Essays in Empirical Labor Economics and Alternative Finance

Manuel Fernández Sierra

Worcester College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2017

Abstract

This thesis consists of three independent chapters. The first two chapters are motivated by an empirical fact: most countries in Latin America experienced sharp changes in the wage structure since the mid-1990s, characterized by a substantial reduction of earnings inequality. At the same time that inequality was falling, the economies in the region experienced a rapid transformation of the age, education, and gender composition of their workforces. To what extent are these two phenomena interconnected? Did the wage structure changed as a consequence of the recomposition of labor supply? To answer these questions I propose two stylized models of the labor market, and estimate their main parameters using household survey data ranging more than two decades from Argentina, Brazil, Chile and Mexico. I provide evidence that supply side trends are the major driver of the observed patterns in relative earnings. While relative demands favoured high-skilled workers during the 1990s, they shifted in favour of low-skilled workers during the 2000s.

In the third chapter the subject matter changes altogether. Here I will focus on the workings of newly developed financial technologies used by entrepreneurs to raise capital for their businesses. In particular, the rapidly growing equity crowdfunding market is analysed. The chapter is motivated by the concerns of analysts and regulators that crowdfunding markets are vulnerable to investment herding, so that prior investors' decisions induce future investors to make similar choices. We develop a stylized model to answer the following question: how should herding be characterized in an equity crowdfunding setting where much of the inter-temporal information about a project is based on the amounts previously invested? The model generates predictions that we test using a unique investment-level dataset from one of the leading UK equity crowdfunding platforms. We provide evidence of a considerable level of immediate inter-temporal herding on an equity crowdfunding platform that, nevertheless, quickly dissipates.

Approximate number of words: 49,274

Declaration

The work in this thesis is based on research carried out by me at the University of Oxford between September 2014 and October 2017. The thesis consists of three independent chapters. Chapter 1 was written jointly with Julian Messina, a Lead Research Economist at the Inter-American Development Bank; Chapter 2 is self-authored; and Chapter 3 was written jointly with three co-authors: Thomas Åstebro, Professor at HEC Paris; Stefano Lovo, Professor at HEC Paris; and Nir Vulkan, Associate Professor of Business Economics at Said Business School.

Manuel Fernández Sierra

(*October 2017*)

Acknowledgements

The value of any academic work consists on the thoughts expressed in it. Over the last three years I have benefited from the acquaintance of many extraordinary people that have stimulated the thoughts contained in this thesis. I am grateful and indebted to them. I would first like to mention my two supervisors, Dr. Brian Bell and Dr. Climent Quintana-Domeque, for their guidance, patience, and support throughout the process, and to my co-authors, Thomas Åstebro, Sonia Bhalotra, Stefano Lovo, Julian Messina, and Nir Vulkan, for many fruitful discussions and for their selfless disposition to help me fulfil this project.

I also benefited from the inputs of many of my fellow graduate peers, either because of the time they took to comment and criticize the work presented here, or for providing the comradeship that was fundamental to making the process enjoyable. The list is too large, but I would like to especially mention Juliana Helo, Sam Jindani, Lukas Kleine-Rueschkamp, Michael Koelle, Julián López, Simone Pedemonte, Viviana Perego, Carlo Pizzinelli, and Felipe Roa-Clavijo.

Finally, I dedicate this work to Ivan Fernández and Luz Elena Sierra, and to the dear memory of my friend Santiago Pardo Escobar.

Contents

List of Figures

List of Tables

Introduction

This thesis consists of three independent and self-contained chapters. The first two are motivated by an empirical fact: most Latin American countries experienced sharp changes in the wage structure since the mid-1990s, characterized by a substantial reduction of earnings inequality. What drivers are behind this reduction in inequality? Why are the observed patterns in Latin America so different from those documented for most high-income countries? I attempt to contribute to the understanding of this phenomenon, emphasizing the effect that changes in the education, age, and gender composition of the workforce had on the structure of wages.

In Chapter 1 the focus is on the cases of Argentina, Brazil, and Chile, three of largest economies in the region. Using recently developed decomposition techniques, we provide evidence that a reduction in the experience premium is a fundamental driver of declines in upper-tail (90/50 interquantile range) inequality, while a decline in the education premium is the primary determinant of the evolution of lower-tail (50/10 interquantile range) inequality. Moreover, we show that the experience and education premiums are falling in a context in which both the average age and education attainment of workers are increasing rapidly.

To what extent are these two trends causally connected? To answer this question, we build a stylized model of the labor market in which workers with different age and schooling are imperfect substitutes in production. The usefulness of the model is that it allows us to disentangle the impact that changes in relative supplies and relative demands had on relative earnings. We estimate the main parameters of the model using household survey data from the three countries spanning more than 25 years. The results from these estimates show that educational upgrading was the key factor for explaining how the skill premiums evolved. Moreover, we show that relative demand trends in these economies favoured high-skilled workers during the 1990s, shifting in favour of low-skilled workers during the 2000s. A combination of changes in the minimum wage, cyclical conditions, and commodity-led terms of trade improvements are behind these relative skill demand trends.

In Chapter 2 the focus moves to one of the most salient compositional changes in Latin American labor markets over the last quarter century: the rapid rise of female labor force participation. Although the massive entry of women into the workforce has fundamentally changed the gender composition of labor supply, there is little understanding of what its impact has been on the wage structure. Moreover, the few studies that have tackled this question provide mixed – and at times contradictory– conclusions, mostly because there is no agreement in the literature of whether male and female labor are close substitutes or not, a key factor underlying the debate.

The chapter analyses the case of Mexico, the country in which female labor force participation increased the most in Latin America since the early 1990s. I show that while the male earnings distribution experienced a similar contraction to the one observed in other Latin American countries, with higher wage growth among low-skilled workers vis-à-vis high-skilled workers, the female earnings distribution remained relatively flat. The result of these contrasting trends was that the gender earnings gap increased among workers in low-paying jobs, and declined among workers in high-paying jobs.

The main proposition of the chapter is that these two phenomena are causally connected, so that the rapid increase of female labor supply has fundamentally changed the wage structure in the Mexican economy. In particular, I argue that the influx of women into the labor market has generated downward pressures to the wages of female workers in low-paying occupations, but to *both* male *and* female workers in high-paying occupations. The hypothesis I put forward is that this heterogeneity arises because the elasticity of substitution between male and female labor is not uniform throughout the occupational distribution: male and female labor are closer substitutes in occupations that rely heavily on abstract analytical skills than in those occupations in which physical or other non-cognitive skills are more prominent in the jobs.

To test this hypothesis I develop an equilibrium model of the labor market that allows me to structurally estimate the elasticities of substitution between male and female labor using Mexican household survey data since 1989. The equilibrium model of supply and demand for labor differs from other work on the subject in two significant ways: First, the model incorporates the ideas of task-based approach (Autor et al., 2003), allowing the elasticity of substitution between male and female labor to vary as a function of the task content of occupations. This feature of the model implies that the effect of female labor supply on relative earnings can vary throughout the pay distribution, a possibility that is suggested by the descriptive evidence. Second, following the work of Johnson and Keane (2013), labor supply is endogenously determined in the model, a feature that addresses one of the major criticisms of the canonical supply and demand framework in the past. The structural estimates from the model provide support to the hypothesis put forward in the chapter.

In the last chapter the subject matter changes altogether. Here we analyse the workings of a newly developed financial technology, equity crowdfunding, that is used by entrepreneurs to raise capital for their businesses.

Crowdfunding is the practice of funding a project or venture by raising small amounts of money from a large number of people, usually through an online platform. Equity crowdfunding, one the fastest growing version of this model, is characterized by the fact that the crowd takes an equity stake in the business in much the same way Venture Capitalist (VC) funding works. This chapter is motivated by the expressed concerns of analysts and regulators that equity crowdfunding backers can be taken advantage of. In particular, the concern is that there will be investment herding in crowdfunding, so that prior investors' decisions induce future investors to take similar decisions. If that is the case, it is possible for the market to fail to aggregate relevant private information, leading to outcomes that differ from those that would have been observed had all investors' private information been taken into account.

In the chapter we analise backers' behavior in crowdfunding campaigns using a unique investment-level dataset from one of the leading UK equity crowdfunding platforms. Our empirical analysis is guided by a stylized model of equity crowdfunding. The model is developed to answer the following questions: first, how should herding be characterized in an equity crowdfunding setting where much of the inter-temporal information about a project is based on the amounts previously invested? Second, what would be the predictions in terms of timing and amounts of future investments based on information about the history of all past investments and investors?

We provide evidence of a considerable level of immediate inter-temporal herding on an equity crowdfunding platform that, nevertheless, quickly dissipates. We show that the observed patterns in the data are consistent with the prediction of the theoretical model. These results suggest that the model can provide an adequate framework to understand what conditions can trigger herding in the equity crowdfunding market.

Chapter 1

Skill Premium, Labor Supply and Changes in the Structure of Wages in Latin America[1]

1.1 Introduction

Inequality declined sharply in Latin American countries after the turn of the century, a contrast with its own history and global trends (Ferreira et al., 2008; Kahhat, 2010; López-Calva and Lustig, 2010; Gasparini and Lustig, 2011; Gasparini et al., 2011; Levy and Schady, 2013; Lustig et al., 2013; **?**). Redistribution through progressive fiscal policy, the emergence of conditional cash transfer programs for the poor, and changes in household demographics played a role in this transition. However, a broad conclusion of previous literature is that the key contribution to inequality reduction was the decline in earnings inequality (Lopez-Calva and Lustig, 2010; Azevedo et al., 2013). Earnings inequality de-

[1]This Chapter was written jointly with Julián Messia, Lead Research Economist at the Inter-American Development Bank.

clined in 16 of the 17 countries in Latin America for which consistent statistics can be calculated (Messina and Silva, 2016), although the intensity and turning points diverged across countries. For example, after a decade of stagnant or slowly increasing inequality, the 90th/10th interquartile range of the labor earnings distribution declined by 20 percent in Argentina and 28 percent in Chile between 2000 and 2013. In Brazil, where earnings inequality started to fall as early as 1990, the reduction has been a remarkable 46 percent since the year 2000.

There is a relatively extensive literature examining the forces behind increasing inequality in Latin American countries during the 1980s and early 1990s, largely focusing on trade liberalization (Goldberg and Pavcnik, 2007; Galiani and Sanguinettu, 2003; Green et al., 2001; Pavcnik, 2003; Robertson, 2004). The literature examining the forces behind the recent inequality decline is much less extensive, and it concludes that traditional trade channels are unlikely to account for a significant fraction of the observed trends. Adao (2015) focuses on shocks due to commodity prices and finds that they can account for only 5 to 10 percent of the fall of earnings inequality in Brazil. Also for Brazil, Costa et al. (2016) examine the local labor market effects of import penetration of manufacturing goods and increasing demand for commodities from China and find that, if anything, the overall impact on inequality was mildly positive. Similarly, Halliday et al. (2015) show that the fall of earnings inequality in Mexico that started in 1995 is inconsistent with traditional trade models.

An often overlooked aspect is that most countries in the region registered a rapid transformation in the age, education and gender composition of their labor forces. Between 1990 and 2013, the share of college-educated workers increased from 16.6 to 26.6 percent in Argentina, virtually doubled in Chile (14.3% to 27.7%) and almost tripled in Brazil (7.5% to 19.5%). The average worker age increased by more than a year in Argentina (37.3 to 39.0), by three years in

Brazil (34.1 to 37.4), and by more than four years in Chile (35.8 to 40.4). The share of females in the labor force increased in all countries with Brazil and Chile ahead of the pack: an increase of more than 8 percentage points. Changes in the composition of the labor force can mechanically affect wage inequality because different types of workers have different levels of within-group wage dispersion (Lemieux, 2006). They can also affect inequality by changing between-group differences in pay, as suggested by the seminal paper of Katz and Murphy (1992) and numerous applications of a simple supply-demand framework to account for changes in the education premium in the US.

This paper investigates how these changes in the demographic and skill structure of the labor force influenced the evolution of the distribution of earnings in Argentina, Chile and Brazil during the last 25 years using household survey data. Our analysis starts by distinguishing the contribution of pure composition changes from changes in the structure of pay. Following the work of Firpo et al. (2007, 2009) we construct counter-factual wage distributions that decompose the observed changes in inequality measures into price and composition effects. The analysis distinguishes between overall inequality and inequality at the bottom and top of the distribution of earnings. This is important because trends have been different. While most of the reduction in earnings inequality in Brazil was the result of a decline in the 50th/10th interquartile range (-31% since 2000), the reduction in inequality in Argentina was mostly driven by a fall of the 90th/50th interquartile range (-23% since 2000). In Chile, inequality fell symmetrically at the bottom and at the top of earnings distribution (-15% since 2000).

We find that falling education and labor market experience premiums are key determinants of the observed changes in inequality. By contrast, the increasing incorporation of women into the labor force had small effects. The declining experience premium had a larger explanatory power in the reduction of inequality in the upper half of the distribution. In contrast, the decline of the returns to schooling explains a larger share of inequality reduction at the bottom. Against these dominating patterns, pure composition changes related to the increase in educational attainment were inequality enhancing, thus contributing to increasing inequality in the early 1990s in Argentina and Chile but working against the recent inequality decline. This may be due to within-group differences in pay as highlighted by Lemieux (2006) in the US, or may reflect a phenomenon previously labeled as the paradox of progress (Bourguignon et al., 2005a), by which increases in educational attainment can be inequality-increasing due to convexity of the returns to education.

Our analysis continues by assessing the role of the aforementioned labor supply changes in the observed education and experience premiums. Have declines in those premiums been driven by increasing educational attainment and aging of the labor force? Following the seminal work of Katz and Murphy (1992), Murphy and Welch (1992) and Card and Lemieux (2001), we build a stylized model of supply and demand for labor in which workers with different skills are imperfect substitutes in production. We then use household-level data from Argentina, Chile and Brazil to estimate the parameters of the model and to derive implications for the role of supply and demand factors in the evolution of experience and education premiums.

We show that a combination of imperfect substitutability between skill groups and the observed movements in relative supplies goes a long way towards explaining the changes in relative returns, especially the declines of both the schooling and experience premiums. Most of the fall in the high school/primary schooling premiums, which we find is a significant factor behind the decline in lower-tail inequality in these countries, can be accounted for by the significant increase of workers with at least a high school degree. According to our model, the observed changes in labor supply should have resulted in an even greater reduction of the high school premium than the observed one, especially during the 1990s. This is because the demand for high school-educated workers increased in this period.

The picture for college-educated workers is slightly different. We show that the rising supply of college-educated workers has also pushed the college premium downwards over the past 25 years. However, relative demand trends did not increase steadily, as was the case for high school graduates. Relative demand favored college-educated workers during the 1990s but started declining at the start of the new millennium. The implication is that demand-side trends attenuated labor supply forces towards a declining college premium during the 1990s, but accentuated the decline after 2003. In other words, the demand for college-educated workers followed an inverse U-shaped pattern that peaked in 2003.

Further, we show that changes in the educational premiums are not the only factors driving the reconfiguration of the wage structure. The experience premiums also declined substantially to contribute to the inequality reduction, especially within groups of workers with similar levels of schooling. We provide novel estimates in the region for the elasticities of substitution between workers with different experience levels. These estimates suggest that aging of the workforce has contributed to changes in the experience premium, and through this channel to changes in inequality.

The supply-demand model used in the second part of the paper is closely related to the framework developed by Manacorda et al. (2010) (henceforth MSPS) to analyze changes in the skill premium in Latin America during the 1990s, when inequality was on the rise in most countries. Our paper extends the data to include the 2000s, a period of marked inequality reduction in Latin America. We also depart from MSPS in two significant ways. First, our model allows for imperfect substitutability across experience groups within schooling levels. We show that this distinction is empirically relevant. MSPS finds workers of different age groups to be perfect substitutes in production, a feature rejected by the data in our model. Second, we extend the model to allow for differential demand trends across education and experience groups, which is crucial for rationalizing the data. Our estimates suggest that relative demand favored more educated workers during the 1990s, but that there was a shift around the 2000s. Beyond these two extensions we reinforce the call in MSPS for differentiating between workers with secondary schooling and those with at most primary education when thinking about labor market outcomes in Latin America. In line with MSPS, our evidence suggests these two groups are not perfect substitutes in production.

The paper concludes by discussing a number of robustness checks and extensions. In particular, we assess what forces may be behind the trend reversal in the demand for high-skilled workers. Real minimum wages increased dramatically during the 2000s in the three countries. This could compress the skill premium by boosting wages of low-skilled workers. Our findings suggest that they had a role in Brazil and Chile, but declining demand for high-skilled workers after 2003 persists after controlling for the evolution of the minimum wage. Other factors including changes in aggregate labor market conditions as represented by changes in the unemployment rate, and above all the rapid terms of trade improvement boosted by rapidly increasing commodity prices had a significant role in the reversal of skills demand.

The rest of the paper is organized as follows. Section 1.2 discusses the data and main stylized facts, reviewing the evolution of inequality and socio-demographic changes in Argentina, Brazil and Chile. Section 1.3 shows how changes in inequality were affected by compositional changes and changes in the wage structure associated with education, experience and gender. In Section 1.4

we develop a stylized model of supply and demand based on the descriptive trends in the data. We provide estimates to the key parameters of the model and discusses its implications for the evolution of the skill premium. We include several extensions and robustness exercises that aim at understanding the sensitivity of our results to the modelling choices in Section 1.5. Finally, Section 1.6 concludes.

1.2 Data and Stylized Facts

We use household surveys for Argentina, Brazil and Chile for the analysis. All surveys include information about general characteristics of the workers (e.g., gender, age, education) and their jobs (type of contract, labor earnings, hours worked). With the exception of Argentina, where information is restricted to urban areas,[2] all other surveys are nationally representative. Earnings refer in the three surveys to total monetary payments from labor in a reference period. Labor earnings are divided by actual worked hours during the same period to obtain hourly earnings. The series are converted into real terms using the Consumer Price Index (CPI).[3] We restrict the sample to individuals between the ages of 16 and 65 and only use earnings of full-time workers (individuals working for more than 35 hours in the reference week).[4] For further details on the characteristics of the surveys and the construction of the variables see Appendix 1.7.2.

Earnings inequality as summarized by the 90/10 log wage differential de-

[2]Urban areas account for almost 90 percent of the total population in the Argentina in 2013.

[3]The official CPI is used in Brazil and Chile. Due to inconsistencies found in the official series in Argentina (see Cavallo (2013)), we use the information from PriceStats (`http://www.statestreet.com/ideas/pricestats.html`) instead. Because the paper focuses on inequality, the use of the price deflator does not make a significant difference in the results.

[4]We show in the robustness section of the paper that our main results are unchanged if we restrict the sample to prime-age workers (between 25 and 55 years of age) or if we include part-time workers. The average share of workers that reported working less than 35 hours per week is 15 percent in Chile and Brazil, and close to 28 percent in Argentina.

clined during the 2000s in Argentina and Chile,[5] reversing the increasing trend documented for the 1980s and the first years of the 1990s (Figure 1.1).[6] Earnings inequality peaks around 1996 in Chile and 2002 in Argentina. Inequality in Brazil declined steadily during our period of analysis that starts in 1990. The contraction of the earnings distribution is quite significant: Since the year 2000, the ratio between the 90th and 10th percentiles contracted by 20 percent in Argentina, 28 percent in Chile, and a remarkable 46 percent in Brazil. Nonetheless, the levels of earnings inequality remain among the highest in the world in 2013. The 90/10 log earnings ratio was close to 1.7 in the three countries, implying that the hourly earnings of a worker at the 90th percentile of the distribution is more than 5.5 times what a worker at 10th percentile gets. As a point of comparison, the OECD average of the 90/10 interdecile ratio in 2012 was 4.[7]

In the three countries we observe a drop in both the 90/50 and 50/10 log wage ratio after inequality peaked. This contrasts with a large body of literature from developed countries that shows that most of the changes in the wage structure have taken place at the top of the income distribution.[8] With the exception of Argentina, which displays a small increase in the 50/10 log wage ratio since the initial levels of 1995, the recent decline of all inequality measures have brought inequality below the levels of the early 1990s.

These changes in the distribution of earnings are taking place when the skill and demographic composition of the workforce is also changing significantly. Major shifts include changes in the education, age and gender composition of the labor force. The percentage of workers with a primary education degree in the

[5]See also Ferreira et al. (2008); Kahhat (2010); López-Calva and Lustig (2010); Gasparini and Lustig (2011); Gasparini et al. (2011); Levy and Schady (2013); Lustig et al. (2013).

[6]See Cragg and Epelbaum (1996); Londoño and Szekely (2000); Sánchez-Páramo and Schady (2003); Behrman et al. (2007); Cornia (2010); Manacorda et al. (2010) among others.

[7]See `http://stats.oecd.org/`.

[8]See Katz and Autor (1999); Autor et al. (2005, 2008); Lemieux et al. (2009); Acemoglu and Autor (2011) and the references therein.

early 1990s was 47.2 percent in Argentina, 49.3 percent in Chile, and 77 percent in Brazil.[9] By 2013 that share had dropped in a range from 17.7 percentage points in Argentina to 34.3 percentage points in Brazil (Table 1.1). The gains in schooling are reflected in an increase of the share of workers with high school education completed, as well as by an increase in the share of workers with a college degree. For example, the share of college-educated workers increased from 16.5 to 26.6 percent in Argentina, almost doubled in Chile (14.3% to 27.7%), and almost tripled in Brazil (7.5% to 19.5%).

The average age of a worker increased by more than a year in Argentina (37.3 to 39), three years in Brazil (34.1 to 37.4), and four years in Chile (35.8 to 40.4). Even with the sharp rise in the levels of schooling, this demographic shift has resulted in a rise in the average level of potential experience.[10] This is especially significant in the case of Chile, were the share of workers with more than 20 years of potential experience increased by 11.9 percentage points.

In the early 1990s female labor force participation was as low as 35 percent in Chile and close to 50 percent in Argentina and Brazil. By 2013 half of the women between the ages of 16 and 65 in Chile were working, and female labor force participation in Argentina and Brazil was close to 60 percent. As a consequence of this shift, the employment share of women rose by more than 8 percentage points in Brazil and Chile.

[9] See Appendix 1.7 for details on the aggregation of workers with incomplete levels of schooling.

[10] We define potential experience as: age-years of education-6.

1.3 Inequality, Workforce Compositional and Wage Structure

Changes in the composition of the labor force will affect inequality across and within labor market skill groups. Perhaps the most studied characteristic is education. Facilitating access to education to the poor is a powerful tool for social mobility and may lead to lower inequality in the long run. However, because the education premium is convex, in the short and medium term educational upgrading may increase between-group inequality. This "paradox of progress" (Bourguignon et al., 2005b) may occur even when changes in educational attainment are moderately in favor of low socio-economic background groups, and was recently confirmed for several Latin American countries by Battistón et al. (2014). Moreover, within-group dispersion is typically much higher among highly educated workers, pushing inequality up when educational attainment increases (Lemieux, 2006).

The role of composition changes associated with labor market experience and gender on inequality is less straightforward. Returns to experience are concave (Murphy and Welch, 1990), a force that would push between-group inequality down when the labor force is aging. However, within-group wage dispersion is higher among high-experience workers than among their low-experienced counterparts, possibly limiting the inequality decline (Lemieux, 2006). Similarly, the importance of composition effects associated with gender depends on the skills distribution of the women that are increasingly accessing the labor market.

A simple decomposition exercise can help disentangle the importance of composition and price effects on inequality. The idea is to exogenously fix the structure of relative wages at the average level across the last two decades and quantify the counterfactual levels of the interquantile wage ratios under the observed compositional changes. Alternatively, we can keep the composition of the labor force fixed at a given point in time and construct counterfactual wage distributions to evaluate how changes in the schooling, experience and male premiums have affected the observed inequality dynamics.

The decomposition we propose follows Firpo et al. (2007, 2009), which have recently shown that using the properties of Recentered Influence Functions (RIF) one can extend the traditional Oaxaca-Blinder decomposition to analyze distributional statistics beyond the mean (e.g., quantiles). Details on the method are found in the Appendix 1.7.3. As a starting point, consider a transformed wage-setting model of the form

$$RIFq_{\tau t} = X_t'\gamma_t + \epsilon_t \ \text{ for } \ t = 1, 0, \tag{1.3.1}$$

where t identifies the initial ($t = 0$) and final ($t = 1$) periods; $RIFq_{\tau t}$ represents the value of the RIF corresponding to the τ'th quantile of the earning distribution at time t; X is a vector of socio-demographic characteristics including quadratic terms in education and experience and a female dummy.[11] We can estimate equation (1.3.1) by OLS, and we express the estimated difference over time of the expected value of the wage quantile $\hat{q}_\tau$ as

$$\Delta\hat{q}_\tau = \underbrace{\left(\overline{X'}_1 - \overline{X'}_0\right)\hat{\gamma}_P}_{\Delta\hat{q}_{X,\tau}} + \underbrace{\overline{X'}_P\left(\hat{\gamma}_1 - \hat{\gamma}_0\right)}_{\Delta\hat{q}_{S,\tau}}, \tag{1.3.2}$$

[11] As in the Oaxaca-Blinder decomposition for the mean, this decomposition is not invariant to the reference variable chosen when covariates are categorical. We limit this problem by using quadratic polynomials in years of education and potential experience. In the case of gender, we repeated the decomposition with a male dummy and a female dummy. The results were qualitatively similar.

where overbars denote averages, and $\hat{\gamma}_P$ and $\overline{X}_P$ correspond to the estimated vectors of parameters and the explanatory variables of a wage-setting model in which observations are pooled across the two periods.[12] Here, $\hat{q}_{X,\tau}$ corresponds to the composition effect, which captures the part of the change in the τ'th wage quantile that is accounted for by changes in the average skill-demographic characteristics of the workforce, given that we set the skill returns at their (weighted) average over the two periods; and $\hat{q}_{S,\tau}$ is the wage structure effect, and captures how changes in returns are affecting wages at the quantile τ, given that the observable characteristics are fixed to be equal to their (weighted) average over time.

Since we are interested in the effects of compositional and price changes on wage inequality, we construct the following measures for the 90/10, 90/50 and 50/10 log wage ratio in each country separately

$$\underbrace{\Delta\hat{q}_{90} - \Delta\hat{q}_{10}}_{\text{Overall}} = \underbrace{(\Delta\hat{q}_{X,90} - \Delta\hat{q}_{X,10})}_{\text{Composition}} + \underbrace{(\Delta\hat{q}_{S,90} - \Delta\hat{q}_{S,10})}_{\text{Wage Structure}} \tag{1.3.3}$$

$$\Delta\hat{q}_{90} - \Delta\hat{q}_{50} = (\Delta\hat{q}_{X,90} - \Delta\hat{q}_{X,50}) + (\Delta\hat{q}_{S,90} - \Delta\hat{q}_{S,50}) \tag{1.3.4}$$

$$\Delta\hat{q}_{50} - \Delta\hat{q}_{10} = (\Delta\hat{q}_{X,50} - \Delta\hat{q}_{X,10}) + (\Delta\hat{q}_{S,50} - \Delta\hat{q}_{S,10}) \tag{1.3.5}$$

The results of the Oaxaca-Blinder decompositions are shown in Table 1.2. In the three countries we observe a very similar pattern: changes in the skill and

[12]This specific counterfactual allows us to analyze composition and wage structure effects relative to a baseline defined by both the (weighted) mean returns and (weighted) mean characteristics over the two periods.

demographic composition of the workforce have had an unequalizing effect on the distribution of earnings as measured by the log 90/10 wage ratio. In Argentina, the counterfactual change is relatively small (5.7% increase), but it is sizeable in Chile (28.3% increase) and Brazil (32.6% increase).[13] These unequalizing effects of compositional changes are observed at both ends of the earnings distribution, but the magnitude tends to be larger in the upper half (90/50 wage ratio). In Brazil and Chile, composition changes would have pushed up the 90/50 wage ratio by 25 and 20 percent, respectively. Thus, changes in the skill-demographic composition alone cannot explain the observed patterns in earning inequality dynamics over the last two decades.

Wage structure effects are key to understanding the evolution of inequality. Earnings inequality would have declined by 13 percent had changes in the composition of the labor force been kept constant in Argentina, and as much as 41 percent in Chile and 67 percent in Brazil. Of course, these are partial equilibrium counterfactuals, which do not take into account the impact compositional changes may have had on the returns to observable characteristics, an aspect to which we will address below.

Among the wage structure effects, changes in the schooling premiums had a prominent role in the observed inequality trends in the three countries, outweighing the observed inequality decline. Thus, changes in the schooling premium more than offset composition changes that were unequalizing during the period, and other unobservable factors that swam against the current of declining inequality. This is particularly remarkable in Brazil and Chile, where, under our counterfactual scenario, changes in the schooling premium would have contributed to a decline in the 90/10 interquartile range of 68 and 81 percent, respectively.

[13] All percent changes are calculated by taking the exponential of the respective values, which are expressed in logarithms, and subtracting one.

But education was not the only aspect of human capital that contributed to the fall of overall inequality during the 2000s. Changes in the experience premium also played a significant role. The contribution of the decline in the experience premium was largest in Argentina (24.6%) and more modest in Chile, where changes in the schooling premium dominated the decline. Although changes in the gender wage gap also had equalizing effects, their impact on overall inequality trends was much smaller.

Interestingly, inequality appears to be driven by different forces in the lower and upper half. The change in the schooling premium is the fundamental factor behind the evolution of the 50/10 interquartile range, but it is only significant in the evolution of inequality in the upper half of the distribution in Chile. In contrast, changes in the experience premium are fundamental to understanding upper-half inequality. In Argentina and Brazil, they alone almost fully explain changes in the 90/50 interquartile range. In Chile, their role with respect to changes in the schooling premium is more modest, but they are still a significant factor.

The decomposition exercise shows that the observed patterns in earnings inequality are mostly driven by how the wage structure changed over time, but it gives no indication as to why those relative returns changed. Moreover, the wage structure effects are calculated under a counterfactual in which the skill-demographic composition of the workforce is held constant, which we know was not the case. A natural hypothesis, then, is that the wage structure is changing because of the compositional changes, not in spite of them. This would be the case if workers with different skill-demographic characteristics are not perfectly substitutable in production, so that changes in relative supplies directly influence relative wages. In the next section we provide descriptive evidence that this simple mechanism is consistent with the observed trends and then proceed to formulate a model that can rationalize the patterns in the data.

1.4 Skill Supply and Demand and the Evolution of Relative Returns

The schooling premiums follow quite closely the evolution of earnings inequality. This is particularly the case with the high-school vs. primary wage gap, as evidenced by Figure 1.2, which shows the evolution of compositionally adjusted high school/primary and college/high school premiums in each country. The composition adjustment holds constant the relative employment shares of the different skill-demographic groups at their average levels across all years of the sample. In particular, we first compute mean (predicted) log real earnings in each country-year for 70 skill-demographic groups (five education levels, seven potential experience categories in five-year intervals, males and females). Mean wages for broader groups shown in the figures are then calculated as fixed-weighted averages of the relevant sub-group means, where the weights are equal to the mean employment share of each sub-group across all years. This adjustment ensures that the estimated premiums are not mechanically affected by compositional shifts.

The peaks in overall (90/10) inequality coincide with the peaks in the high-school vs. primary wage gap in Argentina and Chile. Both series fall since the start of the sample in Brazil. The decline in high school/primary premiums is substantial. After the peak, the high school premium declined by 12 percent in Argentina, 19 percent in Chile, and 46 percent in Brazil. College vs. high-school premiums are also falling during the same period, although at a slower pace in Brazil and Chile. In Argentina, where the expansion of secondary took place earlier and employment of college graduates gains ground with respect to high school quickly (see Table 1.1), the reduction of the college premium is even larger than the one observed for the high-school vs. primary wage gap (22% since 2002).

Returns to experience declined across the board. Table 1.3 shows the change over time of the compositionally adjusted log hourly earnings of high experience (more than 20 years of potential experience) and low experience (less than 20 years) workers. Reductions of the experience premium were larger among high school and college graduates, the two education groups that are rapidly expanding in the three countries. The reduction of the experience premium is largest in Chile, where the employment share of workers with more than 20 years of potential experience grew faster (12 percentage points, as shown in Table 1.1). Thus, in the case of experience there also appears to be a connection between changes in relative supply and the evolution of returns.

Relative quantities and relative prices are moving in opposite directions for the two main drivers of the change in the wage structure: education and potential experience. To what extent can the trends in inequality be explained by this simple mechanism? The answer to this question will depend on the sensitivity of changes in relative wages to movements in relative supplies, that is, to the degree of substitutability between labor types with different skills. We now formalize this idea in a stylized model of supply and demand of skills. We then proceed to estimate the main parameters of the model using the data from the three countries.

1.4.1 A Supply-Demand Model

The basic framework follows the canonical work of Katz and Murphy (1992) and Murphy and Welch (1992) and Katz and Autor (1999). Workers are divided into skill groups, which are allowed to be imperfect substitutes in production. In particular, we assume that aggregate production in this economy can be described by a multilevel nested constant elasticity of substitution (CES) function. At the top level, output is produced as a CES combination of labor with high (college education completed or more) and low (high school degree at most) skills,

$$Y_t = \lambda_t \left(L_{Ut}^{\rho} + \alpha_t L_{St}^{\rho} \right)^{1/\rho}, \tag{1.4.1}$$

where Y_t is total output at time t; L_U is the total supply of low-skill labor; L_S is the total supply of high-skill labor; λ_t is a scale parameter that is allowed to vary in time to capture skill-neutral technological change; α_t is a time-varying parameter that captures both differences in relative productivities between skilled and unskilled labor, and movements in relative demands between these two types; and ρ is a function of the elasticity of substitution (σ_ρ) between skilled and unskilled labor: $\sigma_\rho = \frac{1}{1-\rho}$.

As noted by Katz and Autor (1999), the fact that we model the economy using an aggregate production function means that we have to be careful not to interpret the parameters as if we were dealing with individual firms. For example, the elasticity of substitution σ_ρ reflects not only technical substitution possibilities between workers at the firm level, but also outsourcing and substitution across goods and services in consumption. In a similar way, α_t captures relative productivity changes both at the intensive (workers performing better at the current jobs) and the extensive margins (e.g., a shift in work tasks across workers of different skill groups), changes in relative prices or quantities of non-labor inputs, and shifts in product demands among industries with different skill intensities.

Following Manacorda et al. (2010), we further divide the total supply of unskilled labor (L_{Ut}) into two sub-groups. The first sub-group is formed by labor from workers that have at least obtained a high school degree, but that have not completed any post-secondary education. The second sub-group comprises labor from workers that have at most obtained a primary education degree.[14] The aggregation is done using a productivity-weighted CES combination of the form

$$L_{Ut} = \left(L_{Pt}^{\delta} + \beta_t L_{Ht}^{\delta}\right)^{1/\delta}, \tag{1.4.2}$$

where L_{Pt} is the total supply of labor from workers with at most primary education; L_{Ht} is the total supply of labor from workers with at most secondary education; β_t is a time-varying parameter that captures both differences in relative productivities between the two sub-groups and changes in relative demands; and δ is a function of the elasticity of substitution (σ_δ) between the two low-skill types.

Finally, we divide workers in each of the three schooling categories (primary, high school and college educated) into two potential experience sub-groups. The first sub-group is composed of workers who have fewer than 20 years of potential experience, henceforth denominated as inexperienced workers. The second sub-group comprises workers with 20 years of potential experience or more, henceforth denominated as experienced workers. In practice, we aggregate experience and inexperience workers within schooling levels using a productivity-weighted CES combination. In order to reduce the parameter space, we assume that the elasticities of substitution and the relative productivity parameters within the unskilled group (primary and high school educated) are the same. In particular, we have

[14] Hence, high school dropouts are included in this group.

$$L_{Kt} = \left(L_{KIt}^{\theta_U} + \phi_{Ut} L_{KEt}^{\theta_U}\right)^{1/\theta_U} \quad \text{for} \quad K = P, H \tag{1.4.3}$$

$$L_{St} = \left(L_{SIt}^{\theta_S} + \phi_{St} L_{SEt}^{\theta_S}\right)^{1/\theta_S}, \tag{1.4.4}$$

where I and E index inexperienced and experienced workers, respectively; ϕ_{Ut} and ϕ_{St} are time-varying parameters that capture both differences in relative productivities and changes in relative demands between the potential experience sub-groups; and θ_U and θ_S are both functions of the elasticities of substitution (σ_{θ_U} and σ_{θ_S}) between the two experience sub-groups within the skilled and unskilled labor types.[15]

We make two final assumptions that are key to our identification strategy. First, we assume that the labor supply of each labor type is exogenously determined. This is a strong assumption, especially since we observe a sharp movement of women into the labor market during the last 20 years. Our choices for the relative wage and relative supply series discussed in the next section are made to ameliorate problems of selection arising from endogenous responses of women to changes in market conditions and endogenous responses of labor participation across different skill groups. The robustness section discusses alternative specifications to assess the sensitivity of the results.

Second, we assume that the economy is operating along the competitive equilibrium demand curve. The implication of these assumptions is that the wage of each labor type is fully determined by its marginal productivity. Given that we have six different labor types in the model (3 schooling levels $\times$ 2 potential expe-

[15]Even though modeling choices were made trying to mimic the observed data patterns, there is necessarily some degree of arbitrariness. In the robustness section we estimate an alternative specification to assess the importance of some of the modeling assumptions for the results.

rience groups), we get six equilibrium conditions. Denoting lower-case variables as the natural logarithms of the respective upper-case variables, the four equilibrium conditions for the low-skill types (PI, PE, HI and HE) are summarized in the following expression

$$\begin{aligned} w_{KJt} = \log \zeta_{KJt} + \frac{1}{\sigma_\rho}(y_t - l_{Ut}) + \frac{1}{\sigma_\delta}(l_{Ut} - l_{Kt}) \\ + \frac{1}{\sigma_{\theta_U}}(l_{Kt} - l_{KJt}) \quad \text{for} \quad K = H, P \quad \text{and} \quad J = E, I \end{aligned} \tag{1.4.5}$$

where $\zeta_{PIt} = 1$; $\zeta_{PEt} = \phi_{Ut}$; $\zeta_{HIt} = \beta_t$; and $\zeta_{HEt} = \beta_t \phi_{Ut}$. In a similar way, the two equilibrium conditions for the high-skill types (SI and SE) are

$$w_{SJt} = \log \zeta_{SJt} + \frac{1}{\sigma_\rho}(y_t - l_{St}) + \frac{1}{\sigma_{\theta_S}}(l_{St} - l_{SJt}) \quad \text{for} \quad J = E, I \tag{1.4.6}$$

where $\zeta_{SIt} = \alpha_t$; and $\zeta_{SEt} = \alpha_t \phi_{St}$.

The model has two types of relevant parameters that we wish to estimate: four parameters that are functions of the elasticities of substitution across types (ρ, δ, θ_U and θ_S), and a set of time varying relative productivities/demand shifters parameters (α_t, β_t, ϕ_{Ut} and ϕ_{St}). As shown by Johnson and Keane (2013), we could fit the trends in relative wages perfectly if we did not impose any restrictions on the evolution of the relative demand parameters, but this would mean that we would not be able to identify the parameters capturing the elasticities of substitution. We then restrict these relative productivities to follow a cubic trend in their natural logarithm.[16] For example, the parameter α_t is allowed to change

[16]We also tried quartic time trends without significant changes to our main results (results are available upon request). The estimated parameters associated with the fourth order of the quartic specification were no longer statistically significant.

according to

$$\log \alpha_t = \alpha_0 + \alpha_1 \times t + \alpha_2 \times t^2 + \alpha_3 \times t^3. \tag{1.4.7}$$

As a robustness check, we also report the results from an alternative specification in which the demand trends are estimated using year fixed effects.

1.4.2 Estimation of the Supply-Demand Model

1.4.2.1 Step I

The parameters of the model are estimated sequentially in three stages. In each stage we recover a subset of the parameters, and use them to construct the unobserved productivity weighted CES labor aggregates which are then used as inputs in the next step. For the first stage, we use the equilibrium conditions from equations (1.4.5) to find the expression that characterizes the evolution of relative earnings between experienced and inexperienced labor within the unskilled labor types. In particular, we have

$$w_{KEt} - w_{KIt} = \phi_{Ut} - \frac{1}{\sigma_{\theta_U}}(l_{KEt} - l_{KIt}) \quad \text{for} \quad K = P, H. \tag{1.4.8}$$

The two equations in (1.4.8) show that changes in log relative wages between unskilled experienced labor and unskilled inexperienced labor depend on i) the evolution of the log relative supplies, scaled by the inverse of the elasticity of substitution; and ii) the evolution of relative demand (ϕ_{Ut}), which will be captured by a year-trend polynomial of order three. Note that the relative earnings and relative labor supply series can be constructed directly from the data in each country. In all cases we limit our sample to population between ages 16 and 65. The labor supply of each labor type is drawn from the entire working age population, irrespective of employment status or hours worked. Using working age population to construct labor supply is more appropriate in our context than using the employed population, considering the assumption that labor supply is exogenous in the model.[17]

The labor earnings series only uses full-time workers (reported working 35 hours or more) when estimating the average earnings of the labor types.[18] Moreover, we further restrict the sample to include only male workers when constructing the relative earnings series. This is done to address the concern of sample selection problems regarding female participation in the labor market, especially in a period of rapid movement of women into the workforce.

We estimate the two equations in (1.4.8) by OLS pooling the data from the three countries. In the baseline model we allow the demand trends to be country-specific but restrict the elasticities of substitutions to be common across countries. Both equations in (1.4.8) are estimated in a single regression that includes a skill dummy indicator (P/H). Results of the first step estimates are shown in column 1 of Table 1.4. High and low experience workers within the unskilled group are not perfect substitutes, with the point estimate of the elasticity of substitution around 3.2.[19] Based on the observed changes in relative supplies in each country, the estimated elasticity of substitution implies a predicted fall in the experience

[17] In the robustness section of the paper we show that using total employment or total hours worked has little effect on our estimates.

[18] We report the results when using both full-time and part-time workers in the robustness section of the paper.

[19] Using a different model specification for the United States, Card and Lemieux (2001) provide estimates of this elasticity between 4 and 6, while Johnson and Keane (2013) report estimates of around 10, which are not statistically significant.

premium, absent any demand changes, of -6.6 percent in Argentina, -17 percent in Brazil, and -26 percent in Chile. This relative supply channel, by itself, closely matches the observed change in the experience premium within low-skilled types in Chile (-25.6%), but underestimates the observed fall in Argentina (-12.3%), and slightly overestimates the observed decline in Brazil (-12.6%). Figure 1.3 (Panel A) shows the negative co-movement of relative prices of experience and relative quantities, once the demand trends in Equation (1.4.8) are accounted for.[20]

We can also use the equilibrium conditions from Equations (1.4.6) to arrive at a similar expression for the evolution of relative earnings between experienced and inexperienced workers within the high-skilled types. This expression takes the form

$$w_{SEt} - w_{SIt} = \phi_{St} - \frac{1}{\sigma_{\theta_S}}(l_{SEt} - l_{SIt}). \tag{1.4.9}$$

We proceed symmetrically with the estimation of equation (1.4.9), as we did for the unskilled group. We cannot reject the null hypothesis that experienced and inexperienced workers are perfect substitutes within the high-skilled group, but the precision of the estimation is low, which can be partly explained by the small number of observations available for the regression (Table 1.4). Panel (b) of Figure (1.3) shows a scatter plot of log relative earnings and log relative supplies between experience groups among college-educated workers once the demand trends in equation (1.4.9) are accounted for. In contrast with the unskilled worker case, the regression line is virtually flat.

[20]The log relative earnings series correspond to the residuals of a regression of the observed log relative earnings on country-specific cubic time trends and a skill dummy indicator. Correspondingly, the log relative supply series are obtained as the residuals of a regression of observed log relative supplies on country-specific cubic time trends and a skill dummy indicator.

Figure 1.4 shows the evolution of the cubic demand trends captured by $\log \phi_{Ut}$ and $\log \phi_{St}$.[21] The results are heterogeneous across countries, but, as we will show in the next stages, changes in relative demand trends by experience level are less pronounced than by education, and, in general, are not statistically different from zero. This is particularly true for low-skilled workers, where supply changes explain the bulk of the observed movements in premiums.

