^{1-\chi}_{n,o}$$

$$\frac{E_{n,o}}{L^{\chi}_{n,o}} = \Gamma_n V^{1-\chi}_{n,o}$$

Obtain Γ_n *via fixed effects. Finally, back out* $V_{n,o}$ *conditional on calibrating* χ

- *Data Required:* $[E_{n,o}]_{\forall n,o}, [L_{n,o}]_{\forall n,o}$
- *Calibration Required:* χ
- *Data obtained:* $[\Gamma_n]_{\forall n}$, $[V_{n,o}]_{\forall n,o}$

6. *Back out real hiring costs*

$$E_{n,o,s} = \frac{V_{n,o,s}}{V_{n,o}} \Gamma_{n,o} L^{\chi}_{n,o} V^{1-\chi}_{n,o} \quad \Rightarrow \quad \frac{E_{n,o,s}}{E_{n,o}} = \frac{V_{n,o,s}}{V_{n,o}} \quad \Rightarrow \quad V_{n,o,s} = V_{n,o} \frac{E_{n,o,s}}{E_{n,o}}$$

$$c_{n,o,s} = \frac{\kappa_{n,o,s} E_{n,o,s}}{V_{n,o,s}}$$

- *Data Required:* $[E_{n,o,s}]_{\forall n,o,s}, [E_{n,o}]_{\forall n,o}, [V_{n,o}]_{\forall n,o}, [\kappa_{n,o,s}]_{\forall n,o}$
- *Calibration Required: None*
- *Data obtained:* $[c_{n,o,s}]_{\forall n,o,s}$, $[V_{n,o,s}]_{\forall n,o,s}$

7. *Impute composite hiring cost and implied unemployment benefits*

$$\frac{\tilde{E}_{n,s} \tilde{\kappa}_{n,s}}{\tilde{V}_{n,s}} = \tilde{c}_{n,s}$$

$$b_{n,s} = 0.5 * (\tilde{w}_{n,s} - \tilde{\kappa}_{n,s})$$

$$\tilde{V}_{n,s} = \prod_{o=1}^{O} \left(\left(\frac{V_{n,o,s}}{V_{n,o}} \right)^{\frac{1}{1-\chi}} V_{n,o} \right)^{\beta_{o,s}}$$

- *Data Required:* $[\tilde{E}_{n,s}]_{\forall n,s}$, $[\tilde{\kappa}_{n,s}]_{\forall n,s}$, $[\tilde{w}_{n,s}]_{\forall n,s}$, $[V_{n,o,s}]_{\forall n,o,s}$
- *Calibration Required: None*
- *Data obtained:* $[\tilde{V}_{n,s}]_{\forall n,s}$, $[b_{n,s}]_{\forall n,s}$, $[\tilde{c}_{n,s}]_{\forall n,s}$

8. *We estimate productivity* $[A_{n,s}]_{\forall n,s \in \mathcal{N} \times \mathcal{S}}$ *from Market clearing condition where Total Revenue is taken from Balance sheet data.*

$$(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s} = \left(\sum_{i \in \mathcal{N}} \pi_{n,i,s} \alpha_s Y_i \right) + X^F_{n,s}$$

$$\pi_{n,i,s} = \frac{A^{\theta_s}_{n,s}((\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\zeta_{ni})^{-\theta_s}}{\Phi_{i,s}}$$

- *Data Required:* $[\tilde{E}_{n,s}]_{\forall n,s}$, $[\tilde{\kappa}_{n,s}]_{\forall n,s}$, $[\tilde{w}_{n,s}]_{\forall n,s}$, $[X^F_{n,s}]_{\forall n,s}$, $[\alpha_s]_{\forall s}$, $[Y_n]_{\forall n}$, $[\zeta_{ni}]_{\forall ni}$
- *Calibration Required:* $[\theta_s]_{\forall s}$
- *Data obtained:* $[A_{n,s}]_{\forall n,s}$

9. *Back out* $\lambda_{n,o}$ *defined as*

$$w_{n,o} \sum_s E_{n,o,s} = \lambda_{n,o} \sum_s \beta_{o,s} (\tilde{w}_{n,s} + \tilde{\kappa}_{n,s}) \tilde{E}_{n,s}$$

- *Data Required:* $[\beta_{o,s}]_{\forall o,s}$,$[w_{n,o}]_{\forall n,o}$,$[E_{n,o,s}]_{\forall n,o,s}$,$[\tilde{E}_{n,s}]_{\forall n,s}$, $[\tilde{\kappa}_{n,s}]_{\forall n,s}$, $[\tilde{w}_{n,s}]_{\forall n,s}$,
- *Calibration Required: None.*
- *Data obtained:* $[\lambda_{n,o}]_{\forall n,o}$

Code 2 (Calibration Steady State)**.** *The calibration obtains the remaining parameters of the migration module, that is the amenity values,* $[\varrho_{n,o}]_{\forall n,o}$, *given the estimated switching costs and the data from the static equilibrium.*

- *Use transition equation for steady state and pick amenities such that given the observed allocation of labor the following equations are satisfied,*

$$\bar{L}_{n',o'} = \sum_{n,o} \mu^{o,o'}_{n,n'} \bar{L}_{n,o}$$

$$\mu_{n,n'}^{o,o'} = \frac{\exp\left(1/\nu\left(\beta\mathbb{E}_t V_{n',o'}(\bar{\Omega}) - \tau_{n,n'}^{o,o'}\right)\right)}{\sum_{n'',o''} \exp\left(1/\nu\left(\beta\mathbb{E}_t V_{a,b}(\bar{\Omega}) - \tau_{n,n''}^{o,o''}\right)\right)}$$

$$\bar{V}_{n,o}(\bar{\Omega}) = U(v_{n,o}(\bar{\Omega})) + \nu \log\left[\sum_{n',o'} \exp\left(1/\nu\left(\beta\mathbb{E}_t \bar{V}_{n',o'}(\bar{\Omega}) - \tau_{n,n'}^{o,o'}\right)\right)\right]$$

$$U(v_{n,o}(\Omega_t)) \equiv \frac{\varrho_{n,o} v_{n,o}}{P_n^{1-\delta} r_n^{\delta}}$$

2.10 Algorithm: Counterfactual Static Equilibrium

Algorithm 1 (Solving for the counterfactual static Equilibrium)**.** *Conditional on labor allocations and foreign demand the new equilibrium can be calculated in two steps. First by calculating sector-level outcomes and then by calculating occupation level outcomes. The algorithm implements the following steps:*

1. *Static Equilibrium (sector-level outcomes)*

   (a) *Notice that the static equilibrium can be expressed in terms of $\tilde{V}_{n,s}$ as unknown variables only which we can solve for from,*

$$(2b_{n,s} + 2\tilde{\kappa}_{n,s})\tilde{E}_{n,s} = \left(\sum_{i\in\mathcal{N}} \pi_{n,i,s}\alpha_s Y_i\right) + X_{nF,s}$$

$$\pi_{n,i,s} = \frac{A_{n,s}^{\theta_s}((\tilde{\kappa}_{n,s} + b_{n,s})\zeta_{ni})^{-\theta_s}}{\Phi_{i,s}}$$

$$Y_n = \frac{\bar{Y}}{\bar{Y} + \bar{X}}\left(\sum_s (2b_{n,s} + 2\tilde{\kappa}_{n,s})\,\tilde{E}_{n,s}\right)$$

$$\tilde{\kappa}_{n,s} = \frac{\tilde{c}_{n,s}\tilde{V}_{n,s}}{\tilde{E}_{n,s}}$$

$$\tilde{E}_{n,s} = \Gamma_n \tilde{L}_{n,s}^{\chi}\tilde{V}_{n,s}^{1-\chi}$$

$$\tilde{L}_{n,s} \equiv \prod_{o=1}^{O} (L_{n,o})^{\beta_{o,s}}$$

- *Data Required:* $[b_{n,s}]$, $[\tilde{c}_{n,s}]$, $[L_{n,o}]_{\forall n,o}$, $[\alpha_s]_{\forall s}$, $[\beta_{o,s}]_{\forall o,s}$, $[\Gamma_n]_{\forall n}$

- *Calibration Required:* χ, $[\theta_s]_{\forall s}$
- *Data obtained:* $[\tilde{V}_{n,s}]_{\forall n,s}$, $[\tilde{E}_{n,s}]_{\forall n,s}$, $[\tilde{\kappa}_{n,s}]_{\forall n,s}$, $[\tilde{L}_{n,s}]_{\forall n,s}$, $[\tilde{w}_{n,s}]_{\forall n,s}$, $[Y_n]_{\forall n}$
- *Important Note:* $[\tilde{V}_{n,s}]_{\forall n,s}$ *can only be identified in relative terms, i.e. vis-a-vis a baseline value. However, the whole set has to be consistent with the following condition:*

$$\sum_o w_{n,o} E_{n,o} = \sum_o \lambda_{n,o} \sum_s \beta_{o,s}(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s}$$

(b) *Calculate counterfactual rents from*

$$r_n = \frac{\delta Y_n}{\overline{H}_n}$$

- *Data Required:* $[Y_n]_{\forall n}$, $[H_n]_{\forall n}$
- *Calibration Required:* δ
- *Data obtained:* $[r_n]_{\forall n}$

