IMT School for Advanced Studies – Lucca, Italy

Katholieke Universiteit Leuven – Leuven, Belgium

Essays on the Economics of Labor Markets and Retirement Policies

Ph.D. in System Science – *track in* Economics, Networks and Business Analytics

Ph.D. in Business Economics

Francesco Del Prato

2023

Contents

List of Figures

List of Tables

Acknowledgements

The dissertation owes much to my advisors Paolo Zacchia, Mattia Nardotto and Kenan Huremovic, which I would like to thank for their presence and guidance. I am also indebted to my closest friends and colleagues Edoardo M. Acabbi, Marco Francischello, Luca Mazzone and Matteo Paradisi. Their availability and patience can be matched only by the quality of their comments. The second chapter is joint work with Paolo Zacchia, whom I need to thank twice. The third chapter is joint work with Marco Francischello and Matteo Paradisi, whom I thank for their contributions. Eventually, I would like to thank my family for the unconditional support through this journey: Stefano, Anna, Laura and Maria, I owe you a lot. The generosity and help of many have also helped during these years, I would like in particular to thank Falco J. Bargagli Stoffi, Tatiana Celadin, Salvatore Lattanzio, Andrea Rocchetto, and Nicolò Vallarano. Finally, I would like to thank the community of VisitINPS—particularly Edoardo Di Porto and Paolo Naticchioni—which became, in these years, more than just a workplace.

Vita

Education

2022 Postdoctoral Fellow
Paris School of Economics

2017-2022 Joint Ph.D. Candidate
IMT School for Advanced Studies
Katholieke Universiteit Leuven

2021-2022 External Research collaborator
CERGE-EI

2014-2017 MSc in Economics and Social Sciences
Università Commerciale "L. Bocconi"

2011-2014 BA in Management
Università Commerciale "L. Bocconi"

Scholarship and awards

1. Netspar Comparative Research Grant (2021)
2. VisitINPS Program scholarship type B (2020)
3. VisitINPS Program scholarship type B (2019)
4. Full Ph.D. Scholarship, IMT School for Advanced Studies (2017)

Teaching

Fall 20-21 Microeconometrics (Ph.D.)
CERGE-EI, two lectures on random search models and their estimation

Spring 19-20 Econometrics I (Ph.D.)
IMT School for Advanced Studies, TA for Prof. Paolo Zacchia

Spring 18-19 Econometrics I (Ph.D.)
IMT School for Advanced Studies, TA for Prof. Paolo Zacchia

2014-2017 Microeconomic Theory (Ph.D.)
IMT School for Advanced Studies, TA for Prof. Andrea Canidio and Prof. Kenan Huremovic

Seminars and talks

1. SOLE, Society of Labor Economics (Philadelphia, 2023)
2. 20th Brucchi Luchino Labor Economic Workshop (Napoli, 2022)
3. 63rd SIE Annual Conference (Torino, 2022)
4. VisitINPS Seminar (Rome, 2022, 2021)
5. KU Leuven Doctoral Seminar (Leuven, 2022, 2021)
6. IMT School Workshop on Technology and Labour Marked (Lucca, 2020)

Abstract

This dissertation explores three distinct yet relevant aspects of labor markets, shedding new light the micro- and macroeconomic mechanisms behind them. It comprises three independent essays.

In the first chapter, I explore a novel mechanism through which firms can provide value to their employees: reducing on-the-job search frictions. I build a structural search model where the rate of job offers depends on the current employer. Workers thus value the firms' contribution to accelerating their ascent on the job ladder. Using a reduced-form approach, I demonstrate the existence of this compensating differential and its payoff in terms of future earnings. Finally, I structurally estimate the model, showing a precise fit with the data.

The second chapter offers new evidence of the heterogeneous effects on firm productivity distribution caused by a labor market reform aimed at enhancing labor flexibility, which indirectly reduced labor costs. Specifically, we show that this decrease in labor costs—attributable to the unique features of Italian collective bargaining institutions—suppresses total factor productivity (TFP) among already unproductive firms while increasing it for the most productive ones. We argue that this effect is driven by negative selection at the bottom of the distribution and construct a model that rationalizes this mechanism and provides welfare implications.

The third chapter uses an overlapping generation model to study the implications on optimal taxation of the government's use of a credible set of social security instruments. We reveal that these instruments introduce new redistributive motives and crowd out others in the context of a standard Ramsey problem. We calibrate the model using data from three different economies, showing that current retirement benefits exceed their optimal level and that the implementation of funded social security schemes is desirable.