1.4.2.2 Step II

The second step is aimed at recovering parameter estimates for the elasticity of substitution between the two low-skilled labor types (σ_δ), and for the time trends capturing the evolution of their relative demands ($\log \beta_t$). We use the equilibrium conditions from equations (1.4.5) to derive two equations characterizing the evolution of relative wages between workers with secondary education and those with at most primary education:

$$\begin{aligned} w_{HJt} - w_{PJt} = \beta_t - \frac{1}{\sigma_\delta}(l_{Ht} - l_{Pt})... \\ -\frac{1}{\sigma_{\theta_U}}\left[(l_{HJt} - l_{Ht}) - (l_{PJt} - l_{Pt})\right] \qquad \text{for} \qquad J = E, I, \end{aligned} \tag{1.4.10}$$

where both l_{Ht} and l_{Pt} are productivity-weighted CES labor aggregates. Although neither of the two labor aggregates is observed in the data, we can use the two equations in (1.4.3) and the estimated parameters from step I to calculate them.[22] The second term in equation (1.4.10) is capturing overall (aggregated across experience groups) relative supplies between workers with at most primary education

[21] Each series is scaled so that it takes a value of zero at the first year in which data for the country is available.

[22] This introduces a generated regressors problem in the estimation of the second step. We deal with this problem bootstrapping the standard errors while keeping the sample fixed for all stages.

and workers with at most a high school degree. The last term represents relative changes in the potential experience composition between the two low skill groups. Note that the coefficient associated with this last term is the inverse elasticity of substitution between experience subgroups among unskilled workers, which was already estimated in step I. The estimated results in this second step serve as an internal consistency check.

We estimate both equations in (1.4.10) in a single regression, adding an experience group dummy indicator (E/I). As before, demand trends are allowed to be country-specific and approximated by a cubic trend, but the elasticities of substitution are assumed to be the same across countries. Results of the second step estimates are shown in column 3 of Table 1.4. The estimated elasticity of substitution between workers with at most primary education and workers with at most secondary education is 2.2. This number is in line with the 2.8 estimate found by Manacorda et al. (2010) for a different set of countries in the region during the 1990s, and reinforces the message that within the context of Latin America, there is a meaningful difference in the way the labor market treats the skills supplied by workers with secondary education and workers with at most primary completed. The estimate of σ_{θ_U} is very similar to that obtained in step I.

Panel (a) of Figure 1.5 shows the tight connection between changes in the high school premium vis-á-vis primary education and relative supply. The figure shows the evolution in the three countries of a log earnings and relative supply series that have been purged from country-specific demand trends and changes in the potential experience composition of the labor force. The negative co-movement between changes in labor supply and earnings is apparent, and confirmed by Panel (b) of Figure 1.5, which shows the estimated demand trends as captured by $\log \beta_t$. Relative demand for high school graduates was very stable in Brazil, increased weakly in Argentina, and only increased strongly in Chile during the 1990s. Thus, the observed sharp declines in the high school/primary schooling premiums were fundamentally driven by the educational upgrading of the workforce. Relative demand trends tended to favour high school graduates. This is further illustrated in Figure 1.6, which shows the fit of the model including or excluding demand trends. The exclusion of demand trends does not alter the model fit for Brazil, which is remarkably close to the observed relative wages. Over the whole period the decline of the high-school premium would have been larger in both Argentina and Chile had demand forces not favored high school graduates over those with basic education.

1.4.2.3 Step III

As a last step we obtain an estimate of the elasticity of substitution between skilled and unskilled labor (σ_ρ) to assess the role of relative skill labor supply on the observed changes in the skill premium. After some manipulation of Equations (1.4.5) and (1.4.6) we can derive the following four expressions,

$$\begin{aligned}\log\left(\frac{\widetilde{W_{SJt}}}{W_{KJt}}\right) &= \log\alpha_t - \frac{1}{\sigma_\rho}\log\left(\frac{L_{St}}{L_{Ut}}\right) - \frac{1}{\sigma_\delta}\log\left(\frac{L_{Ut}}{L_{Kt}}\right) - \frac{1}{\sigma_{\theta_S}}\log\left(\frac{L_{SIt}}{L_{St}}\right)... \\ &- \frac{1}{\sigma_{\theta_U}}\log\left(\frac{L_{Kt}}{L_{KJt}}\right) \quad \text{for} \quad K=H,P \quad \text{and} \quad J=E,I,\end{aligned} \tag{1.4.11}$$

where the terms $\log\left(\frac{\widetilde{W_{SJt}}}{W_{KJt}}\right)$ are the log relative earnings of skilled and unskilled workers of a given experience group that has been "demand-detrended" using the time trend estimates from the previous steps.[23] With the exception of L_{Ut} and L_{St}, all of the productivity-weighted CES labor aggregates have been previously used in the estimation. Constructing L_{Ut} and L_{St} is straightforward using the parameters previously estimated and equations (1.4.2) and (1.4.4).

The set of equations in (1.4.11) incorporate all the parameters of the production function. Hence, the estimation of these set of equations provides a third estimate for the elasticity of substitution σ_{θ_U}; a second estimate for both elasticities of substitution σ_{θ_S} and σ_δ; and a first estimate of the elasticity of substitution between skilled and unskilled workers σ_ρ. We estimate the four equations in a single regression, using a country-specific cubic demand trend for $\log \alpha_t$, and including skill-experience dummy variables as covariates.

Results of the third step estimates are shown in column 4 of Table 1.4. The estimated elasticity of substitution between skilled and unskilled labor is 2.1, very close to the elasticity of substitution between the two unskilled sub-groups. Our point estimate for this elasticity is higher than the 1.4 and 1.6 values reported by Katz and Murphy (1992) and Johnson and Keane (2013) respectively for the United States; it is in line with the 2-2.5 range estimated by Card and Lemieux (2001) also in the United States; and is somewhat lower than the 2.5-5 range reported by Manacorda et al. (2010) for the Latin American region. Estimates of the other elasticities of substitution are very similar to those obtained in columns 1-3.

The results show that the large influx of college graduates into the labor market of the last 20 years depressed the college premium significantly. To see this

[23]In particular, $\log\left(\frac{\widetilde{W_{SEt}}}{W_{HEt}}\right) = \log\left(\frac{W_{SEt}}{W_{HEt}}\right) - \log\frac{\hat{\phi}_{st}}{\hat{\beta}_t\hat{\phi}_{Ut}}$; $\log\left(\frac{\widetilde{W_{SEt}}}{W_{PEt}}\right) = \log\left(\frac{W_{SEt}}{W_{PEt}}\right) - \log\frac{\hat{\phi}_{St}}{\hat{\phi}_{Ut}}$; $\log\left(\frac{\widetilde{W_{SIt}}}{W_{HIt}}\right) = \log\left(\frac{W_{SIt}}{W_{HIt}}\right) - \log\frac{1}{\hat{\beta}_t}$; and $\log\left(\frac{\widetilde{W_{SIt}}}{W_{PIt}}\right) = \log\left(\frac{W_{SIt}}{W_{PIt}}\right)$.

more clearly, Panel (a) of Figure 1.7 shows a scatter plot of log relative earnings and log relative supplies once country-specific demand trends, changes in relative potential experience composition, and changes in schooling composition within the unskilled group are taken into account.[24] The negative co-movement between relative supplies and relative earnings is evident. Panel (b) of Figure 1.7 shows the estimated relative demand trends between skilled and unskilled workers, as captured by $\log \alpha_t$ in Equation (1.4.11). In the three countries we observe a similar pattern, albeit with different magnitudes. Relative demand tended to favor college-educated workers during the 1990s, but this trend started to reverse around 2002. By the end of the period relative demand is back to the 1990s level in Argentina and Brazil, but remained higher in Chile.

Figure 1.8 shows the evolution of the observed skill premium and the predictions of the model in the three countries. The skill premium follows an inverted U-shaped pattern in Argentina and Chile, increasing up to the early 2000s and declining thereafter. This is very much in line with the observed evolution of inequality documented in Figure 1.1. Also in line with the evolution of inequality, the skill premium in Brazil declines slowly during the 1990s but more quickly during the 2000s.

Relative labor supply had a strong impact on the evolution of relative earnings by pulling the skill premium down. The dashed line in Figure 1.8 shows the predictions of the model where relative demand trends ($\log \alpha_t$) have been shut down. Labor supply changes alone tend to over-predict the fall of the skill premium during the past two decades. However, they also completely miss the inverted U-shaped dynamics observed in Argentina and Chile, and in Brazil they strongly over-predict the inequality reduction in most years. It is a strong de-

[24]The log earning series is constructed as the residuals of an estimation of Equation (1.4.11) that omits the aggregate relative supply term ($\log(L_{St}/L_{Ut})$). The log relative supply series corresponds to the residuals of an estimation of Equation (1.4.11) in which the aggregate relative supplies ($\log(L_{St}/L_{Ut})$) are used as the dependent variable.

mand for skills in the 1990s that slowly reverses in the 2000s that explains the inverted U shape of the skill premium. Thus, we conclude that while relative supply changes tended to pull the college premium downwards over the past 20 years, relative demand changes ameliorated this effect during the 1990s and magnified the relative supply channel after 2000.

One criticism of the canonical supply and demand framework is the potential sensitivity of the elasticity estimates to how one controls for the unobserved changes in relative demands (Borjas et al., 2012). As a robustness check, we modify the baseline model so that instead of capturing the demand trends with a third order polynomial, we use year fixed effects that are common across the three countries. The results are shown in Table 1.5. The estimates of the elasticities of substitution between skilled and unskilled workers, and between high school and primary educated workers, are very similar to the baseline model (2.1 in both cases). The estimate of the elasticity of substitution between experience and inexperience workers is slightly larger within the unskilled group (4 instead of 2.8), and smaller within the skilled group (3.5 instead 9). Thus, the importance of supply trends for changes in the education and experience premiums in the three countries is, if anything, reinforced in this specification.

Figure 1.9 shows the point estimates of the year fixed effects in each stage, where the year 1990 is used as the base. The patterns are broadly consistent with those of the parametric specifications. Relative demand trends between workers with high and low experience tended to be flat, and are not statistically different from zero. Hence, it is the dynamics in the supply side of the market that explain the changes of the experience premiums. Throughout the period, the demand for workers with at most high school education completed relative to those with at most primary education increased, and it would have pushed the high-school premium up by almost 20 percent in the absence of relative supply changes. Finally, also consistent with our previous estimates, the demand for college-educated workers vis-à-vis those with no tertiary education completed followed an inverted u-shape pattern, increasing during the 1990s and beginning to fall around 2002.

1.4.3 The Role of the Minimum Wage, Unemployment and the Commodity Price Boom

The supply and demand framework presented above is silent about the role of institutional and cyclical conditions in the labor market in explaining changes in the wage structure. However, these economies have experienced sharp changes in external conditions and labor market institutions such as the minimum wage. In particular, the commodity price boom that started in the early 2000s brought about sharp improvements in terms of trade and a period of unprecedented growth in the three countries (Erten and Ocampo, 2013; The World Bank, 2016), possibly affecting relative skill demand.

Figure 1.10 shows the evolution of minimum wages, unemployment and terms of trade in Argentina, Brazil and Chile over the period of analysis. Unemployment rose in the three countries during the late 1990s and early 2000s and then declined steadily after 2002 (Panel a). The pattern is particularly marked in Argentina, which suffered a major economic and financial crisis between the end of 2001 and 2003. The cyclical conditions in these economies, as captured by the unemployment rates, broadly coincide with the movements in the college premium previously documented.

Between 1990 and 2012 the real hourly minimum wage increased by 120 percent in Brazil and by 138 percent in Chile (Panel b). In Argentina it only increased after the 2001 crisis but at a fast pace, more than doubling in the decade that followed. Sharp increases in the minimum wage may result in substantial real wage gains for low-skilled workers, possibly contributing to the reduction of the skill premium.

Lastly, the commodity super cycle that started in 2002 (Panel c) benefited these three net commodity exporter countries, resulting in a rapid improvement in terms of trade (Panel d). In the three countries an expansion of the commodity sector followed the increase in relative prices, possibly reducing the skill premium as this sector tends to be intensive in low-skilled labor. Moreover, the resulting appreciation of the real exchange rate that followed the terms of trade gains boosted the demand in nontradable sectors, where the skill premium is lower (Messina and Silva, 2016). Thus, gains in the terms of trade are likely to be associated with reductions in the skill premium.

We follow Autor et al. (2008) and extend the empirical implementation of the model to examine the sensitivity of our estimates to the inclusion of controls for changes in the minimum wage, unemployment, and terms of trade. In particular, we are interested in understanding if these variables are behind the trend break in the relative demand for skilled workers observed after 2002. The extension of the empirical model is done by including the log of the real hourly minimum wage, the unemployment rate, and the log terms of trade index as covariates in the third step of the estimation of the year fixed effects model (e.g. assuming common demand trends across countries), but the results do not change if we use country-specific cubic polynomials to approximate demand.

Including additional covariates does not alter the main message of previous sections: relative skill supply trends had an important role in the determination of the skill premium. Column I of Table 1.6 replicates the results from the year fixed effects model; columns II-IV show the results when we include the three additional set of covariates one at a time; and column V includes all the covariates in the same specification. The estimated elasticities of subsitution are stable across specifications, and we cannot reject equality of the coefficients.

The estimated coefficients associated to the minimum wage have a negative sign and are statistically significant for Brazil and Chile in column V, with elasticities of -0.30 and -0.43, respectively. Changes in the unemployment rate is only statistically significant in Brazil, with a negative coefficient of around -0.024. Finally, improvements in the terms of trade during the 2000s appear to have contributed to the fall of the skill premium in Argentina and Brazil, while they are not significant in the case of Chile (Column V).

These alternative specifications provide some insights into what factors might be behind the residual demand trend reversal observed in the early 2000s, which is reproduced in Panel (a) of Figure 1.11. Panel (b) of Figure 1.11 shows the same demand trend for a model that includes controls for the log minimum wage and the unemployment rate. The inverted U-shaped pattern of residual demand is preserved. Something similar happens if we include controls for the change in the terms of trade in each country alone (see Panel (c)). Thus, none of these forces in isolation help explain the unobserved demand component.

A different story emerges when we include the full set of controls in the model. In this model we fail to observe the fall in the relative demand for high-skill workers after 2002. These results are suggestive that falling demand for skill labor after 2002 was the result of a combination of favorable external conditions (terms of trade improvements), a tight labor market where unemployment is rapidly declining, and sharp increases of the minimum wage.

1.5 Robustness Exercises

In this section we present a number of exercises aimed at understanding how sensitive our results are to modelling assumptions and sample decisions. Table 1.7 presents the last stage results of the model estimates using alternative supply measures. In our baseline specification (re-estimated in Column I) labor supply is calculated by adding up the total number of individuals of a given skill-demographic cell between the ages of 16 and 65. The second column of Table 1.7 presents the results when labor supply is approximated by the employed population in each cell. There are no significant changes in the estimated elasticities. In column 3 we present the same estimates when labor supply is calculated by adding up the total number of hours worked by each labor type. This is done to account for potential changes in the intensive margin. Since we only have information on hours worked for individuals that are employed, we impute those numbers for individual outside the workforce by assigning them the average number of hours worked by an employed worker with the same education, potential experience, and sex in the respective country-year. The elasticity of substitution between skilled and unskilled workers increases from 2.1 in our main specification to 2.6 when we use hours worked, but the rest of the parameters' point estimates are unchanged.

Table 1.8 replicates the results from Table 1.4, but including only workers aged 25-55. The sharp changes in the educational composition of the workforce might lead to sample selection issues associated with a larger share of younger workers remaining in the educational system, which would affect our wage series. Limiting the sample to prime age workers leads to slightly lower values of the different elasticities of substitution, so the relative supply channel is accentuated. All the relative demand trends have a pattern very similar to that presented in our preferred specification.

Table 1.9 presents the third stage results of an exercise in which we include both part-time and full-time workers in the calculation of the wage series, also separated by the different supply measures. There is a small increase in the point estimate of the elasticity of substitution between high and low skilled workers when we use hours worked (2.7), but beyond this there are no major changes from our baseline results.

The particular structure we use when setting up the demand side of the model is based on the observed patterns in the data. But there is a degree of arbitrariness in modelling decisions, so that it is important to understand how sensitive our results are to alternative specifications. In Appendix 1.7.4 we present a different formulation of the demand side that follows closely the work of Manacorda et al. (2010). There are two main differences with respect to our baseline model that are worth pointing out: first, the ordering of second and third levels of the production technology is reversed. In particular, skilled and unskilled workers are first disaggregated among seven experience groups, and then further divided between the two lowest schooling levels within each potential experience category. This change implies that the elasticity of substitution between potential experience groups is the same for skilled and unskilled workers by construction. Second, the number of potential experience groups is larger (seven in five-year intervals), but the identifying assumption is that the relative productivities/demand shifters between experience sub-groups is time-invariant.

The parameter estimates of this alternative model are presented in Table 1.10.[25] The elasticity of substitution between high school graduates and those who obtained at most a primary degree is larger than the 2.2 value estimated in our baseline result, with a magnitude that fluctuates between 2.68 and 3.78. These numbers are similar to those reported by Manacorda et al. (2010). Although not

[25] See Manacorda et al. (2010) for a description of the estimation procedure.

directly comparable, the point estimates in our model for the two elasticities of substitution between workers with different levels of potential experience ranged between 3 and 9, depending on the skill group. The analogous (single) estimate in the alternative specification is between 5.8 and 7.2. These new estimates are still statistically significant, but they imply that the sensitivity of relative wages to changes in relative supplies across experience groups is smaller in this set-up, more so among unskilled workers. Finally, in this alternative model the elasticity of substitution between skilled and unskilled workers falls, going from 2.1 to 1.3. This implies that the sensitivity of relative earnings to changes in relative supplies is even higher than what we found. The estimated demand trends are virtually unchanged.

1.6 Conclusions

After a decade of stagnant or rising earnings inequality, the distance between top and bottom earners in Latin America fell sharply during the late 1990s and 2000s. This trend was in sharp contrast to the experience of developed countries during the same period. This paper has offered a detailed accounting of the main factors behind the evolution of earnings inequality in three of the largest countries in the region: Argentina, Brazil and Chile.

The first suspect for changes in the wage structure is changes in the composition of the labor force. The three countries studied here were subject to similar employment changes, although in varying degrees: employees are aging and becoming more educated, and they are more likely to be females. We construct counterfactual wage distributions where the returns to labor market characteristics are kept fixed to evaluate how these changes in the composition of employment may have affected the distribution of wages. Our results are unambiguous.

Changes in composition, particularly increased education, were inequality enhancing. Hence, while composition changes may have contributed to increasing wage inequality before the 2000s, they cannot explain the substantial decline of the last decade.

The decomposition also allows us to build counterfactual distributions where composition changes are kept constant to evaluate the role played by changes in the returns to labor market attributes. The analysis suggests a distinct role of education and experience premiums. The decline of the experience premium is key to explaining reductions of upper-tail (90/50 earnings gap) inequality. This is because reductions of the experience premium were stronger among the highly educated, perhaps reflecting some skill obsolescence. In contrast, a falling schooling premium bears a much higher weight in reducing inequality below the median (50/10 earnings ratio). This was driven by a much faster decline of the high school premium vis-à-vis workers who have at most completed primary education than the reduction of the college premium.

To link changes in schooling and experience premiums to the observed changes in labor supply we built a nested CES model where there is imperfect substitution across experience and education groups. A combination of a relative trend in demand that favored high school educated vis-à-vis primary educated workers, and a rapid educational upgrade go a long way toward explaining the high school/primary schooling premiums. A slightly different story can be told for the college premium. A rising supply of college educated workers has pushed the college premium downwards, but this is not enough to explain the reduction that started around 2000. We find instead that the relative demand for college-educated workers fell during the 2000s in the three countries. Changes in the experience premium have also responded to the evolution of relative supplies, suggesting imperfect substitutability. Our estimates of the elasticity of substitution between experienced and inexperienced workers is between 2.8 and 4.0 for unskilled labor, and between 3.3 and 9 for skilled labor.

We show that expanding our empirical model to account for changes in the minimum wage, unemployment and terms of trade does not mitigate the role of labor supply. The simple supply-demand framework retains substantial explanatory power in the evolution of the skill premium. However, the sharp increases of the minimum wage and improvements in terms of trade go a long way toward identifying the residual change in the model. In particular, the fall of the college premium after 2002 is well explained by the evolution of these external and institutional factors. The results are also robust to using different measures of labor supply and to alternative specifications of the underlying theoretical model.

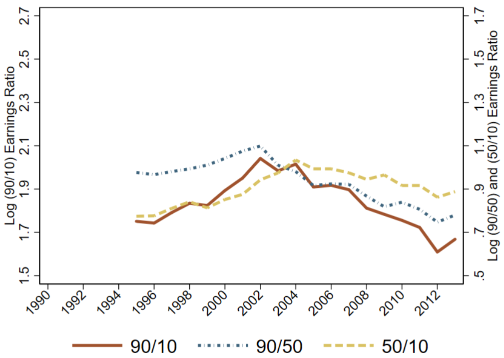
Our exposition focused on the common factors that have driven earnings inequality in the three countries, but in spite of commonalities, substantial heterogeneity remains. Brazil witnessed a much more pronounced reduction in earnings inequality than Argentina and Chile. Top and bottom inequality reductions took place in parallel in Brazil and Chile, while the decline of inequality in Argentina was fundamentally driven by the evolution in the upper half of the distribution. In Chile the demand for high-skilled workers increased much more rapidly during the 1990s than in Brazil and Argentina, and in spite of the recent fall remains above its 1990 level by 2013. The study shows how different relative skill supply trends, and differences in the evolution of minimum wages, unemployment and terms of trade can help explain some, but not all of these heterogeneous trends.

1.7 Appendix

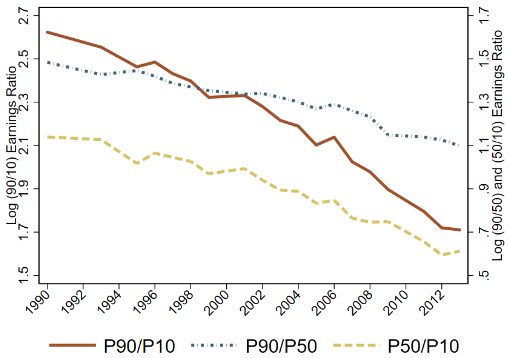
1.7.1 Tables and Figures

Figure 1.1: Interquantile Log Earnings Ratio by Country

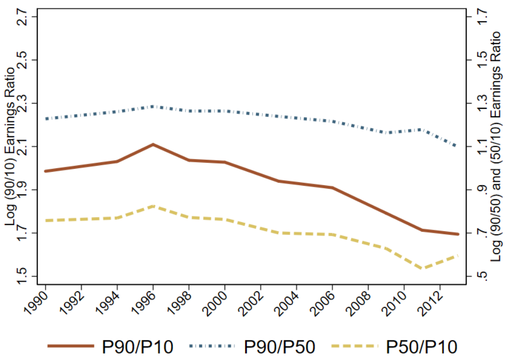
(a) Argentina



(b) Brazil



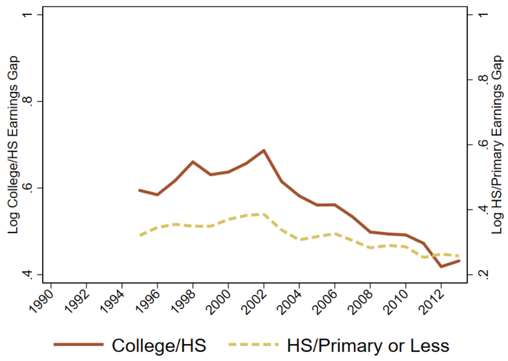
(c) Chile



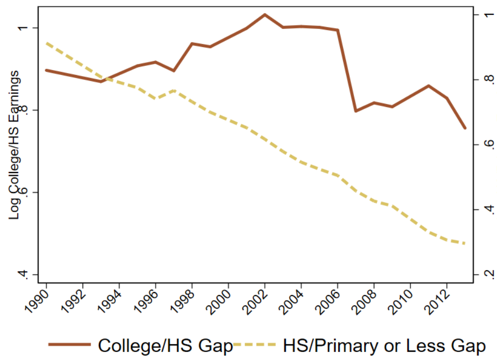
Notes: Sample consists of full-time workers (reported working 35 hours or more) between ages 16 and 65.

Figure 1.2: Composition-Adjusted College/High School and High School/Primary Earnings Gap

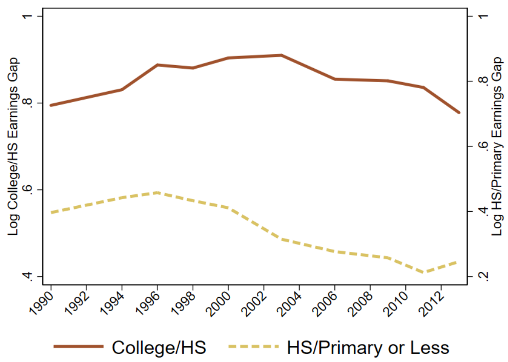
(a) Argentina



(b) Brazil



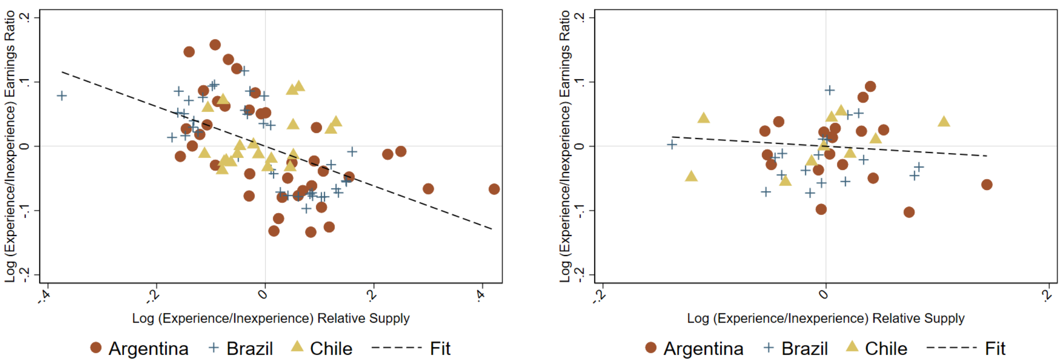
(c) Chile



Notes: Sample consists of full-time workers (reported working 35 hours or more) between ages 16 and 65. See Appendix 1.7 for details on the construction of the compositionally adjusted series.

Figure 1.3: Adjusted Relative Earnings and Relative Supplies by Experience Level

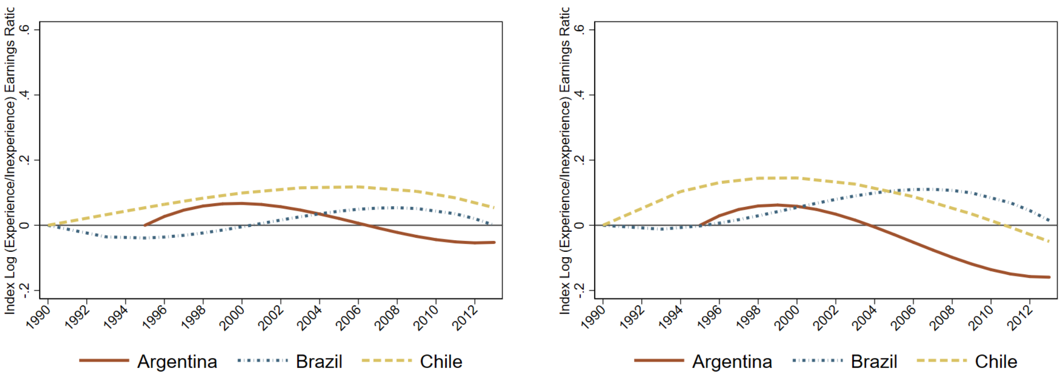
(a) Unskilled Workers (b) Skilled Workers



Notes: Adjusted log relative earnings correspond to the residuals from a regression of observed relative earnings on country-specific cubic time trends and a skill dummy. Adjusted log relative supplies correspond to the residuals from a regression of observed relative supplies on country-specific cubic time trends and a skill dummy.

Figure 1.4: Experienced/Inexperienced Demand Index by Skill Level

(a) Unskilled Workers ($\log \hat{\phi}_{Ut}$) (b) Skilled Workers ($\log \hat{\phi}_{St}$)

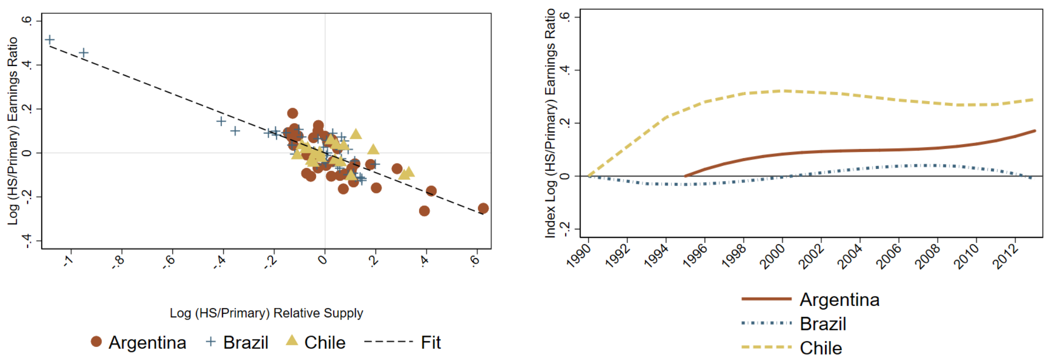


Notes: Relative demand trends between experienced and inexperienced workers are country-specific cubic time trends estimated following Equations (1.4.8) and (1.4.9). Each series is scaled so that it takes a value of zero at the first year in which data for the country are available.

Figure 1.5: Supply and Demand Factors Behind the Fall in the High-School/Primary Earnings Gap

(a) Relative Earnings and Supplies

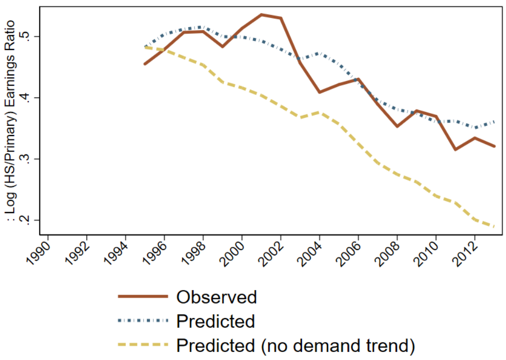
(b) HS/Primary Demand Index ($\log \hat{\beta}_t$)



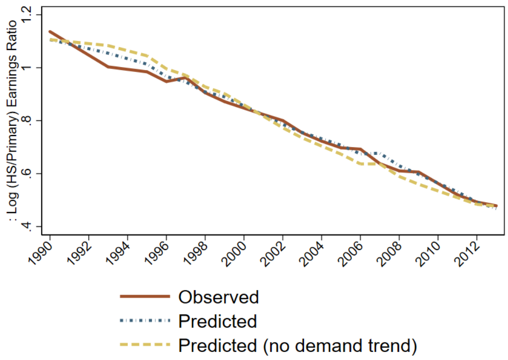
Notes: Panel (a) depicts log relative earnings and log relative supplies of workers with at most a high school degree with respect to those with only primary education, once the country-specific demand trends and changes in the relative potential experience composition are taken into account. The adjusted log earning series is constructed as the residuals of an estimation of Equation (1.4.10) that omits the aggregate relative supply term ($l_{Ht} - l_{Pt}$). The adjusted log relative supply series corresponds to the residuals of an estimation of Equation (1.4.10) in which the aggregate relative supplies ($l_{Ht} - l_{Pt}$) are used as the dependent variable. Panel (b) depicts the estimated relative demand trends between workers with at most a high school degree and those with only primary schooling as captured by the country-specific cubic time trends in Equation (1.4.10). Each series is scaled so that it takes a value of zero at the first year in which data for the country are available.

Figure 1.6: High School/Primary Observed and Predicted Relative Earnings

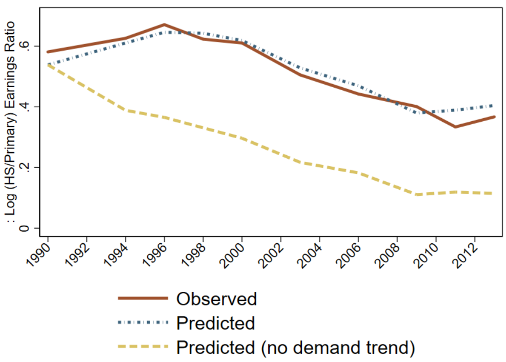
(a) Argentina



(b) Brazil



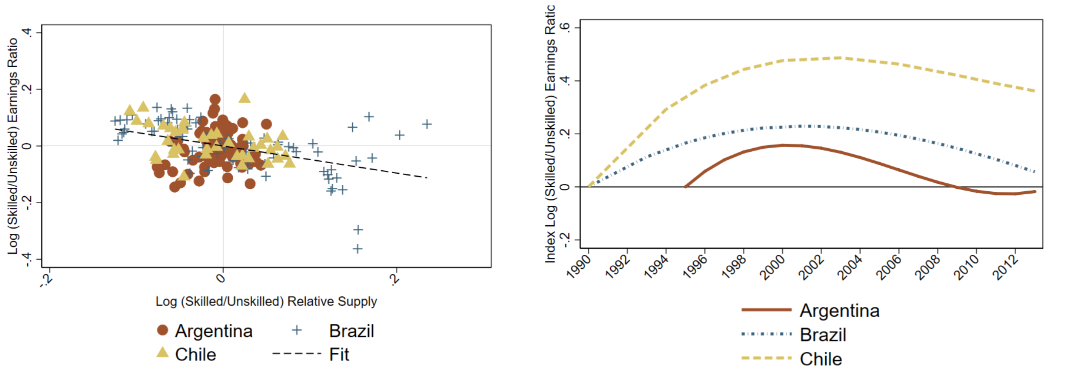
(c) Chile



Notes: "Observed" refers to the Log (HS/Primary) Earnings Ratio observed in the data. "Predicted" refers to the model prediction derived from the estimation of Equation (1.4.10). "Predicted (no demand trend)" is the prediction of a modified version the model in Equation (1.4.10) that omits the country-specific time trends.

Figure 1.7: Supply and Demand Factors Behind the Fall in the Skilled/Unskilled Earnings Gap

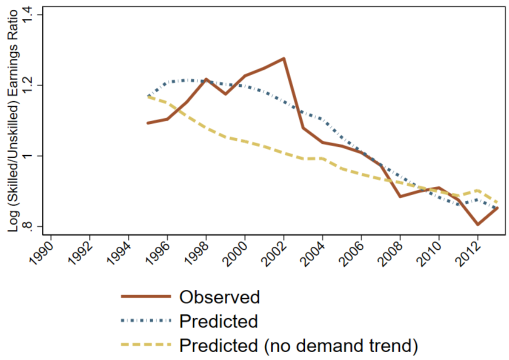
(a) Relative Earnings and Supplies (b) Skilled/Unskilled Demand Index ($\log \hat{\alpha}_t$)



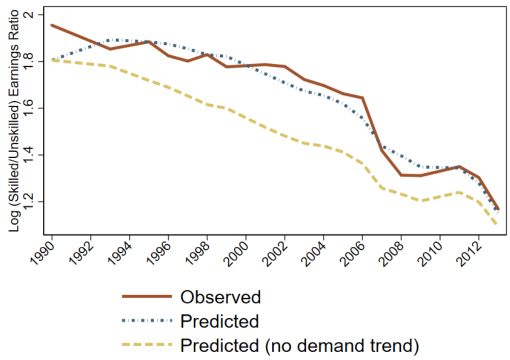
Notes: Panel (a) depicts log relative earnings and log relative supplies of skilled with respect to unskilled workers once the country-specific demand trends, changes in relative potential experience composition, and changes in schooling composition within the unskilled group are taken into account. The adjusted log earning series is constructed as the residuals of an estimation of Equation (1.4.11) that omits the aggregate relative supply term $(l_{St} - l_{Ut})$. The adjusted log relative supply series corresponds to the residuals of an estimation of Equation (1.4.11) in which the aggregate relative supplies $(l_{St} - l_{Ut})$ are used as the dependent variable. Panel (b) depicts the estimated relative demand trends between skilled and unskilled workers as captured by the country-specific cubic time trend in Equation (1.4.11). Each series is scaled so that it takes a value of zero at the first year in which data for the country are available.

Figure 1.8: Skilled/Unskilled Observed and Predicted Relative Earnings

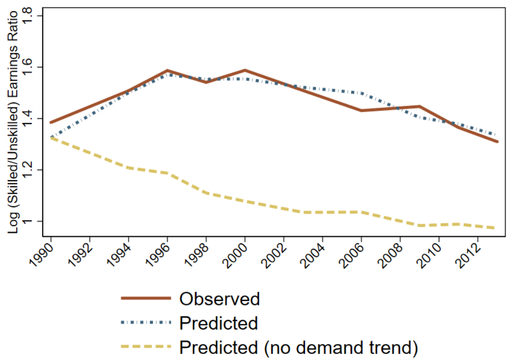
(a) Argentina



(b) Brazil



(c) Chile

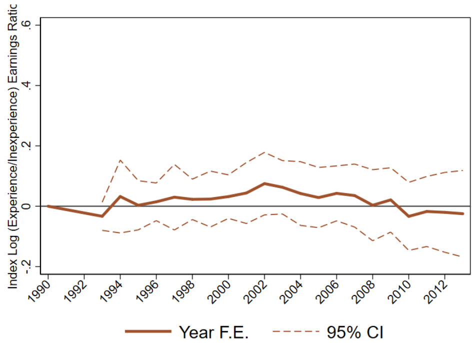


Notes: "Observed" refers to the Log (Skilled/Unskilled) Earnings Ratio observed in the data. "Predicted" refers to the model prediction derived from the estimation of Equation (1.4.11). The observed and predicted unskilled earnings series is constructed as a weighted average between the two low-skill sub-groups, where the weights correspond to the respective labor share. "Predicted (no demand trend)" is the prediction of a modified version the model in Equation (1.4.11) that omits the country-specific time trends.

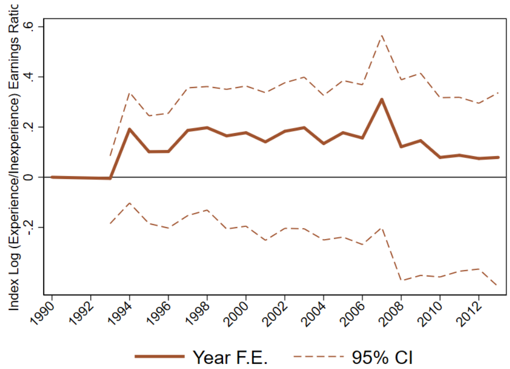
Figure 1.9: Demand Trends Estimates: Common (Across Countries) Year Fixed Effects Model

Experienced/Inexperienced Demand Index by Skill Level

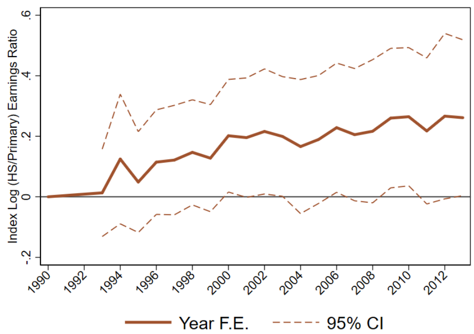
(a) Unskilled Workers ($\log \hat{\phi}_{Ut}$)



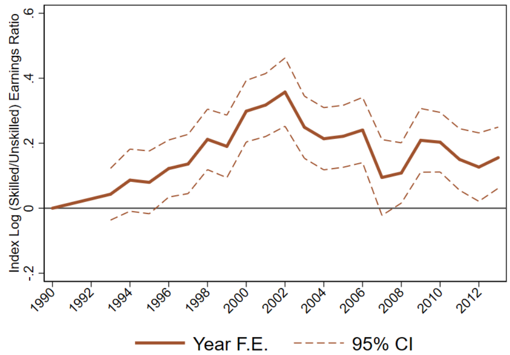
(b) Skilled Workers ($\log \hat{\phi}_{St}$)



(c) HS/Primary Demand Index ($\log \hat{\beta}_t$)

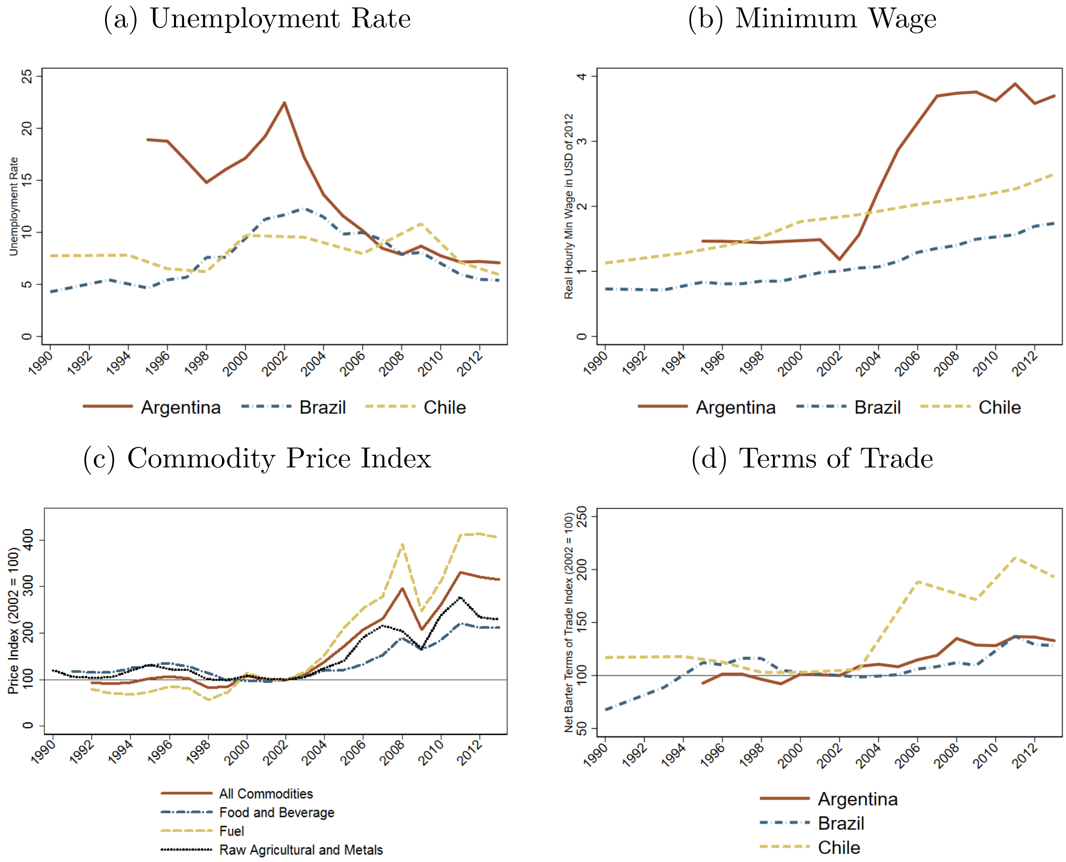


(d) Skilled/Unskilled Demand Index ($\log \hat{\alpha}_t$)



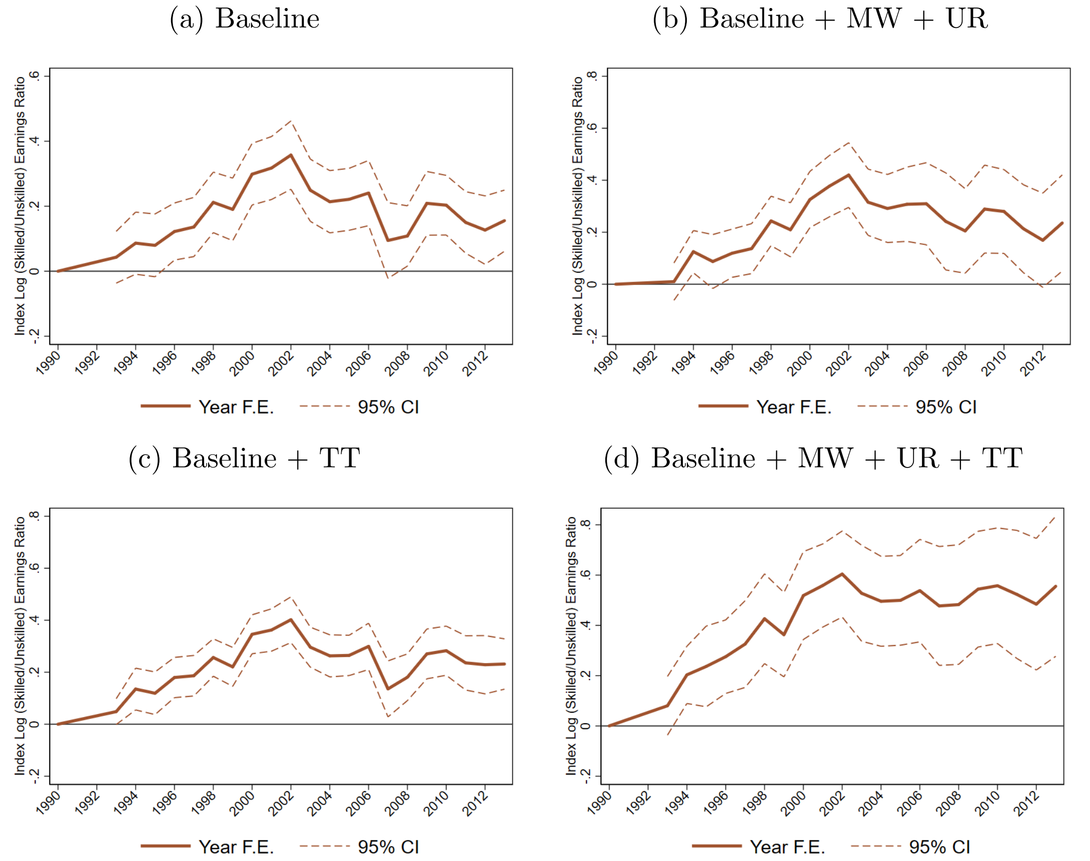
Notes: Each panel depicts the estimated relative demand trends from the model. The estimates correspond to an alternative specification in which we restrict relative demands to be equal across the countries using common year fixed-effects. Panel (a) and (b) shows the relative demand trends between workers of different experience levels within unskilled and skilled groups respectively. Panel (c) shows the relative demand trends between workers with at most a high school degree and those with at most primary education completed. Finally, Panel (d) shows the relative demand trends between skilled and unskilled workers. The demand trends are scaled so that they take a value of zero in 1990.