2. *Static equilibrium (occupation-level outcomes)*

(a) *Use Labor Market clearing equation at the sector and occupation-region level and substitute out* κ *and employment levels to get an equation in* $V_{n,o,s}$ *and* $w_{n,o}$ *only - solve for* $V_{n,o,s}$ *and* $w_{n,o}$

$$(w_{n,o} + \kappa_{n,o,s})E_{n,o,s} = \beta_{o,s}(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s}$$

$$w_{n,o} \sum_s E_{n,o,s} = \lambda_{n,o} \sum_s \beta_{o,s}(\tilde{w}_{n,s} + \tilde{\kappa}_{n,s})\tilde{E}_{n,s}$$

$$\kappa_{n,o,s} E_{n,o,s} = c_{n,o,s} V_{n,o,s}$$

$$E_{n,o,s} = V_{n,o,s} \Gamma_n V_{n,o}^{-\chi} L_{n,o}^{\chi}$$

$$\kappa_{n,o,s} = \frac{\sum_s V_{n,o,s}}{V_{n,o,s}} \frac{c_{n,o,s}}{\Gamma_n} V_{n,o}^{\chi} L_{n,o}^{-\chi}$$

- *Data Required:* $[\tilde{w}_{n,s}]_{\forall n,s}$, $[\tilde{\kappa}_{n,s}]_{\forall n,s}$, $[\tilde{E}_{n,s}]_{\forall n,s}$, $[L_{n,o}]_{\forall n,o}$
- *Calibration Required:* χ
- *Data obtained:* $[V_{n,o,s}]_{\forall n,o,s}$, $[w_{n,o}]_{\forall n,o}$, $[E_{nos}]_{\forall n,o,s}$, $[\kappa_{n,o,s}]_{\forall n,o,s}$

2.11 Algorithm: Counterfactual Dynamics and Steady State

Algorithm 2 (Solving for Dynamics/Steady State)**.** *We solve for the counterfactual dynamics in the following way,*

1. *Guess initial sequence of cross sectional wages* $[w_{n,o,t}]_{\forall n,o,t}$*, rental rates and unemployment probabilities*
2. *Solve value functions*
3. *Calculate implied labor flows and implied labor stocks* $[L_{n,o,t}]$
4. *Calculate static equilibrium and update wages, rental rates and unemployment probabilities*
5. *Go back to step [1]. Repeat until convergence.*

Chapter 3

Globalization and Taste Heterogeneity: Evidence from Hollywood

Konrad Adler and Simon Fuchs

3.1 Abstract

To what extent is the set of products that are available to a country driven by the composition of international markets? We develop a quantitative framework to determine and map the similarity between countries from observed market shares of identical products across markets. We apply our framework to the global movies market where we can abstract from price competition and observe identical products and their market shares across countries. As an application we evaluate the hypothesis that the observed large increase in the revenue share of sequels has been due to shifts in the composition of global demand away from traditional western markets inducing demand risk and with sequels providing insurance. While we find substantial shifts in the profit space and lower risk associated with sequels, our simulations suggest that the risk due to taste heterogeneity in the movies market is quantitatively insufficient to explain the increase in the revenue of sequels, suggesting that other forces such as scale economies might be at play.

3.2 Introduction

2015 was a successful year for Hollywood, with the global box office increasing to 38.3 billion dollars (MPAA 2015) and with the top performing American productions capturing the largest part of the revenue. Creativity, however, seems to have reached low-point: Amongst the top 10 highest performing movies[1] only one movie can be categorized as original work with no preceding media products. Among the rest we find six franchises (James Bond, Hunger Games, Star Wars, Jurassic Park, Avengers/Marvel, Fast and Furious), a book adaptation (The Martian), and a remake (Cinderella). In 2015 sequels or franchises made up almost 60 percent of the yearly revenue up from around 30 percent at the beginning of the 2000s. Simultaneously, the composition of the global movie market has shifted dramatically with the US and Western European markets becoming less important compared to Asia, Eastern Europe and other markets. To what extent and how did these demand shifts influence creativity and the product mix in Hollywood? More generally, how does the composition of global markets with countries affect the product mix?

Traditional trade theory has often abstracted from this question because a measure of taste heterogeneity across countries is needed to provide an answer and this has been difficult to estimate.[2] Yet global market integration and the increasing importance of Emerging countries, in particular the rise of China, most likely influence the product mix and thus affect consumer welfare gains from globalization.

We suggest that the global box office offers a convenient setting to examine the mechanisms that link taste heterogeneity across countries to the product mix. Several reasons make the international movies market particularly attractive for this exercise. Firstly, studios produce and then distribute a single product with fixed observable characteristics that is being released across multiple markets without (major) adjustment. Market specific revenues and thus market shares are readily available and we collected a unique and large dataset combining several online sources. To the extent that the dataset is complete, the available product bundle (that is alternative movies released at the same time)

[1]The top 10 at the global box office in 2015 was as follows: 1. Star Wars: The Force Awakens (937 MM), 2. Jurassic World (652 MM), 3. Avengers: Age of Ultron (459 MM), 4. Inside Out (356 MM), 5. Furious 7 (353 MM), 6. Minions (336 MM), 7. The Hunger Games: Mockingjay - Part 2 (281 MM), 8. The Martian (228 MM), 9. Cinderella (201 MM), 10. Spectre (200 MM)

[2]While market shares of identical products across countries could be informative about the taste for specific characteristics of given products, differences in market structure and available product bundles have made it difficult to directly estimate these differences. Furthermore, while trade data might be available at a relatively high level of disaggregation even at the 8-digit level HS code there can still be substantial heterogeneity in terms of product characteristics and quality.

can be readily constructed while the market structure is such that price competition between movies at the boxoffice is of little importance. This implies that - conditional on controlling for selective release of movies across markets - the covariation of market shares across countries is informative about taste similarity between countries.

Our analysis proceeds in four steps: In a first step we present two stylized facts about the global box office: (1) The share of sequels and adapted content has increased dramatically over recent decades, (2) the global movies market has experienced major shifts away from the traditional 'Western' target audience.

In a second step we use a random utility framework where we decompose the global appeal of a product, i.e. movie, across all markets and the relative appeal between markets by introducing an artificial taste space with fixed positions of countries where market shares are decreasing in the distance between a movie from the location of individual countries. This relative distance of countries is pinned down by the covariation of market shares across movie observations and we estimate a two dimensional space with all country locations from the box office revenues of more than 1000 movie released since 2001. The framework is reminiscent of the address type models explored amongst others by Anderson et al. (1989), but rather than mapping heterogeneity in demand into observed characteristics we focus on unobservable heterogeneity.

In a third step we argue that studios make their production decision by picking locations in this abstract taste space, but face uncertainty in the form of a displacement shock along the two dimensions of the taste space. The variance of the shock can be estimated using a moment inequality approach focusing on movies that are - given their production budget - ex post not profitable and determining the closest position for which the movie would have had positive expected profits - the distance to the observed position can be used to determine a lower bound for the variance of the displacement shock. This procedure can also be implemented across movie types. We find substantial uncertainty across all movie types, but less uncertainty for sequels which tend to be more closely placed in the vicinity of their predecessors. This suggests that in a more multi polar market sequels might be advantageous by reducing the 'spatial' risk in the abstract taste space.

Finally, we conduct counterfactuals by simulating the global revenue share of sequels in the absence of changing market conditions.

This paper contributes to two literatures: The first being the traditional literature on the movies market and cultural economics more broadly speaking as surveyed recently by McKenzie (2012). We abstract from much of the usual features that are the center of attention of other studies and instead focus on the interaction between the global market place for movies, implied demand uncertainty and choice of type of movies in an abstract yet tractable setting.

The second literature is a nascent literature on taste heterogeneity and supply in global markets, where the closest study is a recent examination of the global cars market by Coşar et al. (2018). They examine taste heterogeneity estimated in a BLP framework with a particular focus on the question to what extent preferences for the home brand can account for the home market effect commonly observed in the data. While they focus on the home market effect, our focus is on understanding how shifts in global markets affect products supplied to all markets.

The paper proceeds as follows: The next section introduces the data and stylized facts, section three introduces the quantitative model, the fourth section introduces the estimation and finally section five presents the details of the simulation of movie markets and describes the results for a counterfactual simulation of the movies market in the absence of the rise of China. The final section concludes.

3.3 Data and Stylized Facts

Data We use data from BoxOfficeMojo which has information about the production cost and the boxoffice revenues for a set of countries for each movie. A second data source is TheNumbers which has detailed information about the source of the screenplay of a movie, i.e. original screenplay, book adaptation. We use the "connections" section from The Internet Movie Database to find the title and release year of sequels.

The sample period is 2001-2017. We exlude movies with missing information about the production budget and no information about boxoffice revenue outside the US. Finally we restrict our sample to movies that generated at least $80 million in boxoffice revenue. The final dataset has 1009 movies including boxoffice revenue data from up to 59 countries. Table 3.1 shows summary statistics of our main variables. Out of the 254 sequels in our sample we are able to match 142 to their prequel[3].

The left hand side graph in figure 3.1 shows that a higher production budget results in a larger boxoffice revenue on average but the remaining uncertainty in movie production is considerable. Film studios are using different ways to reduce this uncertainty. One important way is to produce a movie based on a theme or story that has been successful

[3]Most unmatched sequels have a prequel released before our sample period starts

previously. Examples of this "recycling" of content are remakes, sequels and adaptations of books and TV series. In this paper we focus on sequels as a way to reduce uncertainty. The right hand side graph in figure 3.1 compares the distribution of profits for sequels and non-sequels. Sequels indeed reduce the risk for a loss and earn on average a higher profit compared to non-sequels.