Figure 1.10: Relative Skill Demand. Conditioning Factors



Notes: The unemployment rate series in Panel (a) is taken from the World Economic Outlook (WEO) database. The Minimum wage series in Panel (b) is taken from the annual indicators of the International Labour Organization (ILO). The source of the series in Panel (c) is The IMF's Primary Commodity Price System. The Food and Beverage series includes cereal, vegetable oils, meat, seafood, sugar, bananas, oranges, coffee, tea, and cocoa. The Fuel series includes crude oil (petroleum), natural gas, and coal. The Raw Agricultural and Metals series includes timber, cotton, wool, rubber, hides, copper, aluminum, iron ore, tin, nickel, zinc, lead, and uranium. See `http://www.imf.org/external/np/res/commod/index.aspx` for a description of the construction of the indices. The series in Panel (d) are taken from the United Nations Conference on Trade and Development (UNCTAD). Unit value indexes are based on data reported by countries, supplemented by UNCTAD's estimates using the previous year's trade values at the Standard International Trade Classification three-digit level as weights.

Figure 1.11: Skilled/Unskilled Demand Index. The Role of Unemployment (UR), Minimum Wages (MW) and Terms of Trade (TT)



Notes: Each panel depicts the estimated relative demand trends between skilled and unskilled workers ($\log \hat{\alpha}_t$) using different specifications of the last stage of the year fixed effects model. Panel (a) corresponds to the estimates of the year fixed effects of column I in Table 1.6. Panel (b) corresponds to the estimates of the year fixed effects of a model that includes controls for the unemployment rate and the natural logarithm of the minimum wage. Panel (c) corresponds to the estimates of the year fixed effects of column IV in Table 1.6, which includes controls for the log of the terms of trade index. Panel (d) corresponds to the estimates of the year fixed effects of column V in Table 1.6, which includes controls for the log real minimum wage, the unemployment rate of each country, and the log of the terms of trade index. The demand trends are scaled so that they take a value of zero in 1990.

Table 1.1: 100 × Change in Employment Share

	Argentina 1995-2013	Brazil 1990-2013	Chile 1990-2013
Education			
Primary or less	-17.71	-34.53	-24.17
High School	7.70	22.42	10.79
College	10.02	12.11	13.38
Pot. Exper.			
[≥20]	-0.22	2.01	11.93
Sex			
Female	2.32	8.88	8.26
Educ. + Pot. Exper.			
Primary or less [0-19]	-8.16	-20.60	-15.57
Primary or less ≥ 20	-9.55	-13.93	-8.60
High School [0-19]	3.37	12.24	-3.00
High School ≥ 20	4.33	10.18	13.79
College [0-19]	5.01	6.35	6.64
College ≥ 20	5.01	5.76	6.74

Notes: Sample consists of full-time workers (reported working 35 hours or more) between ages 16 and 65. Tabulated numbers are changes in the employment shares for each group.

Table 1.2: Compositional Changes and Inequality Patterns: Oaxaca-Blinder Decomposition Results

	Argentina (1995-2013)		Brazil (1990-2013)		Chile (1990-2013)	
	Est.	[S.E]	Est.	[S.E]	Est.	[S.E]
Log (90/10)						
Overall	-0.091	[0.016]	-0.850	[0.017]	-0.290	[0.021]
Composition	0.056	[0.007]	0.282	[0.016]	0.249	[0.011]
Education	0.054	[0.007]	0.302	[0.016]	0.211	[0.010]
Experience	0.001	[0.001]	-0.003	[0.001]	0.051	[0.002]
Sex	0.002	[0.001]	-0.017	[0.001]	-0.013	[0.001]
Wage Structure	-0.147	[0.018]	-1.132	[0.028]	-0.538	[0.023]
Education	-0.271	[0.113]	-1.153	[0.121]	-1.685	[0.086]
Experience	-0.282	[0.044]	-0.825	[0.095]	-0.497	[0.055]
Sex	-0.049	[0.009]	-0.042	[0.008]	-0.033	[0.007]
Constant	0.454	[0.139]	0.888	[0.230]	1.677	[0.120]
Log (90/50)						
Overall	-0.214	[0.014]	-0.350	[0.011]	-0.149	[0.018]
Composition	0.056	[0.005]	0.222	[0.008]	0.183	[0.009]
Education	0.056	[0.005]	0.229	[0.009]	0.169	[0.009]
Experience	0.001	[0.001]	-0.001	[0.001]	0.030	[0.002]
Sex	-0.001	[0.000]	-0.005	[0.001]	-0.015	[0.001]
Wage Structure	-0.270	[0.015]	-0.572	[0.015]	-0.332	[0.021]
Education	0.084	[0.059]	0.076	[0.041]	-1.021	[0.072]
Experience	-0.253	[0.036]	-0.204	[0.054]	-0.297	[0.053]
Sex	-0.020	[0.005]	-0.013	[0.004]	-0.019	[0.007]
Constant	-0.080	[0.079]	-0.431	[0.090]	1.004	[0.102]
Log (50/10)						
Overall	0.123	[0.018]	-0.500	[0.013]	-0.140	[0.017]
Composition	0.001	[0.005]	0.060	[0.013]	0.065	[0.006]
Education	-0.002	[0.004]	0.073	[0.014]	0.042	[0.006]
Experience	-0.000	[0.001]	-0.002	[0.001]	0.021	[0.001]
Sex	0.003	[0.001]	-0.011	[0.001]	0.002	[0.001]
Wage Structure	0.123	[0.018]	-0.559	[0.024]	-0.206	[0.019]
Education	-0.355	[0.101]	-1.228	[0.095]	-0.664	[0.067]
Experience	-0.029	[0.039]	-0.621	[0.052]	-0.201	[0.027]
Sex	-0.028	[0.008]	-0.029	[0.006]	-0.014	[0.006]
Constant	0.534	[0.124]	1.319	[0.166]	0.672	[0.085]

Notes: Sample consists of full time workers (reported working 35 hours or more) between ages 16 and 65. Standard errors calculated via bootstrap with 100 replications.

Table 1.3: 100 × Changes in Real Composition-Adjusted Log Hourly Earnings

	Argentina	Brazil	Chile
	1995-2013	1990-2013	1990-2013
All	12.23	0.58	37.58
Sex			
Male	12.38	-2.38	33.48
Female	11.87	6.33	45.69
Education			
Primary or less	20.09	28.94	48.83
High School	13.81	-32.71	33.76
College	-2.47	-46.76	32.09
Educ. + Pot. Exper.			
Primary or less [0-19]	21.40	35.55	62.18
Primary or less ≥ 20	19.82	25.61	43.30
High School [0-19]	21.03	-24.02	47.94
High School ≥ 20	4.39	-49.48	16.21
College [0-19]	3.82	-41.01	37.53
College ≥ 20	-13.33	-55.00	22.68

Notes: Sample consists of full time workers (reported working 35 hours or more) between ages 16 and 65. See Appendix 1.7 for details on the construction of the compositionally adjusted series.

Table 1.4: Model Estimation Results

	STEP			
	IA	IB	II	III
Elasticities				
$-1/\sigma_{\theta_U}$	-0.309***		-0.323***	-0.359***
	(0.052)		(0.066)	(0.041)
$-1/\sigma_{\theta_S}$		-0.104		-0.111
		(0.238)		(0.089)
$-1/\sigma_{\delta}$			-0.448***	-0.471***
			(0.034)	(0.018)
$-1/\sigma_{\rho}$				-0.478***
				(0.125)
Demand Argentina				
Time	0.031*	0.034	0.029	0.067***
	(0.016)	(0.028)	(0.017)	(0.013)
Time^2/100	-0.417*	-0.538*	-0.290	-0.830***
	(0.245)	(0.312)	(0.245)	(0.165)
Time^3/1000	0.127	0.165	0.102	0.253***
	(0.102)	(0.109)	(0.099)	(0.059)
Demand Brazil				
Time	-0.019**	-0.011	-0.016	0.045**
	(0.008)	(0.020)	(0.014)	(0.016)
Time^2/100	0.265**	0.246	0.218	-0.251
	(0.083)	(0.238)	(0.136)	(0.181)
Time^3/1000	-0.079***	-0.086	-0.066*	0.029
	(0.024)	(0.078)	(0.035)	(0.051)
Demand Chile				
Time	0.011	0.035	0.076***	0.094***
	(0.012)	(0.022)	(0.014)	(0.017)
Time^2/100	0.016	-0.234	-0.562***	-0.557***
	(0.153)	(0.297)	(0.167)	(0.154)
Time^3/1000	-0.022	0.032	0.124**	0.094**
	(0.049)	(0.101)	(0.052)	(0.042)
N	96	48	96	192
R^2	0.801	0.660	0.943	0.959

*** 1 percent ** 5 percent * 10 percent. Standard errors calculated via bootstrap with 500 replications, where each bootstrap sample remains fixed for all the stages. Each column presents the results of the estimation of the different stages of the model. Column IA shows the OLS estimates of the inverse of the elasticity of substitution between experienced and inexperienced workers within the unskilled group ($\sigma_{\theta U}$) (see Equation (1.4.8)); column IB correspond to the OLS estimates of the inverse of the elasticity of substitution between experience and inexperience workers within the skilled group ($\sigma_{\theta S}$) (see Equation (1.4.9)); column II shows the OLS estimates of the inverse of the elasticity of substitution between the two unskilled sub-groups (σ_{δ}), and a second estimate of the inverse of the elasticity of substitution $\sigma_{\theta U}$ (see Equation (1.4.10)); finally, column III shows the OLS estimates of the inverse of the elasticity of substitution between skilled and unskilled labor (σ_{ρ}), as well as additional estimates from the other elasticities in the model.

Table 1.5: Model Estimation Results: Common Demand Trends Across Countries Using Year Fixed Effects

	Common Year Fixed Effects			
	I	II	III	IV
Elasticities				
$-1/\sigma_{\theta_U}$	-0.249*** (0.073)		-0.330*** (0.082)	-0.245*** (0.036)
$-1/\sigma_{\theta_S}$		-0.304 (0.636)		-0.282*** (0.071)
$-1/\sigma_{\delta}$			-0.549*** (0.081)	-0.546*** (0.015)
$-1/\sigma_{\rho}$				-0.510*** (0.064)
N	96	48	96	192
R2	0.773	0.652	0.935	0.976
Year Fixed Effects	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes

*** 1 percent ** 5 percent * 10 percent. Standard errors calculated via bootstrap with 500 replications, where each bootstrap sample remains fixed for all the stages. Each column presents the results of the estimation of the different stages of the year fixed effects model. Column IA shows the OLS estimates of the inverse of the elasticity of substitution between experienced and inexperienced workers within the unskilled group ($\sigma_{\theta U}$) (see Equation (1.4.8)); column IB correspond to the OLS estimates of the inverse of the elasticity of substitution between experience and inexperience workers within the skilled group ($\sigma_{\theta S}$) (see Equation (1.4.9)); column II shows the OLS estimates of the inverse of the elasticity of substitution between the two unskilled sub-groups (σ_{δ}), and a second estimate of the inverse of the elasticity of substitution $\sigma_{\theta U}$ (see Equation (1.4.10)); finally, column III shows the OLS estimates of the inverse of the elasticity of substitution between skilled and unskilled labor (σ_{ρ}), as well as additional estimates from the other elasticities in the model.

Table 1.6: Model Estimation Results Skilled/Unskilled Premium: Including and Excluding the Minimum Wage, the Unemployment Rate, and the Terms of Trade Index.

	Common Year Fixed Effects				
	I	II	III	IV	V
Elasticities					
$-1/\sigma_{\theta_U}$	-0.245***	-0.262***	-0.255***	-0.262***	-0.269***
	(0.037)	(0.034)	(0.038)	(0.037)	(0.034)
$-1/\sigma_{\theta_S}$	-0.282***	-0.290***	-0.296***	-0.300***	-0.312***
	(0.074)	(0.065)	(0.064)	(0.065)	(0.062)
$-1/\sigma_{\delta}$	-0.546***	-0.545***	-0.548***	-0.547***	-0.546***
	(0.014)	(0.014)	(0.014)	(0.015)	(0.013)
$-1/\sigma_{\rho}$	-0.510***	-0.526***	-0.667***	-0.432***	-0.761***
	(0.064)	(0.079)	(0.112)	(0.077)	(0.119)
Log Real Min. Wage					
Argentina		-0.041			-0.026
		(0.051)			(0.139)
Brazil		-0.083			-0.300**
		(0.085)			(0.145)
Chile		0.083			-0.433**
		(0.123)			(0.191)
Unemployment Rate					
Argentina			-0.003		-0.006
			(0.004)		(0.009)
Brazil			-0.015**		-0.024**
			(0.007)		(0.009)
Chile			0.011		0.018
			(0.010)		(0.011)
Log Terms of Trade					
Argentina				-0.197*	-0.585**
				(0.107)	(0.274)
Brazil				-0.200**	-0.355**
				(0.098)	(0.173)
Chile				0.003	0.040
				(0.050)	(0.134)
N	192	192	192	192	192
R2	0.975	0.977	0.977	0.977	0.981
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes

*** 1 percent ** 5 percent * 10 percent. Standard errors calculated via bootstrap with 500 replications, where each bootstrap sample remains fixed for all the stages. The Table reports the third stage estimates of the parameters of the (fixed effects) model when we include the natural logarithm of the real minimum hourly wage, the unemployment rate, and the natural logarithm of the terms of trade index.

Table 1.7: Model Estimation Results: Alternative Supply Measures

	Supply Measure		
	Working Age Pop.	Occupied Pop.	Total Hours Worked
Elasticities			
$-1/\sigma_{\theta_U}$	-0.359***	-0.337***	-0.321***
	(0.041)	(0.044)	(0.041)
$-1/\sigma_{\theta_S}$	-0.111	-0.128	-0.092
	(0.089)	(0.081)	(0.081)
$-1/\sigma_{\delta}$	-0.471***	-0.457***	-0.451***
	(0.018)	(0.018)	(0.018)
$-1/\sigma_{\rho}$	-0.478***	-0.505***	-0.393***
	(0.125)	(0.117)	(0.105)
Demand Argentina			
Time	0.067***	0.076***	0.075***
	(0.013)	(0.014)	(0.014)
Time^2/100	-0.830***	-0.958***	-0.920***
	(0.165)	(0.163)	(0.173)
Time^3/1000	0.253***	0.290***	0.274***
	(0.059)	(0.057)	(0.061)
Demand Brazil			
Time	0.045**	0.048**	0.055**
	(0.016)	(0.016)	(0.017)
Time^2/100	-0.251	-0.280	-0.289
	(0.181)	(0.176)	(0.176)
Time^3/1000	0.029	0.035	0.031
	(0.051)	(0.050)	(0.049)
Demand Chile			
Time	0.094***	0.088***	0.102***
	(0.017)	(0.015)	(0.014)
Time^2/100	-0.557***	-0.559***	-0.596***
	(0.154)	(0.144)	(0.145)
Time^3/1000	0.094**	0.105**	0.101**
	(0.042)	(0.040)	(0.041)
N	192	192	192
R^2	0.959	0.959	0.957

*** 1 percent ** 5 percent * 10 percent. Standard errors calculated via bootstrap with 500 replications, where each bootstrap sample remains fixed for all the stages. Each column presents the results of the estimation of the third stage of the model using alternative measures to construct total labor supply. The first column corresponds to our baseline results; the second column limits the sample to include only occupied population; and the final column uses total hours worked.

Table 1.8: Model Estimation Results. Prime Age Workers (25-55)

	STEP			
	IA	IB	II	III
Elasticities				
$-1/\sigma_{\theta_U}$	-0.345***		-0.398***	-0.426***
	(0.050)		(0.049)	(0.048)
$-1/\sigma_{\theta_S}$		-0.182		-0.195**
		(0.148)		(0.070)
$-1/\sigma_{\delta}$			-0.439***	-0.462***
			(0.033)	(0.019)
$-1/\sigma_{\rho}$				-0.562***
				(0.098)
Demand Arg.				
Time	0.038**	0.055**	0.027	0.069***
	(0.016)	(0.022)	(0.017)	(0.016)
$\text{Time}^2/100$	-0.595**	-0.815**	-0.346	-0.916***
	(0.249)	(0.288)	(0.259)	(0.190)
$\text{Time}^3/1000$	0.206**	0.265**	0.140	0.293***
	(0.104)	(0.101)	(0.108)	(0.067)
Demand Bra.				
Time	-0.027**	0.009	-0.014	0.042**
	(0.010)	(0.015)	(0.014)	(0.017)
$\text{Time}^2/100$	0.351***	0.087	0.194	-0.233
	(0.102)	(0.199)	(0.141)	(0.177)
$\text{Time}^3/1000$	-0.098***	-0.057	-0.055	0.031
	(0.029)	(0.066)	(0.038)	(0.047)
Demand Chl.				
Time	0.007	0.044**	0.081***	0.099***
	(0.013)	(0.015)	(0.014)	(0.013)
$\text{Time}^2/100$	0.030	-0.392**	-0.614***	-0.661***
	(0.162)	(0.178)	(0.170)	(0.122)
$\text{Time}^3/1000$	-0.020	0.082	0.139**	0.130***
	(0.051)	(0.052)	(0.053)	(0.035)
N	96	48	96	192
R^2	0.801	0.642	0.936	0.960

*** 1 percent ** 5 percent * 10 percent. Standard errors calculated via bootstrap with 500 replications, where each bootstrap sample remains fixed for all the stages. Each column presents the results of the estimation of the different stages of the model, restricting the sample to workers between the ages of 25 and 55. Column IA shows the OLS estimates of the inverse of the elasticity of substitution between experience and inexperience workers within the low-skilled group ($\sigma_{\theta U}$) (see Equation (1.4.8)); column IB corresponds to the OLS estimates of the inverse of the elasticity of substitution between experience and inexperience workers within the skilled group ($\sigma_{\theta S}$) (see Equation (1.4.9)); column II shows the OLS estimates of the inverse of the elasticity of substitution between the two low-skill groups (σ_{δ}), and a second estimate of the inverse of the elasticity of substitution $\sigma_{\theta U}$ (see Equation (1.4.10)); finally, column III shows the OLS estimates of the inverse of the elasticity of substitution between skilled and unskilled workers (σ_{ρ}), as well as additional estimates from the other elasticities in the model.

Table 1.9: Model Estimation Results: Part-Time and Full-Time Workers.

	Supply Measure		
	Working Age Pop.	Occupied Pop.	Total Hours Worked
Elasticities			
$-1/\sigma_{\theta_U}$	-0.389***	-0.373***	-0.359***
	(0.046)	(0.043)	(0.044)
$-1/\sigma_{\theta_S}$	-0.171*	-0.171*	-0.125
	(0.103)	(0.093)	(0.092)
$-1/\sigma_{\delta}$	-0.478***	-0.466***	-0.462***
	(0.021)	(0.021)	(0.020)
$-1/\sigma_{\rho}$	-0.432**	-0.467***	-0.372**
	(0.146)	(0.132)	(0.117)
Demand Arg.			
Time	0.040**	0.049**	0.049**
	(0.015)	(0.016)	(0.016)
$\text{Time}^2/100$	-0.414**	-0.544**	-0.501**
	(0.206)	(0.211)	(0.214)
$\text{Time}^3/1000$	0.125	0.165**	0.145*
	(0.079)	(0.080)	(0.081)
$\text{Time}^3/10000$			
Demand Bra.			
Time	0.049**	0.051**	0.058**
	(0.018)	(0.017)	(0.019)
$\text{Time}^2/100$	-0.301	-0.331*	-0.338*
	(0.214)	(0.191)	(0.201)
$\text{Time}^3/1000$	0.049	0.055	0.052
	(0.061)	(0.054)	(0.057)
$\text{Time}^3/10000$			
Demand Chl.			
Time	0.103***	0.096***	0.110***
	(0.022)	(0.019)	(0.018)
$\text{Time}^2/100$	-0.702***	-0.691***	-0.736***
	(0.187)	(0.164)	(0.179)
$\text{Time}^3/1000$	0.150**	0.156***	0.155**
	(0.051)	(0.045)	(0.051)
$\text{Time}^3/10000$			
Observations	192	192	192
R^2	0.942	0.943	0.943

*** 1 percent ** 5 percent * 10 percent. Standard errors calculated via bootstrap with 500 replications, where each bootstrap sample remains fixed for all the stages. The table reports the third stage estimates of the parameters of the model when we include part-time workers in the construction of the wage series. Each column presents the results using alternative measures of the total labor supply by each group. The first column corresponds to our baseline results; the second column limits the sample to include only employed population; and the final column uses total hours worked.

Table 1.10: Model Estimation Results: Alternative Production Function

	STEP		
	I	II	III
Elasticities			
$-1/\sigma_\delta$	-0.372***	-0.264***	-0.341***
	(0.026)	(0.011)	(0.014)
$-1/\sigma_\theta$		-0.138**	-0.171***
		(0.045)	(0.043)
$-1/\sigma_\rho$			-0.762***
			(0.047)
Demand Argentina			
Time	0.027*		0.108***
	(0.014)		(0.012)
$\text{Time}^2/100$	-0.002		-0.011***
	(0.002)		(0.002)
$\text{Time}^3/1000$	0.000		0.000***
	(0.000)		(0.000)
Demand Brazil			
Time	-0.003		0.007
	(0.012)		(0.010)
$\text{Time}^2/100$	0.001		0.001
	(0.001)		(0.001)
$\text{Time}^3/1000$	-0.000		-0.000
	(0.000)		(0.000)
Demand Chile			
Time	0.062***		0.084***
	(0.010)		(0.009)
$\text{Time}^2/100$	-0.005***		-0.005***
	(0.001)		(0.001)
$\text{Time}^3/1000$	0.000**		0.000**
	(0.000)		(0.000)
N	336	672	672
R^2	0.973	0.934	0.886

*** 1 percent ** 5 percent * 10 percent. Standard errors calculated via bootstrap with 500 replications, where each bootstrap sample remains fixed for all the stages. Each column presents the results of the estimation of the different steps in the alternative model described in the Appendix 1.7.4.

1.7.2 Data and Variable Construction

The household surveys used in Argentina for the period between 1995 and 2003 are waves of the Encuesta Permanente de Hogares (EPH), collected by the Instituto Nacional de Estadística (INDEC). This survey was replaced by the Encuesta Permanente de Hogares Continiua (EPH-C) after 2003, breaking the series. The transition between the EPH and the EPH-C included changes in the questionnaires and the frequency in which the surveys were collected. The geographical coverage in EPH-C was extended to include additional agglomerates. In order to maintain consistency over the period of study we only use the agglomerates that are present in both surveys. The EPH and the EPH-C are representative for urban areas, but close to 90 percent of the population in Argentina live in urban centers.

The survey used in Brazil is the Pesquisa Nacional por Amostra de Domicilios (PNAD), collected by the Instituto Brasilero de Geografía y Estadísticas (IBGE). The PNAD is a nationally representative survey that has been carried out on a yearly basis since 1967. We use the different waves starting from the year 1990. Due to exceptional circumstances the survey was not collected in 1994 and 2000.

The household survey used for Chile is the Encuesta de Caracterización Socioeconómica Nacional (CASEN). The CASEN is a nationally representative household survey collected by the Ministry of Planning through the Department of Economics at Universidad de Chile. The survey was first implemented in 1987 and was carried out every two years from 1990 to 2000, and every three years thereafter. We use all the waves from 1990 to 2013.

We constructed variables capturing the educational attainment and potential experience of all individuals in the sample. Although the countries we analyze differ in the structure of their educational systems, the SEDLAC project has attempted to homogenize the information from the different countries to make it comparable.[26] We use SEDLAC's coding in the construction of the educational attainment series. In particular, we define five possible levels of educational attainment: i) primary education completed or less; ii) high school incomplete; iii) high school completed; iv) college incomplete; and v) college completed or more. Potential experience is defined as the result of subtracting the total number of years of education completed (plus 6) from the age of the individual.

Although we define five possible levels of educational attainment, we mostly work with three categories: primary or less, high school completed and college completed. Individuals with incomplete levels of education are distributed equally between the previous and next completed level. For example, mean real hourly earnings of workers with college education are calculated as a weighted average between the observed mean wages of this group and the observed mean wages of workers with college incomplete. The weight of the latter group is equal to half of their actual number. This also implies that in the labor supplies used in the model, the supply of workers with primary education completed or less includes half of the total supply of workers of the high school incomplete category. The supply of workers with high school education completed includes both half of the supply of workers with high school incomplete and half of the supply of workers with college incomplete. Finally, the supply of workers with college education completed includes half of the total supply of workers with college incomplete.

Each survey includes a question asking workers for the total monetary income from labor in a reference period. This is the variable that we use throughout the Chapter to capture labor earnings. The variable is divided by the total number of hours worked to obtain hourly earnings. The series are converted into real

[26] See CEDLAS and The World Bank (2014) for a detailed description of the SEDLAC database

terms using the consumer price index of the respective countries.[27] In the main specification we restrict the sample to individuals between the ages of 16 and 65, and only use earnings of full-time workers (individuals working 35 hours or more in the reference week).

The composition adjusted earnings of aggregate groups are constructed using a fixed-weighted average of the different sex-education-experience sub-groups. We first run a regression of log hourly earnings on the full set of covariates that include indicators for the five education categories, seven dummies for potential experience in five-year intervals, and all possible interactions. The regression is estimated separately for males and females in each available country-year. The predicted log wages from these regressions are evaluated for the 70 sub-groups, and a weighted average is estimated when aggregating to broader groups. The weights are equal to the mean employment share of each sub-group across all years.

1.7.3 Using RIF to Decompose Changes in Distributional Statistics beyond the Mean

Firpo et al. (2007, 2009) allow extending the traditional Oaxaca-Blinder decomposition to distributional statistics beyond the mean. This is achieved through the use of influence functions (IF). Influence functions measure the effect that an infinitesimal amount of "errors" have on a given estimator (Cowell and Victoria-Feser, 1996), but they also have properties that allows us to model the sensitivity of a given unconditional wage quantile to a change in a set of covariates. To see this, let $q_\tau(F_W)$ be τth quantile of the distribution of wages, expressed in terms

[27]Due to inconsistencies found in the official Consumer Price Index in Argentina (see Cavallo (2013)), we use the information from PriceStats (`http://www.statestreet.com/ideas/pricestats.html`) to deflate nominal wages in this country.

of the cumulative distribution $F_W(w)$. Define the following mixture distribution:

$$G_{W,\epsilon} = (1-\epsilon)F_W + \epsilon H_W \quad for \quad 0 \leq \epsilon \leq 1 \tag{1.7.1}$$

where H_W is some perturbation distribution that only puts mass at the value w. In that case, $G_{W,\epsilon}$ is a distribution where, with probability $(1-\epsilon)$, the observation is generated by F_W, and with probability ϵ, the observation takes the arbitrary value of the perturbation distribution. By definition, the influence function corresponds to:

$$IF(w; q_\tau, F_W) = lim_{\epsilon \to 0} \ \frac{q_\tau(G_{W,\epsilon}) - q_\tau(F_W)}{\epsilon} \tag{1.7.2}$$

where the expression is analogous to the directional derivative of q_τ in the direction of H_W. Analytical expressions for influence functions have been derived for many distributional statistics.[28] The influence function in the case of the τth quantile takes the form:

$$IF(w; q_\tau, F_W) = \frac{\tau - \mathbb{1}[w \leq q_\tau]}{f_W(q_\tau)} \tag{1.7.3}$$

where $\mathbb{1}[\cdot]$ is an indicator function and f_W is the PDF.[29] Using some of the properties of influence functions, a direct link with the traditional Oaxaca-Blinder approach can be established. In particular, a property that is shared by influence functions is that, by definition, the expectation is equal to zero.

[28] Essama-Nssah and Lambert (2011) provides a comprehensive list of influence functions for different distributional statistics.

[29] Note that the influence function in this case depends on the density. In order to obtain the empirical density the authors propose non-parametric kernel density estimation.

$$\int_{-\infty}^{+\infty} IF(w; q_\tau, F_W)dF(w) = 0 \tag{1.7.4}$$

Firpo et al. (2009) propose a simple modification in which the quantile is added back to the influence function, resulting in what the authors call the Recentered Influence Function (RIF).

$$RIF(w; q_\tau, F_W) = q_\tau + IF(w; q_\tau, F_W) \tag{1.7.5}$$

The importance of this transformation lies in the fact that the expectation of the RIF is precisely the quantile q_τ. With this result, Firpo et al. (2009) show that we can model the conditional expectation of the RIF as a linear function of the explanatory variables.

$$E[RIF(w_t; q_\tau, F_{W,t}|X_t)] = X_t'\gamma_t \tag{1.7.6}$$

Moreover, if we apply the law of iterated expectations to Equation 1.7.6, the end result is an expression that directly relates the impact of changes in the expected values of the covariates on the unconditional quantile q_τ. Note that this result is all that is required to extend the Oaxaca-Blinder decomposition to quantiles, since the basic components of the method are all present in Equation (1.7.6).

Estimation of Equation (1.7.6) can be done by OLS, and only requires replacing the dependent variable, $\log w_t$ in the original wage setting model with the RIF of the quantile q_τ. The interpretation of the estimates $\widehat{\gamma}_t$ can be thought of as the effect of a small change in the distribution of X on q_τ, or as linear approximation of the effect of large changes of X on q_τ (Firpo et al., 2007).

1.7.4 Alternative Model Specification

In this section we present an alternative specification of the production function of the model in Section 1.4.1. We mostly follow the work of Manacorda et al. (2010), with some small modifications to allow for comparability with our baseline results. Production in the economy is also modeled using a nested constant elasticity of substitution (CES) function with three levels. The first level is identical to the one we use

$$Y_t = \lambda_t \left(L_{Ut}^{\rho} + \alpha_t L_{St}^{\rho}\right)^{1/\rho} \tag{1.7.7}$$

with the parameters having the same interpretation. In the second level, labor from skilled and unskilled workers is divided into seven potential experience subgroups, aggregating them with a productivity-weighted CES combination of the form

$$L_{Mt} = \left(\sum_{A=1}^{7} \phi_{MA} L_{MAt}^{\theta}\right)^{1/\theta} \quad \text{for} \quad M = S, U \tag{1.7.8}$$

where A indexes the potential experience groups; ϕ_{MA} is a time-invariant parameter capturing differences in relative productivities between potential experience groups; and θ is a function of the elasticity of substitution: $\sigma_\theta = \frac{1}{1-\theta}$. Two key differences with our baseline model are worth pointing out. First, the second level of the production function aggregates labor by experience, not by skill subgroups. The ordering between the second and third levels is then shifted. Second, the model assumes that there are no relative demand/productivity changes between workers with different levels of potential experience. This follows from the assumption that the respective parameters (ϕ_{MA}) are time-invariant, which largely simplifies the estimation.

Finally, the supply of labor from workers with a given potential experience within the unskilled group is composed of a CES combination of labor from the two lower schooling levels

$$L_{UAt} = \left(L_{PAt}^{\delta} + \beta_t L_{HAt}^{\delta}\right)^{1/\delta} \tag{1.7.9}$$

where P and H denote workers with primary education or less and high school completed, respectively; β_t is a time-variant measure of the relative productivity between the two low education levels; and δ is a function of the elasticity of substitution between the two groups: $\sigma_\delta = \frac{1}{1-\delta}$. Note that β_t is constant across potential experience groups, so relative demand shifts are common in this dimension. Finally, the natural logarithm of the two time-variant parameters (α_t and β_t) are estimated using cubic time trends.

Two differentiating factors between this specification and the work of Manacorda et al. (2010) are worth pointing out. First, we allow for differential demand trends within low skilled workers, which they assume to be constant. Second, we allow for a more flexible specification of the demand trends by fitting a cubic polynomial instead of a linear time trend. This allows us to fit the trend reversals in the skill premiums that we observe in the data.

Chapter 2

The Rise of Female Labor Force Participation and Changes in the Wage Structure: Evidence from Structural Estimates in Mexico

2.1 Introduction

The secular increases in the labor force participation of women are one of the most salient features of labor markets in most countries over the last century (Killingsworth and Heckman, 1987; Goldin, 2006; Fogli and Veldkamp, 2011; Goldin and Olivetti, 2013). But although the massive entry of women into the workforce has fundamentally changed the gender composition of labor supply, there is little understanding of what its impact has been on the wage structure. The relative scarcity of empirical studies on this topic reflects the complexity of the phenomena, since a variety of supply and demand factors will influence the decision of women to participate in the labor market, and natural experiments that could provide exogenous variation in female labor supply have been elusive. Moreover, the few studies that have tackled this question provide mixed – and at times contradictory– conclusions, mostly because there is no agreement in the literature of whether male and female labor are close substitutes or not, a key factor underlying the debate.

This chapter studies the effect of the rapid rise of female labor force participation on the wage structure in Mexico. Using an equilibrium model of the labor market, I structurally estimate the elasticities of substitution between male and female labor that are consistent with observed patterns of the gender earnings gap since 1989. The equilibrium model of supply and demand for labor differs from other work on the subject in two significant ways: First, the model incorporates the ideas of task-based approach (Autor et al., 2003), allowing the elasticity of substitution between male and female labor to vary as a function of the task content of occupations. This feature of the model implies that the effect of female labor supply on relative earnings can vary throughout the pay distribution, a possibility that is suggested by the descriptive evidence. Second, following the work of Johnson and Keane (2013), labor supply is endogenously determined in the model, a feature that addresses one of the major criticisms of the canonical framework in the past.

The focus on Mexico is motivated by the extent to which Mexican women have joined the workforce over the last quarter century, which was the largest in the Latin American region (Ñopo, 2012), and one of the largest in the world (The World Bank, 2012): between 1989 and 2014, labor force participation among prime-age women in Mexico increased from 36 to 58 percent, going from 4.7 to 14.7 million female workers in a spam of 25 years – it took women in the U.S. 37 years, from 1956 to 1993, to cover the same ground. Moreover, had the female participation rate remained at the levels of 1989, there would be approximately 5.6 million fewer prime-age females in the labor force, close to 12 percent of its actual size in 2014.

At the same time that women started moving into the workforce, growth in labor earnings in Mexico diverged significantly between men and women across the occupational distribution. Growth in labor earnings among workers in low-paying occupations (e.g. agriculture, services, transportation, and repetitive production occupations) was much higher for males than for females, while the opposite is true among workers in high-paying occupations (e.g. professional, education, technical and managerial occupations). The implication of these diverging dynamics is that the (male/female) gender earnings gap increased for workers below the median of the pay distribution, but declined among workers above the 80th percentile.

The main proposition of this chapter is that these two phenomena are causally connected, so that the rapid increase of female labor supply has fundamentally changed the wage structure in the Mexican economy. In particular, I argue that the influx of women into the labor market has generated downward pressures to the wages of female workers in low-paying occupations, but to *both* male *and* female workers in high-paying occupations. The hypothesis I put forward is that this heterogeneity arises because the elasticity of substitution between male and female labor is not uniform throughout the occupational distribution: male and female labor are closer substitutes in occupations that rely heavily on abstract analytical skills than in those occupations in which physical or other non-cognitive skills are more prominent in the jobs.

To test this hypothesis I propose an extension to the canonical supply and demand framework popularized by Katz and Murphy (1992), Murphy and Welch (1992), and Juhn et al. (1993), which incorporates the ideas of the task-based approach of Autor et al. (2003). In the original formulation of the supply and demand framework, the degree to which changes in the composition of labor supply are associated with changes in the wage structure depends on the magnitude of the elasticity of substitution between the groups being compared, but these elasticities are usually taken to be constant across occupations. Starting from the work of Autor et al. (2003), a large strand of the literature[1] has emphasized that the complementary or substitutability between factors of production is determined by the type of tasks in which they are employed. In all occupations, cognitive, manual, physical, socio-emotional, and interpersonal skills are applied to specific tasks in different intensities, so the relative importance of any subset of skills is mostly determined by the nature of the activities being done. A logical implication of this theory is that as long as there is some difference in the bundle of skills that men and women are supplying to the labor market, compositional changes by gender will have heterogeneous impacts on relative earnings that depend on the task content of occupations.

The chapter contributes to the relatively scarce empirical literature studying the relation between female labor supply and changes in the wage structure. Arguably, the most influential paper on this topic is that of Acemoglu et al. (2004), who exploited State level variation in the U.S. military mobilizations for World War II to investigate the effects of a rise of female labor force participation on the wage structure during the 1940s and 1950s. The authors found that rising female labor supply reduces both male and female wages, and report estimates of the short-run elasticity of substitution between male and female labor of around 3. The authors qualify this finding arguing that this elasticity potentially varies across skill groups, but they are unable to properly test that hypothesis.

[1] See, among others, Autor et al. (2003); Goos and Manning (2007); Autor et al. (2008); Gathmann and Schoenberg (2010); Acemoglu and Autor (2011); Firpo et al. (2011); Autor and Dorn (2013); Altonji et al. (2014); Goos et al. (2014); Michaels et al. (2015).

In earlier studies, both Topel (1994) and Juhn and Kim (1999) examined the relation between the rise in female labor supply and rising inequality in the U.S. during the 1970s and 1980s, reaching opposite conclusions: while Topel finds that the rise in the supply of female labor negatively affected the wages of low-skilled workers, Juhn and Kim challenged this view arguing that, once changes in relative demand are accounted for, there is little evidence that women are close substitutes for men or that the entry of educated women into the labor force contributed to male wage inequality growth in the 1980s. More recently, Johnson and Keane (2013) estimated a dynamic equilibrium model of the labor market fitted to replicate the patterns of the U.S. wage structure from 1968 to 1996. They report a relatively high elasticity of substitution between male and female labor of between 4.76 and 5.6.

The lack of agreement about the range of values that the elasticity of substitution between male and female labor can take is surprising, especially since this is a structural parameter that can be of first-order importance in contexts in which women are transitioning rapidly into the workforce. I argue that the application of the task-based approach provides the right framework to bridge together the mixed evidence on this subject. In the particular case of Mexico, I estimate these elasticities to be between 1.2 in manual and routine task-intensive occupations, and 2.7 in abstract task-intensive occupations. These numbers imply that the change in the gender composition of the workforce has had a significant effect on the gender earnings gap in Mexico, especially at the lower end of the pay distribution.

The chapter also contributes to the literature studying changes in the wage structure in Latin American countries. As shown in Chapter 1, most countries in Latin America experienced a sharp fall in earnings inequality since the late 1990s. I show that Mexico also experienced a significant contraction of the male earnings distribution in the last 25 years, driven by higher wage growth among low-skilled workers, but that this was not the case for women: female wage growth was relatively uniform throughout the earnings distribution, a phenomenon that is partly attributable to the impact of rising female labor supply. The estimates from the model also suggest that part of the sluggish wage growth of high-skilled workers in the region could be explained by a combination of a high elasticity of substitution between male and female labor in high paying occupations, and the observed changes in relative supplies. This specific channel has been mostly overlooked by economists working on earnings inequality in Latin America.

The rest of the chapter is organized as follows: Section 2.2 discusses the data, sample selection, and treatment of the main variables used in the paper. Section 2.3 presents the main stylized facts, reviewing the evolution of male and female labor force participation in Mexico, and documenting the changes in the wage and occupational structure over the last quarter century. In Section 2.4 I formulate an equilibrium model of the labor market, and describe the empirical strategy that will be used to estimate its main parameters. Results are presented in Section 2.5, while robustness exercises using alternative specification of the model and different measures of labor supply are shown in Section 2.6. Finally, Section 2.7 concludes.

2.2 Data, Choice of Variables, and Sample Selection

The data for this study comes from the Mexican Household Income and Expenditure Survey (ENIGH), which is a nationally representative household survey carried out by the Mexican National Institute of Statistics and Geography (IN-

EGI).[2] The first three waves of ENIGH were collected intermittently (1984, 1989, and 1992), but from 1992 onwards the survey was collected biannually, with an exceptional gathering in 2005. I use the 13 waves of ENIGH between the years 1989 and 2014. As a check for consistency, I replicate part of the results using data from the 1990, 2000 and 2010 Mexican CENSUS.[3]

The focus of the chapter is on the wage structure, so the variable measuring individual incomes is restricted to include only remuneration from labor. In particular, I use the monthly monetary remuneration from labor in all occupations,[4] which include wages, salaries, piecework, and any overtime pay, commissions, or tips usually received, but excludes income received from government transfers or profits from self-employment work.[5] Monthly earnings are converted into hourly rates dividing by the worker's total hours of work per week in all jobs multiplied by the usual number of weeks in a month.

Each year, between 77 and 83 percent of workers reported having worked full time (35 hours or more in the previous week), but there are clear differences between males and females: on average, full-time female workers represent between 62 and 67 percent of all female workers, while the share of full-time male workers ranges between 87 and 90 percent. In order to have groups that are more closely comparable, the earnings series is calculated using incomes from full-time

[2]The main labor force survey in Mexico is the Encuesta Nacional de Ocupación y Empleo (ENOE), which replaced the Encuesta Nacional de Empleo (ENE) and the Encuesta Nacional de Empleo Urbano (ENEU) in 2005. The transition from the ENE/ENEU to the ENOE led to methodological changes that difficult comparability over time. To guarantee a time consistent series of employment and earnings I use the ENIGH survey.

[3]The CENSUS data corresponds to the extended questionnaire applied to a 10 percent sample of the population in each year.

[4]Only after the year 2008 it becomes possible to separate the remuneration arising from the main job from that of any secondary jobs. To keep a consistent series, I add up remuneration from labor in all jobs in every year.

[5]The extended questionnaire of the CENSUS only asks about overall remuneration from labor, which can include income from self-employment. Hence, the two series are not fully comparable.

workers only.[6] All earnings are transformed into real U.S. Dollars of 2012 using the Mexican Consumer Price Index and the purchasing power parity adjusted exchange rate estimated by the IMF. There were exceptionally low and high values of hourly earnings in the data, so I only use earnings from workers that reported having hourly rates above \$0.1 and below \$150.[7]

The ENIGH survey uses the Mexican occupation classification system to categorize workers according to the type of tasks they perform in the main job. The system went trough two changes since 1989: first there was an update of the original Clasificación Mexicana de Ocupaciones (CMO) in 1992, and then a full change to the newly introduced Sistema Nacional de Clasificación de Ocupaciones (SINCO) in 2010. These changes make the series incompatible at high levels of disaggregation of the occupational groups, but it is possible to homogenize the SINCO classification to the principal group level of the CMO using the comparability tables produced by INEGI.[8] The principal group division has 18 distinct occupational groups that can be consistently followed throughout the period of analysis.

Following the task-based framework, occupations are classified in three groups defined by whether the activities performed in the jobs are predominantly manual, routine, or abstract/analytical. The division is based on the measures constructed by Autor et al. (2003) from different sets of variables from the U.S. 1977 Dictionary of Occupational Titles (DOT). Details of the construction of the groups are presented in Appendix 2.8.2, while the final division is shown in Table 2.1.

Finally, the sample is limited to workers between the ages of 25 and 55

[6] Results of the main estimates including the earnings of part-time workers, or alternative measures that account for differences in the intensive margin, are presented in the robustness section of the paper.

[7] In no year did the percentage of the observations trimmed exceeded one percent. Moreover, the results in the paper are qualitatively unchanged if I include the outliers in the sample.

[8] `http://www.inegi.org.mx/est/contenidos/proyectos/aspectosmetodologicos/clasificadoresycatalogos/sinco.aspx`

(henceforth prime-age workers). This is done to guarantee I am working with a group that has a strong labor market attachment, and to ameliorate selection problems arising from changes in educational and retirement choices of different cohorts.

2.3 Trends in Female Labor Supply and Relative Earnings by Gender

In 1989, prime-age population in Mexico was just above 25 million people, out of which 64.2 percent were either working or actively searching for a job. Female labor force participation was as low as 36 percent, and women accounted for only 29 percent of the workforce (4.7 million). By 2014, this picture had changed drastically: prime-age population had grown to 48 million people, and women already represented 41 percent of the workforce, with close to 14.7 million females active in the labor market (Panels (a) and (b) of Figure 2.1). Moreover, overall participation rates had reached 76 percent, with all of the increase during this 25 year period accounted by the sharp rise of female labor force participation (58% in 2014) (see Panels (c) and (d) of Figure 2.1).

The extent to which Mexican women have joined the workforce over the last quarter century is remarkable: had the female participation rate remained at the levels of 1989, there would be 5.6 million fewer prime-age females in the labor force, which is close to 12 percent of its actual size in 2014. But the pace of this transition was also exceptional: in 1956, female labor force participation in the U.S. was similar to that of Mexico in the late 1980's, but it took the U.S. until 1993, 37 years later, to reach 58 percent, a rate at which economy stabilized (U.S. Bureau of Labor Statistics, 2017). Cross-country comparisons suggest that the rise of female labor force participation in Mexico during the last 25 years was the largest in the Latin American region (Ñopo, 2012), and one of the largest in the world (The World Bank, 2012).

Panels (a) and (b) of Figure 2.2 show that, on average, Mexican women earn between 16 and 22 percent less than their male counterparts, even after taking into account differences in the levels of education, age, and occupational choice.[9] Interestingly, the mean gender earnings gap in Mexico has remained more or less constant over the past 25 years, a fact that has generated significant debate over the reasons for such lack of convergence. But focusing only on mean relative earnings conceals the substantial heterogeneity in the earnings dynamics between men and women across the distribution. Panels (c) and (d) of Figure 2.2 show the change between C.1990 and C.2013 (1990 and 2010 in the CENSUS)[10] of log hourly earnings at different percentiles of the earnings distribution, calculated separately for men and women. Growth (decline) in earnings among workers at the bottom was significantly higher (lower) for males than for females, while the exact opposite is true among workers at the top. The implication of these divergent dynamics (Panels (e) and (f) of Figure 2.2) is that the gender earnings gap increased between 10 and 22 percent among workers below the median, but declined between 5 and 18 percent among workers above the 80th percentile, with the average gap remaining relatively stable across the years.

I postulate that these diverging dynamics of relative earnings across the distribution are associated to the sharp gender compositional shift that the Mexican economy experienced during this period. In particular, that the influx of

[9] The gender earnings gap corresponds to the estimated coefficient of a male dummy variable in a regression in which log hourly earnings is the dependent variable, and I include controls for education attainment (7 different categories), age (6 categories in five year intervals), occupation indicators (18 groups), and all possible interactions. The regressions are estimated separately in each year.

[10] To increase sample size in the ENIGH survey, I joined together surveys from 1989 and 1992, and from 2012 and 2014.

women into the labor market generated downward pressures to female earnings among workers in low-paying occupations, but to *both* male and female workers in high-paying occupations. The reason for this heterogeneity is that the elasticity of substitution between male and female labor is not homogeneous throughout the occupational distribution, so similar changes in male/female relative supplies in high and low-paying occupations can have very different affects on the wage structure.

The underlying theoretical base for this proposition is the canonical supply and demand framework popularized by Katz and Murphy (1992), Murphy and Welch (1992), and Juhn et al. (1993); and the more recent task-based approach proposed by Autor et al. (2003). Within the supply and demand framework, the degree to which changes in the composition of labor supply are associated with changes in the wage structure fundamentally depends on the magnitude of the elasticity of substitution between the groups being compared. A simple example can help clarify the mechanism. Suppose that workers in an economy were similar in every respect except for their gender, and that aggregate production could be characterized by a Constant Elasticity of Substitution (CES) function of the form:

$$Y_t = \left[\alpha_t L_{k,t}^{\rho} + (1-\alpha_t) L_{f,t}^{\rho}\right]^{1/\rho}, \tag{2.3.1}$$

where Y_t is total output at time t; $L_{k,t}$ is the total supply of labor from male workers; $L_{f,t}$ is the total supply of labor from female workers; α_t is a time-varying "share" parameter that captures differences in the intensity of labor demand between male and female labor, which is allowed to change in time; and $\rho \in (-\infty, 1]$ is a function of the elasticity of substitution (σ_ρ) between male and female labor: $\sigma_\rho \equiv \frac{1}{1-\rho}$. If the economy is operating along the demand curve, wages are equated to marginal productivities and the log (male/female) earnings ratio takes the form:

$$\log\left(\frac{W_{k,t}}{W_{f,t}}\right) = \log\left(\frac{\alpha_t}{1-\alpha_t}\right) - \frac{1}{\sigma_\rho}\log\left(\frac{L_{k,t}}{L_{f,t}}\right), \qquad (2.3.2)$$

where $W_{k,t}$ and $W_{f,t}$ are the wages of male and female workers respectively.

Equation (2.3.2) shows that the evolution of the log (male/female) earnings ratio depends on two factors: (i) how the relative supply of labor between males and females is evolving $\left(\frac{L_{k,t}}{L_{f,t}}\right)$, scaled by the inverse of the elasticity of substitution (σ_ρ); and (ii) how relative demands are moving, as captured by the variation in time of the log ratio of the α_t share. If male and female labor are imperfectly substitutable in production, that is, if σ_ρ has a "low" value, and relative demands trends are constant, large shifts in female labor supply would impose downward pressure on female wages, and to a lesser extent to male wages, leading to an increase of the gender earnings gap: as the supply of female labor becomes more abundant its relative price falls. If, on the other hand, male and female workers are close to perfect substitutes – "high" values of σ_ρ –, a rise in female labor force participation should depress *both* male and female wages, with relative earnings remaining fairly constant.

A direct implication of this simplified framework is that if there is some variation in the substitutability between male and female labor across the earnings distribution, relative changes in the gender composition of labor supply will not have a symmetric impact on the wage structure. But should we expect this type of heterogeneity in the labor market? Starting from the influential work of Autor et al. (2003), a rapidly growing strand of the literature has emphasized that the complementary or substitutability between factors of production is determined by the type of tasks in which they are employed.[11] Any job requires, to a higher or lesser degree, cognitive, manual, physical, socio-emotional, and interpersonal skills. The relative importance of any subset of skills is then a function of the specific activities that workers are performing. Hence, as long as there is some difference in the bundle of skills that men and women supply to the labor market, compositional changes by gender will have an affect on relative earnings, and the extent to which this mechanism matters will depend on the task content of occupations.

Panels (a)-(c) of Figure 2.3 present some suggestive evidence that the sharp change in the gender composition of labor supply in Mexico was negatively correlated with the change in the gender earnings gap – something we might expect from the supply and demand framework – but only in low-paying occupations. Each panel, corresponding to a specific occupational group, depicts two series: one has the evolution of log (male/female) earnings ratio, and the other has the evolution of log (male/female) relative supply. The three groups are chosen following the standard practice in the task-based framework, where occupations are divided based on whether the activities in the jobs are more intensive in manual, routine, or abstract tasks (see Table 2.1).[12]

This division has the advantage that it clearly separates occupations into groups defined by which skills and aptitudes are more relevant in the jobs, but also because it breaks the earnings distribution in three subsequent segments. For example, manual task-intensive occupations have more demands for physical than abstract analytical aptitudes, and require skills like strength and hand, eye, and

[11]This argument holds true when we think about the role that information and communication technologies have played in substituting workers in middle-income routine occupations, the phenomena that has received most attention from academics, but it also holds when we think about differences by gender or any other dimension.

[12]Section 2.2 discusses the details of the division of the principal level occupations along the three groups.

foot coordination; these occupations are mostly found at the lower end of the pay distribution, and include examples like agriculture, services, and transportation. Abstract task-intensive occupations have more demands for analytical aptitudes than for physical or manual labor, and require skills like quantitative reasoning, direction, control, and planning of activities; these occupations are found at the top of the pay distribution, with the most prominent examples being professionals and managers. Finally, routine task-intensive occupations are in a middle ground, with a mixture of physical and analytical demands, and were aptitudes like adaptability to repetitive work and finger dexterity play a significant role; these occupations tend to be located at the middle of the pay distribution, with examples being clerical and crafts and trades jobs.

The first three Panels of Figure 2.3 show that between 1989 to 2014, log (male/female) relative supply declined by 44 log points in both manual and routine task-intensive occupations, and by 69 log points in abstract task-intensive occupations. During the same period, the log (male/female) earnings ratio only increased in the lower-paying jobs: 11 log points in manual task-intensive occupations and 3 log points in routine task-intensive occupations. In the abstract task-intensive group the gender earnings gap fell by 10 log points.

The negative comovement between relative quantities and relative prices suggest that the rise of female labor force participation – the main driver behind the change in relative supply – is potentially a factor behind the rise of the gender earnings gap at the low end of the earnings distribution, but the argument is less clear for high-paying abstract task-intensive jobs, even though the relative shift in supply was of an even larger magnitude. Under the canonical framework, the fact that relative quantities and relative prices are moving in the same direction indicate that men and women are more closely substitutable in those jobs, or that relative demand trends must be favouring female workers at the top. The equilibrium model I propose in Section 2.4 will shed light and quantify the relative importance of these two alternatives.

Panel (d) of Figure 2.3 provides further suggestive evidence that the supply and demand framework, in combination with the task-based approach, might be useful to understand the changes in the gender earnings ratio in Mexico. The figure depicts the change in log mean (male/female) earnings ratio as a function of the change of log mean (male/female) relative supply between C.1990 and C.2013, where each circle represents one of the principal group occupations of the ENIGH. Although some occupations, mostly in the abstract task-intensive group, saw a positive comovement between relative quantities and relative prices, there is a clear negative relation between the magnitude of the compositional shift in labor supply and the magnitude of the change in the gender earnings gap.

2.3.1 Did the Wage Structure Change? Decomposing the Gender Earnings Gap Across Quantiles

The underlying assumption behind the previous descriptive evidence is that the wage structure did in fact change, which is not necessarily the case. The rise of female labor force participation was not the only major change in the Mexican labor market over this period, so pure compositional shifts, absent of any changes in relative pay, could also be consistent with the observed trends.[13] For example, as shown in Table 2.2, the share of prime-age women in the workforce with at least some college education increased from 14.5 to 24.0 percent between C.1990 and C.2013, while that of men went from 15.6 to 20.8 percent: the larger share

[13]A similar argument was made by Lemieux (2006) in the context of the debate about the rise of income inequality in the United States. Lemieux shows evidence that a substantial share of the the rise in residual earnings inequality in the U.S. – the largest component of overall earnings inequality – can be accounted by the fact that the earnings of workers that are older and have more schooling –both of which have increased their share in the workforce – tends to be more dispersed. This argument was later controverted by Autor et al. (2008).

of women with more schooling can be an alternative explanation for the observed convergence in earnings at the top of the distribution. Also, the average Mexican worker is becoming older, with the share of workers between 25 and 34 years of age falling between C.1990 and C.2013 from 46.5 to 34.9 percent in the case of women, and from 43.5 to 36 percent in the case of men: if the gender earnings gap increases with age (Barth et al., 2017), this age compositional change can also be a factor behind the divergence in earnings at the bottom of the distribution. Finally, women saw a slight relative shift towards the low-paying manual task-intensive occupations (from 34.4% to 35.3%), which is mostly explained by an increase of female workers in service occupations, while the participation of males in that occupational group declined (from 40.4% to 37.1%), which is mostly explained by a movement away from agricultural occupations.

Has the wage structure changed or is there a simple re-composition of the workforce? A variation of the decomposition method used in Chapter 1 can help disentangle the importance of these two potential drivers. Here, the idea is to exogenously fix the structure of relative earnings at the average level across the last 25 years, and then quantify the counterfactual levels of the gender earnings ratio at different percentiles under the observed compositional changes. Alternatively, I can keep the composition of the labor force fixed at a given point in time and construct counterfactual earning ratios to evaluate how changes in schooling, age and occupational premiums have affected the observed dynamics. Again, these are partial equilibrium counterfactuals, something that I will return to in the next section.

The decomposition methodology is based on Firpo et al. (2007, 2009), and the specific details can be found in Appendix 1.7.3. As a starting point, consider a transformed wage-setting model of the form:

$$RIFq_{\tau,gen,t} = X'_{gen,t}\gamma_{gen,t} + \epsilon_{gen,t}, \tag{2.3.3}$$

where subscript gen indicates if the worker is male ($gen = k$) or female ($gen = f$); the subscript t indicates the period, either initial ($t =$C.1990) or final ($t =$C.2013); $RIFq_{\tau,gen,t}$ represents the value of the RIF corresponding to the τ'th quantile of the earning distribution at time t and for gender gen; X is a vector of socio-demographic characteristics including quadratic terms in education and age, and a set of indicator variables for the 18 principal group level occupations of the ENIGH;[14] and $\epsilon_{gen,t}$ is the error term assumed to have zero conditional mean. We can estimate Equation (2.3.3) for each gender and period separately by OLS, and then express the estimated difference over time of the expected value of the earnings quantile $\hat{q}_\tau$ as:

$$\Delta_t\hat{q}_{\tau,gen} = \underbrace{\left(\overline{X'}_{gen,\text{C.2013}} - \overline{X'}_{gen,\text{C.1990}}\right)\hat{\gamma}_{gen,P}}_{\Delta_t\hat{q}_{X,\tau,gen}} + \underbrace{\overline{X'}_{gen,P}\left(\hat{\gamma}_{gen,\text{C.2013}} - \hat{\gamma}_{gen,\text{C.1990}}\right)}_{\Delta_t\hat{q}_{S,\tau,gen}}, \tag{2.3.4}$$

where overbars denote averages, and $\hat{\gamma}_{gen,P}$ and $\overline{X}_{gen,P}$ correspond to the estimated vectors of parameters and the explanatory variables of a wage-setting model in which observations are pooled across the two periods.[15] Here, $\Delta_t\hat{q}_{X,\tau,gen}$

[14] As in the Oaxaca-Blinder decomposition for the mean, this decomposition is not invariant to the reference variable chosen when covariates are categorical. We limit this problem by using quadratic polynomials in years of education and age instead of levels. In the case of occupations, I use the more disaggregate principal group level, and repeat the decomposition changing the reference groups, but results were qualitatively similar.

[15] This specific counterfactual allows us to analyze composition and wage structure effects relative to a baseline defined by both the (weighted) mean returns and (weighted) mean characteristics over the two periods.

corresponds to the composition effect, which captures the part of the change in the τ'th earnings quantile that is accounted for by changes in the average skill-demographic and occupational composition of workers, given that we set the returns at their (weighted) average over the two periods; and $\Delta_t \hat{q}_{S,\tau,gen}$ is the wage structure effect, and captures how changes in returns are affecting earnings at the quantile τ, given that the observable characteristics are fixed to be equal to their (weighted) average over time.

Since I am interested in the effects of composition and price changes on the gender earnings gap, I construct the following measures at 19 different percentiles:

$$\underbrace{\Delta_t \hat{q}_{\tau,k} - \Delta_t \hat{q}_{\tau,f}}_{\text{Overall}} = \underbrace{(\Delta_t \hat{q}_{X,\tau,k} - \Delta_t \hat{q}_{X,\tau,f})}_{\text{Composition}} + \underbrace{(\Delta_t \hat{q}_{S,\tau,k} - \Delta_t \hat{q}_{S,\tau,f})}_{\text{Wage Structure}}. \tag{2.3.5}$$

The results of the decompositions for 5 selected percentiles are shown in Table 2.3. Each cell reports the contribution of the row variable to the observed change in the log (male/female) earnings ratio between C.1990 and C.2013 at the given percentile. The first conclusion from the decompositions is that wage structure effects are quantitatively more important than pure compositional effects in explaining the change in the gender earnings gap across the distribution. Estimated wage structure effects contribute 72 percent of the observed rise in the gender earnings gap at the 10th percentile, and close to 98 percent at the median. Moreover, they over-predict the fall in the gap at the 90th percentile (-0.097 log points observed vs. -0.119 log points attributed to the wage structure).

The results also show that if the wage structure had remained constant at the average levels over the two periods, compositional effects would lead to a larger counterfactual gender earnings gap at all points of the distribution, which is indicative that pure compositional shifts are contributing to the expansion of the gap at the lower tail, but have impede further convergence at the top of the distribution.

The second conclusion from the decomposition is that, out of the three set of variables in the wage setting model, changes in relative pay across occupations are the most important factor behind the overall wage structure effects, particularly at the two tails of the distribution. Wage structure effects associated with the occupation variables account for 83 percent of the overall wage structure effects at the 10th percentile, and also over-predict the contribution of changes in relative pay at the top (-0.151 log points attributed to occupations vs. -0.119 log points attributed to the overall wage structure). This result can be seen more clearly in Figure 2.4, where I show both the overall change in the log (male/female) earnings ratio between C.1990 and C.2013, and the estimated amount of this change that is attributed to changes in the occupation wage structure. It is clear from the figure that the trends in the two series mimic each another.

The evidence from the decompositions support the claim that the structure of relative pay between males and females in Mexico has in fact changed, but the question remains if this phenomenon was driven by the rise in female labor force participation. In the next section I develop an equilibrium model of supply and demand for labor to try to answer this question. The objective of the model is threefold: first, to explore if a general equilibrium model combining the task-based approach with the canonical supply and demand framework is able to recreate the patterns and changes in the Mexican wage structure over the past 25 years; second, to estimate the key structural parameters to test the hypothesis of differential degrees of substitutability between male and female labor across the occupational distribution; finally, to run counterfactual exercises and get quantitative estimates of the effect of the rise of female labor force participation on the gender earnings gap.

2.4 Theoretical Model

2.4.1 Demand Side

The model assumes that aggregate production in the economy is a function of the amount of labor that imperfectly substitutable types of workers supply to the market. Agents are divided into four types according to their gender – male or female – and their level of schooling – secondary education at most, referred as unskilled, or at least some college education, referred as skilled.[16] Each type of agent chooses between entering the workforce in one of three possible market occupations: abstract, routine or manual task-intensive; or, alternatively, the agents can opt to stay in home production.

Furthermore, the model assumes that the aggregate production technology can be described by a three-level nested constant elasticity of substitution (CES) function, where each nest of the production technology corresponds to a given dimension: occupation, education, and gender. At the top level, output is produced by a CES combination of labor in the three types of market occupations:

$$Y_t = Z_t \left[\alpha_{1,t} L_{a,t}^{\rho_1} + (1-\alpha_{1,t}) \Big(\alpha_{2,t} L_{r,t}^{\rho_2} + (1-\alpha_{2,t}) L_{m,t}^{\rho_2} \Big)^{\rho_1/\rho_2} \right]^{1/\rho_1}, \tag{2.4.1}$$

where Y_t is total output at time t; Z_t is a scale parameter that is allowed to vary

[16] In order to maintain a tractable number of parameters in the model, I omit the age dimension from this analysis. The decision is also supported by the fact that the age dimension played a secondary role for explaining the patterns of the gender earnings gap in the decomposition.

in time to capture skill-neutral technological change; $L_{a,t}$, $L_{r,t}$, and $L_{m,t}$ are the total supply of labor in abstract, routine, and manual task-intensive occupations respectively; $\rho_1 \in (-\infty, 1]$ is a function of the elasticity of substitution (σ_{ρ_1}) between labor in non-abstract – manual and routine – and abstract task-intensive occupations ($\sigma_{\rho_1} \equiv \frac{1}{1-\rho_1}$); and $\rho_2 \in (-\infty, 1]$ is a function of the elasticity of substitution (σ_{ρ_2}) between labor in routine and manual task-intensive occupations ($\sigma_{\rho_2} \equiv \frac{1}{1-\rho_2}$).[17] Finally, $\alpha_{1,t}$ is a time-varying share parameter that captures both differences in the intensity of labor used between non-abstract – routine and manual – and abstract task-intensive occupations, as well as movements in relative demands between them. Similarly, $\alpha_{2,t}$ captures both differences in the relative intensity of labor used between routine and manual task-intensive occupations, and movements in relative demands between them. General examples of possible sources of shifts in relative demand include non-neutral technical change (e.g. routine-biased technical change as in Goos et al. (2014)); variations in non-labor input demands (e.g. through capital skill complementarity as in Krusell et al. (2000)); product market demand shifts (e.g. through changes in the external demand for commodities as in Fernandez and Messina (2017)); and trade and outsourcing (e.g. generating incentives to modernize the production processes as in Juhn et al. (2014), and increasing competition with local industries as in Autor et al. (2016)).

In the second level of the production technology, labor in each occupation is divided in two groups based on the schooling level of the workers. In particular, labor in each occupation consists of a productivity weighted CES combination of labor from skilled workers, indexed by s, and labor from unskilled workers,

[17] Note that by assumption, the elasticity of substitution between labor in abstract and manual task-intensive occupations is the same as the elasticity of substitution between abstract and routine task-intensive occupations. We consider this a natural way of organizing the three occupational groups since I observe they are align this way in the earnings distribution – low vs. high paying occupations –, but I present results using alternative specifications in the robustness section of the paper.

indexed by u. That is:

$$L_{occ,t} = \left[\alpha_{3,occ,t} L_{s,occ,t}^{\rho_{3,occ}} + (1-\alpha_{3,occ,t}) L_{u,occ,t}^{\rho_{3,occ}}\right]^{1/\rho_{3,occ}} \quad \text{for} \quad occ = a, r, m, \quad (2.4.2)$$

where the parameters have an analogous interpretation to those in Equation (2.4.1).

Finally, in the third level of the production technology labor is disaggregated in each occupation-education group according to the gender of the workers. This is done using a productivity weighted CES combination of female workers, indexed by f, and male workers, indexed by k. That is:

$$L_{edu,occ,t} = \left[\alpha_{4,edu,occ,t} L_{k,edu,occ,t}^{\rho_{4,occ}} + (1-\alpha_{4,edu,occ,t}) L_{f,edu,occ,t}^{\rho_{4,occ}}\right]^{1/\rho_{4,occ}} \quad \text{for} \quad edu = s, u, \quad \text{and} \quad occ = a, r, m, \quad (2.4.3)$$

where the parameters have an analogous interpretation to those in Equation (2.4.1). Note that elasticities of substitution between male and female labor are allowed to vary between occupations, but not within occupation and education. Any differences in substitutability between labor by education group is captured by the relevant parameters in the second level of the production technology.[18]

The demand side of the model has two types of relevant parameters that

[18]To test how sensitive are our results to the selection of the ordering of the levels in the production technology, I report results using alternative model specifications in the robustness section.

we need to estimate: eight parameters that are functions of the elasticities of substitution (ρ_1, ρ_2, $\rho_{3,a}$, $\rho_{3,r}$, $\rho_{3,m}$, $\rho_{4,a}$, $\rho_{4,r}$, and $\rho_{4,m}$); and a set of time varying relative productivities/demand shifters parameters (Z_t, $\alpha_{1,t}$, $\alpha_{2,t}$, $\alpha_{3,a,t}$, $\alpha_{3,r,t}$, $\alpha_{3,m,t}$, $\alpha_{4,s,a,t}$, $\alpha_{4,s,r,t}$, $\alpha_{4,s,m,t}$, $\alpha_{4,u,a,t}$, $\alpha_{4,u,r,t}$, and $\alpha_{4,u,m,t}$). As argued by Johnson and Keane (2013), it is possible to fit the trends in relative wages perfectly if we did not impose any restrictions on the evolution of the relative demand parameters, but this would mean that we would not be able to identify the parameters capturing the elasticities of substitution. I then restrict these relative productivities to follow a cubic trend in their natural logarithm.[19] For example, the parameter $\alpha_{1,t}$ is allowed to change according to:

$$\log \alpha_{1,t} = \alpha_{1,0} + \alpha_{1,1}t + \alpha_{1,2}t^2 + \alpha_{1,3}t^3. \tag{2.4.4}$$

A simple example is useful to understand how the different set of parameters are identified in the model. Note that under the assumption that the economy is operating along the demand curve, log relative earnings between male and female labor within a given occupation and education group can be expressed as:

$$\log\left(\frac{W_{k,edu,occ}}{W_{f,edu,occ}}\right) = \log\left(\frac{\alpha_{4,edu,occ,t}}{1-\alpha_{4,edu,occ,t}}\right) - \frac{1}{\sigma_{\rho_{4,occ}}}\log\left(\frac{L_{k,edu,occ,t}}{L_{f,edu,occ,t}}\right) \quad \text{for} \quad edu = s,u, \quad \text{and} \quad occ = a,r,m, \tag{2.4.5}$$

where $W_{k,edu,occ}$ and $W_{f,edu,occ}$ are the wages of the respective groups. In the

[19]The cubic trends provided the best fit of the model to the data. Quadratic polynomials were not flexible enough, while the coefficients associated to the quartic polynomials were not statistically significant in most cases. Results using alternative specification of the polynomials are available upon request.

model, the elasticities of substitution are identified by the movement in relative supply, something I specify in the next section, but the demand trends, as captured by the log ratio of the third order polynomials, are identified residually. In particular, any change in relative wages that is not explained by movements in relative quantities is then absorbed by the relative demand parameters.

In total, the demand side of the model has 56 parameters that I need to estimate.

2.4.2 Occupational Choice

Male and female workers sort themselves into different market occupations based on preferences about job flexibility and earnings profiles (Goldin, 1984, 1986; Adda et al., 2017); responding to societal expectations and attitudes towards female work (Brown et al., 1980; Goldin, 1984, 2006); and as a function of gender specific comparative advantages associated to differences in physical, sensory, motor, and spatial aptitudes (Galor and Weil, 1996; Black and Juhn, 2000; Rendall, 2010; Welch, 2000; Rendall, 2013; Baker and Cornelson, 2016). The model attempts to incorporate both individual preferences and equilibrium returns to labor in the decision problem of the agents.

Following the work of Johnson and Keane (2013), I model occupational choice using a random utility framework where agents of different type choose either to enter the workforce in any of the three market occupations or remain in home production. Each alternative generates an utility for the worker, and agents choose the alternative that provides the highest utility. I model these utilities as linear functions that depend only on pecuniary and nonpecuniary rewards from each choice. In particular, the utility that a worker of a given type receives from choosing to enter the workforce in one of the three market occupations at time t

is

$$U(occ \mid gen, edu, t) = \psi_{gen,edu,occ} + \psi_1 W_{gen,edu,occ,t} + \epsilon_{gen,edu,occ,t}, \tag{2.4.6}$$

where $\psi_{gen,edu,occ}$ are time invariant parameters that capture nonpecuniary rewards that a worker gets from choosing occupation occ at time t; and ψ_1 measures the weight (in utility terms) that a worker gives to labor earnings ($W_{gen,edu,occ,t}$). $\epsilon_{gen,edu,occ,t}$ is an idiosyncratic taste shock assumed to be independent and identically distributed extreme value. The assumption about the distribution of the taste shock generates a tractable Multinomial Logit form to the choice probabilities.

The utility from home production is modelled in a symmetric way. The economic literature has linked movements of women into the labor market – or, more specifically, into market occupations – to changes in fertility and contraceptive technology (Katz and Goldin, 2000; Costa, 2000; Cruces and Galiani, 2007); changes in the marriage market (Grossbard-Shechtman and Neuman, 1988; Fernández and Wong, 2014; Greenwood et al., 2016); changes in social norms and attitudes towards women's work (Ronald R. Rindfuss, 1996; Costa, 2000; Fernández et al., 2004; Goldin, 2006; Fernández, 2013); and improvements in capital and technologies used for home production activities (Costa, 2000; Greenwood et al., 2005; de V. Cavalcanti and Tavares, 2008; Coen-Pirani et al., 2010). I do not specify how the underlying mechanism that explain the rise of female labor force participation interact with the demand side of the model. What I do is to condition the decision to remain in home production on variables linked to fertility choice, marriage patterns, changes in preferences, and technical change that is specific to home production activities (e.g. appliances). In particular, the utility from choosing home production, denoted by h, takes the form:

$$\begin{aligned} U(h \mid gen, edu, t) = &\pi_{1,gen} + \pi_{2,gen} t + \pi_{3,gen,edu} Pr(\text{CHL5} = 1 \mid gen, edu, t) \\ &+ \pi_{4,gen,edu} Pr(\text{MARR} = 1 \mid gen, edu, t) + \epsilon_{gen,edu,h,t}, \end{aligned} \tag{2.4.7}$$

where $\pi_{1,gen}$ and $\pi_{2,gen}$ are the intercept and slope of a gender specific linear trend that captures both changes in preferences for home production over time, and changes in the technology used in home production activities; $Pr(\text{CHL5} = 1 \mid gen, edu, t)$ is the probability that the agent has a child under the age of 5, which, since the level of aggregation is at the gender-education level, corresponds to the proportion of the population from a given group that has at least one children under the age of five at time t; and $Pr(\text{MARR} = 1 \mid gen, edu, t)$ is the probability that the worker is married or has permanent partner. Finally, $\epsilon_{gen,edu,h,t}$ is a idiosyncratic taste shock assumed to be independent and identically distributed extreme value.

Panel (a) and (b) of Figure 2.5 show the trends of the two variables I incorporate in the home production utility function – the proportion of individuals with children under the age of 5, and the proportion of individuals that are married or with a permanent partner – both conditional on education and gender. The two variables can only be calculated for a sample restricted to the household head, and, if applicable, to his/her spouse or partner, so the trends should be interpreted with this caveat in mind.[20] The percentage of each group with children under the age of 5 was between 45 and 55 percent in 1989, but these

[20] The ENIGH survey started asking the question on marital status to all member of the household in 1996, and for the number of children since 2004. For these reason, the two variables can only be constructed for the household head, and, if applicable, to his/her spouse or partner.

numbers fell in all cases by close to 20 percentage points, so the same probability in 2014 ranged between 25 and 35 percent. The change in the fertility decisions of the Mexican population is expected to be a significant driver behind the rise in female labor force participation. Changes in the marriage market were less pronounced: in 1989, the percentage of each group that was either married or had a permanent partner was between 86 and 94 percent; by 2014, these numbers had fallen between 3 and 8 percentage points, with the largest decline observed among the college educated (7.4 and 8 percentage points decline for males and females respectively).

Given the assumed distribution of the idiosyncratic taste shocks, the probability that a worker chooses one of the market occupations or home production is

$$Pr(d_O = 1 \mid gen, edu, t) = \frac{\exp(U(O \mid gen, edu, t))}{\sum_{occ=a,r,m,h} \exp(U(occ \mid gen, edu, t))} \quad \text{for} \quad O = a, r, m, h. \tag{2.4.8}$$

We can use these probabilities to find the total labor supply of each type in each occupation. For example, the total supply of female workers with college education in abstract task-intensive occupations is

$$L^s_{f,s,a,t} = L_{f,s,t} \times Pr(d_a = 1 \mid f, s, t) \tag{2.4.9}$$

Where $L_{f,s,t}$ is the total number of female workers with college education at time t, which I take as given. As the example shows, I condition on the schooling level of the agents, but I assume that the educational choice was taken before the age of 25, the starting age for an agent to become part of the sample.

The supply side of the model has a total of 25 parameters that I need to estimate.

2.4.3 Equilibrium and Estimation

The amount of labor demanded from agents of a given type in a specific market occupation and moment in time, denoted by $L^d_{gen,edu,occ,t}$, is fully determined by the equilibrium condition that wages are equated to marginal productivities:

$$W_{gen,edu,occ,t} = \frac{\partial Y_t}{\partial L^d_{gen,edu,occ,t}}. \tag{2.4.10}$$

Also, in equilibrium, wages are set to equate the supply and demand for the different labor types:

$$L^d_{gen,edu,occ,t} = L^s_{gen,edu,occ,t} \quad \text{for} \quad occ = a, r, m, \tag{2.4.11}$$

while the total supply in home production can be recovered residually.

A solution to the model is then obtained by finding the vector of wages for which Equations (2.4.10) and (2.4.11) are satisfied by each type of agent in each occupation. For a given set of parameters, this can be accomplish in an iterative way using a fixed-point algorithm: (i) start with an arbitrary wage vector W^0 and find the total supply of workers from each type in each occupation, which can be calculated using Equations (2.4.8) and (2.4.9). (ii) Plug the estimated supply of workers of each type into the marginal productivity function defined by Equation (2.4.10) and calculate the new vector of wages W^1. (iii) If $W^0 = W^1$, we have a solution for this set of parameters. If $W^0 \neq W^1$, set $W^0 = W^1$ and go back to step (i).

The model generates a prediction of the wage and labor supply of the four worker types in the four occupations (including home production) at every time period. With 13 years of data there are $(12+16) \times 13 = 364$ predictions in total[21] that are a function of the 81 parameters. I fit the model to the ENIGH data using the method of moments, targeting observed labor shares and wages. Given the dimensionality of the problem, and the different scales of the moments being targeted, I assume a simplified error structure to facilitate estimation. Details on the estimation technique are presented in Appendix 2.8.3.

2.5 Results

2.5.1 Model Fit to the Data

Figure 2.6 and Table 2.4 show the fit of the model with respect to the targeted moments. The figure shows that the dynamics predicted by the model for both log (male/female) relative earnings and log (male/female) relative supply follow closely what is observed in the ENIGH data. The model is not able to accurately capture short-term variations in these series, but it is able to replicate the

[21]Wages: 4 types of workers in three market occupation leads to 12 predictions per year. Supplies: 4 types of workers in 4 possible occupations leads to 16 predictions per year.

long-term trends by occupation groups across the 25 year period. Importantly, the model also does a good job capturing the level and changes in labor force participation for both males and females.

Table 2.4 shows the observed and predicted occupation shares and mean wages for each type of agent in the model in two periods: C.1990 and C.2013. There is a close fit of the model with respect to the occupational distribution, with no major differences between model predictions and data in any of the cells. In the case of mean wages, the model predictions are also close to the data, except in the cases in which the occupational shares of the groups being compared are relatively small. For example, the model predicts higher wages for college-educated workers in manual and routine task-intensive occupations C.1990, especially among females, but these groups only represent between 0.05 and 0.63 percent of the prime-age population according to the ENIGH.

Having established that the model can recreate the level and changes of the occupational and wage structure of the Mexican economy over the last 25 years, I now turn to the estimates of the structural parameters that give rise to these patterns.

2.5.2 Parameter Estimates

Table 2.5 presents the point estimates and standard errors of the different elasticities of substitution in the production technology. The elasticities of substitution between male and female labor are estimated to be around 1.2 in the low-paying manual and routine task-intensive occupations, and close to 2.6 in the high-paying abstract task-intensive occupations. These results provide support to the hypothesis that male and female labor are closer substitutes in occupations with stronger requirements for abstract analytical skills.

To get a sense of what these values represent in the Mexican context, note that the log (male/female) relative supply fell by 44 log points between 1989 and 2014 in both manual and routine task-intensive occupations, so an elasticity of 1.2 implies that the log (male/female) earnings ratio should have increased – holding everything else constant – by 36 log points (43%) in both cases. These numbers are much higher than the observed 3 and 11 log point increase of the gender earnings gap in the respective occupations, so that relative demand trends favouring women in those occupations must be part of the story. For abstract task-intensive occupations, a similar calculation shows that the 69 log points decline in the relative supply should have resulted in a larger earnings gap by close to 26 log points (30%), instead of the 10 log points decline observed in the data. Again, these results suggest that relative demand trends in high-paying occupations were strongly favourable to women.

To my knowledge, these are the first estimates of the elasticity of substitution between male and female labor for any country in Latin America, so there are no points of comparison. For the case of the U.S., Acemoglu et al. (2004) estimate this elasticity to be around 3, while Johnson and Keane (2013) report values for this parameter in a range between 4.76 and 5.6. Although not directly comparable, the differences in the estimates between the U.S. and Mexico are substantial. It is possible that the technologies of production in less developed economies are such that the differences in the skills supplied in the labor market by men and women are accentuated (e.g. due to a lower technification of tasks and higher reliance on "brawn" instead of "brain"). Moreover, social norms might be such that women are restricted from entering certain occupations, so that the elasticity of substitution between men and women not only captures technical possibilities, but also cultural traits. It is not possible to disentangle between these and other alternative explanations, but this is a fertile area for future research.

On the educational dimension, the point estimates of the elasticities of substitution between skilled and unskilled labor are 1.4 in abstract, 1.6 in routine, and 3.6 in manual task-intensive occupations. The fact that the estimated elasticities of substitution between these groups are lower in occupations that are more intensive in cognitive than in physical labor is encouraging. Moreover, these results suggest that the sharp educational upgrading in Mexico has contributed to the fall of overall earnings inequality by generating downward pressures on wages of workers with more schooling, an outcome that is consistent with the results reported in Chapter 1.

The three point estimates of the elasticities of substitution between skilled and unskilled labor are in a range that incorporates the values that other studies have reported for Latin American. For example, using data from five Latin American countries during the 1990s, Manacorda et al. (2010) report estimates of this elasticity of around 3. In Chapter 1 we expanded the model used by Manacorda et al. (2010) finding a value close to 2.1. For the United States, there is a general convergence among different studies showing that the elasticity of substitution between skilled and unskilled labor is close to 1.5 (Katz and Murphy, 1992; Ciccone and Peri, 2005; Johnson and Keane, 2013).

The previous discussion shows that the rise of female labor force participation is having a substantial influence on the Mexican wage structure, but they also suggest that the demand for female labor must have increased considerably during the period, mitigating the compositional effects, otherwise we should have observed a much larger rise in the gender earnings gap. In Figure 2.7 I present the model predictions of the evolution of both the relative demand trends and total labor productivity.[22] To facilitate interpretation, each series capturing changes in relative demands is normalized to take a value of zero in 1989. Panels (a) and

[22]See Equations (2.4.4) and (2.4.5) for an example of how the relative demands trends are specified.

(b) show that during the last 25 years, the demand for female labor in Mexico increased relative to that of males. This result holds within the three occupational groups, and also among workers with low and high levels of schooling. For example, in the case of workers with at most high school education, changes in the demand side of the economy imply that – holding everything else constant – the log (male/female) earnings ratio should have fallen by 14 log points in abstract and routine task-intensive occupations, and by 25 log points in manual task-intensive occupaitons. Among workers with at least some years of tertiary education completed, these effects are even stronger: in abstract task-intensive occupations, the model predicts that log (male/female) earnings ratio should have fallen by 58 log points.

The coefficients of the relative demand polynomials are estimated residually, so we are not able to pinpoint exactly what are the main drivers behind the observed patterns. But the predictions of the model are consistent with previous studies that have shown that the Mexican labor market has experienced an increase in the relative demand for female labor. Juhn et al. (2014) show that as a consequence of the reduction of import tariffs, the more productive firms in the Mexican economy were induced to enter the export markets and modernize their technologies. The authors argue that these new technologies lowered the need for physically demanding skills in blue-collar occupations, which translated in a rise of the relative demand for female labor. A similar argument emphasizing how structural change can lower the need for physically intensive tasks, favouring female labor, has been recurrent in other contexts (Galor and Weil, 1996; Rendall, 2010; Black and Spitz-Oener, 2010; Lup Tick and Oaxaca, 2010; Akbulut, 2011; The World Bank, 2012; Aguayo-Tellez et al., 2013; Rendall, 2013). It is also possible that social attitudes regarding female labor are changing, and that employers are more willing to hire women in positions that were previously exclusive to men. We can only speculate about possible mechanisms, so more research is needed to understand the drivers behind these changes in demand.

Relative demand trends are also favouring workers with more schooling and labor in abstract task-intensive occupations (see Panels (c) and (d)).[23] Some potential drivers behind this skill specific demand shift in Mexico have been associated to the trade and investment liberalization of the Mexican economy since the late 1980s (Feenstra and Hanson, 1997; Hanson, 2003; Sánchez-Páramo and Schady, 2003; Behrman et al., 2007; Caselli, 2012); and to the growth of foreign direct investment (Feenstra and Hanson, 1997). Interestingly, the results provide evidence that non-neutral technical change – the most widely accepted explanation for the rise in income inequality in developed economies – is also a factor in developing countries, but that the educational upgrading of the workforce is strong enough to counteract this force, so we don't observe such a sharp rise in the wages of high-skilled workers. Quoting the famous Tingbergian analogy, the race between education and technology is being won by the former. Finally, the evolution of total labor productivity provide a grim picture for the Mexican economy, with the estimated trends falling by as much as 27 log points between 1989 and 2014.

The estimates of the parameters from the supply side of the model are shown in Table 2.6. Each row reports the point estimate and standard error of a different parameter, and in the more relevant cases, the table also reports the average marginal effects. Average marginal effects of the fertility and marital status variables are calculated by taking the numerical derivative of the probability of choosing home production with respect to the given variable. I calculate this derivative in each year separately and then take the average across all years. In the case of the pecuniary rewards (ψ_1), the reported average marginal effect

[23] Other studies showing that relative demand for skilled labor has been growing in Mexico include Sánchez-Páramo and Schady (2003); Binelli and Attanasio (2010); Manacorda et al. (2010); Gasparini and Lustig (2011)

is calculated by finding the numerical derivative of the probability that a labor type chooses a given market occupation with repect to the wage. I calculate this derivative in each year and for every possible labor type and occupation combination separately, and then take the average across all the values.

As expected, an increase in the wage in a given market occupation raises the probability that an agent will choose that occupation, but the wage elasticity of labor supply is relatively low. For example, take the case of a college-educated female worker choosing among the three market occupations or home production. If the hourly wage for her group in the abstract task-intensive occupation increased from 6.5 to 7.15 – a ten percent increase of the sample average of that group –, the probability of her choosing that occupation is predicted to increases by 1.43 percentage points. Since the individual probabilities are in one-to-one correspondence with the occupation shares, the share of college-educated female workers in abstract task-intensive occupation would also be predicted to increase by 1.43 percentage points. This result suggest that there is substantial persistence in the part of workers when choosing an occupational group, and that there has to be very large changes in relative earnings across occupations for workers to switch between them.

The model predicts that having children in pre-school ages and being married or with a permanent partner are important determinants of the decision to participate in market occupations, but only among workers with lower levels of schooling. For example, the percentage of prime-age females with at most a high school degree and with children under the age of 5 fell by 18.9 percentage points from 1989 to 2014. This decline is associated with a rise in labor force participation of 3.3 percentage points, which is close to 17 percent of the total observed change. The percentage of females in the same educational group that are either married or with a permanent partner fell by 6 percentage points, and this decline is associated with a rise of labor force participation of 1 percentage point, close to 5 percent of the observed change. Together, the fertility transition and changes in marital status explain close to 22 percent of the overall rise in female labor force participation among unskilled females, the group for which we observe the largest movement into the workforce. The other 78 percent is explained by either changes in the demand side of the economy, or by changes in preferences for home production and home production technology as captured by the linear trends in Equation (2.4.7).

Among men with at most a high school degree, both marital status and fertility decisions influence the decision to participate in the labor market, but the direction of the effects is the opposite to that of females. Having children under the age of five and being married or with a permanent partner is associated with higher labor force participation, although the average marginal effects are small. Interestingly, neither of the two variables is statistically significant in the case of college educated men and women.