Stylized facts Our first stylized fact is shown on the left hand side graph in figure 3.2: between 2001 and 2017 the share of sequels in total boxoffice revenue has increased from around 30 to almost 50%. When adding remakes and other non-original content the increase is even more pronounced. At the same time the share of non-US boxoffice revenue in total revenue increased from 40 to 65%. This is our second stylized fact. The increase in the share of non-US boxoffice revenue comes from emerging countries. All regions except Western Europe and other developped countries (Japan, South Korea, Australia, New Zealand) become more important between 2001 and 2013 as shown in figure 3.3. After 2013 the increase in the non-US revenue share is mainly due to the increasing importance of Asia.

3.4 Model

We present a random utility framework where we decompose the global appeal of a product, i.e. movie, across all markets and the relative appeal between markets by introducing an artificial taste space with fixed positions of countries where market shares are decreasing in the distance between a movie from the location of individual countries. This approach is closely related to the address based approach where both consumers and products are represented by a location in a characteristic space and where the consumers location pins down his optimal product as explored by - among others - Anderson et al. (1989). On the supply side, firms choose their location in that characteristics space. Rather than defining the dimensions of this space in terms of actually observed characteristics, we map the observed demand patterns into an unobserved heterogeneity space, which we call taste space.

3.4.1 Demand side

We posit a random utility model (RUM), where the utility of consumer i who chooses product j (in our case a specific movie), is given by,

$$U_{ij} = \alpha_i(w_i - p_j) + \xi_j + \epsilon_{ij}$$

$$\epsilon_{ij} = \sum_l \gamma_l |c_j^l - c_i^l| + u_{ij}$$

where w_i is the income of the consumer, p_j the price of product, ξ_j an unobservable that is constant across all consumers. The error term is structurally decomposed into a mean zero, double exponentially distributed error term , u_{ij} and a term that measures the distance in a n-dimensional taste space whose dimensions are denoted by l, where the vector c_j denotes the position of the product and the vector c_i the position of the consumer. Consumers will be distributed around a midpoint for each given country as described below. The utility is a decreasing function in the distance - that is the $L1$ norm - between consumer and product location, where the sensitivity to the distance along each dimension is measured by γ. The probability for consumer i to choose product j is given by the following,

$$Pr(i,j) = Pr(u_{ij} > u_{ik}, \forall k \neq i)$$

which under symmetric prices (p_i=p_k) translates into,

$$= Pr(\xi_j + \sum_l \gamma_l |c_j^l - c_i^l| + u_{ij} > \xi_k + \sum_l \gamma_l |c_k^l - c_i^l| + u_{ik}$$

$$= Pr\left(\xi_j + \sum_l \gamma_l |c_j^l - c_i^l| - (\xi_k + \sum_l \gamma_l |c_k^l - c_i^l|) > u_{ik} - u_{ij}\right)$$

which assuming logit errors gives us the familiar reduced form for the probability of consumer i choosing product j,

$$Pr(i,j) = \frac{\exp(\xi_j + \sum_l \gamma_l |c_j^l - c_i^l|)}{\sum_k \exp(\xi_k + \sum_l \gamma_l |c_k^l - c_i^l|)}$$

The market share of product j across all the consumers in a given country c, that is the set of consumers I_c, is given by the integral across all consumers in that country,

$$s_j^c = \frac{p_j q_j}{\sum_k p_k q_k} = \int_{i \in I_c} \frac{\exp(\xi_j + \sum_l \gamma_l |c_j^l - c_i^l|)}{\sum_k \exp(\xi_k + \sum_l \gamma_l |c_k^l - c_i^l|)} di$$

While for each individual consumer the independence of irrelevant alternatives property holds, for the aggregate market share that is not the case. Specifically, the relative market share of two products, depends on how they affect different consumers in the consumer space individually, and how that aggregates. If a certain group of consumers is already well provided for, with many products in their close vicinity, locating an additional product there might bring lower revenues, than serving a less competitive section in the consumer space.

3.4.2 Supply side

There is a large number of entrepreneurs each endowed with a movie script. A movie script, j, is defined by a triplet consisting of an expected location in the taste space, $\tilde{c}_j$, the production cost, b_j, and the content type, that is a variable s_j which assumes the value 1 if the script is a sequel. Uncertainty comes from a taste shock for each dimension where the vector of disturbances is denoted by ε_j which is the difference between the expected taste location of the script $\tilde{c}_j$ and the actualized - ex-post - taste location c_j of the movie, i.e.

$$c_j = \tilde{c}_j + \varepsilon_j \quad \text{where} \quad \begin{cases} \varepsilon \sim G(0, \Sigma_{\text{sequel}}) & \text{iff} \quad s_j = 1 \\ \varepsilon \sim G(0, \Sigma_{\text{no sequel}}) & \text{iff} \quad s_j = 0 \end{cases}$$

where the disturbance vector ε is drawn from a zero mean distribution that features a lower variance if the script is a sequel rathern than a original script. In addition a sequel's taste location is linked to its prequel. We assume entrepreneurs to make no mistake about the taste location of their script on average for both sequels and non-sequels, which implies a zero mean disturbance. Entrepreneurs are risk-neutral and maximize profits by making a discrete choice to produce a script if and only if expected profits are positive, i.e.

$$\mathbb{E}\pi(\tilde{c}_j, b_j, s_j) \geq 0$$

where $\mathbb{E}\pi((\tilde{c}_j, b_j, s_j)$ refers to the expected profits of a script with expected location $\tilde{c}_j$, budget b_j and of type s_j. Each entrepreneur takes the location and production choice of other entrepreneurs as given.

3.5 Estimation

The estimation of the quantitative model proceeds in two steps. In a first step we will exploit co-variance of revenue shares of movies across countries together with the framework introduced above to estimate the relative location of countries towards each other. In a second step we will then use a moment inequality approach to exploit the assumption that only movies that are in expected terms profitable would have been produced to back out the uncertainty in the taste space associated with the production of different types of movies (notably, sequels vs non sequels).

3.5.1 Demand side

Estimation procedure We assume that consumers within a given country c, are distributed according to a normal distribution, with a fixed mean and variance, independently across all dimensions, that is,

$$c_i^l \sim N(\mu_c^l, \sigma_c^l) \quad for \quad i \in I_c$$

The model can then be estimated by simulated methods of moments. That is we simulate markets by drawing consumer locations given the mean and variance, and then choose ξ_j to minimise the distance between observed and market shares for the products within countries and across countries, that is we target s_j^c for all j and for all c that we observe. The objective to be minimized is as follows,

$$\eta_{j,c,t}\left(\mu_c, \sigma_c, \xi_j, \gamma, c_j\right) = \frac{q_j}{\sum_k q_k} - \int_{i \in I_c} \frac{\exp(\xi_j + \sum_l \gamma_l |c_j^l - c_i^l|)}{\sum_k \exp(\xi_k + \sum_l \gamma_l |c_k^l - c_i^l|)} di$$

Since we can only determine the relative distance between countries and thereby the relative location of countries and movies in the taste space, some normalization is necessary to obtain a well defined taste space and all positions. We normalize the space by calibrating the γ parameters and choosing a location for the most important market - that is the US.[4]

[4]For computational convenience we also impose lower and upper bounds and choose a value for γ such that all countries are contained within that space. Effectively we estimate a two dimensional space between 0 and 1 for both dimensions and assign the US the midpoint position at (.5, .5).

Results The estimated taste space is depicted in figures 3.4 and 3.5. The figures present the location of individual countries as well as simulated revenue contours for 2001 and 2017. Revenue contours are obtained by introducing a grid of location at which we place homogenous movies with identical ξ_j and production budget. The contours can then be obtained by calculating the profit given the observed market size for each country (i.e. the total observed box office revenues in a given year and country) for each location along the grid and smoothing across them. The traditional and more established markets are tightly clustered together in close proximity to the US market at the mid point of the space. Asian markets cluster in the North-West quadrant of the space. Asian markets and particularly China exert a pull on the profit space that becomes visible at the end of the sample in 2017.

In figures 3.6 and 3.7 we furthermore map the revenue difference between a movie at precisely the indicated location and the expected profit of a movie that is facing the probability of a small disturbance into all directions. In the baseline year it is particularly important to have precision in the center of the taste space to match exactly the most important markets and not to lose market shares to competing movies - this creates an advantage for sequels with potentially lower production uncertainty. Towards the end of the sample in 2017 we observe that there are additional zones where lower variance movies are advantageous, creating more demand for sequels. While this illustrates the possibility for taste heterogeneity to create demand for lower variance products, the quantitative impact depends on the precise size of different parameters, such as the distance penalty parameter as well as the relative and absolute size of the shocks across product groups.

We also obtain the taste locations of movies. We regress taste space coordinates of movies on their observable characteristics. Table 3.3 shows that movies with high values in the second taste dimension for example, tend to be IMAX format, Horror and Science-fiction movies.

3.5.2 Supply side

We estimate Σ, the variance-covariance matrix of the taste shock, separately for sequels and non-sequels. For sequels we compute the variance of the distance in the taste space between each sequel and its prequel:

$$\Sigma_{\text{sequel}} = E[(c_{\text{sequel}} - c_{\text{prequel}})^2]$$

For non-sequels we search for a sequence of taste shocks such that: first, the likelihood of observing the actual taste location is maximized and second, expected profits at the script location, i.e. actual location plus taste shock, are positive. We assume taste shocks to be jointly normally distributed and to have zero covariance between taste dimensions. Because only the absolute value of the distance between expected and actual taste location matters for Σ we maximize the likelihood with respect to the distance d and compute expected profits for all combinations of signs of d.

$$\max_{\{d^2\}} \sum_i logL(|d|, \Sigma_{\text{no sequel}})$$
$$s.t.\pi(\tilde{c}_i) \geq 0$$

where $\pi(\tilde{c}_i)$ refers to the expected profit at the initially chosen location, the constraint requires the initial location to have expected positive profits[5], and where $\Sigma = E[d^2]$ is the implied variance-covariance matrix to be estimated.