2.5.3 Counterfactual Exercises

To quantify the impact of the different factors behind the decision to participate in the labor market on the occupation and wage structure, I do a series of counterfactual exercises. The idea is to "turn off" the different drivers, one at a time, and then compute the model predictions of occupational shares and mean wages, taking advantage of the general equilibrium set-up. In particular, I construct counterfactual distributions under four scenarios: (CF1) setting the linear, quadratic, and cubic coefficients of the demand trends (α shares) to zero; (CF2) setting the share of each group with children under the age of 5 to be equal to the level in 1989, and constant across the years; (CF3) setting the share of each group that is married or with a permanent partner to be equal to the level in 1989, and constant across the years; and (CF4) setting the coefficient of the linear trends ($\pi_{2,f}, \pi_{2,k}$) in the home production utility function to zero.

Table 2.7 presents the results from these experiments, with each column corresponding to a different scenario and each cell reporting the difference between males and females of the difference between C.1990 and C.2013 of the occupational shares and wages, that is, the change over time in occupation and wage gaps. The results from the first counterfactual (CF1) reinforce our previous discussion about the role of changes in the demand side of the economy: relative demand trends in Mexico tended to favour female labor, and they helped mitigate the impact of the gender compositional change in labor supply. In the absence of these shifts in demand, the model predicts that the (male/female) gender earnings gap would be significantly higher in all occupations and skill groups. The differences in occupational choices between men and women were mostly unaffected by how the demand evolved, which is probably due to the fact that the wage elasticity of labor supply is low.

The second (CF2) and third counterfactuals (CF2) capture the equilibrium effects of the changes in fertility and marital status of the Mexican population. The results show that these changes did not have a meaningful impact on the occupational or wage structure. Even tough we found that both variables were significant factors behind the decision to enter the workforce among unskilled workers, the magnitude of the effect is not strong enough to generate a change in relative earnings.

But the story is different when we look at the final counterfactual (CF4). The linear trends in the home production utility, which can be capturing either changes in preferences for this choice or improvements in the technology used in home production activities, have a strong impact on the occupational and wage structure. In particular, once we switch off the trends in the utility for home production, the rise in female labor force participation is much lower (6% predicted instead of the 22% observed). This in turn implies that the gender compositional shift in the workforce is small, and that downward pressures on female wages are attenuated. In fact, in this scenario there is strong convergence in (male/female) relative earnings in all occupations and skill groups, including in the manual and routine task-intensive occupations.

2.6 Robustness Exercises

The earnings series used for the baseline estimates of the parameters of the model included only incomes from full-time workers – those that reported working 35 hours or more in the previous week. This was done to have groups that were more closely comparable, since the share of part-time workers differs markedly between males and females: part-time female workers represent between 33 and 38 percent of all female workers, while the share of part-time male workers ranged between 10 and 13 percent. Table 2.8 reports the point estimates and standard errors of the different elasticities of substitution in the production technology using alternative earnings series and measures of labor supply. The table reports the baseline estimates; the estimates once we include income from part-time workers; and the estimates if we measure labor supply by the total number of hours worked of each group instead of the head-count. Since we don't have a measure of hours worked for people that are in home production, we have to impute those values. We assign each person in home production the average number of hours worked by workers in market occupation with the same level of schooling, sex, and age.

There are no major difference between the baseline estimates of the elasticities of substitution and the ones obtained using the alternative earnings and labor supply series. The rank-size of the values in each of the levels is maintained. Including income from par-time workers leads to a lower estimate of the elasticity of substitution between male and female labor in manual and routine task-intensive occupations, which are now 0.8 and 0.97 respectively. This result only reinforces our conclusions from the previous section about the strong impact that female labor supply is having on the wages of female workers in low-paying occupations. Using a labor supply measure that captures changes in the intensive margin doesn't change our analysis in any meaningful way.

The corresponding estimates of the parameters from the supply side of the model under the alternative earnings and supply measures are shown in Table 2.9. Results are essentially unchanged compared to the baseline.

When modelling the structure of the production technology, two decisions were made that could influence the results but are not grounded on a solid theoretical basis: first, in the three nests, labor is first divided by education and then by gender. This division changes the number of relative demands that are estimated in each dimension, but it should not alter the main results in a significant way. Second, the model assumes that the elasticity of substitution between abstract and routine task-intensive occupations is the same to that between abstract and manual-task intensive occupations. The way the model is set-up implies that at least one occupational group will have a common elasticity with the other two; we choose the abstract task-intensive occupation since it leads to a natural division between high and low-paying jobs, but it did not have to be that way.

Table 2.10 presents point estimates and standard errors of the elasticities of substitution in the production technology under four different model specifications: (i) the baseline estimates; (ii) the estimates if we switch the order of second and third nests of the production technology; (iii) the estimates if the occupational group that has the common elasticity with the other two is the routine task-intensive; and (iv) the estimates if the occupational group that has the common elasticity with the other two is the manual task-intensive. Once again, we find that the rank-size of the values of the elasticities of substitution between male and female labor is maintained in all cases: in manual and routine task-intensive occupations these elasticities are between 0.7 and 1.2, while in abstract task-intensive occupations the number is between 1.9 and 2.6.

The corresponding estimates of the parameters from the supply side of the model under the alternative model specifications are shown in Table 2.11. Results are essentially unchanged compared to the baseline.

2.7 Conclusions

This chapter studies the effect of the rapid rise of female labor force participation on the Mexican occupational and wage structure over the last 25 years. It argues that the increase in female labor supply in Mexico, one of the largest in the world during this period, has fundamentally changed the patterns of relative earnings between men and women.

The chapter shows that the change in the gender earnings gap in Mexico over the last quarter century has not been uniform throughout the pay distribution. In particular, the gender earnings gap increased in low-paying manual and routine task-intensive occupations, but it fell in high-paying abstract task-intensive occupations. Using recently developed decomposition techniques, I provide evidence that the shift in relative earnings was not the result of a recomposition of the socio-demographic characteristics of the workforce, but that it reflects actual changes in relative pay between men and women.

The chapter develops a equilibrium model of the labor market to study if rising female labor supply can explain the peculiar evolution of the wage structure. The model follows the task-based approach, allowing the elasticity of substitution between male and female labor to vary depending on the task content of occupations. This feature implies that it is possible for changes in female labor supply to have heterogeneous impacts on relative earnings throughout the pay distribution.

I structurally estimate the model using Mexican household survey data covering the last 25 years. The estimates from the model support the hypothesis that the elasticity of substitution between male and female labor is lower in low-paying manual and routine task-intensive occupations, than in high-paying abstract task-intensive occupations. In particular, in the preferred specification, I find that these elasticities are between 1.2 in manual and routine task-intensive occupations, and 3 in abstract task-intensive occupations. The implication is that the movement of women into the labor market has imposed a large burden on the wage growth of female workers, especially those with low levels of schooling.

The estimates from the model also show that the demand for female labor relative to that of males has been increasing during the last 25 years in Mexico. I find that this was the case across the board: both within broadly defined occupational groups, and for workers with high and low levels of schooling. The counterfactual exercises show that these favourable trends in the demand side of the economy have attenuated the supply-driven downward pressure on female wages at low-paying occupations, and fully counteract it at high-paying occupations.

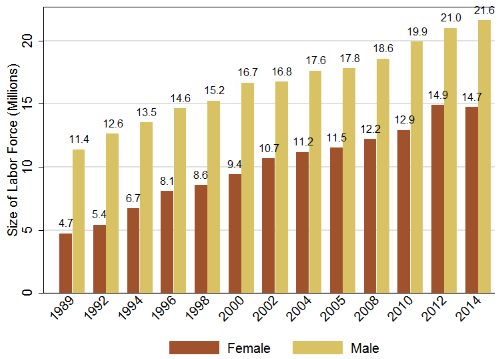
2.8 Appendix

2.8.1 Tables and Figures

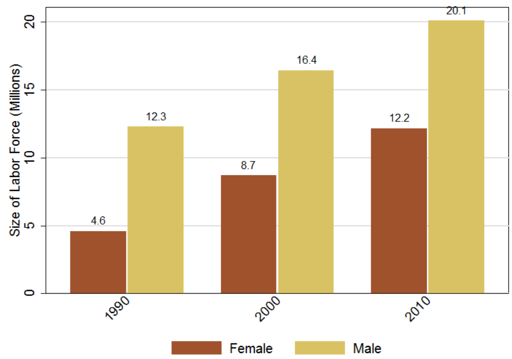
Figure 2.1: Labor Force Participation by Sex

Absolute Numbers

(a) ENIGH



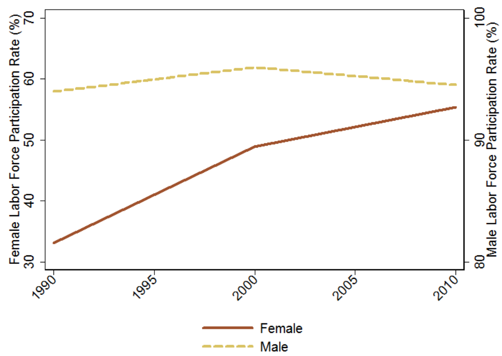
(b) CENSUS



Participation Rates

(c) ENIGH

(d) CENSUS

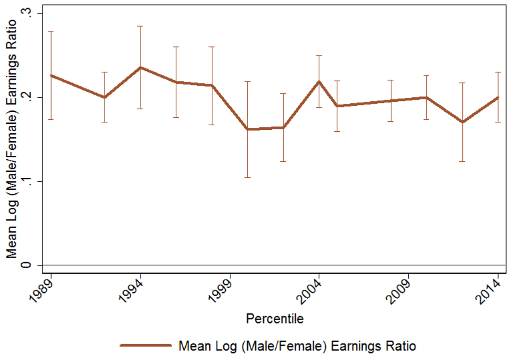


Notes: Panels (a) and (b) show the total number of prime-age population that are either working or actively searching for a job by sex. The series in Panels (c) and (d) show the share of prime-age males/females that are either working or actively searching for a job by sex. Sample weights used in all calculations.

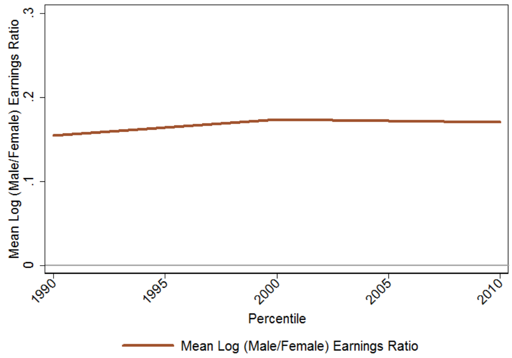
Figure 2.2: Changes in the Log (Male/Female) Earnings Ratio

Mean Log (Male/Female) Earnings Ratio

(a) ENIGH

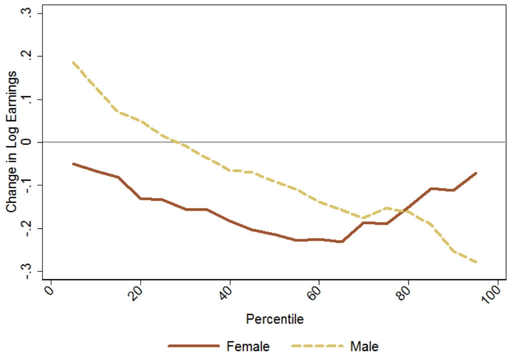


(b) CENSUS

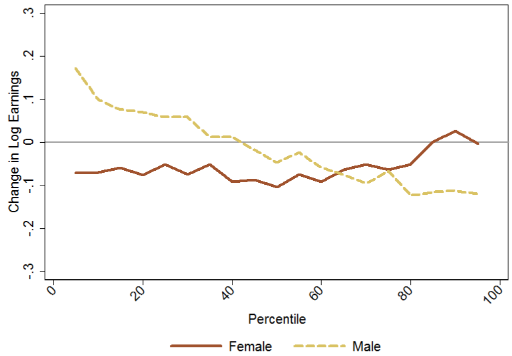


Change in Log Hourly Earnings by Percentile

(c) ENIGH: C.1990-C.2013

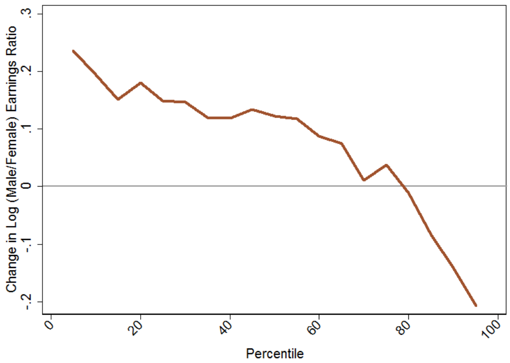


(d) CENSUS: 1990-2010

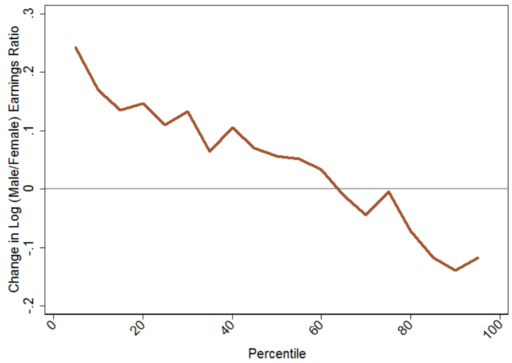


Change in Log (Male/Female) Earnings Ratio by Percentile

(e) ENIGH: C.1990-C.2013



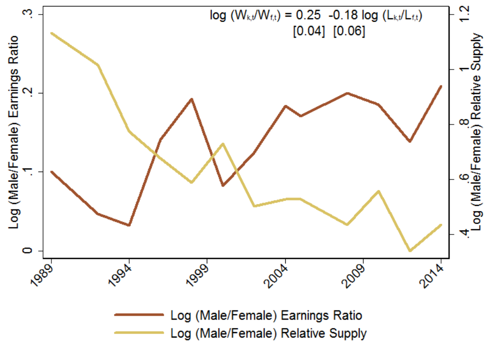
(f) CENSUS: 1990-2010



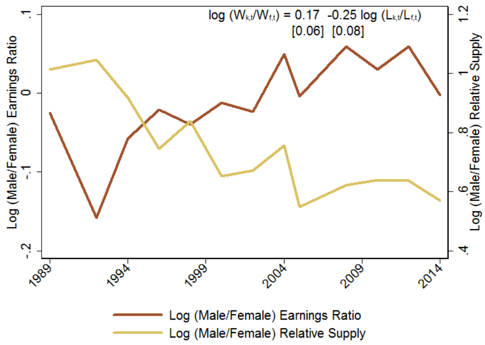
Notes: Panel (a) and (b) show the estimated (male/female) gender earnings gap. The gender earnings gap corresponds to the estimated coefficient of a male dummy variable in a regression in which log hourly earnings is the dependent variable, and we include controls for education attainment (7 different categories), age (6 categories in five year intervals), occupation indicators (18 groups), and all possible interactions. The regressions are estimated separately in each year. Panels (c) and (d) show the change in log hourly earnings by percentile between C.1990 and C.2013 (1990 and 2010 in CENSUS), calculated for each gender separately. To increase sample size in the ENIGH survey, we joined together surveys from 1989 and 1992, and from 2012 and 2014. Panels (e) and (f) show the change in log (male/female) earnings ratio by percentile between C.1990 and C.2013 (1990 and 2010 in CENSUS). Sample is restricted to prime-age population that reported working for more than 35 hours a week. Vertical bars correspond to 10 percent confidence intervals. Sample weights used in all calculations.

Figure 2.3: Comovement Between Log (Male/Female) Earnings Ratio and Log (Male/Female) Relative Supply by Occupation Groups

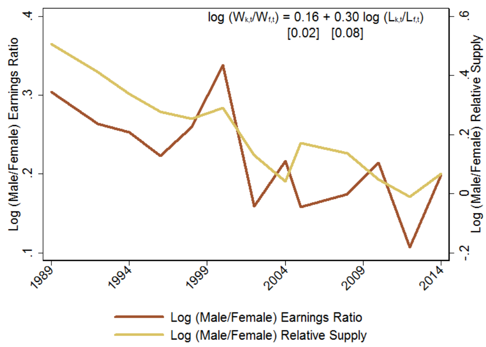
(a) Manual



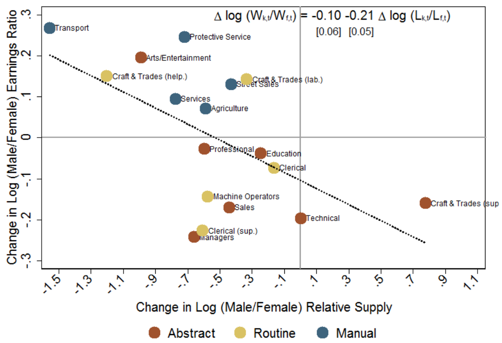
(b) Routine



(c) Abstract

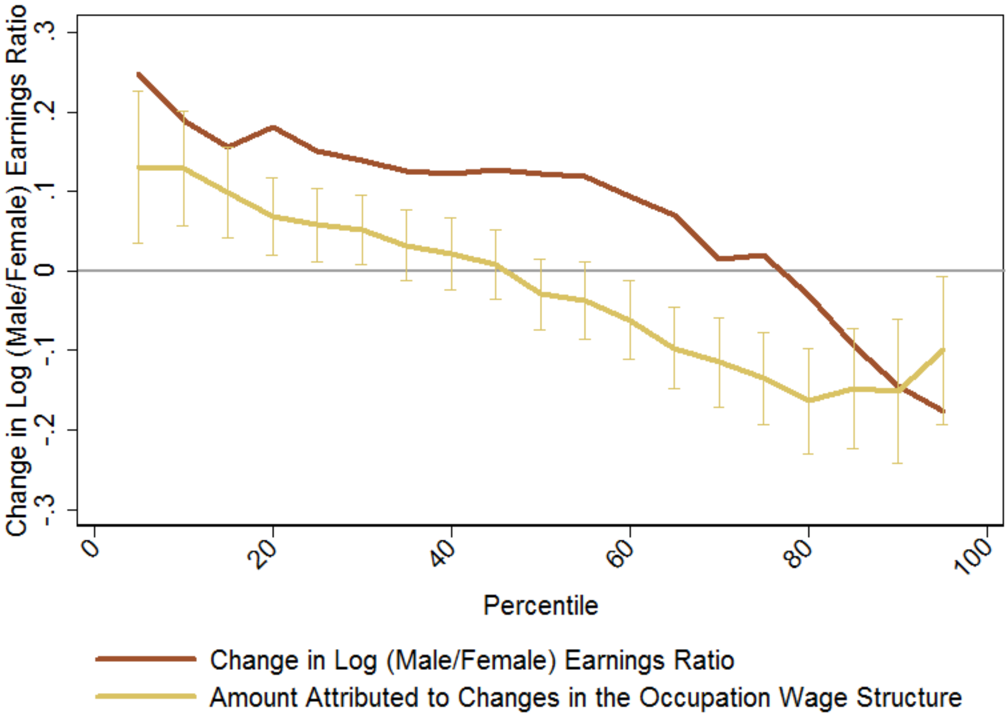


(d) Change in the Earnings Gap and Change in Relative Supplies: C.1990 - C.2013



Notes: Panels (a)-(c) plot log (male/female) earnings ratios and log (male/female) relative supplies over the period of analysis, one panel for each occupational group. Panel (d) plots the change in log (male/female) earnings ratio and the change in log (male/female) relative supply between C.1990 and C.2013, using the principal group occupational division of the ENIGH as the unit of analysis. The regression line is estimated weighting each occupation by the average employment share between the two periods. The sample for the earnings series is restricted to full-time workers between the ages of 25 and 55.The sample for the relative supply series includes all workers in the labor force between the ages of 25 and 55. Sample weights used in all calculations.

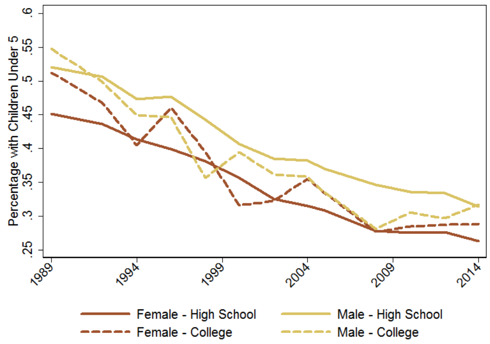
Figure 2.4: Decomposition Results: Change in Log (Male/Female) Earnings Ratio Between C.1990 and C.2013 Attributed to Changes in Occupational Wage Structure



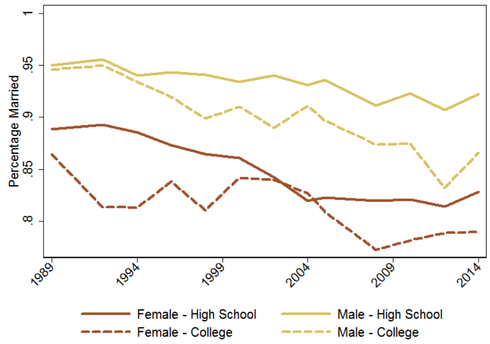
Notes: The occupation wage structure series depicts the unexplained component of the detailed Oaxaca-Blinder decomposition – defined in Equation (2.3.5) – that is attributed to the 18 principal group occupations of the ENIGH. The estimation is done separately for 19 percentiles. Confidence intervals are estimated via bootstrap with 500 replications. Sample weights used in all calculations.

Figure 2.5: Determinants of Movements into the Labor Market

(a) Changes in Fertility



(b) Changes in the Marriage Market

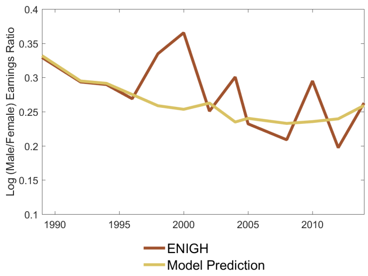


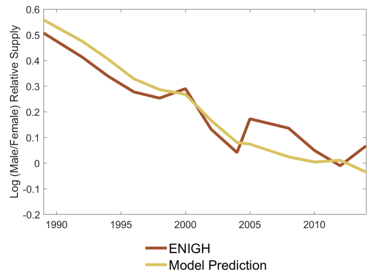
Notes: Panel (a) depicts the proportion of the sample with children under the age of 5, and Panel (b) depicts the proportion of the sample that is married, both conditional on education and gender. The ENIGH survey started asking the question on marital status to all member of the household in 1996, and for the number of children since 2004. For these reason, the two variables can only be constructed for the household head, and, if applicable, to his/her spouse or partner. The sample is restricted to prime-age population. Sample weights used in all calculations.

Figure 2.6: Log (Male/Female) Relative Earnings, Log (Male/Female) Relative Supplies, and Participation Rates: ENIGH and Model Predictions

Analytical Occupations

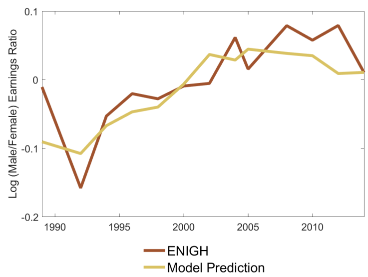
(a) Log (Male/Female) Earnings Ratio (b) Log (Male/Female) Relative Supply

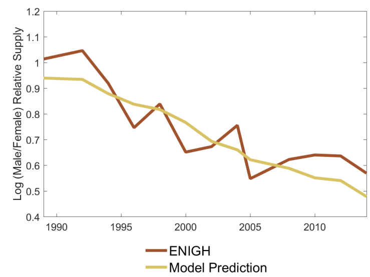




Routine Occupations

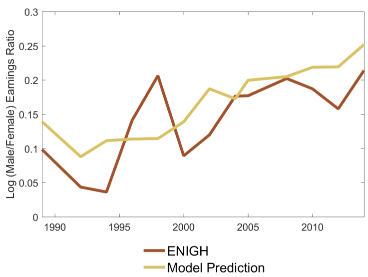
(c) Log (Male/Female) Earnings Ratio (d) Log (Male/Female) Relative Supply

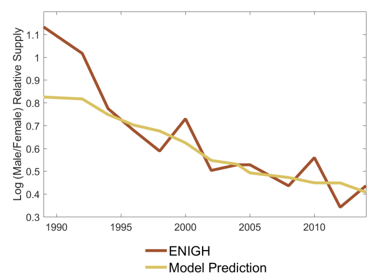




Manual Occupations

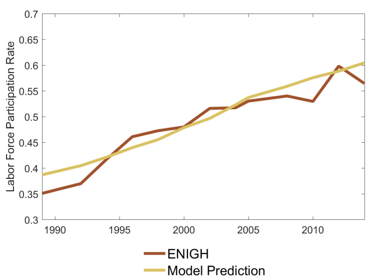
(e) Log (Male/Female) Earnings Ratio (f) Log (Male/Female) Relative Supply

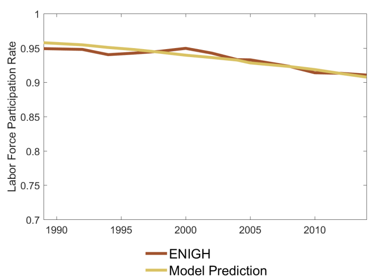




Participation Rates

(g) Female (h) Male



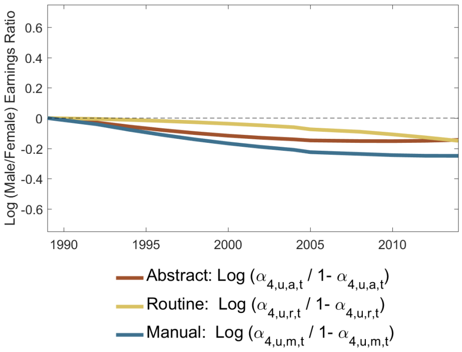


Notes: The different panels depict the series of (male/female) relative earnings, (male/female) relative supplies, and labor force participation rates, both from the ENIGH and predicted from the model.

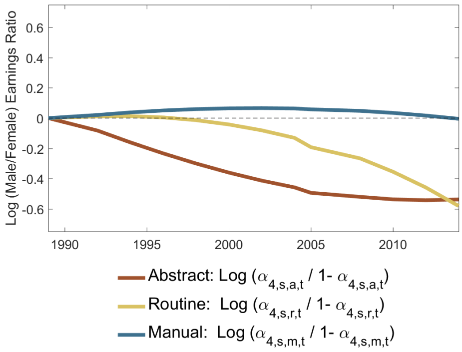
Figure 2.7: Estimates of the Relative Demand Indexes and Total Factor Productivity

Production Technology: Level III

(a) Male vs. Female (High School)

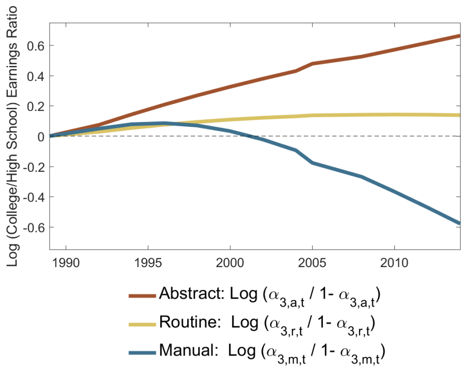


(b) Male vs. Female (College)



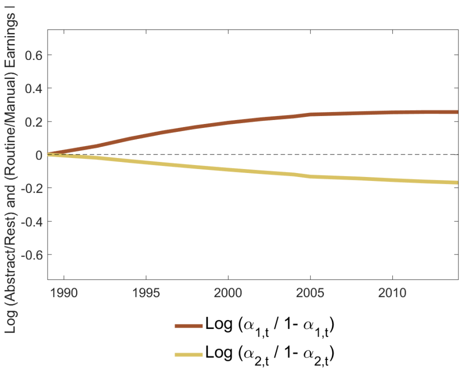
Production Technology: Level II

(c) College vs. High School

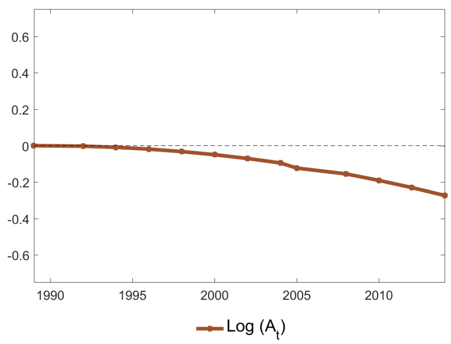


Production Technology: Level I

(d) Analytical vs. Routine and Manual; and Routine vs. Manual



(e) Total Factor Productivity



Notes: Panels (a)-(d) show the estimated change in the relative demand indexes captured by the log ratio of the α shares. Panel (e) shows the estimated change of log total factor productivity. The changes in total factor productivity and the α shares are estimated using a cubic time trend in their natural logarithm (see Equations (2.4.4) and (2.4.5) for an example). To facilitate interpretation, each series is normalized to zero in 1989.

Table 2.1: Occupation Groups

	Median Percentile of the Task Measure			
ENIGH Principal Group	Abstract	Routine	Manual	Group
Managers	90.0	17.0	27.5	Abstract
Crafts and Trades (Supervisors)	84.0	42.0	62.0	Abstract
Education	83.0	11.0	65.0	Abstract
Professional	83.0	42.0	46.0	Abstract
Technical	71.0	69.0	43.0	Abstract
Arts/Entertainment	66.0	35.0	48.0	Abstract
Sales	61.0	22.5	15.0	Abstract
Crafts and Trades (Laborers)	40.0	82.0	73.0	Routine
Clerical (Supervisors)	61.0	63.0	51.5	Routine
Crafts and Trades (Helpers)	10.5	62.0	60.5	Routine
Machine Operators	16.0	62.0	51.0	Routine
Clerical (Laborers)	41.5	53.0	12.0	Routine
Transport	19.5	21.0	96.0	Manual
Agriculture	32.0	27.0	82.0	Manual
Protective Services	24.5	5.5	76.5	Manual
Domestic Service	9.0	8.0	76.0	Manual
Street Sales	38.0	13.0	64.0	Manual
Service	28.0	25.0	63.0	Manual

Notes: The three task measures were originally constructed for three-digit occupational codes of the U.S. CENSUS by Autor et al. (2003). For each measure, I first organize the three-digit occupations by percentiles, and then calculate the median percentile within the broader 18 occupational groups of the ENIGH. Each of the 18 occupations is assigned to the group in which the median percentile was highest.

Table 2.2: Changes in Composition between C.1990 and C.2013

	C.1990			C.2013			Dif. in Dif.
	Female	Male	Δ	Female	Male	Δ	
Overall Population							
Participation Rate	38.59	96.49	57.89	59.59	95.82	36.24	-21.66
Education							
High School	92.09	84.27	-7.82	81.24	79.05	-2.18	5.64
College	7.91	15.73	7.82	18.76	20.95	2.18	-5.64
Age							
25-34	44.63	43.61	-1.02	35.74	36.36	0.62	1.64
35-44	32.34	32.84	0.50	34.16	33.97	-0.20	-0.69
$\geq$ *45*	23.02	23.55	0.52	30.09	29.67	-0.42	-0.95
Workforce							
Education							
High School	85.48	84.37	-1.11	76.00	79.19	3.19	4.29
College	14.52	15.63	1.11	24.00	20.81	-3.19	-4.29
Age							
25-34	46.48	43.58	-2.90	34.96	36.03	1.08	3.97
35-44	33.31	33.38	0.07	36.18	34.72	-1.46	-1.53
$\geq$*45*	20.21	23.04	2.83	28.87	29.25	0.39	-2.44
Occupations (Emp. Share by Gender (%))							
Abstract	38.19	26.27	-11.93	37.45	27.47	-9.98	1.95
Routine	27.37	33.35	5.98	27.22	35.43	8.21	2.23
Manual	34.44	40.39	5.95	35.32	37.09	1.77	-4.18
Occupations (Emp. Share Overall (%))							
Abstract	11.91	18.08	6.17	15.59	16.04	0.45	-5.72
Routine	8.53	22.95	14.42	11.33	20.69	9.36	-5.06
Manual	10.74	27.79	17.06	14.70	21.66	6.96	-10.10
Overall	31.18	68.82	37.65	41.62	58.38	16.77	-20.88

Notes: The table shows the composition by sex, education, age, and occupation of both overall prime-age population and of prime-age population in the workforce. Each cell reports the share of the respective column group. In the specific case of the overall employment by occupations, the cells report the shares including both males and females. Any individual that has at least completed one year of post-secondary education is included in the college group. We joined together surveys from 1989 and 1992, and from 2012 and 2014 to increase sample size of the ENIGH survey. The sample is restricted to prime-age population. Sample weights used in all calculations.

Table 2.3: Compositional Changes and The Gender Earnings Gap: Oaxaca-Blinder Decomposition Results by Selected Percentiles

	P10		P25		P50		P75		P90	
	Est.	[S.E.]	Est.	[S.E.]	Est.	[S.E.]	Est.	[S.E.]	Est.	[S.E.]
Observed Change	0.214	[0.034]	0.131	[0.020]	0.150	[0.023]	0.022	[0.025]	-0.097	[0.032]
Overall Wage Structure	0.155	[0.036]	0.108	[0.021]	0.148	[0.021]	0.006	[0.023]	-0.119	[0.033]
Occupation	0.129	[0.037]	0.058	[0.023]	-0.029	[0.023]	-0.135	[0.029]	-0.151	[0.046]
Education	0.007	[0.040]	0.021	[0.020]	0.019	[0.012]	0.004	[0.014]	0.052	[0.019]
Age	0.009	[0.014]	-0.004	[0.008]	0.002	[0.008]	0.029	[0.009]	0.049	[0.014]
Constant	0.009	[0.071]	0.033	[0.039]	0.157	[0.033]	0.108	[0.037]	-0.069	[0.059]
Overall Composition	0.059	[0.019]	0.023	[0.013]	0.002	[0.014]	0.016	[0.016]	0.022	[0.022]
Occupation	0.001	[0.002]	-0.001	[0.002]	-0.001	[0.004]	-0.003	[0.005]	-0.011	[0.006]
Education	0.079	[0.017]	0.037	[0.011]	0.022	[0.011]	0.040	[0.013]	0.055	[0.018]
Age	-0.021	[0.005]	-0.013	[0.003]	-0.019	[0.003]	-0.020	[0.004]	-0.022	[0.007]

Notes: The table shows results of the aggregate and detailed Oaxaca-Blinder decomposition defined in Equation (2.3.5). The Standard errors in brackets are calculated via bootstrap with 500 replications. Sample weights used in all calculations.

Table 2.4: Model Fit: ENIGH and Model Predictions of the Occupation Shares and Wages

	Female C.1990		Male C.1990		Female C.2013		Male C.2013		Dif. in Dif.	
	ENIGH	Model	ENIGH	Model	ENIGH	Model	ENIGH	Model	ENIGH	Model
Occupation Shares										
College										
Abstract	2.26	1.72	5.30	4.58	5.22	5.32	6.14	6.18	-2.13	-2.01
Routine	0.63	0.77	1.27	1.38	1.66	1.81	1.90	1.67	-0.39	-0.76
Manual	0.05	0.61	0.39	1.12	0.49	0.83	0.92	1.05	0.09	-0.29
Home Production	1.26	1.09	0.45	0.33	2.55	1.95	0.93	0.99	-0.81	-0.20
High School										
Abstract	5.45	5.36	6.39	6.77	6.27	6.44	5.69	5.43	-1.53	-2.42
Routine	4.89	5.64	13.58	14.63	6.69	7.00	13.34	12.98	-2.04	-3.00
Manual	6.90	7.34	17.59	16.47	10.35	10.08	15.04	15.68	-5.99	-3.53
Home Production	31.49	30.37	2.12	1.83	19.59	19.35	3.23	3.25	13.02	12.44
Mean Wages										
College										
Abstract	6.99	6.44	10.04	9.51	6.30	6.15	8.11	7.93	-9.39	-12.13
Routine	5.98	6.98	7.98	7.70	4.90	4.81	5.25	5.46	-15.99	2.06
Manual	3.17	8.06	5.11	6.37	3.16	2.31	3.30	2.45	-35.05	9.84
High School										
Abstract	3.89	3.58	4.54	4.26	2.67	2.68	3.12	3.29	0.20	2.63
Routine	3.33	3.14	3.06	2.91	2.41	2.44	2.52	2.60	9.74	11.94
Manual	1.89	1.78	2.14	2.09	1.73	1.74	2.18	2.24	10.52	9.55

Notes: The Table shows the occupation shares and average wages in C.1990 and C.2013, both from the ENIGH and predicted by the model. The last two columns report the difference between males and females of the difference between C.1990 and C.2013 of occupational shares and wages. In the case of wages the difference in difference corresponds to the percent change.

Table 2.5: Parameter Estimates: Production Technology

	Elasticities of Substitution		
	Estimate	[SE]	Implied Elasticity $(1/(1-\rho))$
Gender			
$\rho_{4,m}$: female, male (manual)	0.175	[0.181]	1.212
$\rho_{4,r}$: female, male (routine)	0.179	[0.129]	1.219
$\rho_{4,a}$: female, male (abstract)	0.622	[0.099]	2.646
Education			
$\rho_{3,m}$: college, high school (manual)	0.722	[0.067]	3.594
$\rho_{3,r}$: college, high school (routine)	0.355	[0.041]	1.549
$\rho_{3,a}$: college, high school (abstract)	0.276	[0.121]	1.382
Occupation			
ρ_1: abstract, routine and manual	0.031	[0.094]	1.032
ρ_2: routine, manual	-0.141	[0.183]	0.877

Notes: The table shows the point estimates and standard errors of the elasticities of substitution from the production technology.

Table 2.6: Parameter Estimates: Occupational Choice

	Occupational Choice		
	Estimate	SE	Average Marginal Effect
Earnings			
ψ_1: earnings	0.154	[0.009]	0.022
Fertility/Children			
$\pi_{3,f,u}$: female, unskilled	0.712	[0.127]	0.174
$\pi_{3,f,s}$: female, skilled	-0.139	[0.229]	-0.025
$\pi_{3,k,u}$: male, unskilled	-0.367	[0.182]	-0.022
$\pi_{3,k,s}$: male, skilled	0.122	[0.219]	0.008
Marriage			
$\pi_{4,f,u}$: female, unskilled	0.589	[0.107]	0.144
$\pi_{4,f,s}$: female, skilled	0.267	[0.134]	0.047
$\pi_{4,k,u}$: male, unskilled	-0.524	[0.155]	-0.032
$\pi_{4,k,s}$: male, skilled	0.026	[0.196]	0.002
Non Pecuniary Rewards/Tastes			
$\pi_{1,f}$: female, home production	0.961	[0.078]	
$\pi_{2,f}$: female, home production trend	-0.059	[0.003]	
$\pi_{1,k}$: male, home production	-0.716	[0.127]	
$\pi_{2,k}$: male, home production trend	0.054	[0.005]	
$\psi_{f,u,m}$: female, unskilled, manual	0.036	[0.048]	
$\psi_{f,u,r}$: female, unskilled, routine	-0.437	[0.046]	
$\psi_{f,u,a}$: female, unskilled, abstract	-0.555	[0.044]	
$\psi_{f,s,m}$: female, skilled, manual	-0.756	[0.130]	
$\psi_{f,s,r}$: female, skilled, routine	-0.362	[0.088]	
$\psi_{f,s,a}$: female, skilled, abstract	0.508	[0.080]	
$\psi_{k,u,m}$: male, unskilled, manual	0.541	[0.063]	
$\psi_{k,u,r}$: male, unskilled, routine	0.296	[0.064]	
$\psi_{k,u,a}$: male, unskilled, abstract	-0.682	[0.066]	
$\psi_{k,s,m}$: male, skilled, manual	-0.347	[0.073]	
$\psi_{k,s,r}$: male, skilled, routine	-0.347	[0.092]	
$\psi_{k,s,a}$: male, skilled, abstract	0.581	[0.083]	

Notes: The table shows the point estimates, standard errors and average marginal effects of the main parameters from the supply side of the model. Average marginal effects of the fertility and marital status variables are calculated by taking the numerical derivative of the probability of choosing home production with respect to the given variable. I calculate this derivative in each year separately and then take the average across all years. In the case of the pecuniary rewards (ψ_1), the reported average marginal effect is calculated by finding the numerical derivative of the probability that a labor type chooses a given market occupation with repect to the wage. Ie calculate this derivative in each year and for every possible labor type-occupation combination separately and then take the average across all the values.

Table 2.7: Counterfactual Excercises

	Change in Occupation and Wage Gaps: Difference in Difference.					
	ENIGH	Model	CF1 (α's)	CF2 (CHL5)	CF3: (MARR)	CF4: ($\pi_{2,f}$, $\pi_{2,k}$)
Occupation Shares						
College						
Abstract	-2.13	-2.01	-0.40	-2.06	-2.00	-0.95
Routine	-0.39	-0.76	-0.65	-0.76	-0.74	-0.46
Manual	0.09	-0.29	-0.34	-0.29	-0.29	-0.10
Home Production	-0.81	-0.20	-0.35	-0.14	-0.22	-1.75
High School						
Abstract	-1.53	-2.42	-2.09	-2.06	-2.30	-0.48
Routine	-2.04	-3.00	-2.94	-2.62	-2.88	-0.88
Manual	-5.99	-3.53	-3.75	-2.96	-3.34	-0.40
Home Production	13.02	12.44	12.40	11.13	12.01	5.26
Mean Wages (% Change)						
College						
Abstract	-9.39	-12.13	20.01	-11.81	-12.13	-18.47
Routine	-15.99	2.06	34.51	2.06	1.68	-7.58
Manual	-35.05	9.84	18.45	10.04	9.79	3.87
High School						
Abstract	0.20	2.63	10.94	0.97	2.07	-7.91
Routine	9.74	11.94	19.78	8.50	10.79	-9.94
Manual	10.52	9.55	21.78	5.03	8.03	-19.37

Notes: The Table shows the difference between males and females of the difference between C.1990 and C.2013 of the occupational shares and wages under alternative scenarios. The first two columns correspond to the ENIGH and and model predictions. Column CF1 corresponds to the counterfactual estimates once all the linear, quadratic, and cubic coefficients of the α shares are set to zero. Column CF2 correspond to the counterfactual estimates once we set the probability of having a children under the age of 5 to the values of 1989, and constant across the years. Column CF3 correspond to the counterfactual estimates once we set the probability of being married or having a permanent partner to the values of 1989, and constant across the years. Finally, column CF4 correspond to the counterfactual estimates once we set the coefficients of the linear trends $(\pi_{2,f}, \pi_{2,k})$ in the home production utility function to zero.

Table 2.8: Parameter Estimates: Production Technology. Alternative Supply Measures

	Full-Time Workers			Part-Time Workers			Hours Worked		
	Estimate	SE	Elasticity	Estimate	SE	Elasticity	Estimate	SE	Elasticity
Gender									
$\rho_{4,m}$: female, male (manual)	0.175	[0.181]	1.212	-0.258	[0.152]	0.795	0.161	[0.138]	1.192
$\rho_{4,r}$: female, male (routine)	0.179	[0.129]	1.219	-0.030	[0.110]	0.971	0.355	[0.146]	1.551
$\rho_{4,a}$: female, male (abstract)	0.622	[0.099]	2.646	0.607	[0.121]	2.543	0.666	[0.108]	2.990
Education									
$\rho_{3,m}$: college, high school (manual)	0.722	[0.067]	3.594	0.771	[0.083]	4.371	0.803	[0.120]	5.081
$\rho_{3,r}$: college, high school (routine)	0.355	[0.041]	1.549	0.364	[0.073]	1.572	0.342	[0.122]	1.519
$\rho_{3,a}$: college, high school (abstract)	0.276	[0.121]	1.382	0.151	[0.197]	1.177	0.173	[0.211]	1.209
Occupation									
ρ_1: abstract, routine and manual	0.031	[0.094]	1.032	0.688	[0.167]	3.206	0.621	[0.186]	2.639
ρ_2: routine, manual	-0.141	[0.183]	0.877	-0.519	[0.146]	0.658	-0.246	[0.192]	0.803

Notes: The table shows the point estimates and standard errors of the elasticities of substitution from the production technology using different labor supply measures. Columns 1-3 report the baseline estimates; columns 4-6 report the estimates if income from part-time workers is also included in earnings series; and columns 7-9 report the estimates if we measure labor supply by the total number of hours worked of each group instead of the head-count. Since there is no measure of hours worked for people that are in home production, those values are imputed. We assign each person in home production the average number of hours worked by workers in market occupation with the same level of schooling, sex, and age.

Table 2.9: Parameter Estimates: Occupational Choice. Alternative Supply Measures

	Full-Time Workers			Part-Time Workers			Hours Worked		
	Estimate	SE	Av. MFX	Estimate	SE	Av. MFX	Estimate	SE	Av. MFX
Earnings									
ψ_1: earnings	0.154	[0.009]	0.022	0.090	[0.009]	0.013	0.138	[0.012]	0.023
Fertility/Children									
$\pi_{3,f,u}$: female, unskilled	0.712	[0.127]	0.174	0.647	[0.214]	0.159	0.661	[0.227]	0.162
$\pi_{3,f,s}$: female, skilled	-0.139	[0.229]	-0.025	0.158	[0.172]	0.029	0.126	[0.207]	0.022
$\pi_{3,k,u}$: male, unskilled	-0.367	[0.182]	-0.022	-0.242	[0.259]	-0.015	-0.238	[0.206]	-0.015
$\pi_{3,k,s}$: male, skilled	0.122	[0.219]	0.008	0.022	[0.235]	0.002	0.023	[0.263]	0.001
Marriage									
$\pi_{4,f,u}$: female, unskilled	0.589	[0.107]	0.144	0.567	[0.146]	0.139	0.574	[0.17]3	0.141
$\pi_{4,f,s}$: female, skilled	0.267	[0.134]	0.047	0.127	[0.187]	0.023	0.294	[0.179]	0.051
$\pi_{4,k,u}$: male, unskilled	-0.524	[0.155]	-0.032	-0.537	[0.182]	-0.034	-0.535	[0.191]	-0.033
$\pi_{4,k,s}$: male, skilled	0.026	[0.196]	0.002	-0.043	[0.234]	-0.003	-0.045	[0.166]	-0.003
Non Pecuniary Rewards/Tastes									
$\pi_{1,f}$: female, home production	0.961	[0.078]		0.811	[0.116]		0.935	[0.167]	
$\pi_{2,f}$: female, home production trend	-0.059	[0.003]		-0.051	[0.004]		-0.057	[0.004]	
$\pi_{1,k}$: male, home production	-0.716	[0.127]		-0.707	[0.135]		-0.609	[0.165]	
$\pi_{2,k}$: male, home production trend	0.054	[0.005]		0.047	[0.004]		0.048	[0.004]	
$\psi_{f,u,m}$: female, unskilled, manual	0.036	[0.048]		-0.041	[0.057]		-0.126	[0.096]	
$\psi_{f,u,r}$: female, unskilled, routine	-0.437	[0.046]		-0.420	[0.059]		-0.361	[0.096]	
$\psi_{f,u,a}$: female, unskilled, abstract	-0.555	[0.044]		-0.504	[0.061]		-0.455	[0.094]	
$\psi_{f,s,m}$: female, skilled, manual	-0.756	[0.130]		-0.641	[0.099]		-0.723	[0.107]	
$\psi_{f,s,r}$: female, skilled, routine	-0.362	[0.088]		-0.263	[0.106]		-0.066	[0.082]	
$\psi_{f,s,a}$: female, skilled, abstract	0.508	[0.080]		0.721	[0.105]		0.622	[0.089]	
$\psi_{k,u,m}$: male, unskilled, manual	0.541	[0.063]		0.577	[0.062]		0.687	[0.124]	
$\psi_{k,u,r}$: male, unskilled, routine	0.296	[0.064]		0.385	[0.063]		0.404	[0.123]	
$\psi_{k,u,a}$: male, unskilled, abstract	-0.682	[0.066]		-0.497	[0.063]		-0.542	[0.129]	
$\psi_{k,s,m}$: male, skilled, manual	-0.347	[0.073]		-0.484	[0.116]		-0.355	[0.100]	
$\psi_{k,s,r}$: male, skilled, routine	-0.347	[0.092]		-0.378	[0.111]		-0.318	[0.102]	
$\psi_{k,s,a}$: male, skilled, abstract	0.581	[0.083]		0.873	[0.115]		0.727	[0.106]	

Notes: The table shows the point estimates, standard errors and average marginal effects of the main parameters from the supply side of the model using different labor supply measures. Average marginal effects of the fertility and marital status variables are calculated by taking the numerical derivative of the probability of choosing home production with respect to the given variable. We calculate this derivative in each year separately and then take the average across all years. In the case of the pecuniary rewards (ψ_1), the reported average marginal effect is calculated by finding the numerical derivative of the probability that a labor type chooses a given market occupation with repect to the wage. We calculate this derivative in each year and for every possible labor type-occupation combination separately and then take the average across all the values.

Table 2.10: Parameter Estimates: Production Technology. Alternative Model Specifications

	Baseline			Nests Order Swap			Routine			Manual		
	Estimate	SE	Elasticity	Estimate	SE	Elasticity	Estimate	SE	Elasticity	Estimate	SE	Elasticity
Gender												
$\rho_{4,m}$: female, male (manual)	0.175	[0.181]	1.212	-0.246	[0.098]	0.802	-0.427	[0.177]	0.701	-0.029	[0.104]	0.972
$\rho_{4,r}$: female, male (routine)	0.179	[0.129]	1.219	-0.278	[0.102]	0.782	-0.095	[0.153]	0.913	0.007	[0.023]	1.007
$\rho_{4,a}$: female, male (abstract)	0.622	[0.099]	2.646	0.466	[0.115]	1.872	0.529	[0.099]	2.121	0.551	[0.099]	2.225
Education												
$\rho_{3,m}$: college, high school (manual)	0.722	[0.067]	3.594	0.564	[0.045]	2.292	0.454	[0.104]	1.831	0.581	[0.058]	2.385
$\rho_{3,r}$: college, high school (routine)	0.355	[0.041]	1.549	0.382	[0.054]	1.618	0.380	[0.051]	1.614	0.189	[0.047]	1.233
$\rho_{3,a}$: college, high school (abstract)	0.276	[0.121]	1.382	0.012	[0.040]	1.012	0.008	[0.048]	1.008	0.446	[0.150]	1.805
Occupation												
ρ_1: abstract, routine and manual	0.031	[0.094]	1.032	0.441	[0.109]	1.788						
ρ_2: routine, manual	-0.141	[0.183]	0.877	-1.816	[0.135]	0.355						
ρ_1: routine, abstract and manual							-0.784	[0.231]	0.560			
ρ_2: abstract, manual							0.332	[0.171]	1.496			
ρ_1: manual, abstract and routine										0.411	[0.134]	1.697
ρ_2: abstract, routine										-0.714	[0.099]	0.583

Notes: The table shows the point estimates and standard errors of the elasticities of substitution from the production technology using different model specification. Columns 1-3 report the baseline estimates; columns 4-6 report the estimates after switching the order of the second (education) and third (gender) nests of the production technology; columns 7-9 show the estimates if the occupational group that has the common elasticity with the other two groups is the routine task-intensive; and columns 10-12 report the estimates if the occupational group that has the common elasticity with the other two groups is the manual task-intensive.

Table 2.11: Parameter Estimates: Occupational Choice. Alternative Model Specifications

	Baseline			Nests Order Swap			Routine			Manual		
	Estimate	SE	Av. MFX	Estimate	SE	Av. MFX	Estimate	SE	Av. MFX	Estimate	SE	Av. MFX
Earnings												
ψ_1: earnings	0.154	[0.009]	0.022	0.109	[0.009]	0.018	0.085	[0.008]	0.014	0.098	[0.010]	0.016
Fertility/Children												
$\pi_{3,f,u}$: female, unskilled	0.712	[0.127]	0.174	0.805	[0.118]	0.197	0.832	[0.162]	0.206	0.664	[0.050]	0.162
$\pi_{3,f,s}$: female, skilled	-0.139	[0.229]	-0.025	-0.042	[0.116]	-0.008	-0.126	[0.330]	-0.022	0.022	[0.150]	0.004
$\pi_{3,k,u}$: male, unskilled	-0.367	[0.182]	-0.022	-0.162	[0.098]	-0.010	-0.174	[0.240]	-0.009	-0.246	[0.164]	-0.015
$\pi_{3,k,s}$: male, skilled	0.122	[0.219]	0.008	0.073	[0.113]	0.005	0.061	[0.310]	0.004	-0.050	[0.196]	-0.003
Marriage												
$\pi_{4,f,u}$: female, unskilled	0.589	[0.107]	0.144	0.767	[0.115]	0.188	0.500	[0.107]	0.124	0.588	[0.056]	0.144
$\pi_{4,f,s}$: female, skilled	0.267	[0.134]	0.047	0.318	[0.105]	0.059	0.103	[0.176]	0.018	0.193	[0.093]	0.035
$\pi_{4,k,u}$: male, unskilled	-0.524	[0.155]	-0.032	-0.598	[0.115]	-0.037	-0.635	[0.244]	-0.034	-0.577	[0.144]	-0.035
$\pi_{4,k,s}$: male, skilled	0.026	[0.196]	0.002	-0.145	[0.124]	-0.009	-0.142	[0.299]	-0.010	-0.129	[0.154]	-0.009
Non Pecuniary Rewards/Tastes												
$\pi_{1,f}$: female, home production	0.961	[0.078]		0.822	[0.088]		0.712	[0.105]		0.848	[0.049]	
$\pi_{2,f}$: female, home production trend	-0.059	[0.003]		-0.051	[0.003]		-0.026	[0.003]		-0.054	[0.002]	
$\pi_{1,k}$: male, home production	-0.716	[0.127]		-0.646	[0.098]		-0.971	[0.221]		-0.672	[0.124]	
$\pi_{2,k}$: male, home production trend	0.054	[0.005]		0.045	[0.002]		0.081	[0.005]		0.045	[0.003]	
$\psi_{f,u,m}$: female, unskilled, manual	0.036	[0.048]		0.196	[0.050]		0.101	[0.076]		0.019	[0.021]	
$\psi_{f,u,r}$: female, unskilled, routine	-0.437	[0.046]		-0.228	[0.052]		-0.208	[0.076]		-0.391	[0.024]	
$\psi_{f,u,a}$: female, unskilled, abstract	-0.555	[0.044]		-0.356	[0.052]		-0.525	[0.078]		-0.492	[0.027]	
$\psi_{f,s,m}$: female, skilled, manual	-0.756	[0.130]		-0.699	[0.063]		-0.804	[0.148]		-0.720	[0.092]	
$\psi_{f,s,r}$: female, skilled, routine	-0.362	[0.088]		-0.235	[0.061]		-0.126	[0.120]		-0.265	[0.063]	
$\psi_{f,s,a}$: female, skilled, abstract	0.508	[0.080]		0.739	[0.058]		0.863	[0.123]		0.757	[0.064]	
$\psi_{k,u,m}$: male, unskilled, manual	0.541	[0.063]		0.625	[0.025]		0.717	[0.062]		0.631	[0.022]	
$\psi_{k,u,r}$: male, unskilled, routine	0.296	[0.064]		0.423	[0.029]		0.527	[0.063]		0.424	[0.023]	
$\psi_{k,u,a}$: male, unskilled, abstract	-0.682	[0.066]		-0.522	[0.026]		-0.403	[0.063]		-0.489	[0.033]	
$\psi_{k,s,m}$: male, skilled, manual	-0.347	[0.073]		-0.591	[0.069]		-0.582	[0.069]		-0.474	[0.064]	
$\psi_{k,s,r}$: male, skilled, routine	-0.347	[0.092]		-0.149	[0.063]		-0.269	[0.065]		-0.316	[0.058]	
$\psi_{k,s,a}$: male, skilled, abstract	0.581	[0.083]		0.783	[0.063]		0.895	[0.063]		0.749	[0.071]	

Notes: The table shows the point estimates, standard errors and average marginal effects of the main parameters from the supply side of the model using different model specifications. Columns 1-3 report the baseline estimates; columns 4-6 report the estimates after switching the order of the second (education) and third (gender) nests of the production technology; columns 7-9 show the estimates if the occupational group that has the common elasticity with the other two groups is the routine task-intensive; and columns 10-12 report the estimates if the occupational group that has the common elasticity with the other two groups is the manual task-intensive. Average marginal effects of the fertility and marital status variables are calculated by taking the numerical derivative of the probability of choosing home production with respect to the given variable. We calculate this derivative in each year separately and then take the average across all years. In the case of the pecuniary rewards (ψ_1), the reported average marginal effect is calculated by finding the numerical derivative of the probability that a labor type chooses a given market occupation with repect to the wage. We calculate this derivative in each year and for every possible labor type-occupation combination separately and then take the average across all the values.

2.8.2 Division of Occupations into Manual, Routine, and Abstract Task-Intensive Groups

Following the task-based framework, the 18 principal level occupations from the ENIGH are classified in three groups defined by whether the activities done in the jobs are predominantly manual, routine (repetitive and easily codifiable tasks), or abstract intensive. The division is based on the measures constructed by Autor et al. (2003) from different sets of variables of the 1977 Dictionary of Occupational Titles (DOT) of the U.S., and then linked to the three-digit occupation codes of the CENSUS. The DOT evaluated highly detailed occupations along 44 objective and subjective dimensions that include physical demands and required worker aptitudes, temperaments and interests. Autor et al. (2003) used a subset of those dimensions to generate a simple typology consistent of three aggregates for abstract, routine, and manual tasks. The abstract task measure corresponds to the average from two variables of the DOT: DCP, which measures direction, control, and planning of activities; and GED-MATH, which measures quantitative reasoning requirements. The routine task measure corresponds to an average from two variables of the DOT: STS, which measures adaptability to work requiring set limits, tolerances, or standards; and FINGDEX, measuring finger dexterity. Finally, the manual task measure uses a single variable, EYEHAND, which measures eye, hand, foot coordination.[24]

In practice, I first create a cross-walk between three-digit CENSUS codes in the U.S. and the 18 categories of the principal group occupational division of the ENIGH. This task is facilitated by the fact that both the ENIGH and the U.S. CENSUS follow similar international standards when constructing their own occupation classifications. Since the three task measures are ordinal, there is no

[24]See the online Appendix in Dorn (2009) for further details. Other papers that have used this measures include Autor et al. (2006); Goos and Manning (2007); Dorn (2009); Rendall (2013); Autor and Dorn (2013); Adda et al. (2017).

direct way to use the actual magnitude of the variables to compare occupations across the three dimensions. For each task measure I first organize the three-digit occupations by percentiles, and then calculate the median percentile of the measure within the broader 18 occupational groups of the ENIGH. Each of the 18 occupations is assigned to the group in which the median percentile was highest (see Table 2.1).

This procedure generated a balanced division with respect to the overall employment share of each group, and it is also consistent with the broad classification of aggregate occupations used in the literature that follow the task-based framework. Two important caveats should be stressed: First, any attempt to homogenize occupation classification systems from different countries involves some subjective choices. In the cases where I found that an occupation didn't have an immediate correspondence between the two systems, I had to use my judgement, based on documentation about the description of the occupation, to select a corresponding match. Second, the task measures were created specifically for U.S. economy, and it is likely that there are differences in the intensity in which certain skills are used in given occupations between the U.S. and Mexico. Results should be interpreted with this two caveats in mind.

2.8.3 Estimation of the Model: Error Structure, Weight Matrix, and Standard Errors

I assume a simplified error structure to facilitate estimation. Let Θ be the 81 $\times$ 1 vector of parameters to be estimated. Let $p(\Theta)$ by the 364 $\times$ 1 vector of wage and supply predictions of the model as function of the parameters. Finally, let q be the observed vector of wages and labor shares taken directly from the ENIGH data. For any given prediction i, we assume that the error term, e_i, at the true parameter vector, Θ^*, follows a normal distribution centred at zero that is independent across i. That is,

$$e_i = q_i - p_i(\Theta^*), \tag{2.8.1}$$

where $f(e_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} exp(\frac{e_i^2}{2\sigma_i^2})$.

The log-likelihood function takes the form

$$\log \mathcal{L}(\Theta) = \sum_i \log f(e_i) = \sum_i \log f(q_i - p_i(\Theta)), \tag{2.8.2}$$

and the respective score function, $s(\Theta)$, is:

$$s(\Theta) = \frac{\partial \log \mathcal{L}(\Theta)}{\partial \Theta} = \sum_i \frac{\partial \log f(q_i - p_i(\Theta))}{\partial \Theta} = \sum_i \frac{1}{\sigma_i^2} \frac{\partial p_i(\Theta)}{\partial \Theta}(q_i - p_i(\Theta)), \tag{2.8.3}$$

which we can write more compactly in vector form as

$$s(\Theta) = W'(\Theta)(q - p(\Theta)). \tag{2.8.4}$$

Here, $W(\Theta)$ is 364×81 weight matrix that depends on the derivatives of the vector of predictions with respect to each of the parameters, and the variance of each prediction error σ_i^2. At the maximum likelihood estimate, $\hat{\Theta}_{ml}$, the score vector of the log likelihood is set to zero:

$$s(\hat{\Theta}_{ml}) = W'(\hat{\Theta}_{ml})(q - p(\hat{\Theta}_{ml})) = 0. \tag{2.8.5}$$

We use $m = q - p(\Theta)$ as a vector of population moments such that $E(q - p(\Theta)) = 0$, and obtain a a consistent estimator of Θ^* by GMM:

$$g(\hat{\Theta}_{gmm}) = W'(q - p(\hat{\Theta}_{gmm})) = 0, \tag{2.8.6}$$

where W' is a fixed positive definite matrix of instruments. Efficient GMM estimator can be obtained by choosing instruments that are asymptotically equivalent to the weights $W'(\hat{\Theta}_{ml})$ in Equation (2.8.4). The problem is that we would need to have a consistent initial estimate of Θ^*. Given that we do not have those initial consistent estimates, we follow a iterative process. We start from a plausible set of initial values of the parameters (Θ_0), and use them to estimate the vector of partial derivatives $\frac{\partial \hat{p}_i(\Theta_0)}{\partial \Theta_0}$. The estimates of the variance of each error, $\hat{\sigma}_{i,0}^2$, are calculated as the square of the estimated error from this initial set of parameter values. Both of these estimates are then used to construct an initial weight matrix, which allows us to solve the minimization problem.[25] The estimates obtained after this first iteration[26] are used to update the weight matrix, and the process continues until the parameter vector converges to a stable point.

[25]The parameter search is done using the interior-point algorithm in Matlab.

[26]Note that even though the weight matrix is a function of the parameters, it remains fixed during the parameter search.

Since it is usually not possible to satisfy Equation (2.8.6), we estimate the parameters of the model using the quadratic form:

$$\hat{\Theta}_{gmm} = \text{argmin}[q - p(\Theta)]'W(\Theta)W'(\Theta)[q - p(\Theta)]. \tag{2.8.7}$$

Finally, the standard errors of the parameter estimates are calculated applying the standard method of moments formula. Let Γ be the matrix of partial derivatives of the sample moments $\bar{m}(\hat{\Theta}_{gmm})$ with respect to the parameters. The ith row correpsonds to:

$$\Gamma_i(\hat{\Theta}_{gmm}) = \frac{\partial \bar{m}_i(\hat{\Theta}_{gmm})}{\partial \hat{\Theta}_{gmm}}, \tag{2.8.8}$$

so the variance-covariance matrix can be calculated using:

$$V\hat{a}r(\hat{\Theta}_{GMM}) = \left[\Gamma(\hat{\Theta}_{gmm})'V\hat{a}r[\bar{m}(\Theta_{gmm})]^{-1}\Gamma(\hat{\Theta}_{gmm})\right]^{-1}. \tag{2.8.9}$$

Chapter 3

Herding in Equity Crowdfunding[1]

3.1 Introduction

In recent years crowdfunding, the practice of funding a project or venture by raising small amounts of money from a large number of people, usually via an online dedicated platform, has emerged as a popular alternative channel for entrepreneurs. Funds raised via crowdfunding expanded by 167 percent in 2014 to reach $16.2 billion, up from $6.1 billion in 2013 (Massolution, 2015). A number of high profile campaigns and an increasing appetite to "cut out the middleman" mean crowdfunding is likely to remain an important part of early stage finance for some years to come.

Broadly speaking, crowdfunding can be divided into four main categories: donations, rewards-based (also called pre-selling), lending, and equity crowdfunding. The focus of this paper is on equity crowdfunding, where the crowd take an equity stake in the business in much the same way Venture Capitalist (VC) funding works. Although the typical individual investment ("pledge")

[1]This Chapter was written jointly with Thomas Åstebro (HEC Paris); Stefano Lovo (HEC Paris); and Nir Vulkan (Said Business School).

is much smaller in equity crowdfunding than for business angels and VCs, the equity raised can be as substantial. Moreover, this model has already become a significant financing vehicle for start-ups in the UK, and is growing rapidly:[2] around 21 percent of all early stage investment and as much as 35.5 percent of all seed-stage investment deals in the UK went through equity crowdfunding sites in 2015 (Beauhurst, 2016).

This chapter is motivated by the expressed concerns of analysts and regulators that crowdfunding backers can be taken advantage of by investment promoters. For example, the recently (October 2015) announced details of the regulatory framework for equity crowdfunding in the U.S.[3] has already been criticized by Shiller (2015) for not protecting backers enough. Yet, there exists little information on the actual investment behaviour of equity crowdfunding backers, so it is hard to provide strong guidance at this stage.[4]

One of the most common concerns is that there will be herding in crowdfunding. Herding is said to occur when prior investors' decisions induce future investors' to take similar decisions. When future investors decisions are orthogonal to their private information, a so called information cascade occurs and the market fails to aggregate relevant private information (e.g. Banerjee (1992)).

[2]The UK is the fastest growing country for equity crowdfunding campaigns in the world, both in terms of the number of campaigns and their sizes. This is because the UK has had a clear regulatory framework for the equity crowdfunding business model since the end of 2011. Backers of start-ups in the U.K. also benefit from a very generous tax incentive via the Seed Enterprise Investment Scheme, SEIS, and the Enterprise Investment Scheme, EIS. Both schemes are designed to help small UK-based companies raise finance by offering tax relief on new shares in those companies. The EIS is aimed at wealthier backers who receive 30 percent tax relief but whose pledges cannot be sold or transferred for a minimum lock-in period of three years. The SEIS is more generous and provides tax relief of up to 50 percent on pledges of up to £100,000, and capital gains tax exemption. The maximum investment that can be raised by a company under this scheme is limited to £150,000

[3]See the Security and Exchange Commission (SEC) final draft at: `http://www.sec.gov/rules/final/2015/33-9974.pdf`

[4]Most analysis of crowdfunding has been limited to rewards-based platforms (for reviews see e.g. Belleflamme et al. (2014); Agrawal et al. (2014)) which avoid equity investment regulation by allowing backers to pre-purchase a product, obtain rewards such as a t-shirt, or simply to donate funds.

This could lead to decisions that differ from those that would have been taken had all investors' private information been taken into account. Regulators may find such behaviour of great concern if weak or even incorrect signals get amplified because information verification is difficult and because of a herd response overtaking private signals.

Investment opportunities presented on the world wide web may be particularly sensitive to such herding as signals in this medium can be propagated easily to a large number of recipients with little quality control or verification. Moreover, on most crowdfunding platforms the funds that investors pledge are invested only if the total amount pledged reaches a pre-specified threshold by a given date. This conditional-funding-clause is popularly called "all-or-nothing". The clause introduces the additional worry that trusting that only profitable projects will get enough financing to reach the threshold, uninformed investors might choose to invest in the absence of any public or private information about the project's quality and in this way will trigger herding.

In this chapter we analyze backers' behavior in crowdfunding campaigns using a unique investment-level dataset from one of the leading UK equity crowdfunding platforms which uses an all-or-nothing funding scheme. Our empirical analysis is guided by a stylized model of equity crowdfunding. The model is developed to answer the following questions: First, how should herding be characterized in an equity crowdfunding setting where much of the inter-temporal information about a project is based on the amounts previously invested? Second, what would be the predictions in terms of the timing and amounts of future investments based on information about the history of all past investments and investors?

Our model features a sequence of risk averse backers who visit the crowdfunding platform following a Poisson process. Each backer is either uninformed or has some information that is positive or negative. All past strictly positive pledges are observable. We show that in equilibrium, negatively informed backers do not pledge, uninformed and positively informed backers' pledges depend on the past history of pledges, and positively informed backers pledge on average more than uninformed backers. Because positively (negatively) informed investors are more (less) numerous for a better than for a worse project, large investments provide positive public information about the project's quality, whereas periods of absence of investment provide negative information.

We show that while investment herding can occur, in the senses that past backer's pledges can trigger future pledges, it is never associated with an information cascade. The reason is that, although noisy, a pledge always provides some information about the project. Furthermore, investment herding does not last in the sense that a long enough period of absence of pledges will result in no further investment, so-called abstention herding. Different from investment herding, abstention herding is always associated with an information cascade. This simple model provides us with some testable predictions which we can then bring to the data.

We use a unique dataset kindly shared with us by SEEDRS, one of the leading UK equity crowdfunding platforms, to test the predictions of our model. We analyse 69,699 investments ("pledges") by 22,615 investors ("backers") made between October 2012 and March 2016. Our data analysis represents a novel attempt to surmount omitted variable bias present in much of prior literature using field data on financial decision making. First, our data details pledges by the second, making it possible to order all pledges and examine Granger causality. Second, we have information from the crowdfunding platform on the characteristics of backers that are not revealed to potential follow-on backers. In particular, for a given pledge an investor may decide to maintain anonymity on the platform.

We instrument the amount pledged by a backer in a specific campaign using information which the econometrician can observe but which subsequent investors cannot.

We examine the correlation between the sizes of subsequent pledges and the decay in the sizes of the pledges over subsequently arriving backers. More precisely, we show that pledges are strikingly affected by immediate prior pledges and the timing between subsequent investments. This behaviour is consistent with our model of rational herding containing a mix of informed and uninformed arriving investors, where large prior pledges from informed investors provide information to future arriving uninformed investors which dissipates over their arrival.

The results suggest a considerable level of immediate inter-temporal herding on an equity crowdfunding platform that is, nevertheless, quickly dissipating. The effect sizes are moderate. For example, a doubling of the size of the immediate prior pledge increases the size of the subsequent pledge by between 5.8 percent and 20.3 percent, depending on the combination of characteristics of backers, with a typical effect size of 12.3 percent.

The theoretical model also proposes a mechanism to explain why the final success of a campaign is so closely correlated with the support it gets at the very early stage of fund-raising (Kuppuswamy and Bayus, 2015; Colombo et al., 2015; Vulkan et al., 2016). The absence of pledges at the start of a campaign are indicative that the project is of bad quality. As the public belief about the quality of the project becomes more pessimistic, private signals of arriving investors have to be larger in order to induce a positive pledge. This reinforces the negative dynamics and leads to the possibility of an abstention information cascade occurring from the onset, so that failed campaigns are not able to get any traction.

Our paper is most closely related to Zhang and Liu (2012) and Bursztyn et al. (2014). Zhang and Liu (2012) study pledges made on a peer-to-peer lending web site, while Bursztyn et al. (2014) examine investor behaviour in a randomized control experiment working closely with a large financial brokerage in Brazil. Zhang and Liu (2012) use panel field data on daily (and hourly) lending amounts as a function of the cumulative amount of funding up to $t-1$, and its interaction with observable "listing" attributes, while controlling for listing fixed effects and observable listing attributes. Rational herding is said by the authors to exist if the coefficient for the interaction between the cumulative amount and listing attributes is significant and takes the opposite sign of the listing attribute's main effect. The argument made is that for poor (good) listing attributes, such as a low (high) credit rating, the incremental increase in cumulative prior funding must signal higher (lower) unobserved quality the lower (higher) the observed attribute. The authors provide evidence that this specific type of rational herding is observed in the data.

Bursztyn et al. (2014) investigate the effect on private investment decisions from a) knowing whether a peer –a colleague from work, a friend, or a family member– had a desire to purchase the asset and b) whether the peer actually became in possession of the asset. Both a) and b) were randomized. This set-up allows the authors to disentangle herding based on social learning a), and social utility b), and they find both to be at play with large differences in take-up rates compared to the backers not informed about peers investing preferences or behaviour.

Our work is also related to the theoretical literature of rational herding (Banerjee, 1992; Bikhchandani et al., 1992; Smith and Sorensen, 2000) and rational herding in the financial market (Avery and Zemsky, 1998; Park and Sabourian, 2011). In the rational herding literature, traded quantities are discrete and both buy and sell herding and information cascades can potentially occur. The presence of prices, on the other hand, has been shown to prevent herding and/or information cascades. In the equity crowdfunding we consider, the price is fixed so the disciplining effect of prices is absent. However, tradable quantities are continuos, and the disciplining effect of price is at least partially replaced by this continuity.

Finally, our paper also contributes to the broader empirical literature on herding behaviour in financial decisions. Several papers have tried to study this using observational data (Hong et al., 2004, 2005; Ivkovic and Weisbenner, 2007; Brown et al., 2008; Banerjee et al., 2012; Li, 2014) and experimental data (Duflo and Saez, 2003; Bursztyn et al., 2014; Beshears et al., 2015). Our paper offers insights into a relatively new type of financial decision that seems to be rapidly growing in size and importance, and for which regulators are showing a keen interest.

The rest of the chapter is organized as follows. In Section 3.2 we present the theoretical model and derive its empirical implications. Section 3.3 provides a description of the data used in the analysis. In Section 3.4 we test if the theoretical predictions are in line with what is observed in the data. Finally, Section 3.5 concludes.

3.2 A Model of Herding in Equity Crowdfunding

Firm: A firm seeks financing from investors (henceforth backers) to implement a risky project. The project can be of two qualities: "good" or "bad". Each dollar invested in the project generates ρ dollars, where $\rho = \alpha > 1$, for a good project, whereas $\rho = 0$ for a bad project. We denote with π_0 the ex-ante probability that the project is good.

Investors: Backers are risk averse with log utility function. There are two populations of backers: uninformed and informed. Uninformed backers have no private information about the project's quality. Informed backers possess some private information regarding the project's quality. Each informed backer i receives a private signal, $s_i \in \{g, b\}$, indicating whether the project is "good" ($s_i = g$) or "bad" ($s_i = b$). We assume that

$$\begin{aligned} \mathbb{P}[s_i = g | \text{ good project}] &= 1, \\ \mathbb{P}[s_i = b | \text{ good project}] &= 0, \\ \mathbb{P}[s_i = g | \text{ bad project}] &= 1 - q, \\ \mathbb{P}[s_i = b | \text{ bad project}] &= q, \end{aligned} \tag{3.2.1}$$

where $q \in (0, 1)$. Thus, a signal g is a partially informative positive signal about the project's quality, whereas b is a perfectly informative signal that the project is of bad quality. Informed backers' signals are assumed to be conditionally i.i.d. We shall say that a backer's type is $\theta \in \{g, b, u\}$ specifying whether she has received a positive signal g, a negative signal b, or is uninformed, respectively. As a starting point, we assume all bakers have the same initial wealth W, but we extend the model in Subsection 3.2.3 to show that the qualitative predictions do not change if backers' initial wealth are independently distributed or if we allow for multiple signals.

The firm tries to raise funds from backers using a crowdfunding platform. Time is continuous. The crowdfunding campaign starts at $t = 0$ and closes at $t = T$, finite. At $t = 0$ a target amount Y is announced. If at $t = T$

the total amount of funds committed is at least Y the campaign is successful and committed funds are invested. If not the campaign fails and backers are returned their funds. Backers arrive at the platform following an exogenous Poisson processes with intensity 1. A fraction λ of backers is informed whereas the remaining $1-\lambda$ is uninformed. Considering the conditional distribution of signals, the arrival rate of positively informed backers is λ if the project is good, and $\lambda(1-q)$ if the project is bad. The arrival rate of negatively informed backers is 0 if the project is good and λq if the project is bad.

Denote y_t the total net amount of funds committed until time t. We will say that a campaign is in the *underfunding phase* as long as $t < T$ and $y_t < Y$. The campaign enters the *overfunding phase* if $y_t > Y$. Note that if the overfunding phase does not start before T, the campaign fails.

Histories and strategies: Pledges are public so that a backer arriving at time t observes all past strictly positive pledges. Importantly, they do not observe backers who visited the platform but chose not to make any pledge. Denote with $h_t = \{(t_1, x_{t_1}), \ldots, (t_n, x_{t_n}), t\}$ the history of pledges before time t, where t_i is the time of the i-th pledge, $x_{t_i} > 0$ is its size, and n is the number of pledges before t. Let $\mathcal{H}$ be the set of all possible histories.

A (pure) strategy σ for a backer is a mapping from his type and public history into a non-negative pledge, i.e., $\sigma : \{g, b, u\} \times \mathcal{H} \to \mathbb{R}^+$.

Beliefs and success probabilities: The public belief that the project is good, given h_t, is denoted $\pi_t := \mathbb{P}(\rho = \alpha | h_t)$. From Bayes' rule, the belief of a type g backer arriving at time t is $p(\pi_t) := \frac{\pi_t}{\pi_t + (1-\pi_t)(1-q)}$.[5] Clearly the way h_t affects π_t depends on the strategy used by backers until t. We denote with π_t^θ the belief

[5] $\mathbb{P}(\rho = \alpha | g; h_t) := \frac{\mathbb{P}(g|\rho=\alpha;h_t)\times\mathbb{P}(\rho=\alpha|h_t)}{\mathbb{P}(g|\rho=\alpha;h_t)\times\mathbb{P}(\rho=\alpha|h_t)+\mathbb{P}(g|\rho=0;h_t)\times\mathbb{P}(\rho=0|h_t)} = \frac{\pi_t}{\pi_t+(1-\pi_t)(1-q)}$

of a type θ backer at time t. From the assumption on the distribution of private signals, we have

$$0 = \pi_t^b \leq \pi_t = \pi_t^u \leq p(\pi_t) = \pi_t^g, \tag{3.2.2}$$

where the inequalities are strict for $\pi_t \in (0, 1)$. We denote by $\pi_t(x)$ the time t public belief that results from the observation of a pledge $x_t = x > 0$ at t. Given any history $h_t \in \mathcal{H}$, any $\rho \in \{0, \alpha\}$, and any additional pledge x at time t, let

$$S_\rho(x, h_t) := \mathbb{P}(y_T \geq Y|\rho, x, h_t)$$

denote the probability that the campaign succeeds given h_t, $x_t = x$, and the project's quality ρ. This probability depends on the strategies of the current and future backers.

Objective function and equilibrium conditions: We can express the expected payoff for a type θ backer arriving at time t with wealth W and pledging x as

$$\pi_t^\theta S_\alpha(x, h_t)(\ln(W + (\alpha - 1)x) - \ln(W)) \quad + \quad (1 - \pi_t^\theta)S_0(x, h_t)(\ln(W - x) - \ln(W)). \tag{3.2.3}$$

An equilibrium strategy profile is $\hat{\sigma}$ such that for any backer i, of any type θ and wealth W, after every history h_t, she chooses the x that maximizes (3.2.3), where π_t^θ, $S_\alpha(.)$ and $S_0(.)$ are computed taking into account other backer's equilibrium strategies.

Herding and information cascades: We define herding and information cascades similarly to Avery and Zemsky (1998). A backer engages in *investment herding* if she would have pledged 0 at time 0 but if she arrives after enough other backers have made pledges, she pledges a strictly positive amount. On the contrary, a backer engages in *abstention herding* if she would have pledged a strictly positive amount at time 0 but if she arrives after a long enough period of absence of pledges she does not pledge. We will say that the economy enters an *information cascade* after history h_t if (on the equilibrium path) for all $t' > t$ one has $\pi_{t'} = \pi_t$, i.e. the additional history of pledges provide no information about the project's quality. That is, whereas herding can be transitory, once an information cascade starts it never ends.

3.2.1 Equilibrium Pledging Behavior

Expression (3.2.3) shows that a backer's optimal pledge depends on, first, her belief π_t^θ, that is a function of π_t that in turn depends on the past history of pledges and on the strategy used by past backers. Second, it also depends on the conditional probabilities $S_\alpha(x, h_t)$ and $S_0(x, h_t)$ that the campaign succeeds. This second element is absent during the overfunding phase because $y_t \geq Y$ implies $S_\alpha(x, h_t) = S_0(x, h_t) = 1$. As a result, equilibrium in the overfunding phase is unique and a backer's pledge only depends on her belief π_t^θ.

Things are different for the underfunding phase. Before the campaign succeeds, a backer has to take into account how his pledge will affects future backers pledges and through this the probabilities of success $S_\alpha(x, h_t)$ and $S_0(x, h_t)$. This implies that the underfunding phase has multiple equilibria, including equilibria where pledges that are above some arbitrary threshold lead the campaign to fail. Here we are interested in equilibria that satisfy a rather compelling regularity condition, that is, by increasing her pledge a backer cannot make the campaign strictly less likely to succeed. Formally:

Definition 3.2.1. *An equilibrium is said to be* regular *if for any x, x' with $0 \leq x < x'$, any history $h_t \in \mathcal{H}$ and any project quality $\rho \in \{0, \alpha\}$ one has $S_\rho(x, h_t) \leq S_\rho(x', h_t)$.*

Rather than focusing on one of the many possible regular equilibria, in the next proposition we detail the properties that are common to all regular equilibria of the underfunding phase. We then characterize the unique equilibria of the overfunding phase and show that it is regular.

Proposition 3.2.2. *Let $\hat{\sigma}$ be a regular equilibrium of the crowdfunding campaign and let h_t be such that $\mathbb{P}(y_T \geq Y | h_t) > 0$. Then,*

1. *Pledging behavior satisfies the following properties:*

   (a) *A negatively informed investor never invest*

   $$\hat{\sigma}(b, h_t) = 0$$

   (b) *Backers' pledges are weakly increasing in π_t, furthermore: if $\pi_t > \alpha^{-1}$, one has*
   $$0 < \frac{\pi_t \alpha - 1}{\alpha - 1} W \leq \hat{\sigma}(u, h_t) \leq \hat{\sigma}(g, h_t)$$

   (c) *There is $\underline{\pi} > 0$ such that no backer ever invest as soon as $\pi_t \leq \underline{\pi}$.*

2. *The public belief evolves according to the following rule:*

   (a) *As soon as $\pi_t < \underline{\pi}$, for all $t' > t$, $\pi_{t'} = \pi_t$, That is an abstention informational cascade occurs.*

(b) If between t and $t' > t$ there is no pledge and an information cascade has not started, then

$$\pi_{t'} = \phi(\pi_t, t'-t) := \frac{\pi_t e^{-\lambda(t'-t)}}{\pi_t e^{-\lambda(t'-t)} + (1-\pi_t)e^{-\lambda(1-q)(t'-t)}} < \pi_t,$$

(c) The public belief $\pi_t(x)$ is non-decreasing in x. In particular, if $\hat{\sigma}(u, h_t) \neq \hat{\sigma}(g, h_t)$, then $\pi_t(x) = p(\pi_t)$ if $x = \hat{\sigma}(g, h_t)$, whereas $\pi_t(x) = \pi_t$ if $x = \hat{\sigma}(u, h_t)$.

What the proposition says is that in all regular equilibria as long as the probability of success is not nil, negatively informed backers never pledge (1.a). The other backer's pledge is increasing in their belief so that if an uninformed backer pledges, then a positively informed backer would pledge strictly more (1.b). For any pledge to occur at all the public belief that the project is good should not be too small, as for π_t below a given positive threshold $\underline{\pi}$ not even a positively informed backer would pledge (1.c).

This pledging behaviour has an effect on the evolution of the public belief. First, because for $\pi_t < \underline{\pi}$ no backer pledges, the fact of observing no pledge when π_t is small provides no information about the project quality. Thus, the public belief cannot evolve and an information cascade occurs. Because in such cascade backers are abstaining from pledging we call it an *abstention information cascade* (2.a). Suppose that π_t is large enough, so that at t the equilibrium is not in an abstention information cascade. How does π_t evolve? First, because positively informed backers are more optimistic than uninformed backers (see inequalities (3.2.2)) and pledges are increasing in beliefs (see 1.b), a large pledge is more likely to come from a positively informed backer than for an uninformed backer, thus the reaction of the public belief to a pledge is not decreasing in pledge size (2.c). Second, because negatively informed backers do not invest and they are more common when the project is bad, a period of absence of pledges is more likely if the project is bad than if the project is good. Hence before an abstention information cascade starts, a periods of absence of pledges makes backers more pessimistic the longer this period lasts (2.b). When this period last long enough, the public belief will reach $\underline{\pi}$ and an abstention information cascade starts.

The proof of the proposition is in Appendix 3.6.2.1 where we first show that for any regular equilibrium and any history that does not lead the campaign to fail with certainty, a backer's pledge is an increasing function of her belief. Because there are more positively informed backers if the project is good than if the project is bad, the first result implies that a campaign for bad project is not more likely to succeed than a campaign for a good project. This implies that a backer would invest at least as much as what she would invest if the probability that the campaign succeeds did not depend on the project's quality. Because the likelihood ratio on good and bad project success probability is bounded, the positive information given by the fact that the campaign succeeds cannot overwhelm negative enough prior, so that abstention information cascade are always possible. Given these properties of the pledging strategies the evolution of belief immediately follows.

It is not possible to find a closed form solution for a regular equilibrium during the underfunding phase. However in the overfunding phase, i.e., as soon as the campaign succeeds, the equilibrium is unique, is regular, and pledging strategies can be explicitly computed.

Proposition 3.2.3. *As soon as $y_T \geq Y$, the crowdfunding campaign has a unique equilibrium. This equilibrium is regular and such that*

$$\hat{\sigma}(\theta, h_t) = \max\left\{0, \frac{\alpha\pi_t^\theta - 1}{\alpha - 1}W\right\}$$

The threshold $\underline{\pi} = (1-q)/(\alpha - q) \in (0, \alpha^{-1})$.

The proof of the Proposition is in Appendix 3.6.2.2.

3.2.2 Herding and Cascades

What do propositions 3.2.2 and 3.2.3 tell us about the possibility of herding and cascades? First notice that results (2.b) of Proposition 3.2.2 implies that for any $\pi_t < 1$ and for any given $\underline{\pi} > 0$ there is a time interval τ long enough such that if between t and $t+\tau$ no pledge is observed, then $\pi_{t+\tau} = \underline{\pi}$. Thus, from (1.c) we can conclude that a long enough period without pledges will induce a backer to engage in abstention herding. The time without pledges required to induce abstention herding for a type u backer is shorter than the one required to induce a type g backer to engage in abstention herding. However, if enough time passes without a pledge both types will engage in abstention herding and an abstention information cascade occurs. Hence we have

Corollary 3.2.4. *In a regular equilibrium,*

1. *A long enough but finite period without pledges will induce backers to abstain from pledging.*
2. *Uninformed backers engage in abstention herding earlier than positively informed backers.*
3. *An information cascade occurs as soon as positively informed backer engage in abstention herding.*

Differently from abstention herding, pledge herding can only affect uninformed backers and never generate an information cascade. Recall that the definition of pledge herding requires that at $t = 0$ the backer would not invest. Now if at $t = 0$ an informed backer does not invest, then because pledges are non decreasing in belief, neither an uninformed backer would invest, but then an abstention information cascade starts at 0. Hence consider the case where at time 0 a positively informed backer would pledge, but not an uninformed backer. Then initially the only pledges can come from informed backers and by result (2.c) of proposition 3.2.2 the public belief π_t increase at each pledge. Result (1.b) implies that as soon as $\pi_t > \alpha^{-1}$, also an uninformed backer will start pledging. Thus, pledge herding. Note however that pledge herding cannot generate an information cascade. This for two reasons, first the size of the pledge is indicative of the fact that the backer is informative or not, and second, in the time between two pledges the public belief decreases. Hence we have

Corollary 3.2.5. *In a regular equilibrium,*

1. *Informed backers do not engage in pledge herding.*
2. *If π_0 is such that at time 0 only type g backers would pledge, then after enough pledges a type u backers engage in pledge herding.*
3. *Pledge herding cannot generate an information cascade.*

3.2.3 The Effect of Heterogenous Wealth and Multiple Signals

In the baseline model we have assumed that there are only three types of backers and that all backers have the same wealth. In this section we discuss why the predictions of the model would not qualitatively change if we relax these assumptions. For this purpose we focus on the unique equilibrium of the overfunding phase. First, suppose that bakers' wealth are i.i.d. on the interval $[0, 1]$ with density z, and that a backer's wealth is not correlated with the project's quality or the backer's private information. Second, suppose that backers receive conditionally i.i.d. private signals that are drawn from a "smooth" density f on the interval $[\underline{\theta}, \overline{\theta}]$. Without loss of generality we can order signals so that the monotone likelihood ratio property holds:

$$L(\theta) := \frac{f(\theta|\rho = \alpha)}{f(\theta|\rho = 0)} \text{ is increasing in } \theta$$

Thus $\theta < \theta'$ implies $\pi_t^{\theta} < \pi_t^{\theta'}$, where $\pi_t^{\theta} = \frac{\pi_t L(\theta)}{\pi_t L(\theta) + 1 - \pi_t}$. We further assume that $L(\underline{\theta}) = 0$ and $L(\overline{\theta})$ is bounded. This implies that for any $\pi_t \in (0, 1)$ we have that $\pi_t^{\underline{\theta}} = 0$ and $\pi_t^{\overline{\theta}} < 1$. Let's denote with $\theta^*(\pi) > 0$ the backers type θ such that if h_t leads to a public belief π then $\pi_t^{\theta} = \alpha^{-1}$. It is easy to verify that $\theta^*(\pi)$ satisfies $L(\theta^*(\pi)) = \frac{1-\pi}{\pi(\alpha-1)}$ and is decreasing in θ. Let $\underline{\pi} > 0$ be such that $\theta^*(\underline{\pi}) = \overline{\theta}$.