Results The estimated variance along each dimension for both sequels and non sequels is reported in table 3.2. Sequels have substantially lower estimated production uncertainty along both dimensions.

3.6 Simulation & Counterfactual

Sequels offer a trade off: They promise lower production uncertainty but at the price of locating a movie close to the predecessor at what is potentially a less than optimal location in the profit space. We argue that when market shares shift and the profit space becomes riskier, then sequels might become more attractive than original productions, shifting the global product mix and the observed revenue share for that type of product.

[5]More precisely, because the pdf is symmetric the sign of the difference between the realized taste and the script location does not matter for the log likelihood, that is the actual constraint is as follows,

$$max(\pi(\tilde{c}_i^+, \Sigma_{\text{no sequel}})], \pi(\tilde{c}_i^-, \Sigma_{\text{no sequel}})]) \geq 0$$
$$\tilde{c}_i^+ = c_i + d$$
$$\tilde{c}_i^- = c_i - d$$
$$0 \leq d^2$$

In this section we employ the quantitative model and estimated taste space to examine that quesiton. We first describe how to simulate the model for a specific period, using the estimated location of consumers of individual countries as well as the uncertainty of sequels and non sequels as inputs and determining a set of movies that clears populates the taste space and clears the market conditional on the market size and the arrival frequency of sequels as script ideas. Secondly, we show how to use this approach to obtain a simulated time series for the sequel revenue share that tracks the revenue share in the data from the early 2000s to 2017. Finally, we demonstrate the counterfactual revenue share if the distribution of income across countries would have remained constant at 2001 levels.

3.6.1 Static Simulation and Time Series

To simulate the movie market for a given period we take the country locations and budgets as given. We start by drawing a large number[6] of movie scripts which are identical in their global appeal, and budget, i.e.

$$\xi_j = \bar{\xi} \qquad \forall j$$
$$b_j = \bar{b} \qquad \forall j$$

Scripts do however differ in their location across the taste space. The location of different scripts is drawn uniformly along both dimensions of the space. We then iterate over two steps: We first calculate the profits under the assumption that all scripts under consideration are being produced and then we drop the movie with the largest losses. We continue the iteration until all scripts have positive expected profits. Finally, we draw the taste shocks and calculate ex post locations and profits.

With regard to sequels, we introduce a paramters μ_t that determines the share of scripts that are sequels. In practice, this parameter is calibrated to match the observed revenue share of sequels. For sequels, rather than drawing the location randomly, we select the location from a previous successful (i.e. positive ex post profits) movie.

Using the evolution of country specific market sizes (that is the total box office revenue) throughout the years, we generate year-by-year static simulations and backout a time series for revenue share of sequels for the period under consideration.

[6]In practice, we take the observed number of movies and add 60 additional scripts.

3.6.2 Counterfactual: Hollywood without China

To examine the impact of the changing composition of the global movies market we simulate a counterfactual where we keep the market sizes at 2001 levels but feed in the calibrated sequel script arrival share. The resulting time series is presented in figure 3.8. Surprisingly but consistent with the profit spaces depicted before, the shifts in market sizes have not induced a higher production of sequels. While these movements did tilt the product supply towards the asian markets, they did not increase spatial risk and therefore did not shift production towards sequels. Sequels instead must have other benefits that are orthogonal to the between country taste heterogeneity explored here.

3.7 Conclusion

We propose a methodology to estimate taste similarity between countries and estimate a taste space using a newly assembled data on international box office. As an application we investigate how a dramatic change in the movie market structure, namely the increasing importance of Emerging countries for the international boxoffice revenue, can affect the mix of product types, in our case original movies and sequels. We show first, that sequels have a lower production uncertainty and second, that the revenue map has become steeper between 2001 and 2017 because of the the change in market structure.

We also simulate the revenue share of sequels once with the actual and once holding the movie market structure constant at the 2001 level. Our current results show only a small influence of market structure on the share of sequels.

3.8 Tables

Table 3.1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max	P25	P50	P75
Production budget	1009	79.8	55.5	0	300	37	65	110
Total boxoffice	1009	272.1	248.8	52.8	2655.7	115.3	180	323
Profit	1009	192.3	218	-84.2	2418.7	67.1	117.9	227.2
Sequel	1009	.3						

Notes: All variables in $ million. Production budget and total boxoffice from BoxOfficeMojo. Profit is defined as Total boxoffice revenue minus production budget. Sequel is the revenue share of sequels.

Table 3.2: Relative Variance of Sequels vs Non-Sequels

	$\sigma^2_{\epsilon,1}$	$\sigma^2_{\epsilon,2}$	N
non sequels	0.23	0.22	755
sequels	0.02	0.04	254

Table 3.3: Taste space coordinates regression

	(1) taste dim 1	(2) taste dim 2
PG	-0.0700**	0.00921
	(-2.20)	(0.26)
R	-0.0644*	-0.0219
	(-1.90)	(-0.59)
Nominated	0.0193	0.00886
	(1.30)	(0.55)
Win	0.0000832	-0.0205
	(0.00)	(-1.00)
IMAX	-0.0150	0.0586***
	(-0.74)	(2.66)
Normal Image	0.0336*	-0.0177
	(1.74)	(-0.84)
Sequel=1	-0.0427***	0.00293

	(-3.67)	(0.23)
Runtime	-0.000933***	0.0000302
	(-2.88)	(0.09)
Adventure	-0.0833***	0.0399
	(-3.52)	(1.55)
Animation	-0.149***	-0.0693**
	(-5.40)	(-2.30)
Comedy	-0.0292	-0.0900***
	(-1.24)	(-3.49)
Documentary	-0.0179	-0.121
	(-0.19)	(-1.18)
Drama	-0.0150	-0.0897***
	(-0.59)	(-3.25)
Family	-0.0274	-0.125***
	(-0.79)	(-3.32)
Fantasy	-0.109***	0.0440
	(-3.72)	(1.37)
Foreign	-0.143**	-0.0829
	(-1.99)	(-1.05)
Horror	-0.0722***	0.0537*
	(-2.72)	(1.85)
Musical	-0.142***	-0.0755
	(-2.89)	(-1.40)
Romance	-0.0454	-0.0615
	(-1.08)	(-1.35)
Romantic Comedy	-0.100***	-0.0961***

	(-3.14)	(-2.76)
Sci-Fi	-0.0766***	0.0557*
	(-2.87)	(1.91)
Thriller	-0.000778	-0.0465*
	(-0.03)	(-1.68)
Western	-0.0613	-0.0511
	(-1.05)	(-0.80)
Year FE	Yes	Yes
r2	0.252	0.206
N	981	981

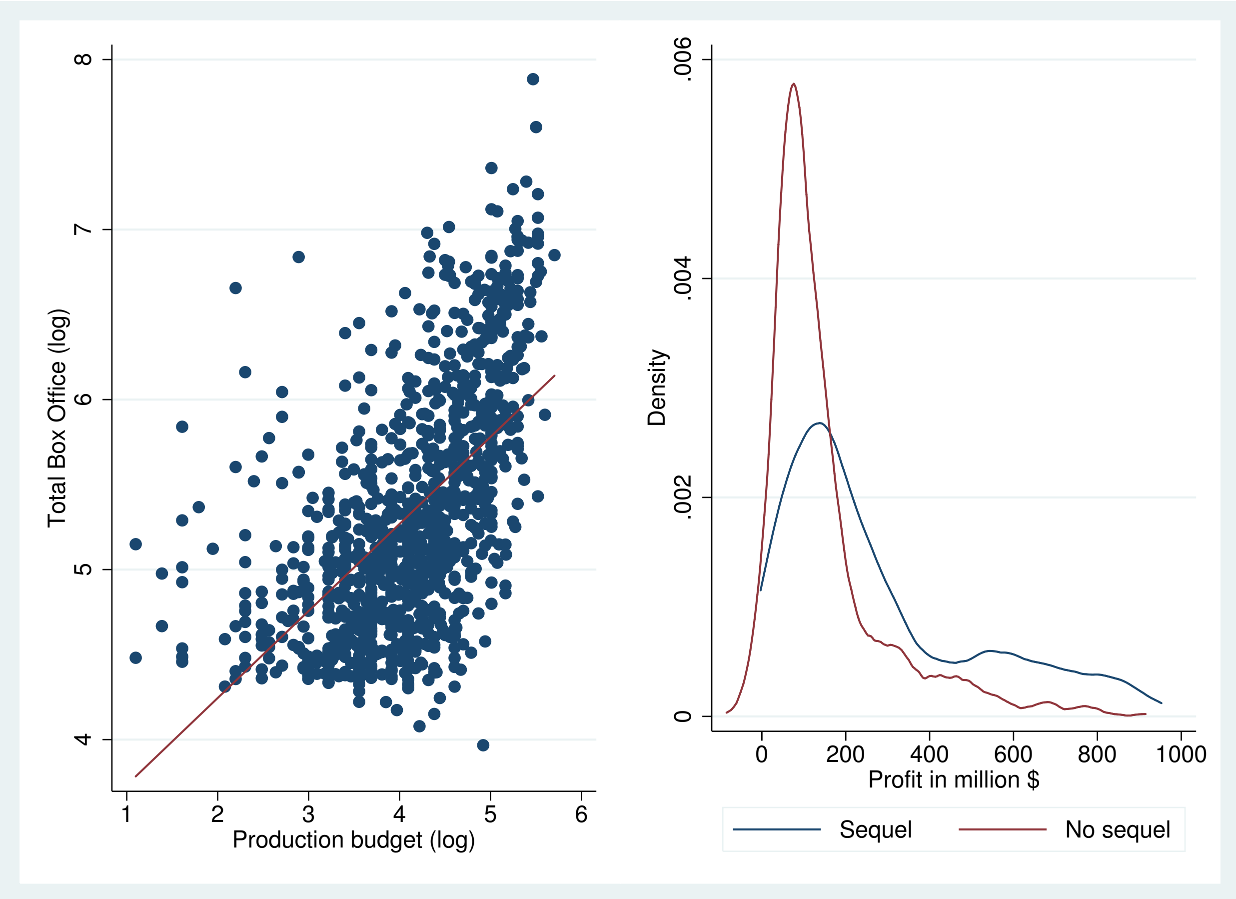
t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Regression of taste coordinates on observable movie characteristics: movie rating (PG: parental guidance suggested, R: restricted), Oscar nominated or win, screen format (IMAX, 3D, normal image), runtime in minutes and the genre of the movie.