For any $x > 0$, π_t and $W \in \{0, 1\}$, let $\theta(x, W, \pi_t)$ be the θ such that $x = \max\left\{0, \frac{\pi_t^{\theta}\alpha - 1}{\alpha - 1} W\right\}$ and if no such θ exists set $\theta(x, W, \pi_t) > \overline{\theta}$.

Then we have:

Proposition 3.2.6. *During the overfunding phase:*

1. *Pledging behaviour satisfies the following properties:*

   (a) *A backer arriving at time t with signal θ and wealth W will pledge*

   $$\hat{\sigma}(\theta, h_t) = \max\left\{0, \frac{\pi_t^{\theta}\alpha - 1}{\alpha - 1} W\right\}$$

   (b) *No backer ever invests as soon as $\pi_t \leq \underline{\pi}$*

2. *The public belief evolves according to the following rules:*

   (a) *During the periods of absence of pledges*

   $$\frac{\partial \pi_t}{\partial t} = \frac{\pi_t(\lambda F(\theta^*(\pi_t)|\rho = \alpha) + 1 - \lambda)}{\lambda(\pi_t F(\theta^*(\pi_t)|\rho = \alpha) + (1 - \pi_t)F(\theta^*(\pi_t)|\rho = 0)) + 1 - \lambda} - \pi_t$$

that is strictly negative for $\pi_t > \underline{\pi}$ and nil for $\pi_t \leq \underline{\pi}$.

(b) The public belief $\pi_t(x)$ is

$$\pi_t(x) = \frac{\pi_t \int_0^1 z(W) f(\theta(x, W, \pi_t)|\rho = \alpha) dW}{\pi_t \int_0^1 z(W) f(\theta(x, W, \pi_t)|\rho = \alpha) dW + (1 - \pi_t)\pi_t \int_0^1 z(W) f(\theta(x, W, \pi_t)|\rho = 0) dW}$$

that is non-decreasing in x.

The proof of the Proposition is in Appendix 3.6.2.3.

3.2.4 Empirical Predictions

In this section we discuss some empirical implications of the theoretical model. Although it is possible to derive sharp predictions from the model assuming constant wealth and a dual signal structure, we focus on two empirical predictions that follow from the extended model.

Prediction 1 (The effect of past pledges): The amount pledged by an investor to a campaign is increasing in the amounts pledged by previous backers to that campaign (see Proposition 3.2.6, 1.a and 2.b), but decreases with the time elapsed since the last pledge occurs (see Proposition 3.2.6, 1.a and 2.a). Moreover, since new information is being incorporated in the public belief with each pledge, the positive correlation between the amount pledged among investors in a campaign should be larger the closer in time the investors are to each other in the sequence of investments.

We can contrast this prediction with two alternative models: First, a model in which each backer bases her pledge solely on her (conditionally i.i.d.) private signal. In that case, past pledges should have no effect on current pledges and the distribution of pledges should be time and past-history independent. Second, a model in which backers base their pledge on their private signal *and* the *total* accumulated investment in the campaign up to that moment. In that case pledges should be increasing in the cumulative amount invested up that point in time, but, conditional on this value, they should be independent of the specific path of the sequence of pledges.

Prediction 2 (The effect of absence of pledges): The probability of success of a campaign decreases with the time elapsed without pledges, specially at the start of the campaign. Because periods without pledges are more likely when agents have worst signals, and these are more common when the project is of bad quality, these periods negatively affect the public belief and hence the expected future pledge. Moreover, if a campaign doesn't have a good start, the public belief will tend to fall from the onset and an abstention information cascade is likely to occur. The success of a project, then, will be strongly dependent on the dynamics of investments early on.

As a corollary, the model predicts that the probability that a backer finds it optimal not to pledge a positive amount increases if the public belief declines (see Proposition 3.2.6, 1.a), so the probability of observing a pledge at any moment should be correlated with the factors that are affecting the public belief. In particular, the probability of observing a pledge increases with the amount pledged in previous periods (see Proposition 3.2.6, 2.b), and declines with the time elapsed between subsequent pledges (see Proposition 3.2.6, 2.a).

3.3 Data, Variables, and Some Descriptives

The data used for the analysis comes from the equity crowdfunding platform SEEDRS. The information was made available directly to us by SEEDRS and comprises the full universe of campaigns from October 2012 up until March 2016. In total, there are 710 campaigns, 22,615 unique backers and 69,699 pledges.[6] A campaign has 60 days to raise funds on the platform. If it does not reach its target all pledges are null and void. Entrepreneurs may accept pledges beyond the campaign target. All shares in a campaign are priced equally.

For each project, the raw data contains information about the date the campaign started raising funds, the declared investment target, the pre-money valuation of the company, the number of entrepreneurs, and the timing and value of each of the pledges received while the campaign was running. Each pledge is also matched to a specific investor associated with some descriptive investor data so we can analyse the behaviour of both individual campaigns and individual backers. Variable definitions are displayed in Table 3.1.

Descriptive statistics at the campaign level are provided in Table 3.2.[7] Out of the 710 campaigns for which information is available, 243 (34.2%) were successful in raising the declared investment goal. The average campaign goal was £174,216, but there is large heterogeneity in the amounts asked by individual projects, with values that range from £2,500 to more than £1,600,000. Moreover, the amount of capital sought in campaigns has grown considerably as SEEDRS has consolidated since their beginning in 2012. During the first year of operation, the average desired investment was close to £68,000, but this number was more than £200,000 during the last 12 months for which data is available. This

[6] These numbers correspond to the final sample after depurating the data. We started with 727 campaigns and 84,761 pledges. We dropped 12 campaigns that didn't have information on the valuation of the projects. More importantly, SEEDRS allows investors to cancel pledges before a campaign closes. There were 12,373 regretted investments in our sample, and we had 5 campaigns in which all investments are reported as cancelled. We dropped all pledges that were made but later cancelled. Although we do not have information on the time at which an investment is cancelled, the data management department at SEEDRS communicated that the majority of cancellations happen within several minutes after a pledge is made. All the estimations shown in the paper have been replicated including the cancelled pledges and the results remain qualitatively unchanged.

[7] Vulkan et al. (2016) analyze cross-campaign data from the same platform. For this paper we report updated figures for a longer data series.

desired investment corresponds to an average equity offered (in pre-money valuation terms) of 12 percent. The level of the investment target and pre-money valuations of the campaigns in SEEDRS present a sharp contrast with other non-equity crowdfunding schemes. For example, Mollick (2014), in a study of more than 48,500 projects raising funds in Kickstarter, shows that the average goal is less than $10,000, much lower than what is observed in our sample.

Investor profiles vary in their information content, but they include their geographic location, previous pledges in other campaigns, and, on some occasions, social media contacts or short biographic descriptions. About half of the backers making a pledge to a project decide not to allow their profiles to be seen by others. Most backers in a campaign (79%) are "authorized", meaning that they first have to pass a simple financial knowledge test in order to make pledges on SEEDRS. The rest are either "sophisticated" (7%) backers, or "high net worth" (14%) backers as defined in regulations made pursuant to the UK Financial Service and Market Act of 2000 (see Table 3.1 for the definitions). Approximately only 23 percent of investors in SEEDRS are recurrent, meaning that they have made pledges in more than one campaign, but those investors represent on average 73 percent of the pledges made to a campaign. The average sized pledge is £1,189. It is much smaller for authorized backers (£931), than for high net worth backers (£3,696), while sophisticated backers pledge about twice the average amount of authorized backers (£1,894), and recurrent investors pledge £897 on average, which is 3 times smaller than for one-time investors.

Some suggestive patterns appear in the descriptives. First, early pledges appear to be important in determining the chance that a campaign has of reaching the funding goal. Successful campaigns accumulate, on average, 21 percent of the total amount at the end of the first day, and this number increases to 75.5 percent after the first week. Failed campaigns, on the other hand, never really get started. Halfway through the time limit these projects have only covered about 15.7 percent of the total sought. Campaigns that fail to raise the desired capital tend to do so by a large margin, while most successful campaigns overfund, going up to an average among overfunded of 110 percent of the target.

Second, a few large pledges appear to have a major role in driving the success of a campaign. The largest pledge in an average campaign represent fully 15 percent of the total, and for the average successful campaign it accounts for about 31 percent of the total investment sought. As discussed in the previous section, large pledges can be important in two ways: by contributing to the accumulated capital stock of the campaign, and by directly influencing other backers into investing in a particular project.

3.4 Empirical Tests of the Theoretical Model

In this section we test if the model is consistent with the observed patterns in the data. We have detailed information on the timing in which decisions were made, and we can also reconstruct the available information to all backers in the platform at the moment of every investment. We use these features of the data to analyse if investors act as predicted by our model.

3.4.1 Empirical Tests Prediction 1: The Effect of Past Pledges

We start by providing scatter-plots of the relation between size and timing of a pledge and the size of subsequent investments. The first prediction of the model states that a pledge size should be *increasing* in the size of the previous pledges, but *decreasing* on the time since the previous pledge. Figure 3.1 shows suggestive evidence to support this prediction. To construct the figure we first organize all pledges in bins of size 5 log points according to the size of (Panel (a)), and time since (Panel (b)), the most recent pledge. We then compute the average amount pledged within each bin. The figure reports the scatter-plot and the correlation between the median point of the bin and the respective averages.

The figure shows that there is a positive correlation between the amounts pledged by subsequent backers, with the slope of the linear fit of the variables (in logs) estimated to be around 0.32. Also consistent with the model is the negative correlation between the time since the last pledge and amounts invested, where the slope of the linear fit (in logs) is estimated to be around -0.07.

In the model, the positive correlation between pledges decreases as new information arrives and the public belief evolves. Figure 3.2 shows supporting evidence for this prediction. The figure is constructed in a similar way as Panel (a) of Figure 3.1, but each panel correspond to a different lagged distance between the pledges. For example, Panel (a) replicates the results when we look at correlations between immediate pledges, Panel (b) looks at the correlation between $n-2$ and n-th pledge in a campaign, continuing all the way until the fifth lag. Consistent with the model, the size of the positive correlation between pledges declines as the pledges are further apart from each other. The slope of the linear fit of the variables (in logs) goes from 0.32 between immediate pledges to 0.16 between the $n-5$ and the nth pledge.

We now make the analysis more rigorous exploiting the richness of the data. Let $I_{n,c}$ be the amount pledged by the nth investor after the start of a campaign c. Let $T_{n-1,c}$ be the time (in hours) since the previous pledge. We use a distributed lag model of the form:

$$\log I_{n,c} = \beta_0 + \sum_{k=1}^{5} \beta_k \log I_{n-k,c} + \beta_6 \log T_{n-1,c} + \alpha W_{n,c} + \gamma Z_{n,c} + \eta_c + \epsilon_{n_c}, \tag{3.4.1}$$

where our interest lies in the estimates of the beta coefficients accompanying the values of the investment lags, $\log I_{n-k,c}$, and the time since previous investment, $T_{n-1,c}$.

The econometric model includes a full set of controls to account for differences in the characteristics of the backers, and for the information available to them when the decisions were made. In particular, $W_{n,c}$ is a vector of dummy variables indicating if the investor self-reported being high net worth, sophisticated, or authorized (authorized backers are used as the base). The vector also includes a dummy variable that takes the value of one if the investor is recurrent, and zero otherwise. $Z_{n,c}$ is a vector of campaign and time-varying variables capturing the information available to the investor when she makes the pledge. Among these variables are: the natural logarithm of the total amount funded in the campaign up that point; the square of the natural logarithm of the total amount funded in the campaign; the total number of bids in the campaign up to that point; the number of days since the start of the campaign; the SEEDRS' index between 0 and 100 of campaign "hotness" at the beginning of the day;[8] and an indicator that takes the value one if the campaign hotness index increased during the day, and zero otherwise. Finally, η_c is a campaign fixed effect capturing all the time-invariant observed and unobserved campaign characteristics, and $\epsilon_{n,c}$ is the error term.

[8]The hotness indicator is used by SEEDRS to decide the order in which the campaigns are presented on its landing page (see Table 3.1 for further details). Campaigns with a higher hotness index are listed earlier in presentation pages of all current ongoing campaigns.

Observing a positive (negative) correlation between the size (time) of subsequent pledges in a campaign does not imply the relation is causal. For example, backers might be responding to some specific characteristic of the investor making a pledge, and not necessarily to its size. If backer's characteristics are observed, we might be confounding the effect of observing a large pledge with that of observing that an investor with certain characteristics made that pledge. Moreover, it is also possible that the arrival of several large pledges is driven by common information shocks. An example of this would be some positive news about a campaign, which could simultaneously explain several large sized pledges arriving in a short interval of time.

Our identification strategy is designed to address these concerns. We do this in two ways: first, we can identify in our dataset if the investor making each pledge allowed its SEEDRS' profile to be visible to other members of the platform. The value of each pledge made to a project is displayed in one of the campaign's tabs for other members to see, but the profile of the investor associated to the pledge is only shown if she allows it to be public. Second, although other members of the platform are not able to see an investor's profile if she decides to remain anonymous, we can see that profile in our data. We can then use information from the unobserved profiles as instruments of the value of the pledges. Given that this profile information is only observable to us, the instrument is orthogonal to the potential response by subsequent backers.

We use two pieces of information as instruments for each investment lag ($\log I_{n-k,c}$): (1) the total number of pledges made by the investor in all campaigns interacted with the anonymous indicator; and (2) the largest single amount pledged by the investor in previous campaigns interacted with the anonymous indicator. Since recurrent investors tend to pledge smaller amounts than single-campaign investors, the first instrument is expected to have a negative correlation with size of a pledge. On the other hand, investors that have previously pledged large amounts are expected to continue doing so, so the second instrument is expected to have a positive correlation with pledge size. Finally, the time since the previous pledge ($\log T_{n-1,c}$) is instrumented with the (log) difference in hours between the hour in the day in which the previous pledge is made and twelve midday. We argue that if a pledge is made very early or late in the day, the time for a new investor to arrive is potentially longer because most of the decisions are made within traditional working day hours. This fact, on the other hand, is unrelated to the amount invested by subsequent backers.

Results from the first exercise are shown in Table 3.3. The table presents the estimates of Equation 3.4.1 for three specifications: in the first column we show the OLS estimates without the inclusion of the additional controls; in the second column we show the OLS estimates once controls for differences in the characteristics of the backers, and for the information available to them when the decisions were made, are included; the last column shows the Two Stage Least Squares (2SLS) estimates when we instrument each of the lagged pledges and time since the previous pledge as described above. We report the standard LM test for instrument relevance at the bottom of the table; given that we have two instruments for each investment lag, we also report Hansen's overidentification test for instrument validity. Instruments are found to be both relevant and valid. The reported standard errors are clustered by campaign.

We find that backers that immediately follow pledges of a larger value invest on average higher amounts in the campaigns. Moreover, we find that this relation is stronger the closer the pledges are to each other. For example, in our preferred IV specification in the last column, a doubling (100% increase) of the value of a prior pledge is associated with a rise of 12.3 percent in the subsequent amount pledged. Our findings also show that the magnitude of the effect dissipates rapidly: for the backer corresponding to the second lag, the size of the estimated coefficient declines to 4.6 percent for a similar change. For higher order lags we don't find any statistically significant effect in the IV specification.

The results also show that the amount of time since the last pledge occur is negatively correlated with the size of a pledge. In our preferred IV specification, doubling (100% increase) the time since the last pledge is associated to a fall in the pledge size of 9.2 percent.

Both descriptive and econometric evidence suggest that backers do respond to the size of previous investments, and to the time schedule in which they happened, in a way that is consistent with our theoretical model.

We now analyse whether the reactions are similar across all types of backers. Table 3.4 shows the results of estimating Equation 3.4.1 separately for four different potential types of backers: (i) sophisticated or high-net worth; (ii) authorized; (iii) recurrent; and (iv) single campaign investors.[9] Our baseline model predicts that since uninformed backers have no private information, their pledges follow more closely the evolution of the public belief. Positively informed backers weight their own private signals with the public belief, while negatively informed backers never pledge, so both are relatively less influenced by the past history of pledges. Although there is no direct mapping between the proposed division of investors and whether they are more or less informed about the quality of a campaign, we do expect a priori that both sophisticated/high net worth and recurrent investor will on average be more informed about the quality of investment opportunities.

The estimates shown in Table 3.4, all corresponding to our preferred IV specification, are in line with our expectations. All types of investor appear to react to the size of the previous pledge by investing a larger amount, but the

[9]Note that the first two and last two types are mutually exclusive, but not the four altogether.

magnitude of the effect is stronger for authorized (0.14% after a one percent increase in the previous pledge) and single-campaign investors (0.20% after a one percent increase in the previous pledge), than for sophisticated/high net worth (0.058% after a one percent increase in the previous pledge) and recurrent investors (0.091% after a one percent increase in the previous pledge). In terms of the time between subsequent pledges, we only find a statistically significant negative effect among authorized and recurrent investors.

Upon discussions with representatives of SEEDRS it came to our attention that some campaigns are run in two phases: the first being a "private" phase, and the second a "public" phase. In the private phase the entrepreneur is given a private *url* for the campaign which they can share privately with anyone. Entrepreneurs are urged to increase the chances that the public campaign will succeed by showing enough traction with private investments, using various methods for raising funds in the private phase from "the business's personal network of customers, followers, fans, friends and family."[10] Some common approaches used by entrepreneurs is to have a local launch party, and private meetings with select potential investors during the private phase. The private phase lasts between one to two weeks.

It is not clear what effects the private phase would have on our estimates of herding. It could be that potential investors more easily get to know each other and therefore coordinate their investments off-line during the private phase. It could also be that informed investors choosing to be anonymous when listed on the campaigns' web-site on SEEDRS are more likely to be known to other investors during the private phase. Finally, it could be that in the early stages of a campaign there is less public information, and investments from informed investor matter even more. All three mechanisms would suggest that the coefficients for

[10]See for example `https://www.seedrs.com/learn/help/what-is-private-launch`.

inter-temporal correlation from the private phase may be larger. This type of social influence might be just part and parcel of the mechanisms suggested by Banerjee (1992) and others and identified by for example Bursztyn et al (2014) to be driving herding in financial decision making.

It is simple for us to separately estimate coefficients for the private and public phases of the campaigns. We do this by separating data for the first two weeks and the rest of the campaign days. Results are provided in Table 3.5. As the table shows, the results are not remarkably different between the two phases of the campaign. If anything the magnitudes of coefficients are larger for the private phase.

In addition, our theory was precise to the overfunding phase, but there were potentially multiple equilibria for the underfunding phase. We can estimate whether empirically it makes any difference for the herding coefficients whether a campaign is in the underfunding phase or overfunding phase. We do so by separately analysing the data from campaigns in the overrfunding and the underfunding phase. Results are reported in Table 3.6. As the table shows, the relevant coefficients are the same in the two phases. The estimate of the coefficient capturing the effect of the time passed since the previous pledge is not statistically significant once we condition on being in the overfunding phase. Intuitively, this makes sense. In the early phase of a campaign investments are likely to convey more information than in the later stage when the campaign is wrapping up and most private information has already been transmitted.

3.4.2 Empirical Tests Prediction 2: The Effect of Absence of Pledges

In the model, the absence of pledges to a campaign is indicative that investors are not arriving with good private signals. If pledges don't arrive, the public belief about the project's quality will decline, and the chances of being successful in raising the funds fall. Since the quality of a project is unchanged during the campaign, the model predicts that bad projects will have a poor performance from the onset.

Figure 3.3 shows the average and median (across campaigns) number of backers (Panel(a) and (b)), and the average and median (across campaigns) cumulative amount invested (Panel(c) and (d)), for each day a campaign is active. We report two series: one for successful and one for unsuccessful campaigns. The figure shows that there is a clear difference between successful and unsuccessful campaigns in the support during the early stages of a campaign. On average, campaigns that end up raising the funds are able to attract both more backers and more capital during the first days. Moreover, as predicted by the theory, failed campaigns never get any traction.

The figure also shows that the number of backers arriving each day is larger during the first two weeks, which potentially reflects the differences between the private and public stages of a campaign. This two stage structure makes it difficult to interpret the evidence causally, since the herding effect is confounded with differences in overall quality of projects and pre-campaign marketing. It is encouraging, though, that after week 2, where most campaigns have already finished their private stage, the difference in terms of support between successful and unsuccessful campaigns remains. Moreover, the fact that arrival of investment each day remains fairly constant for both groups after week 2 is also supportive that, at least during the public stage, the assumption of an exogenous arrival rate where the signals depend on the underlying project's quality could hold.

We now formalize the graphical evidence in a regression framework. Table 3.7 reports the average marginal effect of a change in a set of measures of early campaign support on the probability that a campaign is successful. In particular, we are interested in how the probability of being successful changes with the number of backers and cumulative investments in the first week of a campaign. The table reports three specification: one without any additional controls; one controlling for predetermined characteristics of the campaign (pre-money valuation, desired investment, number of entrepreneurs, and access to tax incentives for investors); and one controlling for both predetermined characteristics of the campaigns and other variables describing campaign dynamics while it was active (share of recurrent, authorized and high net worth investors, total number of pledges, average amount pledged, and maximum amount invested by a single backer). The econometric model corresponds to a probit model, and standardized effects[11] are reported at the bottom of the table.

The econometric results are in line with the prediction that early campaign support is strongly correlated with the probability of success. For example, an increase of one standard deviation in the percentage of the desired investment covered during the first week is associated with a probability of success that is larger by 24 percentage points. Interestingly, the number of backers has a lower impact on the probability of success, and the effect becomes negative in the specification that includes all the controls. This result is suggestive that the quantity of investments early on is not sufficient to improve the chances of success, but that the "quality" of those initial investments matters.

[11]The standardized effect is calculated by multiplying each variable's standard deviation by the respective average marginal effect. Of course, this is only an approximation since the effects are non-linear by assumption, so they should be read with that caveat in mind.

The fact that we can only use campaign level variation to explore the determinants of the probability of success implies that we are unable to control for campaign specific characteristics. This is a clear limitation that impedes any causal interpretation of the results. We interpret the previous evidence with caution, and simply state that it is consistent with the predictions of theoretical model.

The last prediction derived from the theoretical model states that the probability of observing a pledge at any given point in time directly depends on the current public belief. For example, if the public belief is declining because pledges are absent or of a low value, the private signals have to be larger in order to induce a positive investment. But the arrival distribution of all type of backers to a campaign is fixed, so as the public belief falls the probability of observing any activity in the campaign also falls.

In order to test this prediction we need to change the structure of the data. Since we only observe a backer if she pledges a positive amount, we expand our dataset in such a way that the total duration of a campaign is divided into one hour bins, we then assign variables to the bins depending on whether there was any activity in the period or not. In particular, let $DI_{t,c}$ be a dichotomous indicator of activity in campaign c at the hourly bin t after the first investment. Let $I_{t,c}$ be the amount invested in campaign c at the hourly bin t, where the value is either zero if no investments were made, or the sum of all positive investments within the respective hour. We use a linear probability model that takes the form

$$DI_{t,c} = \beta_0 + \sum_{k=1}^{5} \beta_k \log I_{t-k,c} + \gamma Z_{t,c} + \eta_c + \nu_h + \epsilon_{t_c}, \tag{3.4.2}$$

where we are interested in the estimates of the beta coefficients capturing the amounts invested in those lagged periods.[12] From the model, we expect that some activity in previous hours, specially if it reflects large investments, should be associated with a higher probability of observing a pledge.

The model also includes the controls in $Z_{t,c}$, which is a vector of campaign and time-varying variables capturing how the campaign has evolved up to the period t. Among these variables are: the natural logarithm of the total amount funded in the campaign up that point; the square of the natural logarithm of the total amount funded in the campaign; the total number of bids in the campaign up to that point; and the number of days since the start of the campaign. Finally, η_c is a campaign fixed effect capturing all the time-invariant observed and unobserved campaign characteristics, ν_h is a hour of the day fixed effect capturing differences in frequency of arrivals during the day (e.g. working hours vs. non-working hours), and $\epsilon_{t,c}$ is the error term.

Results are shown in Table 3.8. The table presents the estimates of Equation 3.4.2 for four specifications: excluding and including the controls, and separating the private and the public stages of the campaigns. The results support the prediction that the probability of observing a pledge in a campaign at any point in time is positively correlated with how active the campaign was in previous periods. For example, the likelihood of observing a pledge at any given hour increases by 1.9 percentage points after a doubling of the total amount invested in the previous hour. Since the unconditional probability of observing a pledge at any given hour is around 5.5 percent, the magnitude of the effect is considerable. The results also show that this prediction holds for both the private and public stages of the campaign, but the estimated coefficients are of a larger magnitude

[12] Given the large number of observation in which the amount invested in the time period is zero, the log amount pledged is calculated using an inverse hyperbolic sine transformation, which can be interpreted in the same way as the standard logarithmic transformation.

during the private stage. This is consistent with the idea that the herding effect should be larger at the early stages of a campaign.

3.5 Conclusions

In this paper we provide a detailed study using micro-level data of herding on a major equity crowdfunding platform. Equity crowdfunding is an important and fast growing economic phenomenon. It has already had a significant impact on early stage funding in the UK, and is likely to become an important avenue for entrepreneurial finance in years to come in the U.S. as regulation for its provision was recently introduced.

Herding is likely common in all types of crowdfunding. It is what we expect in a situation with so much uncertainty: the decisions of the crowd provide some information in the absence of much else. We developed a model which captures what we believe to be the main information asymmetries in these investment platforms. The model is able to predict much of the dynamics of campaign funding based on the random arrival of investors with different private information about the projects.

We show that the amount pledged by an investor to a campaign is affected by the sizes of prior pledges. This is because a large investment signals to the public that the investor making the pledge potentially knows something about the project that others might not. This in turn may cause follow-on backers to alter their investment strategies, even though they don't actually observe the information of the backer making the large pledge. We find that a doubling of the size of a pledge is associated with a subsequent pledge that is larger in between 5.8 percent and 20.3 percent.

The model also predicts that the time elapsed between investments has a negative affect on the amounts pledged. This is because the absence of pledges is indicative that investors are not arriving to the campaign with sufficiently good private signals. We find that a doubling of the time between subsequent investments is associated to a pledge that is lower in 9.2 percent.

Consistent with other studies about campaign dynamics in crowdfunding, we show that the probability that a campaign has of being successfull depends largely on the support it gets at the early stages of fund-raising. The model also rationalises this observation by the effect that low or absent investments have on the public belief about the project. Lack of support to a campaign is indicative that only a few investors are arriving with positive signals. Having a bad start makes potential backers more pessimistic that the project is of good quality, so that they either pledge lower amounts or decide not to invest at all. In this context an abstention information cascade is likely to occur from the onset, and failed campaigns end up missing the mark by a large margin.

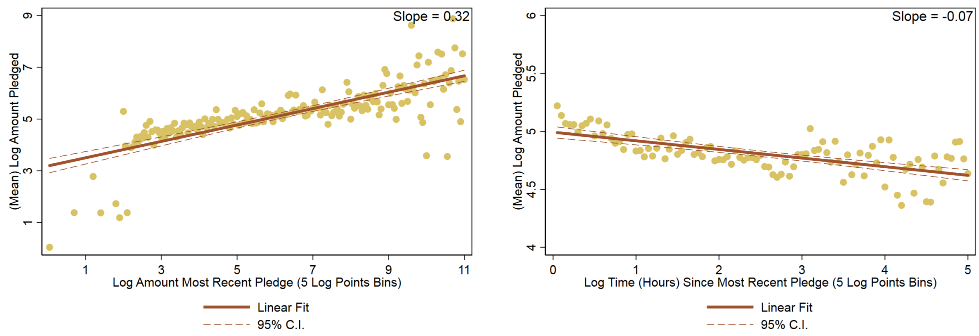
Equity crowdfunding is a new and important phenomena which is already having a large impact on early stage financing. It is part of what is sometimes known as the democratisation of finance. Here the crowds make their investment decisions directly and without the help (and without paying commission) of professional investors. The success of this movement depends largely on there being wisdom in the crowds. This paper provides a first step at understanding this important phenomena.

3.6 Appendix

3.6.1 Tables and Figures

Figure 3.1: Correlations Between the Amounts Pledged and the Time and Size of the Previous Pledge

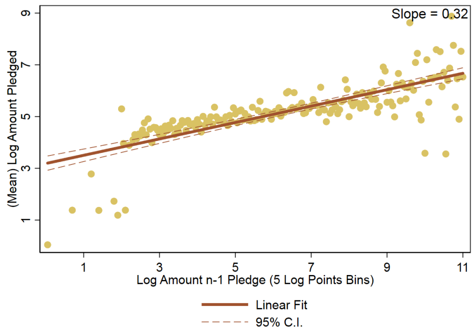
(a) Size of Previous Pledge (b) Time Since Previous Pledge



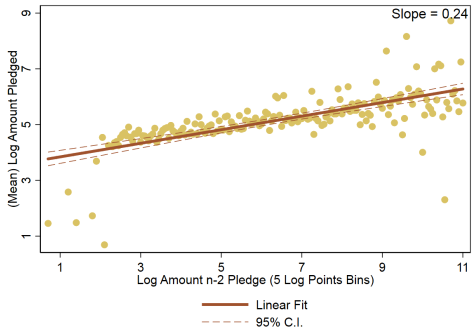
Notes: All pledges are organized in bins of size 5 log points according to the size of the previous pledge (Panel (a)), and the time elapsed since the previous pledge (Panel (b)). Each panel shows the relation between the median value of the respective bin and the average amount invested by the subsequent backers.

Figure 3.2: Correlations Between the Amounts Pledged by Following Backers in a Campaign

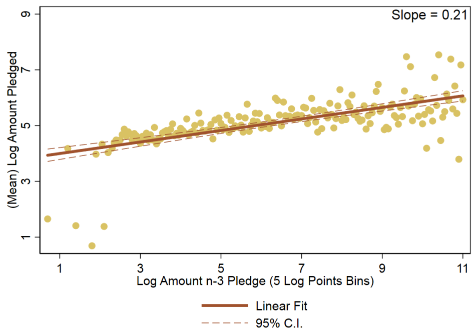
(a) $(n, n-1)$



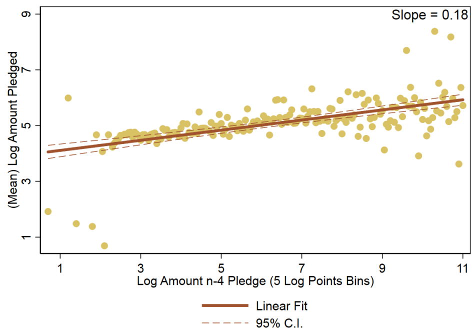
(b) $(n, n-2)$



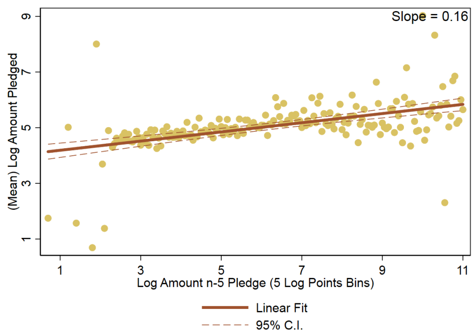
(c) $(n, n-3)$



(d) $(n, n-4)$



(e) $(n, n-5)$

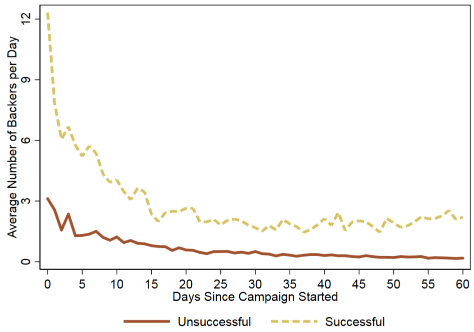


Notes: All pledges are organized in bins of size 5 log points according to the size of the previous $n-k$ pledge, where $k = \{1, 2, 3, 4, 5\}$. Each panel shows the relation between the median value of the respective bin and the average amount invested by the subsequent $n+k$ backer.

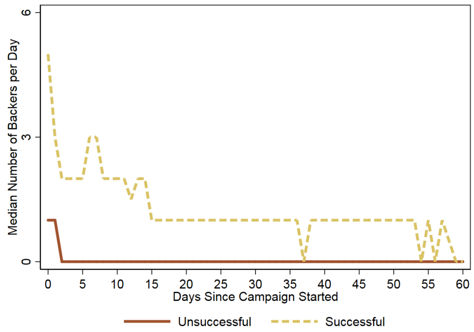
Figure 3.3: Evolution of Backers and Cumulative Investments to the Campaigns: Successful and Unsuccessful Campaigns

Average and Median Number of Backers per Day

(a) Mean

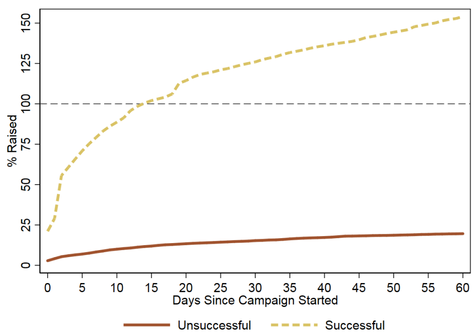


(b) Median

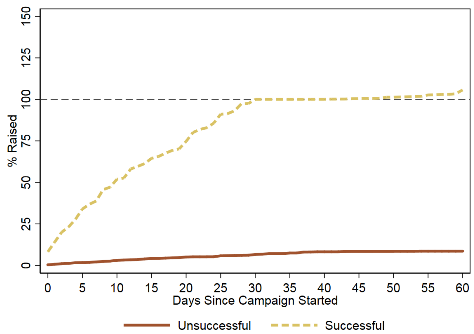


Number of Days to Reach a Given Percentage of the Investment Target

(c) Mean



(d) Median



Notes: Panel (a) and (b) depict the average and median number of backers making pledges to a campaign each day, conditional on whether they end up being successful or not. Panels (b) and (c) depicts the average and median number of days that a campaign needs to reach a given percentage of the overall desired investment, conditional on whether they end up being successful or not.

Table 3.1: Variable Description

Variable	Definition
Successful campaign	=1 if the campaign goal was met, zero otherwise. SEEDRS is an "all or nothing" platform in which projects have up to 60 days to raise investment, so companies only receive funding if they reach the declared investment goal within the time limit.
Pre-money valuation	Self-reported pre-money valuation of the project.
Equity offered	Percentage of equity that the campaign managers are offering.
Campaign goal	Declared desired investment by the campaign promoters.
SEIS tax relief	=1 if investors in the campaign have access to the Seed Enterprise Investment Scheme (SEIS) tax relief, zero otherwise. The SEIS Scheme encourages investment in qualifying new seed-stage startups companies by providing individuals with 50 percent of their investment back in income tax relief. Investors can also benefit from 50 percent capital gains tax relief on gains which are reinvested in SEIS eligible shares. Any gain arising on the disposal of the shares may also be exempt from capital gains tax, and loss relief is available if the disposal results in a loss.
EIS tax relief	=1 if investors in the campaign have access to the Enterprise Investment Scheme (EIS) tax relief, zero otherwise. The EIS scheme is designed to encourage investment in qualifying slightly later-stage companies than the SEIS by providing investors with up to 30% of their investment back in income tax relief. Investors can also defer any capital gains tax on gains which are reinvested in EIS eligible shares, gains arising on the disposal of the shares may be exempt from capital gains tax, and loss relief is available if the disposal results in a loss.
% Raised	Total amount raised by the campaign divided by the campaign goal. SEEDRS allows campaign promoters to accept more capital than what they had originally asked for, so they can "overfund" the projects once the target is reached. In cases in which there is overfunding, the variable takes a value that is greater than 100.
# Entrepreneur	Number of entrepreneurs in charge of the project.
# Backers	Number of different investors that have made pledges to the campaign.
# Pledges	Number of different pledges made to the campaign.
% Anonymous pledges	Investors can choose to share their SEEDRS' profile with other members of the platform. Each profile includes information about the investor location, the amount they have invested in different projects within the platform, campaigns in which they are promoters, and, occasionally, social media contacts or short biographic descriptions. Each pledge made is recorded in the campaign's page in order of magnitude, and investors are asked if they want their profiles to be seen next to the value of the investment. The variable is then constructed as the ratio between investments that are not public, that is, investments in which the backer profile is not available to the public, and total investments made in a given campaign.
Hotness indicator	Seedrs has an automatic algorithm to rank how much interest a campaign is generating at any given point in time. The algorithm measures four factors across the last three days: (i) amount invested; (ii) number of investors; (iii) investment traction; and (iv) days since the start of the campaign. The index takes values between [0,100], and is constructed using a weighted average of the four factors.
Intraday increase hotness indicator	=1 if hotness indicator increased during the day; =0 otherwise.
Authorized, High net worth and Sophisticated	Seedrs uses a classification scheme in which all individuals that subscribe to the platform have to self-select into one of three groups: high net worth, sophisticated, or authorized. High-net-worth corresponds to individuals who had annual incomes of at least £100,000 and/or held net assets to of at least £250,000 in the preceding financial year, as defined in regulations made pursuant to the UK Financial Services and Markets Act 2000. A sophisticated investor is an individual who has been an angel investor for at least the last six months, or for at least the last two years has made at least one investment in an unlisted company, has worked in private equity or corporate finance and/or has been a director of a company with an annual turnover of at least £1 million, as defined in regulations made pursuant to the UK Financial Services and Markets Act 2000. The rest of authorized individuals are those that do not fit in the previous categories, and need to fill out a questionnaire and score all questions correct in order to qualify as investors.
Recurrent investor	=1 if investor has made pledges in more than one campaign; =0 otherwise.
Mean Pledge	Average value in pounds of the pledges made to the campaign.
Median pledge	Median value in pounds of the pledges made to the campaign.
Max pledge	Maximum single pledge made in each campaign.
Max pledge / goal	Maximum single pledge made divided by campaign goal.
% Covered	The share of the campaign goal that was raised during a given period of time.
Mean time between pledges	Average time in hours between adjacent pledges in a campaign.

Table 3.2: Summary Statistics

	All	Successful (34.2%)	Unsuccessful	Difference
Campaigns				
Pre-money valuation (£)	1,845,449 (5,028,629)	2,793,642 (7,834,238)	1,352,064 (2,426,423)	1,441,578***
Equity offered	11.95 (7.66)	8.82 (6.43)	13.57 (7.75)	-4.76***
Campaign goal (£)	174,216 (327,598)	176,630 (252,718)	172,959 (360,711)	3,670
% EIS tax relief	34.51 (47.57)	46.50 (49.98)	28.27 (45.08)	18.24***
% SEIS tax relief	57.18 (49.52)	45.68 (49.92)	63.17 (48.29)	-17.49***
% Raised	76.37 (195.13)	178.95 (306.45)	22.99 (28.57)	155.96***
# Entrepreneurs	3.28 (1.96)	3.74 (2.07)	3.04 (1.87)	0.70***
# Backers	83.44 (126.38)	169.46 (174.62)	38.68 (50.99)	130.78***
# Pledges	96.48 (146.15)	199.03 (200.68)	43.13 (56.99)	155.90***
% Anonymous pledges	51.40 (18.60)	54.22 (9.50)	49.93 (21.75)	4.29***
Hotness indicator (start of day)	11.13 (13.08)	21.39 (14.80)	5.79 (7.95)	15.60***
Intraday increase in hotness indicator	0.76 (0.24)	0.82 (0.17)	0.74 (0.27)	0.08**
# Days the campaign is active	54.70 (38.46)	58.19 (38.56)	52.89 (38.32)	5.30*
Type of Investor				
% Authorized	79.21 (15.57)	76.50 (11.45)	80.63 (17.17)	-4.13***
% High-net-worth	13.55 (12.73)	14.82 (10.31)	12.89 (13.79)	1.92*
% Sophisticated	7.23 (8.24)	8.68 (4.91)	6.48 (9.44)	2.20***
% Recurrent investors	72.64 (27.54)	79.16 (19.99)	69.25 (30.22)	9.91***
Investments				
Mean pledge (£)	1,202.71 (2,949.49)	1,745.98 (3,472.27)	920.03 (2,596.30)	825.95***
Median pledge (£)	354.32 (2,336.11)	571.29 (3,146.97)	241.42 (1,767.18)	329.87*
Max pledge (£)	38,201.94 (158,251.56)	81,341.40 (259,065.82)	15,754.64 (42,113.87)	65,586.76***
Max pledge / goal	0.15 (0.19)	0.31 (0.22)	0.07 (0.10)	0.23***
Timing				
Mean time between pledges (hours)	56.57 (109.83)	9.96 (8.98)	82.04 (129.56)	-72.08***
% Covered in day 1	9.07 (20.94)	21.07 (30.77)	2.83 (7.84)	18.24***
% Covered in week 1	30.79 (165.41)	75.47 (276.97)	7.55 (14.45)	67.92***
% Covered in month 1	53.33 (187.77)	126.66 (306.94)	15.17 (21.33)	111.49***
Observations	710	243	467	

Notes: Each cell is computed by taking the average across the campaigns. The mean time between pledges corresponds to the average across all pledges. Standard deviation in parenthesis. The last column reports the difference of means between successful and unsuccessful campaigns for each variable, and the result from a mean comparison test at standard levels of statistical significance: *** 1 percent ** 5 percent * 10 percent.

Table 3.3: Prediction 1: The Effect of Past Pledges

	Dependent Var: log amount pledged (£)		
	Model	Model + Controls	IV
Prior pledges			
Log amount pledged (n-1)	0.097***	0.087***	0.123***
	(0.004)	(0.006)	(0.019)
Log amount pledged (n-2)	0.041***	0.036***	0.046**
	(0.004)	(0.004)	(0.019)
Log amount pledged (n-3)	0.030***	0.027***	0.016
	(0.004)	(0.004)	(0.017)
Log amount pledged (n-4)	0.021***	0.017***	-0.006
	(0.004)	(0.004)	(0.018)
Log amount pledged (n-5)	0.018***	0.016***	0.010
	(0.004)	(0.004)	(0.016)
Log time (hours) since most recent pledge	-0.028**	0.019	-0.092**
	(0.014)	(0.017)	(0.035)
Cummulative amount			
Log amount funded		0.262**	0.364***
		(0.087)	(0.092)
Log amount funded squared		-0.015***	-0.020***
		(0.005)	(0.005)
Investor type			
Dummy high-net-worth		1.166***	1.163***
		(0.038)	(0.038)
Dummy sophisticated		0.421***	0.411***
		(0.034)	(0.035)
Dummy recurrent investor		-0.576***	-0.562***
		(0.047)	(0.049)
Campaign history			
Campaign hotness at start of the day		0.003***	-0.000
		(0.001)	(0.001)
Dummy campaign hotness intraday rise		0.173***	0.107***
		(0.020)	(0.027)
Total pledges (/100)		0.006	0.016*
		(0.009)	(0.009)
Days from start of campaign		0.004***	0.005***
		(0.001)	(0.001)
Observations	66,347	66,347	61,426
Average pledge (£)	1,247	1,247	1,244
SD pledge (£)	12,151	12,151	11,859
Average time (hours) since most recent pledge	11.2	11.2	11.4
S.D. time (hours) since most recent pledge	40.4	40.4	40.2
Kleibergen and Paap rk statistic			207.24
Hansen J statistic P-Val			0.55
Campaign FE	Yes	Yes	Yes

*** 1 percent ** 5 percent * 10 percent

Notes: Robust standard errors, clustered by campaign. Each lagged pledge in the IV setting has two instruments: (i) total number of pledges made by the investor in all campaigns interacted with the anonymous indicator; and (ii) the largest single amount pledged by the investor in previous campaigns interacted with the anonymous indicator. The time since the previous pledge is instrumented with the (log) difference in hours between the hour in the day in which the previous pledge is made and twelve midday.