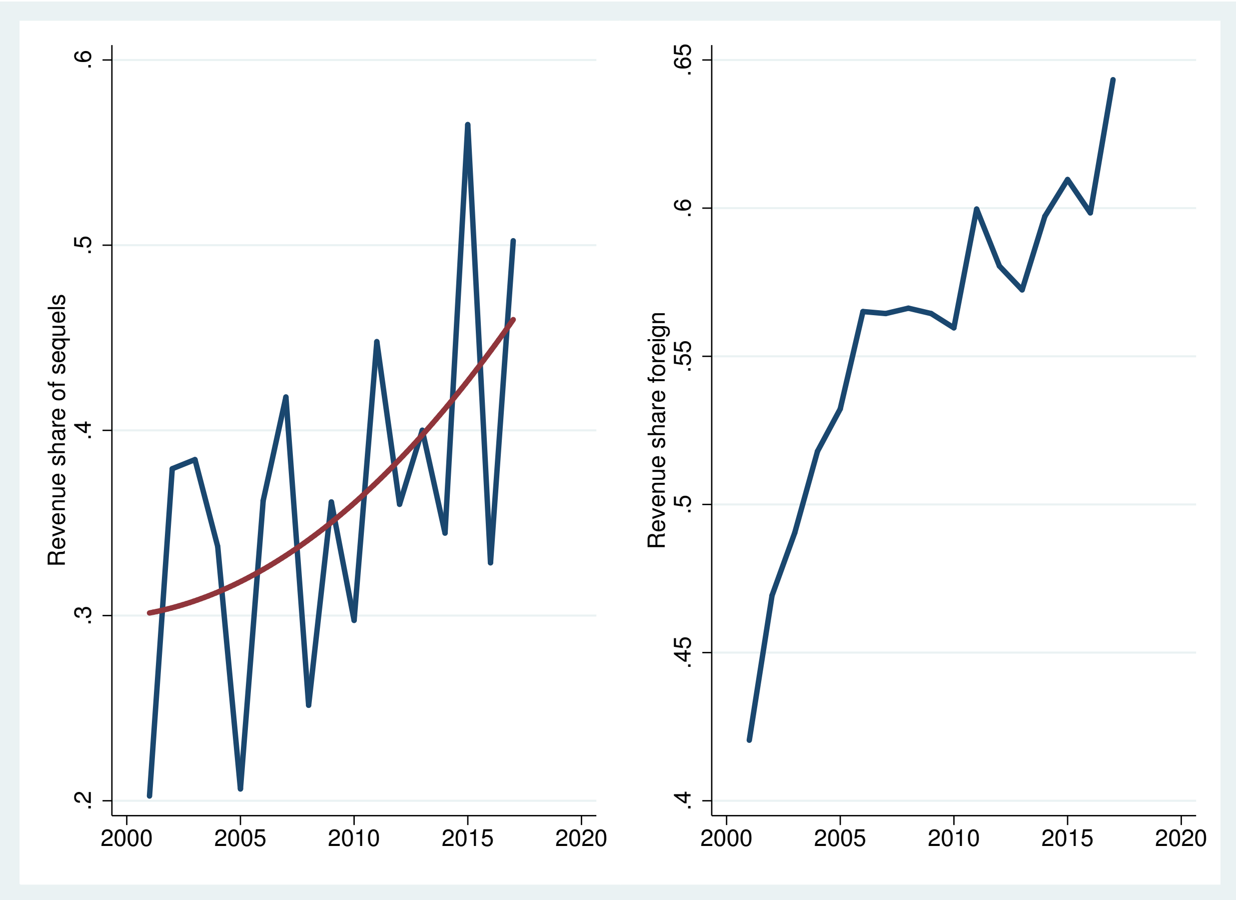
3.9 Figures

Figure 3.1: Production Budget and Profitability



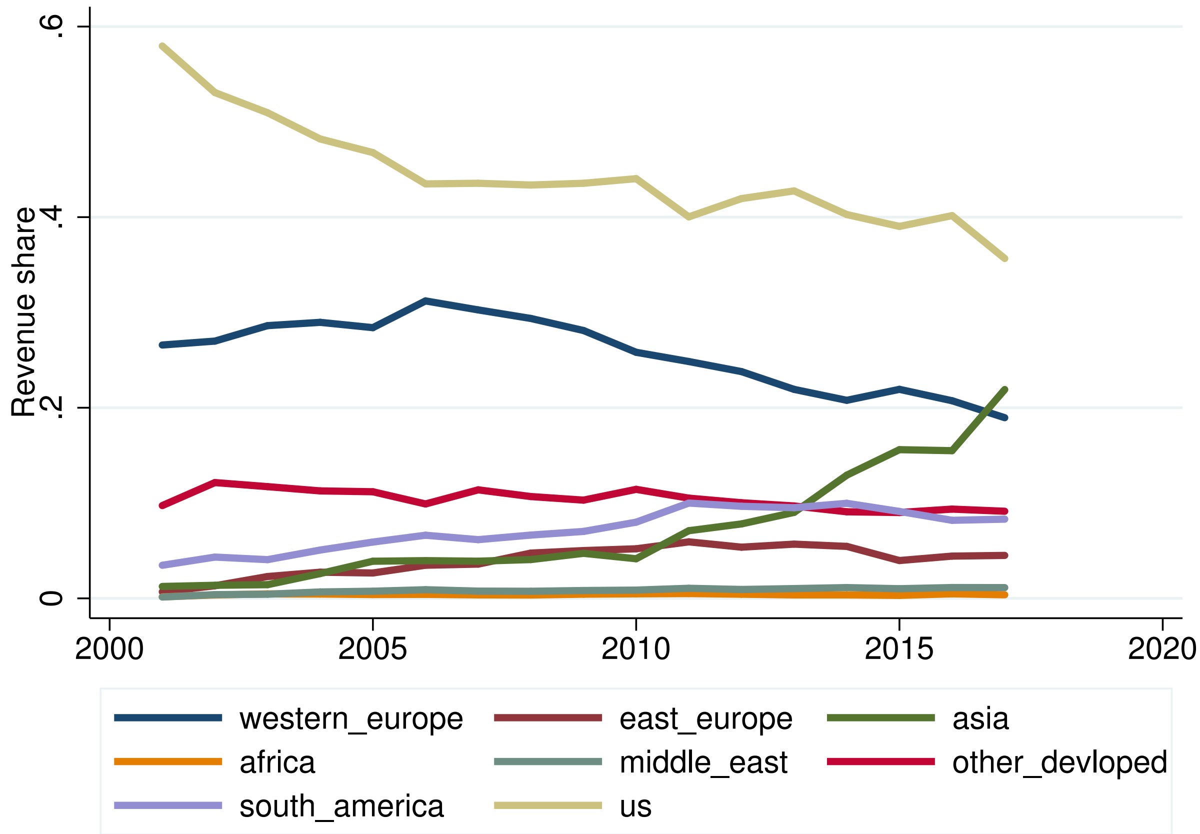
Notes: Left hand side: Relation between log production budget and log total boxoffice revenue with a linear fit in red *Right hand side:* Density of profits for sequels and non-sequels.

Figure 3.2: Sequel Revenue Share



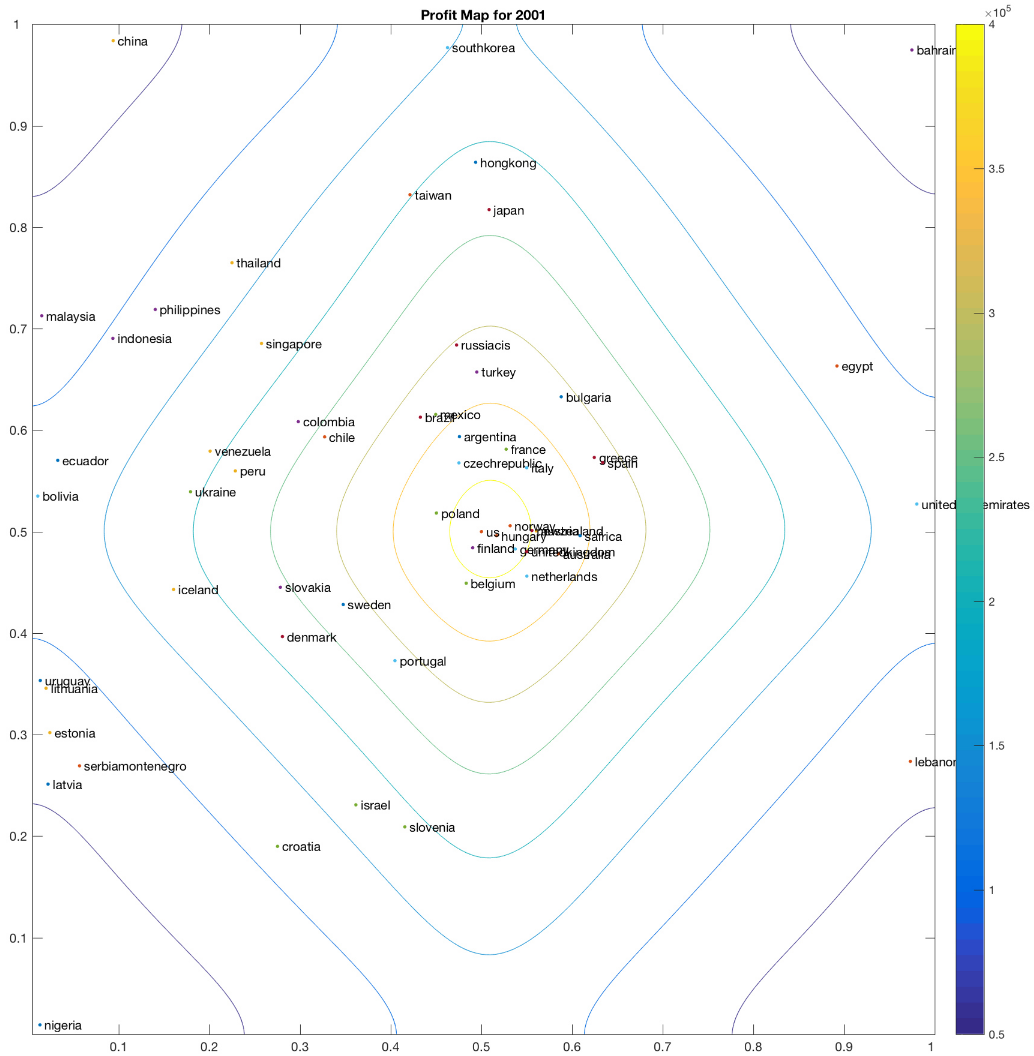
Notes: Left hand side: boxoffice revenue share of sequels *Right hand side:* share of non-US boxoffice revenue in total boxoffice revenue

Figure 3.3: Global Box Office



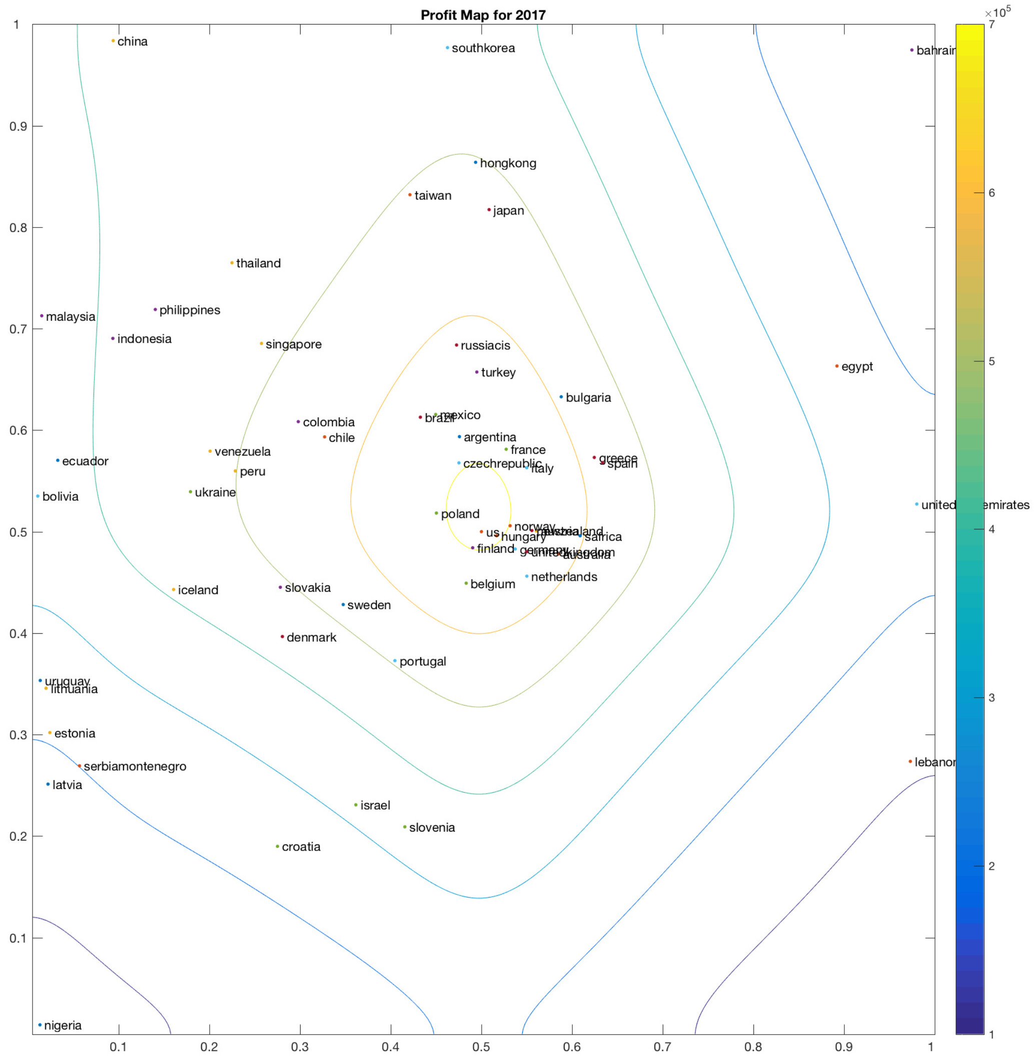
Notes: Share of worldwide boxoffice revenue by region. Other developed countries are: Japan, South Korea, Australia and New Zealand.

Figure 3.4: Estimated revenue space (2001)



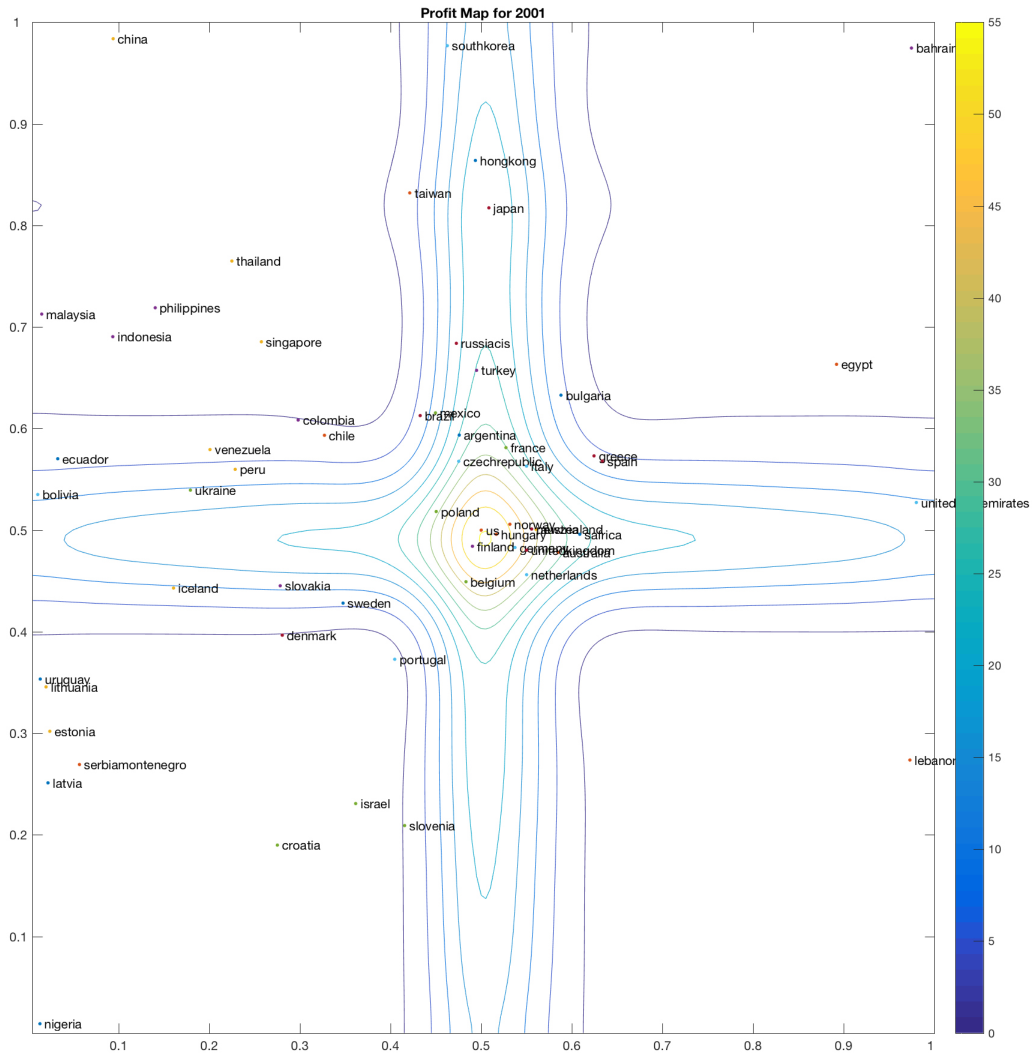
Notes: Revenue map for 2001 in the taste space. Dots are the average taste locations of countries, assumed to be constant over time, relative to the US which is normalized to be at taste position (0.5 0.5). Lines are iso-revenue lines for the year 2001 assuming competitor movies on an evenly spaced grid.