Table 3.4: Prediction 1. The Effect of Past Pledges: Heterogeneous Effects by Investor Type

	Dependent Var: log amount pledged (£)				
	All	Sophisticated/High Net Worth	Authorized	Recurrent	Single Campaign Investor
Prior pledges					
Log amount pledged (n-1)	0.123***	0.058	0.140***	0.091***	0.203**
	(0.019)	(0.037)	(0.026)	(0.018)	(0.068)
Log amount pledged (n-2)	0.046**	0.097**	0.031	0.050**	0.032
	(0.019)	(0.036)	(0.021)	(0.019)	(0.050)
Log amount pledged (n-3)	0.016	0.037	0.012	0.014	0.016
	(0.017)	(0.046)	(0.018)	(0.019)	(0.065)
Log amount pledged (n-4)	-0.006	0.012	-0.015	-0.008	0.011
	(0.018)	(0.044)	(0.019)	(0.019)	(0.052)
Log amount pledged (n-5)	0.010	-0.043	0.031*	0.024	-0.105
	(0.016)	(0.038)	(0.018)	(0.016)	(0.074)
Log time (hours) since most recent pledge	-0.092**	-0.102	-0.088**	-0.070**	0.020
	(0.035)	(0.070)	(0.041)	(0.035)	(0.088)
Observations	61,426	13,236	48,132	45,477	15,888
Average pledge (£)	1,244	2,610	867	814	2,470
SD pledge (£)	11,859	12,564	11,629	10,741	14,515
Average time (hours) since most recent pledge	11.4	11.1	11.5	12.0	9.7
S.D. time (hours) since most recent pledge	40.2	42.5	39.5	40.1	40.3
Kleibergen and Paap rk statistic	207.24	126.12	195.78	216.53	74.99
Hansen J statistic P-Val	0.55	0.49	0.47	0.37	0.00
Campaign FE	Yes	Yes	Yes	Yes	Yes

*** 1 percent ** 5 percent * 10 percent

Notes: Robust standard errors, clustered by campaign. All the controls from Table 3.3 are included but not reported. Each lagged pledge has two instruments: (i) total number of pledges made by the investor in all campaigns interacted with the anonymous indicator; and (ii) the largest single amount pledged by the investor in previous campaigns interacted with the anonymous indicator. The time since the previous pledge is instrumented with the (log) difference in hours between the hour in the day in which the previous pledge is made and twelve midday. See Table 3.1 for the definitions used to classify investors.

Table 3.5: Prediction 1. The Effect of Past Pledges: Private and Public Stages of a Campaign

	Dependent Var: log amount pledged (£)		
	All	Weeks 1-2	Weeks > 2
Prior pledges			
Log amount pledged (n-1)	0.123***	0.133***	0.106***
	(0.019)	(0.039)	(0.022)
Log amount pledged (n-2)	0.046**	0.039	0.039*
	(0.019)	(0.029)	(0.021)
Log amount pledged (n-3)	0.016	0.030	-0.002
	(0.017)	(0.026)	(0.024)
Log amount pledged (n-4)	-0.006	-0.018	-0.002
	(0.018)	(0.026)	(0.026)
Log amount pledged (n-5)	0.010	0.010	0.008
	(0.016)	(0.023)	(0.022)
Log time (hours) since most recent pledge	-0.092**	-0.107**	-0.050
	(0.035)	(0.045)	(0.046)
Observations	61,426	28,593	32,796
Average pledge (£)	1,244	1,255	1,234
SD pledge (£)	11,859	14,428	9,041
Average time (hours) since most recent pledge	11.4	4.3	17.6
S.D. time (hours) since most recent pledge	40.2	11.6	53.1
Kleibergen and Paap rk statistic	207.24	179.66	143.96
Hansen J statistic P-Val	0.55	0.70	0.45
Campaign FE	Yes	Yes	Yes

*** 1 percent ** 5 percent * 10 percent

Notes: Robust standard errors, clustered by campaign. All the controls from Table 3.3 are included but not reported. Each lagged pledge has two instruments: (i) total number of pledges made by the investor in all campaigns interacted with the anonymous indicator; and (ii) the largest single amount pledged by the investor in previous campaigns interacted with the anonymous indicator. The time since the previous pledge is instrumented with the (log) difference in hours between the hour in the day in which the previous pledge is made and twelve midday.

Table 3.6: Prediction 1. The Effect of Past Pledges, Underfunding and Overfunding Stages of a Campaign

	Dependent Var: log amount pledged (£)		
	All	Underfunding	Overfunding
Prior pledges			
Log amount pledged (n-1)	0.123***	0.118***	0.127***
	(0.019)	(0.026)	(0.035)
Log amount pledged (n-2)	0.046**	0.033	0.060*
	(0.019)	(0.021)	(0.036)
Log amount pledged (n-3)	0.016	0.026	-0.005
	(0.017)	(0.022)	(0.025)
Log amount pledged (n-4)	-0.006	0.004	-0.034
	(0.018)	(0.018)	(0.046)
Log amount pledged (n-5)	0.010	0.010	0.009
	(0.016)	(0.019)	(0.025)
Log time (hours) since most recent pledge	-0.092**	-0.126**	0.021
	(0.035)	(0.039)	(0.079)
Observations	61,426	45,385	16,026
Average pledge (£)	1,244	1,128	1,571
SD pledge (£)	11,859	7,144	19,851
Average time (hours) since most recent pledge	11.4	13.0	6.8
S.D. time (hours) since most recent pledge	40.2	43.7	27.1
Kleibergen and Paap rk statistic	207.24	193.56	66.42
Hansen J statistic P-Val	0.55	0.78	0.43
Campaign FE	Yes	Yes	Yes

*** 1 percent ** 5 percent * 10 percent

Notes: Robust standard errors, clustered by campaign. All the controls from Table 3.3 are included but not reported. A campaign is said to be in the overfunding phase if it has already raised the target amount, but has not reached the time limit. Each lagged pledge has two instruments: (i) total number of pledges made by the investor in all campaigns interacted with the anonymous indicator; and (ii) the largest single amount pledged by the investor in previous campaigns interacted with the anonymous indicator. The time since the previous pledge is instrumented with the (log) difference in hours between the hour in the day in which the previous pledge is made and twelve midday.

Table 3.7: Prediction 2: Variables Associated with the Probability that a Campaign is Successful

	Average Marginal Effects after Probit		
	I	II	III
Early Campaign Dynamics			
Number of backers in week 1	0.000	0.000	-0.001***
	(0.001)	(0.001)	(0.000)
% Covered in week 1	0.008***	0.007***	0.001***
	(0.001)	(0.001)	(0.000)
Predetermined Campaign Controls			
Log pre-money valuation (£)		0.085***	0.016
		(0.022)	(0.015)
Log campaign goal (£)		-0.093***	-0.219***
		(0.024)	(0.017)
# Entrepreneurs		0.015*	0.004
		(0.008)	(0.005)
% EIS tax relief		0.000	-0.000
		(0.000)	(0.000)
% SEIS tax relief		0.000	0.000
		(0.000)	(0.000)
Campaign Variable Controls			
% Recurrent investors			0.001
			(0.001)
% Authorized			-0.000
			(0.001)
% High-net-worth			-0.000
			(0.002)
Total # Pledges			0.002***
			(0.000)
Log mean pledge (£)			0.113***
			(0.023)
Log max pledge (£)			0.059***
			(0.014)
Observations	710	710	710
Standardized Effect			
Number of backers in week 1	0.01	0.01	-0.04
% Covered in week 1	1.31	1.14	0.24
Log pre-money valuation		0.09	0.02
Log campaign goal		-0.10	-0.24
# Entrepreneurs		0.03	0.01
% EIS tax relief		0.02	-0.00
% SEIS tax relief		0.00	0.01
% Recurrent investors			0.02
% Authorized			-0.00
% High-net-worth			-0.00
Total # Pledges			0.28
Log mean pledge			0.16
Log max pledge			0.14

*** 1 percent ** 5 percent * 10 percent

Notes: Standard errors calculated using the delta-method. The standardized effect is calculated by multiplying each variable's standard deviation by the respective average marginal effect.

Table 3.8: Prediction 2: Probability of Observing a Pledge at Any Given Hour

	Dependent Var: Dummy Investment in the Period (Hour)			
	Full Sample	Full Sample	Weeks 1-2	Weeks > 2
Prior pledges				
Log amount pledged (t-1)	0.019***	0.019***	0.021***	0.015***
	(0.001)	(0.001)	(0.001)	(0.001)
Log amount pledged (t-2)	0.015***	0.014***	0.015***	0.012***
	(0.000)	(0.000)	(0.001)	(0.001)
Log amount pledged (t-3)	0.011***	0.011***	0.011***	0.010***
	(0.000)	(0.000)	(0.001)	(0.001)
Log amount pledged (t-4)	0.010***	0.009***	0.008***	0.008***
	(0.000)	(0.000)	(0.001)	(0.000)
Log amount pledged (t-5)	0.008***	0.008***	0.007***	0.006***
	(0.000)	(0.000)	(0.001)	(0.001)
Campaign history				
Log amount funded		-0.003	0.005	-0.006
		(0.005)	(0.005)	(0.012)
Log amount funded squared		0.001*	0.001**	0.001
		(0.000)	(0.000)	(0.001)
Total pledges (/100)		-0.011**	-0.006	-0.000
		(0.005)	(0.006)	(0.006)
Days from start of campaign		-0.000***	-0.004***	-0.000***
		(0.000)	(0.000)	(0.000)
Observations	866,912	866,912	209,363	657,549
R2	0.067	0.069	0.093	0.045
Frequency of Investments per Hour	0.055	0.055	0.088	0.044
SD of Frequency of Investments per Hour	0.227	0.227	0.284	0.205
Campaign FE	Yes	Yes	Yes	Yes
Hour of Day FE	Yes	Yes	Yes	Yes

*** 1 percent ** 5 percent * 10 percent

Notes: Robust standard errors, clustered by campaign. The total time for which a campaign is running is divided into bins of length one hour. The dataset is then organized as a panel in which the time dimension corresponds to the hours passed since the start of the campaign. Log amount funded corresponds to the total amount invested in the campaign up to previous hour; total bids correspond to total number of investments in the camapign up to the previous hour. Given the large number of observation in which the amount invested in the time period is zero, the log amount pledged is calculated using an inverse hyperbolic sine transformation, which can be interpreted in the same way as the standard logarithmic transformation.

3.6.2 Proofs

3.6.2.1 Proof of Proposition 3.2.2

(1.a) First observe that because $\pi_t^b = 0$ (see expression (3.2.2)), not pledging is a weakly dominant strategy for a type b backer, and it is the strictly dominant strategy if the campaign succeeds with positive probability.

(1.b) Note that (1.a) implies that a campaign may succeed only if uninformed and/or positively informed backers arrive. Let's denote with s a history of arrival times of the uninformed and positively informed backers, but not of the negatively uninformed backers, that is $s := \{(t_1, s_1), \ldots, (t_n, s_n), \ldots\}$, where for all $n > 0$, $0 \leq t_n < t_{n+1} \leq T$ and $s_n \in \{g, u\}$. Let $\mathcal{S}$ denote the set of all possible such histories. Because arrival time does not depend on the project quality, and $q \in (0, 1)$, for any subset $S \subseteq \mathcal{S}$ we have that

$$\begin{aligned} \mathbb{P}(S|\text{bad project}) \in (0,1) \quad &\Leftrightarrow \quad \mathbb{P}(S|\text{good project}) \in (0,1) \Rightarrow \\ \mathbb{P}(S|\text{bad project}) \quad &\neq \quad \mathbb{P}(S|\text{good project}) \end{aligned} \tag{3.6.1}$$

In particular the following property applies:

Property 1: *There is a positive and bounded M such that for any $S \subseteq \mathcal{S}$ such that $\mathbb{P}(S|\text{good project}) > 0$ one has:*

$$0 < \frac{\mathbb{P}(S|\text{good project})}{\mathbb{P}(S|\text{bad project})} < M \tag{3.6.2}$$

Now consider a history of pledges h_t completed with a pledge x_t and let $\mathcal{X}(h_t, x_t) \subset \mathcal{H}$ denote the set of histories that start with h_t, x_t, lead the campaign to succeed, and are compatible with the equilibrium strategies. To any history $h \in \mathcal{X}(h_t, x_t)$ corresponds a history $s(h) \in \mathcal{S}$ that leads to the observation of h. Then the probability that the campaign succeeds conditional on h_t, x_t and the project's quality ρ can be written as

$$S_\rho(x, h_t) = \mathbb{P}(y_T \geq Y|h_t, x_t = x, \rho) = \mathbb{P}(\cup_{h \in \mathcal{X}(h_{t-1}, x_t = x)} s(h)|h_{t-1}, x_t = x, \rho),$$

that in a regular equilibrium is a non-decreasing function of x.

We can now prove that backer's pledges are increasing in their belief. Namely for a baker arriving with belief π the optimal pledges solves

$$x \in \arg\max \pi A(x) + B(x)$$

where we defined

$$\begin{aligned} A(x) &:= S_\alpha(x,h_t)(\ln(W+(\alpha-1)x)-\ln(W)) + S_0(x,h_t)(\ln(W)-\ln(W-x)) \\ B(x) &:= S_0(x,h_t)(\ln(W-x)-\ln(W)) \end{aligned}$$

Observe that $\mathbb{P}(y_T \geq Y|h_t, x_t) > 0$ and (3.6.1) imply $S_\rho(x,h_t) > 0$ for $\rho \in \{0,\alpha\}$. Thus, because the equilibrium is regular, $A(x)$ is a strictly increasing function. Now take two backers, one with belief π and the the other with belief $\pi' > \pi$ and let x and x' be their respective optimal pledges. We want to show $x' \geq x$. Observe that x and x' must satisfy

$$\begin{aligned} \pi A(x) + B(x) &\geq \pi A(x') + B(x') \\ \pi' A(x') + B(x') &\geq \pi' A(x) + B(x). \end{aligned}$$

Summing-up these two inequality and rearranging we get $(\pi'-\pi)(A(x')-A(x)) \geq 0$, thus, $\pi' > \pi$ and the monotonicity of $A(.)$ imply $x' \geq x$.

Because $\pi_t = \pi_t^u < \pi_t^g$ it immediately follows that $\hat{\sigma}(u,W,h_t) \leq \hat{\sigma}(g,W,h_t)$. Because there are more positively informed backers when the project is good it immediately follows that the probability that the campaign succeeds is not smaller for a good project than for a bad project:

$$S_0(x, h_t) \leq S_\alpha(x, h_t) \tag{3.6.3}$$

Also note that x affects S_ρ either by triggering the success, if $x > Y - y_t$, or by affecting the behavior of future uninformed and positively informed backers. Now, the probability of uninformed bakers arrival does not depend on the project quality but positively informed bakers are more likely to arrive if the project is good. Thus, we have that if $x' > x$, then $S_\alpha(x', h_t) - S_\alpha(x, h_t) \geq S_0(x', h_t) - S_0(x, h_t) \geq 0$, that is the probability of success is more sensitive to current pledges if the project is good than if the project is bad. If $S_\rho(x, h_t)$ is piecewise differentiable, then $S'_\alpha(x, h_t) > S'_0(x, h_t) \geq 0$.

To show that for $\pi_t > \alpha^{-1}$ a backer invests at least $x^*(\pi_t) := \frac{\pi_t \alpha - 1}{\alpha - 1} > 0$, let focus on the case where $S\rho(x, h_t)$ is piecewise differentiable in x. Then the first order condition for an uninformed backer is

$$\begin{aligned}
\pi_t S_\alpha(x, h_t) U'(W + (\alpha - 1)x)(\alpha - 1) \quad &- \quad (1 - \pi_t) S_0(x, h_t) U'(W - x) + \\
\pi_t S'_\alpha(x, h_t)(U(W + (\alpha - 1)x) - U(W)) \quad &+ \quad (1 - \pi_t) S'_0(x, h_t)(U(W - x) - U(W)) = 0
\end{aligned}$$

It is sufficient to show that the above expression is non-negative for $x = x^*(\pi_t)$. Observe that one has $\pi_t U'(W + (\alpha - 1)x^*(\pi_t))(\alpha - 1) - (1 - \pi_t) U'(W - x^*(\pi_t)) = 0$ and $U^* := \pi_t (U(W + (\alpha - 1)x^*(\pi_t)) - U(W)) + (1 - \pi_t)(U(W - x^*(\pi_t)) - U(W)) >$ 0,for $\pi_t > \alpha^{-1}$. Thus evaluating the f.o.c. at $x = x^*(\pi_t)$ one obtains

$$\begin{aligned}
(S_\alpha(x, h_t) - S_0(x, h_t))(1 - \pi_t) S_0(x, h_t) U'(W - x) \quad &+ \\
\pi_t S'_\alpha(x, h_t)(U(W + (\alpha - 1)x) - U(W)) + (1 - \pi_t) S'_0(x, h_t)(U(W - x) - u(W)) \quad &\geq \\
(S_\alpha(x, h_t) - S_0(x, h_t))(1 - \pi_t) S_0(x, h_t) U'(W - x) + S'_0(x, h_t) U^* \quad &> \quad 0.
\end{aligned}$$

(1.c) Consider a type θ backer arriving at time t and pledging x and let consider the expected net cash flow of the project conditional on the campaign succeeding. This is equal to

$$ECF_t(x) := \frac{\pi_t^\theta S_\alpha(x,h_t)}{\pi_t^\theta S_\alpha(x,h_t) + (1-\pi_t^\theta)S_0(x,h_t)}\alpha - 1$$

If $ECF_t(x) \leq 0$ for all $x > 0$, then a risk averse backer will strictly prefer abstention to investing. Equation (3.6.1) implies that $S_\alpha(x,h_t) > 0$ if and only if $S_0(x,h_t) > 0$. If for all $x < Y - y_t$ one has $S_\alpha(x,h_t) = 0$ then the campaign fails with certainty unless the backer triggers success by pledging at least $Y - y_t$. If π_t is such that $\pi_t/(\pi_t + (1-\pi_t)(1-q)) < \alpha$, then $ECF_t(x)$ is negative even to a positively informed backer and hence not investing is an optimal strategy for all types of backers. For this case, the statement is satisfied by setting $\underline{\pi} > 0$ such that $\underline{\pi}/(\underline{\pi} + (1-\underline{\pi})(1-q) = \alpha^{-1}$. Now, suppose that $S_\alpha(x,h_t) > 0$ for some $x < Y - y_t$ and take any of such x. Observe that $ECF_t(x)$ is an increasing function of the likelihood ratio $\frac{S_\alpha(x,h_t)}{S_0(x,h_t)}$ that is strictly positive a bounded by M because of (3.6.1) and (3.6.2). But this implies that that

$$ECF_t(x) \leq \frac{\pi_t^\theta M}{\pi_t^\theta M + 1 - \pi_t^\theta}\alpha - 1$$

Let $\underline{\underline{\pi}} > 0$ be such that the r.h.s of the above expression is nil for $\pi_t^\theta = \underline{\underline{\pi}}$ and let $\underline{\pi} > 0$ be such that $\underline{\pi}/(\underline{\pi} + (1-\underline{\pi})(1-q) = \underline{\underline{\pi}}$. Then for $\pi_t < \underline{\pi}$, one has that $ECF_t(x) < 0$ for all θ, that is, backer t will strictly prefer not to pledge no matter what his type is.

(2.a) If $\pi_t \leq \underline{\pi}$, then in equilibrium no backer pledges. In this case absence of pledges provides no additional information about the project quality implying $\pi_{t'} = \pi_t$.

(2.b) Consider now the case π_t is such that no abstention cascade has started. Because pledges are increasing in belief, a positively informed backer would invest some amount. If no investment was observed between t and t', this simply means that no positively informed backer arrived between t and t'. This event occurs with probability $e^{-\lambda(t'-t)}$ if the project is good and with probability $e^{-\lambda(1-q)(t'-t)} > e^{-\lambda(t'-t)}$ if the project is bad. The expression $\phi(\pi, z)$ follows from Bayes' rule.

(2.c) Suppose that just before a pledge of x_t arrives at time t the public belief is π_t. This pledge either arrives from an uninformed backer and in this case it has no information content, or it arrives from a positively informed baker whose information is endowed with a positive signal. Let define

$$L(x_t, h_t) := \frac{\mathbb{P}(\text{type } g \text{ backer pledges } x_t|h_t)}{\mathbb{P}(\text{type } u \text{ baker pledges } x_t|h_t)}$$

This likelihood ratio is non decreasing in x_t because pledges are non-decreasing in belief. Then we have:

$$\pi_t(x_t) = \frac{\pi_t\mathbb{P}(x_t|h_t, \rho = \alpha)}{\pi_t\mathbb{P}(x_t|h_t, \rho = \alpha) + (1 - \pi_t)\mathbb{P}(x_t|h_t, \rho = 0)} = \frac{\pi_t(\lambda q L(x_t, h_t) + 1 - \lambda)}{\lambda L(x_t, h_t)(\pi_t q + (1 - \pi_t)(1 - q)) + 1 - \lambda}$$

That is increasing in $L(.)$, thus the result. Q.E.D.

3.6.2.2 Proof of Proposition 3.2.3

Observe that during the overfunding phase $S_\rho(x, h_t) = 1$ for all x and all $\rho\{0, \alpha\}$. Substituting this expression in (3.2.3), we can take first order conditions with respect to x which provides the expression for $\hat{\sigma}(\theta, h_t)$. Observe that as soon as $\pi_t < \underline{\pi}$, one has that $\hat{\sigma}(g, h_t) = 0$. Because of the monotonicity of pledges with respect to beliefs, this implies that when $\pi_t < \underline{\pi}$ no backer ever invests and hence an abstention cascade occurs. To sustain this equilibrium it is sufficient to focus on off-path belief guaranteeing that a positive pledge during an abstention cascade does not induce too large off-path posterior beliefs. Q.E.D.

3.6.2.3 Proof of Proposition 3.2.6

1.a The proof is identical to the proof of Proposition 3.2.3

1.b Observe that $\underline{\pi}$ is defined such that given the pledging strategy in (1.a) the most optimistic backer, i.e. type $\overline{\theta}$, would not invest if $\pi_t \leq \underline{\pi}$. The the result follows from the monotonicity of backer pledging in their beliefs.

2.a Let consider the instantaneous probability of observing no pledges between t and $t+dt$. This corresponds to the chance of no informed backer arriving, $1-\lambda$, plus the chance of one informed backer arriving , λ, times the probability that the informed backer does not pledge. Given (1.a), a backer does not pledge only if $\pi_t^\theta < \alpha^{-1}$, that is equivalent to $\theta < \theta^*(\pi_t)$. The probability that $\theta < \theta^*(\pi_t)$ given ρ, is $F(\theta^*(\pi_t)|\rho)$. Applying Bayes' rule one has the first term on the r.h.s of the expression that is π_t because $F(.|\rho = \alpha) < F(.|\rho = 0)$.

2.b The density of observing a pledge of x at t given a wealth W and a project quality ρ is the density that a time t backer is of type $\theta(x, W, \pi_t)$, that is, $f(\theta(x, W, \pi_t)|\rho)$. To obtain the probability of observing x given ρ and a pledge in t, one has to take the expectation over all possible wealth to have $\int_0^1 z(W)f(\theta(x, W, \pi_t)|\rho)dW$. The formula then follows by Bayes' rule. Q.E.D.

Bibliography

Acemoglu, D. and Autor, D. (2011). *Skills, Tasks and Technologies: Implications for Employment and Earnings*. Elsevier, Handbook of Labor Economics *4*, chap. 12. 1043–1171.

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: 497–551.

Adao, R. (2015). Worker Heterogeneity, Wage Inequality, and International Trade: Theory and Evidence from Brazil. Unpublished paper, MIT.

Adda, J., Dustmann, C. and Stevens, K. (2017). The Career Costs of Children. *Journal of Political Economy* 125: 293–337.

Agrawal, A., Catalini, C. and Goldfarb, A. (2014). Some Simple Economics of Crowdfunding. *NBER Innovation Policy & the Economy* 14(1): pp. 63–97.

Aguayo-Tellez, E., Airola, J., Juhn, C. and Villegas-Sanchez, C. (2013). Did Trade Liberalization Help Women? The Case of Mexico in the 1990s. In Polachek, S. and Tatsiramos, K. (eds), *New Analysis of Worker Well-Being*. Research in Labor Economics. Bingley, UK: Emerald Group Publishing Limited, *vol. 38*.

Akbulut, R. (2011). Sectoral Changes and the Increase in Wome's Labor Force Participation. *Macroeconomic Dynamics* 15: pp. 240–264.

Altonji, J. G., Kahn, L. B. and Speer, J. D. (2014). Trends in Earnings Differentials across College Majors and the Changing Task Composition of Jobs. *American Economic Review* 104: 387–393.

Autor, D. and Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review* 103(5): pp. 1553–1597.

Autor, D., Katz, L. and Kearney, M. (2005). Rising Wage Inequality: The Role of Composition and Prices. NBER Working Paper Series 11628.

Autor, D., katz, L. and Kearney, M. (2006). The Polarization of the U.S. Labor Market. *American Economic Review* 96(2): pp. 189–194.

Autor, D., Levy, F. and Murnane, R. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics* 118(4): pp. 1279–1333.

Autor, D. H., Dorn, D. and Hanson, G. H. (2016). The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade. *Annual Review of Economics* 8: 205–240.

Autor, D. H., Katz, L. F. and Kearney, M. S. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics* 90: 300–323.

Avery, C. and Zemsky, P. (1998). Multidimensional Uncertainty and Herd Behavior in Financial Markets. *American Economic Review* 88: 724–748.

Azevedo, J. P., Inchauste, G. and Sanfelice, V. (2013). Decomposing the Recent Inequality Decline in Latin America. Policy Research Working Paper Series 6715, The World Bank.

Baker, M. and Cornelson, K. (2016). Gender Based Occupational Segregation and Sex Differences in Sensory, Motor and Spatial Aptitudes. NBER Working Papers 22248, National Bureau of Economic Research, Inc.

Banerjee, A. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics* 107: pp. 797–817.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2012). The Diffusion of Microfinance. NBER Working Papers 17743, National Bureau of Economic Research, Inc.

Barth, E., Kerr, S. P. and Olivetti, C. (2017). The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data. Working Paper 23381, National Bureau of Economic Research.

Battistón, D., García-Domench, C. and Gasparini, L. (2014). Could an Increase in Education Raise Income Inequality? Evidence for Latin America. *Latin American Journal of Economics-formerly Cuadernos de Economía* 51: 1–39.

Beauhurst (2016). The Deal. Review of Equity Investment Landscape. [Online Report] `http://about.beauhurst.com/the-deal-h1-16`.

Behrman, J., Birdsall, N. and Szekely, M. (2007). Economic Policy Changes and Wage Inequality in Latin America. *Economic Development and Cultural Change* 56(1): pp. 57–97.

Belleflamme, P., Lambert, T. and Schwienbacher, A. (2014). Crowdfunding: Tapping the Right Crowd. *Journal of Business Venturing* 29(14): pp. 585–609.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C. and Milkman, K. L. (2015). The Effect of Providing Peer Information on Retirement Savings Decisions. *Journal of Finance* 70: 1161–1201.

Bikhchandani, S., Hirshleifer, D. and Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* 100(5): pp. 992–1026.

Binelli, C. and Attanasio, O. (2010). Mexico in the 1990s: the Main Cross-Sectional Facts. *Review of Economic Dynamics* 13: 238–264.

Black, S. and Juhn, C. (2000). The Rise of Female Professionals: Are Women Responding to Skill Demand? *The American Economic Review* 90(2): pp. 450–455.

Black, S. and Spitz-Oener, A. (2010). Explaining Women's Success: Technological Change and the Skill Content of Women's Work. *The Review of Economics and Statistics* 92(1): pp. 187–194.

Borjas, G. J., Grogger, J. and Hanson, G. H. (2012). Comment: On Estimating Elasticities Of Substition. *Journal of the European Economic Association* 10: 198–210.

Bourguignon, F., Ferreira, F. H. and Lustig, N. (2005a). *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*. No. 14844 in World Bank Publications. The World Bank.

Bourguignon, F., Ferreira, F. H. and Lustig, N. (2005b). *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*. No. 14844 in World Bank Publications. The World Bank.

Brown, J. R., Ivkovic, Z., Smith, P. A. and Weisbenner, S. (2008). Neighbors Matter: Causal Community Effects and Stock Market Participation. *Journal of Finance* 63: 1509–1531.

Brown, R., Moon, M. and Zoloth, B. (1980). Occupational Attainment and Segregation by Sex. *Industrial and Labor Relations Review* 33(4): pp. 506–517.

Bursztyn, L., Ederer, F., Ferman, B. and Yuchtman, N. (2014). Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions. *Econometrica* 82: 1273–1301.

Card, D. and Lemieux, T. (2001). Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis. *The Quarterly Journal of Economics* 116(2): pp. 705–746.

Caselli, M. (2012). Trade, Skill-Biased Technical Change and Wages in Mexican Manufacturing. Working Paper.

Cavallo, A. (2013). Online and Official Price Indexes: Measuring Argentina's Inflation. *Journal of Monetary Economics* 60(2): pp. 152–165.

CEDLAS and The World Bank (2014). A Guide to SEDLAC: Socio-Economic Database for Latin America and the Caribbean.

Ciccone, A. and Peri, G. (2005). Long-Run Substitutability Between More and Less Educated Workers: Evidence from U.S. States, 1950-1990. *The Review of Economics and Statistics* 87: 652–663.

Coen-Pirani, D., León, A. and Lugauer, S. (2010). The Effect of Household Appliances on Female Labor Force Participation: Evidence from Microdata. *Labour Economics* 17: 503–513.

Colombo, M. G., Franzoni, C. and Rossi-Lamastra, C. (2015). Internal Social Capital and the Attraction of Early Contributions in Crowdfunding. *Entrepreneurship Theory and Practice* 39: 75–100.

Cornia, A. (2010). Income Distribution Under Latin America's New Left Regimes. *Journal of Human Development and Capabilities* 11(1): pp. 85–114.

Costa, D. L. (2000). From Mill Town to Board Room: The Rise of Women's Paid Labor. *Journal of Economic Perspectives* 14: 101–122.

Costa, F., Garred, J. and Pessoa, J. P. (2016). Winners and Losers from a Commodities-for-Manufactures Trade Boom. *Journal of International Economics* 102: 50–69.

Cowell, F. A. and Victoria-Feser, M.-P. (1996). Robustness Properties of Inequality Measures. *Econometrica* 64: 77–101.

Cragg, M. I. and Epelbaum, M. (1996). Why has wage dispersion grown in Mexico? is it the incidence of reforms or the growing demand for skills? *Journal of Development Economics* 51(1): pp. 99–116.

Cruces, G. and Galiani, S. (2007). Fertility and Female Labor Supply in Latin America: New Causal Evidence. *Labour Economics* 14: 565–573.

Dorn, D. (2009). Essays on Inequality, Spatial Interaction, and the Demand for Skills. Ph.D. thesis, University of St. Gallen, Graduate School of Business Administration, Economics, Law and Social Sciences (HSG).

Duflo, E. and Saez, E. (2003). The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment. *The Quarterly Journal of Economics* 118: pp. 815–842.

Erten, B. and Ocampo, J. A. (2013). Super Cycles of Commodity Prices Since the Mid-Nineteenth Century. *World Development* 44: 14–30.

Essama-Nssah, B. and Lambert, P. (2011). Influence Functions for Distributional Statistics. ECINE Working Paper 236.

Feenstra, R. C. and Hanson, G. H. (1997). Foreign Direct Investment and Relative Wages: Evidence from Mexico's Maquiladoras. *Journal of International Economics* 42: 371–393.

Fernandez, M. and Messina, J. (2017). Skill Premium, Labor Supply and Changes in the Structure of Wages in Latin America. IZA Discussion Papers 10718, Institute for the Study of Labor (IZA).

Fernández, R. (2013). Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century. *American Economic Review* 103: 472–500.

Fernández, R., Fogli, A. and Olivetti, C. (2004). Mothers and Sons: Preference Formation and Female Labor Force Dynamics. *The Quarterly Journal of Economics* 119: 1249–1299.

Fernández, R. and Wong, J. C. (2014). Divorce Risk, Wages and Working Wives: A Quantitative Life Cycle Analysis of Female Labour Force Participation. *Economic Journal* 0: 319–358.

Ferreira, F., Leite, P. and Litchfield, J. (2008). The Rise and Fall of Brazilian Inequality. *Macroeconomic Dynamics* 12(2): pp. 199–230.

Firpo, S., Fortin, N. and Lemieux, T. (2007). Decomposing Wage Distributions using Recentered Influence Functions. Unpublished manuscript, PUC-Rio and UBC.

Firpo, S., Fortin, N. and Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica* 77(3): pp. 953–973.

Firpo, S., Fortin, N. and Lemieux, T. (2011). Occupational Tasks and Changes in the Wage Structure. IZA Discussion Paper No. 5542.

Fogli, A. and Veldkamp, L. (2011). Nature or Nurture? Learning and the Geography of Female Labor Force Participation. *Econometrica* 79: 1103–1138.

Galiani, S. and Sanguinettu, P. (2003). The Impact of Trade Liberalization on Wage Inequality: Evidence from Argentina. *Journal of Development Economics* 72(2003): pp. 497–513.

Galor, O. and Weil, D. (1996). The Gender Gap, Fertility, and Growth. *The American Economic Review* 86(3): pp. 374–387.

Gasparini, L., Cruces, G. and Tornarolli, L. (2011). Recent Trends in Income Inequality in Latin America. *Economía* 11(2): pp. 147–201.

Gasparini, L. and Lustig, N. (2011). The Rise and Fall of Income Inequality in Latin America. In Ocampo, J. A. and Ros, J. (eds), *The Oxford Handbook of Latin American Economics*. Oxford University Press.

Gathmann, C. and Schoenberg, U. (2010). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics* 28: 1–49.

Goldberg, P. K. and Pavcnik, N. (2007). Distributional Effects of Globalization in Developing Countries. *Journal of Economic Literature* 45: 39–82.

Goldin, C. (1984). The Historical Evolution of Female Earnings Functions and Occupations. *Explorations in Economic History* 21: 1–27.

Goldin, C. (1986). Monitoring Costs and Occupational Segregation by Sex: A Historical Analysis. *Journal of Labor Economics* 4: 1–27.

Goldin, C. (2006). The Quiet Revolution That Transformed Women's Employment, Education, and Family. *American Economic Review* 96: 1–21.

Goldin, C. and Olivetti, C. (2013). Shocking Labor Supply: A Reassessment of the Role of World War II on Women's Labor Supply. *American Economic Review* 103: 257–62.

Goos, M. and Manning, A. (2007). Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *The Review of Economics and Statistics* 89(1): pp. 118–133.

Goos, M., Manning, A. and Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review* 104: 2509–2526.

Green, F., Dickerson, A. and Saba Arbache, J. (2001). A Picture of Wage Inequality and the Allocation of Labor Through a Period of Trade Liberalization: The Case of Brazil. *World Development* 29: 1923–1939.

Greenwood, J., Guner, N., Kocharkov, G. and Santos, C. (2016). Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment, and Married Female Labor-Force Participation. *American Economic Journal: Macroeconomics* 8: 1–41.

Greenwood, J., Seshadri, A. and Yorukoglu, M. (2005). Engines of Liberation. *The Review of Economic Studies* 72: pp. 109–133.

Grossbard-Shechtman, S. A. and Neuman, S. (1988). Women's Labor Supply and Marital Choice. *Journal of Political Economy* 96: 1294–1302.

Halliday, T., Lederman, D. and Robertson, R. (2015). Tracking Wage Inequality Trends with Prices and Different Trade Models : Evidence from Mexico. Policy Research Working Paper Series 7471, The World Bank.

Hanson, G. (2003). What has Happened to Wage in Mexico since NAFTA? Implications of Hemispheric Free Trade. National Bureau of Economic Research. NBER Working Paper 9563.

Hong, H., Kubik, J. D. and Stein, J. C. (2004). Social Interaction and Stock-Market Participation. *Journal of Finance* 59: 137–163.

Hong, H., Kubik, J. D. and Stein, J. C. (2005). Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers. *Journal of Finance* 60: 2801–2824.

Ivkovic, Z. and Weisbenner, S. (2007). Information Diffusion Effects in Individual Investors' Common Stock Purchases: Covet Thy Neighbors' Investment Choices. *The Review of Financial Studies* 20: pp. 1327–1357.

Johnson, M. and Keane, M. (2013). A Dynamic Equilibrium Model of the US Wage Structure, 1969-1996. *Journal of Labour Economics* 31(1): pp. 1–49.

Juhn, C. and Kim, D. I. (1999). The Effects of Rising Female Labor Supply on Male Wages. *Journal of Labor Economics* 17: 23–48.

Juhn, C., Murphy, K. and Pierce, B. (1993). Wage Inequality and the Rise in Returns to Skill. *The Journal of Political Economy* 101(2): pp. 410–442.

Juhn, C., Ujhelyi, G. and Villegas-Sanchez, C. (2014). Men, Women, and Machines: How Trade Impacts Gender Inequality. *Journal of Development Economics* 106: 179–193.

Kahhat, J. (2010). Labor Earnings Inequality: The Demand for and Supply of Skills. In López-Calva, L. F. and Lustig, N. (eds), *Declining Inequality in Latin America: A Decade of Progress?*. Brookings Institution Press, chap. 2.

Katz, L. and Autor, D. (1999). Changes in the Wage Structure and Earnings Inequality. In Ashenfelter, O. and Card, D. (eds), *Handbook of Labor Economics*. Amsterdam: North-Holland, *vol. 3A*, chap. 26.

Katz, L. and Murphy, K. (1992). Changes in Relative Wages, 1963-1987: Supply and Demand Factors. *The Quarterly Journal of Economics* 107(1): pp. 35–78.

Katz, L. F. and Goldin, C. (2000). Career and Marriage in the Age of the Pill. *American Economic Review* 90: 461–465.

Killingsworth, M. R. and Heckman, J. J. (1987). Female Labor Supply: A Survey. In Ashenfelter, O. and Layard, R. (eds), *Handbook of Labor Economics*. Elsevier, Handbook of Labor Economics *1*, chap. 2, 103–204.

Krusell, P., Ohanian, L., Ríos-Rull, J. and Violante, G. (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica* 68(5): pp. 1029–1053.

Kuppuswamy, V. and Bayus, B. (2015). Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter. UNC Kenan-Flagler Research Paper No. 2013-15.

Lemieux, T. (2006). Increasing Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? *The American Economic Review* 96(3): pp. 461–498.

Lemieux, T., MacLeod, B. and Parent, D. (2009). Performance Pay and Wage Inequality. *Quarterly Journal of Economics* 124(1): pp. 1–49.

Levy, S. and Schady, N. (2013). Latin America's Social Policy Challenge: Education, Social Insurance, Redistribution. *Journal of Economic Perspectives* 27: 193–218.

Li, G. (2014). Information Sharing and Stock Market Participation: Evidence from Extended Families. *The Review of Economics and Statistics* 96: 151–160.

Londoño, J. L. and Szekely, M. (2000). Persistent Poverty and Excess Inequality: Latin America, 1970-1995. *Journal of Applied Economics* 3(1): pp. 93–134.

Lopez-Calva, L. F. and Lustig, N. (2010). *Declining Inequality in Latin America: A Decade of Progress?*. Brookings Institution Press.

López-Calva, L. F. and Lustig, N. (2010). Explaining the Decline in Inequality in Latin America: Technological change, Educational Upgrading, and Democracy. In López-Calva, L. F. and Lustig, N. (eds), *Declining Inequality in Latin America: A Decade of Progress?*. Brookings Institution Press, chap. 1.

Lup Tick, S. and Oaxaca, R. (2010). Technological Change and Gender Wage Gaps in the US Service Industry. *Annals of Economics and Statistics* 99/100: pp. 47–65.

Lustig, N., Lopez-Calva, L. and Ortiz-Juarez, E. (2013). Deconstructing the Decline in Inequality in Latin America. Tulane Economic Working Paper Series.

Manacorda, M., Sánchez-Paramo, C. and Schady, N. (2010). Changes in Returns to Education in Latin America: The Role of Demand and Supply of Skills. *Industrial and Labor Relations Review* 63(2): pp. 307–326.

Massolution (2015). The Crowdfunding Industry Report. [Online Report] `http://reports.crowdsourcing.org/index.php?route=product/product&product_id=54`.

Messina, J. and Silva, J. (2016). Wage Inequality in Latin America. Understanding the Past to Prepare the Future. World Bank Unpublished Manuscript.

Michaels, G., Natraj, A. and Van Reenen, J. (2015). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over 25 Years. *The Review of Economic and Statistics* 86(1): pp. 60–77.

Mollick, E. (2014). The Dynamics of Crowdfunding: An Exploratory Study. *Journal of Business Venturing* 29(14): pp. 1–16.

Murphy, K. and Welch, F. (1992). The Structure of Wages. *The Quarterly Journal of Economis* 107(1): pp. 285–326.

Murphy, K. M. and Welch, F. (1990). Empirical Age-Earnings Profiles. *Journal of Labor Economics* 8: 202–229.

Ñopo, H. (2012). *New Century, Old Disparities : Gender and Ethnic Earnings Gaps in Latin America and the Caribbean*. No. 11953 in World Bank Publications. The World Bank.

Park, A. and Sabourian, H. (2011). Herding and Contrarian Behavior in Financial Markets. *Econometrica* 79: 973–1026.

Pavcnik, N. (2003). What Explains Skill Upgrading in Less Developed Countries? *Journal of Development Economics* 71: 311 – 328.

Rendall, M. (2010). Brain Versus Brawn: The Realization of Women's Comparative Advantage. Institute for Empirical Research in Economics University of Zurich. Working Paper 1424-0459.

Rendall, M. (2013). Structural Change in Developing Countries: Has it Decreased Gender Inequality? *World Development* 45(1): pp. 1–16.

Robertson, R. (2004). Relative Prices and Wage Inequality: Evidence from Mexico. *Journal of International Economics* 64: 387 – 409.

Ronald R. Rindfuss, A. L. K., Karin L. Brewster (1996). Women, Work, and Children: Behavioral and Attitudinal Change in the United States. *Population and Development Review* 22: 457–482.

Shiller, R. (2015). Crowdfunding or Crowdphishing?. [Op-ed]. `https://www.project-syndicate.org/commentary/sec-crowdfunding-regulations`.

Smith, L. and Sorensen, P. (2000). Pathological Outcomes of Observational Learning. *Econometrica* 68: 371–398.

Sánchez-Páramo, C. and Schady, N. (2003). Off and Running? Technology, Trade, and the Rising Demand for Skilled Workers in Latin America. The World Bank. Policy Research Working Paper 3015.

The World Bank (2012). *World Development Report 2012*. No. 4391 in World Bank Publications. The World Bank.

The World Bank (2016). The Commodity Cycle in Latin America: Mirages and Dilemas. Semiannual Report, Office of the Regional Chief Economist.

Topel, R. H. (1994). Regional Labor Markets and the Determinants of Wage Inequality. *American Economic Review* 84: 17–22.

U.S. Bureau of Labor Statistics (2017). Civilian Labor Force Participation Rate: Women [LNS11300002], Retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/LNS11300002`, May 2, 2017.

V. Cavalcanti, T. V. de and Tavares, J. (2008). Assessing the "Engines of Liberation": Home Appliances and Female Labor Force Participation. *The Review of Economics and Statistics* 90: 81–88.

Vulkan, N., Åstebro, T. and Fernandez, M. (2016). Equity Crowdfunding: A New Phenomena. *Journal of Business Venturing Insights* 5: 37 – 49.

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics* 58: 369–396.

Zhang, J. and Liu, P. (2012). Rational Herding in Microloan Markets. *Management Science* 58(5): pp. 892–912.