Figure 3.5: Estimated revenue space (2017)



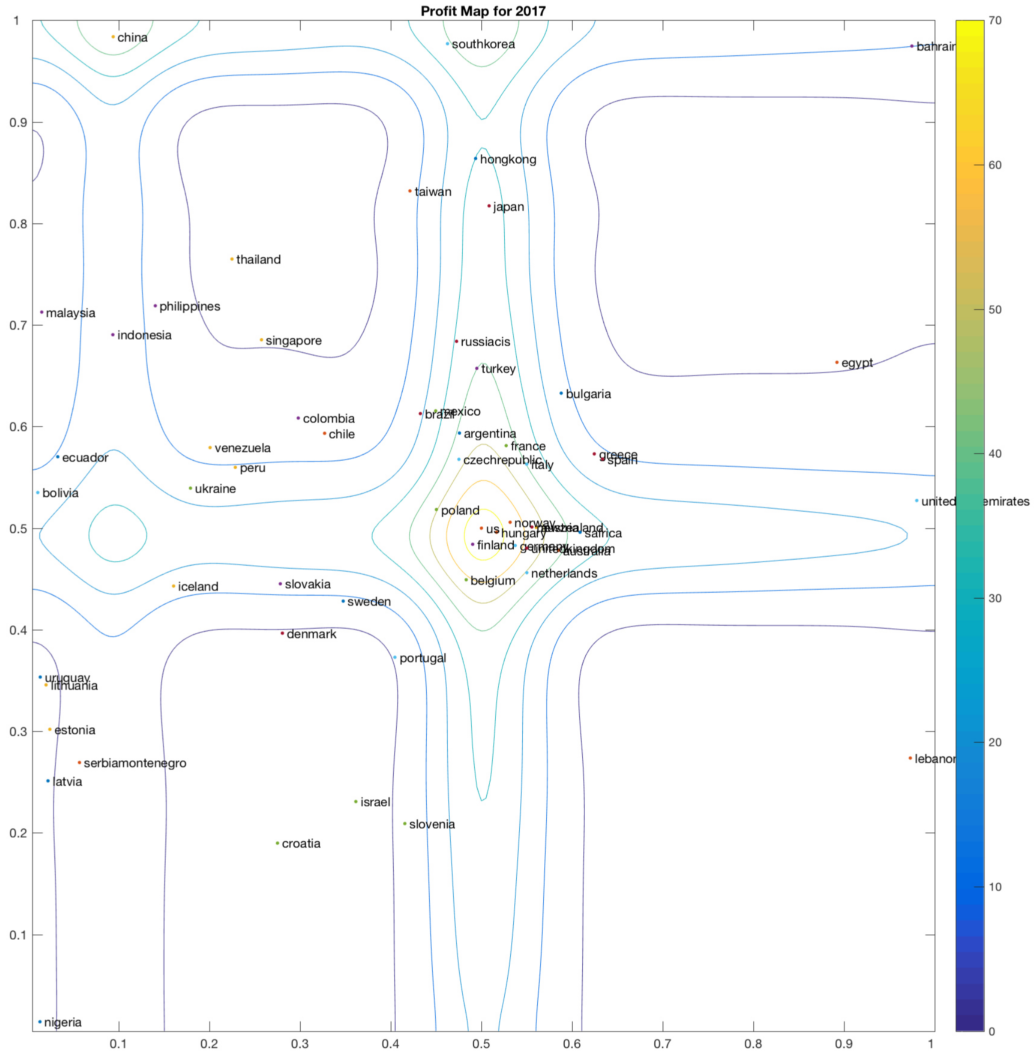
Notes: Revenue map for 2017 in the taste space. Dots are the average taste locations of countries, assumed to be constant over time, relative to the US which is normalized to be at taste position (0.5 0.5). Lines are iso-revenue lines for the year 2017 assuming competitor movies on an evenly spaced grid.

Figure 3.6: Expected revenue: Small vs large variance (2001)



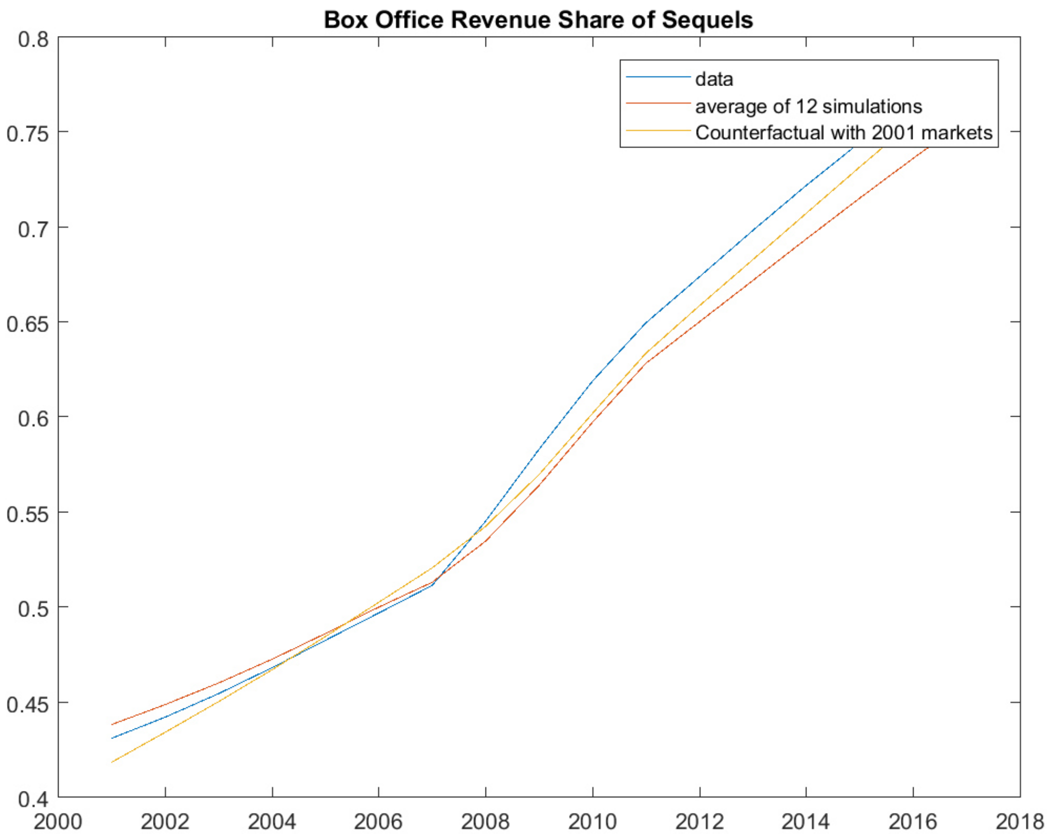
Notes: Revenue map for 2001 in the taste space. Lines represent the difference between the revenue for a movie precisely located at the taste location compared to a movie that faces the risk of a small "taste shock" for the year 2001 assuming competitor movies on an evenly spaced grid. Dots are the average taste locations of countries, assumed to be constant over time, relative to the US which is normalized to be at taste position (0.5 0.5).

Figure 3.7: Expected revenue: Small vs large variance (2017)



Notes: Revenue map for 2017 in the taste space. Lines represent the difference between the revenue for a movie precisely located at the taste location compared to a movie that faces the risk of a small "taste shock" for the year 2017 assuming competitor movies on an evenly spaced grid. Dots are the average taste locations of countries, assumed to be constant over time, relative to the US which is normalized to be at taste position (0.5 0.5).

Figure 3.8: Counterfactual: Sequel revenue share



Notes: Simulation results: The *blue line* shows the smoothed sequel share of total revenue, the *red line* corresponds to the sequel share of total revenue when market structure changes as in the data but using simulated taste locations for movies (average over simulations), the *yellow line* shows the sequel share of total revenue holding the market structure constant at the 2001 level and using simulated taste locations for movies.

Bibliography

Acemoglu, D., Autor, D. H. and Lyle, D. (2004). Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury, *Journal of Political Economy* **112**(3): 497–551.
URL: *https://ideas.repec.org/a/ucp/jpolec/v112y2004i3p497-551.html*

Adao, R., Costinot, A. and Donaldson, D. (2017). Nonparametric counterfactual predictions in neoclassical models of international trade, *American Economic Review* **107**(3): 633–89.
URL: *http://EconPapers.repec.org/RePEc:aea:aecrev:v:107:y:2017:i:3:p:633-89*

Anderson, S., de Palma, A. and Thisse, J. (1989). Demand for differentiated products, discrete choice models, and the characteristics approach, *CORE Discussion Papers RP 824*, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).
URL: *https://EconPapers.repec.org/RePEc:cor:louvrp:824*

Arcidiacono, P. and Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity, *Econometrica* **79**(6): 1823–1867.
URL: *https://EconPapers.repec.org/RePEc:ecm:emetrp:v:79:y:2011:i:6:p:1823-1867*

Blanchard, O. J. and Katz, L. F. (1992). Regional evolutions, *Brookings Papers on Economic Activity* **1**: 1–75.

CGDD (2011). Comptes du Logement: Premiers Résultats 2010, le Compte 2009., *Technical report*, Ministère de l'Ecologie, du Développement Durable, des Transports et du Logement.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade, *American Economic Review* **98**(4): 1707–21.
URL: *http://ideas.repec.org/a/aea/aecrev/v98y2008i4p1707-21.html*

Chodorow-Reich, G. and Wieland, J. (2017). Secular Labor Reallocation and Business Cycles, *Working Paper 313261*, Harvard University OpenScholar.
URL: *https://ideas.repec.org/p/qsh/wpaper/313261.html*

Combes, P.-P., Duranton, G. and Gobillon, L. (2018). The costs of agglomeration: House and land prices in french cities, *Technical report*.
URL: *https://www.gate.cnrs.fr/IMG/pdf/cepr-dp9240updated-3.pdf*

Coşar, A. K., Grieco, P. L., Li, S. and Tintelnot, F. (2018). What drives home market advantage?, *Journal of International Economics* **110**: 135 – 150.
URL: *http://www.sciencedirect.com/science/article/pii/S0022199617301356*

de Aduanas, D. G. (1910-1930). *Estadística general del comercio exterior de España con sus posesiones de*

Ultramar y potencias extranjeras, Imprenta Nacional.
URL: *https://books.google.fr/books?id=1aPIsshlkmYC*

Dix-Carneiro, R. and Kovak, B. K. (2017). Trade liberalization and regional dynamics, *American Economic Review* **107**(10): 2908–46.
URL: *http://www.aeaweb.org/articles?id=10.1257/aer.20161214*

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade, *Econometrica* **70**(5): 1741–1779.
URL: *http://ideas.repec.org/a/ecm/emetrp/v70y2002i5p1741-1779.html*

Ford, R. (1845). *A Hand-book for Travellers in Spain: And Readers at Home : Describing the Country and Cities, the Natives and Their Manners, the Antiquities, Religion, Legends, Fine Arts, Literature, Sports, and Gastronomy : with Notices on Spanish History*, number v. 1 in *A Hand-book for Travellers in Spain, and Readers at Home*, J. Murray.
URL: *https://books.google.com.mx/books?id=PNExAQAAMAAJ*

Fujita, M. and Ogawa, H. (1982). Multiple equilibria and structural transition of non-monocentric urban configurations, *Regional Science and Urban Economics* **12**(2): 161–196.
URL: *https://EconPapers.repec.org/RePEc:eee:regeco:v:12:y:1982:i:2:p:161-196*

Hanlon, W. W. (2014). Temporary Shocks and Persistent Effects in the Urban System: Evidence from British Cities after the U.S. Civil War, *NBER Working Papers 20471*, National Bureau of Economic Research, Inc.
URL: *https://ideas.repec.org/p/nbr/nberwo/20471.html*

Harrison, J. (1978). *An Economic History of Modern Spain*, Manchester University Press.
URL: *https://books.google.com/books?id=1hfpAAAAIAAJ*

Head, K. and Mayer, T. (2013). Gravity equations: Workhorse, toolkit, and cookbook, *CEPR Discussion Papers 9322*, C.E.P.R. Discussion Papers.
URL: *http://ideas.repec.org/p/cpr/ceprdp/9322.html*

Henderson, J. V. (1974). The Sizes and Types of Cities, *American Economic Review* **64**(4): 640–656.
URL: *https://ideas.repec.org/a/aea/aecrev/v64y1974i4p640-56.html*

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models, *The Review of Economic Studies* **60**(3): 497–529.
URL: *http://www.jstor.org/stable/2298122*

Instituto de Reformas Sociales, L. (1916). *Informes de los Inspectores del Trabajo sobre la Influencia de la Guerra Europea en las Industrias Espanolas*.
URL: *https://books.google.com/books?id=z1cPAQAAIAAJ*

Instituto Geográfico, C. y. d. E. S. (1912). *Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1910...*, Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1910, los hijos de M.G. Hernández.
URL: *https://books.google.fr/books?id=7ikxAQAAMAAJ*

Instituto Geográfico, C. y. d. E. S. (1922). *Censo de la población de España según el empadronamiento hecho*

en la península e islas adyacentes el 31 de diciembre de 1920..., Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1920, los hijos de M.G. Hernández.
URL: *https://books.google.fr/books?id=7ikxAQAAMAAJ*

Instituto Geográfico, C. y. d. E. S. (1932). *Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1930...*, Censo de la población de España según el empadronamiento hecho en la península e islas adyacentes el 31 de diciembre de 1930, los hijos de M.G. Hernández.
URL: *https://books.google.fr/books?id=7ikxAQAAMAAJ*

Instituto Nacional de Prevision Social, L. (1930). *Clasificación general de industrias, oficios y comercios.*, 2 edn, Madrid. 006583.

Juhasz, R. (2017). Temporary protection and technology adoption: Evidence from the napoleonic blockade, *CentrePiece - The Magazine for Economic Performance 438*, Centre for Economic Performance, LSE.
URL: *https://ideas.repec.org/p/cep/cepcnp/438.html*

Kennan, J. and Walker, J. R. (2011). The Effect of Expected Income on Individual Migration Decisions, *Econometrica* **79**(1): 211–251.
URL: *https://ideas.repec.org/a/ecm/emetrp/v79y2011i1p211-251.html*

Krugman, P. R. (1979). Increasing returns, monopolistic competition, and international trade, *Journal of International Economics* **9**(4): 469–479.
URL: *http://ideas.repec.org/a/eee/inecon/v9y1979i4p469-479.html*

Kucheryavyy, K., Lyn, G. and Rodríguez-Clare, A. (2016). Grounded by Gravity: A Well-Behaved Trade Model with Industry-Level Economies of Scale, *NBER Working Papers 22484*, National Bureau of Economic Research, Inc.
URL: *https://ideas.repec.org/p/nbr/nberwo/22484.html*

Martin, R. and Sunley, P. (2015). On the notion of regional economic resilience: conceptualization and explanation, *Journal of Economic Geography* **15**(1): 1–42.
URL: *http://dx.doi.org/10.1093/jeg/lbu015*

McFadden, D. (1977). Modelling the Choice of Residential Location, *Technical report*.

McKenzie, J. (2012). The Economics Of Movies: A Literature Survey, *Journal of Economic Surveys* **26**(1): 42–70.
URL: *https://ideas.repec.org/a/bla/jecsur/v26y2012i1p42-70.html*

McLaren, J. (2017). Globalization and Labor Market Dynamics, *Annual Review of Economics* **9**(1): 177–200.
URL: *https://ideas.repec.org/a/anr/reveco/v9y2017p177-200.html*

Ministerio de Trabajo, C. e. I. (1927). *Estadística de los salarios y jornadas de trabajo referida al período 1914-1925*, number v. 1914-1930 in *Estadística de los salarios y jornadas de trabajo referida al período 1914-1925*, Sobrinos de sucesora de M. Minuesa de los Ríos, M. Servet.
URL: *https://books.google.fr/books?id=TJ1CAAAAIAAJ*

Monte, F., Redding, S. and Rossi-Hansberg, E. (2017). Commuting, Migration and Local Employment

Elasticities, *Working Papers 2017-073*, Human Capital and Economic Opportunity Working Group.
URL: *https://ideas.repec.org/p/hka/wpaper/2017-073.html*

Nadal, J. (1975). *El fracaso de la revolución industrial en España, 1814-1913*, Ariel historia, Editorial Ariel.
URL: *https://books.google.com/books?id=xqAYSxmeIJMC*

Prados de la Escosura, L. (2016). Spain's Historical National Accounts: Expenditure and Output, 1850-2015, *Working Papers 0103*, European Historical Economics Society (EHES).
URL: *https://ideas.repec.org/p/hes/wpaper/0103.html*

Prados de la Escosura, L. (2017). *Spanish Economic Growth, 1850–2015*, Palgrave Studies in Economic History, Springer International Publishing.
URL: *https://books.google.com/books?id=Jho0DwAAQBAJ*

Redding, S. J. and Rossi-Hansberg, E. (2016). Quantitative Spatial Economics, *NBER Working Papers 22655*, National Bureau of Economic Research, Inc.
URL: *https://ideas.repec.org/p/nbr/nberwo/22655.html*

Reguant, M. (2016). Bounding Equilibria in Counterfactual Analysis, *Technical report*.

Roldan, S. and Delgado, J. (1973). *La formación de la sociedad capitalista en España, 1914-1920*, number v. 2 in *Publicaciones del Fondo para la Investigación Económica y Social de la Confederación Española de Cajas de Ahorros: Fondo para la Investigación Económica y Social*, Confederacion Española de Cajas de Ahorro.
URL: *https://books.google.com/books?id=SG7iXwAACAAJ*

Scott, P. (2014). Dynamic Discrete Choice Estimation of Agricultural Land Use, *TSE Working Papers 14-526*, Toulouse School of Economics (TSE).
URL: *https://ideas.repec.org/p/tse/wpaper/27568.html*

Silvestre, J. (2005). Internal migrations in spain, 1877-1930, *European Review of Economic History* **9**(2): 233–265.
URL: *http://www.jstor.org/stable/41378422*

Silvestre, J., Ayuda, M. I. and Pinilla, V. (2015). The occupational attainment of migrants and natives in barcelona, 1930, *The Economic History Review* **68**(3): 985–1015.
URL: *http://dx.doi.org/10.1111/ehr.12090*

Traiberman, S. (2017). Occupations and Import Competition, *Technical report*.

Yagan, D. (2017). Employment Hysteresis from the Great Recession, *NBER Working Papers 23844*, National Bureau of Economic Research, Inc.
URL: *https://ideas.repec.org/p/nbr/nberwo/23844.html*