The dissertation contributes to various branches of labor economics and macro-public finance literature: *i.* it investigates a brand new

compensating differential channel for high-skilled workers that explains a significant component of employees' transitions behavior; *ii.* it presents new empirical and theoretical evidence on the heterogeneous effects of labor market reforms on productivity; *iii.* it characterizes optimal distortionary labor and capital taxation for the first time in the context of a rich set of social security instruments, bridging the gap between social security and traditional Ramsey policy instruments.

Introduction

This dissertation consists of three independent essays and focuses on the economics behind multiple aspects of how labor markets work and react to policy interventions.

In the first chapter, I investigate the relationship between workers' wage dynamics and employers' role in mitigating on-the-job search frictions for their employees. In this setting, workers might sacrifice current salaries for an increased likelihood of successful future matches and a faster climb up the job ladder. I introduce a random search model of the labor market with two-sided heterogeneity. The innovative aspect lies in the integration of a 'connectivity' parameter, allowing the offer arrival rate to an employee to be contingent on the employer's characteristics. In this context, forming a match with a highly connected firm provides workers with an opportunity to increase the likelihood of finding better matches sooner. The connectivity parameter is linked to the degree centrality of a job mobility network, in which firms represent nodes and job transitions are edges. The model's two primary predictions are validated through reduced-form evidence. First, I show that the hiring salaries of young employees are inversely correlated with the network centrality of the hiring company, indicating that young workers are prepared to trade off initial salary for increased centrality. Second, I show that job switches at highly central firms yield a wage premium of 6-to-8 percent compared to less central firms. Finally, I structurally estimate the model on the data, revealing a close match with the empirical moments—particularly those deriving from the mobility network.

Chapter two, co-authored with Paolo Zacchia, explores the link between labor market institutions and firm productivity. We study an Italian labor reform aimed at easing temporary contract restrictions, assessing its impact on total factor productivity (TFP) through an indirect labor cost reduction. We highlight unique aspects of Italian collective bargaining institutions, which help us isolate the relationship between changes in contract types and labor cost across industries. Indeed, the reform did not affect temporary contract use in manufacturing but did lower labor costs. It also reduced TFP among the least productive firms while increasing it in the most productive ones. These distributional effects grow progressively within the distribution, suggesting the more productive the firm, the greater the reform's impact. We suggest the reform triggered a negative selection mechanism at the TFP distribution's lower end. Our general equilibrium model, featuring monopolistic competition and firm heterogeneity, rationalize this mechanism. Financial frictions require firms to make upfront investments, supplied by financial intermediaries, leading to an information asymmetry problem. The model shows that higher labor costs lead to fewer entries at the productivity distribution's lower end. This aligns with our empirical evidence, where lower labor costs and fewer exits among unproductive firms explain the observed TFP effect. The model also explains productivity gains due to investment incentives from labor cost savings.

The third and final chapter is a collaborative work with Marco Francischello and Matteo Paradisi. In this paper, we develop an overlapping generations (OLG) model with heterogeneous agents and aggregate uncertainty to study optimal Ramsey taxation when the government has access to a robust collection of social security instruments. Our model reveals that social security minimizes the income effect in labor tax smoothing and, in conjunction with heterogeneity, introduces new redistributive motives while eliminating others in labor and capital taxes. To calibrate the model, we employ data from three distinct economies—the United States, the Netherlands, and Italy. We propose that transitioning from the current allocations to those recommended by a utilitarian Ramsey planner could lead to a range of improvements in efficiency and redistribution for these three countries. Additionally, our simulations show that retirement benefits in the existing economies surpass their Ramsey-optimal levels. However, we also suggest that the implementation of funded social security programs, which are currently overlooked in real policy considerations, could have a beneficial impact on welfare.

Chapter 1

'The Importance of Working for Earnest': Firms' Reputation Network and Wage Dynamics[1]

Abstract. This paper explores the impact of employers' contributions to reducing on-the-job search frictions on workers' mobility decisions and earnings dynamics. I build a search model of the labor market, which introduces a novel perspective: firms vary in their ability to increase the chances of their employees receiving outside job offers. This shifts the heterogeneity in search behavior from workers to firms, suggesting that workers trade off their present salary for a higher probability of better future matches and faster job ladder climb. By linking the model's primitives to the degree centrality of a job-to-job network, I leverage comprehensive administrative data on white-collar workers in Italy to document two key implications of my theory. First, the hiring earnings of

[1]I express my gratitude to Paolo Zacchia, Mattia Nardotto, and Kenan Huremovic for their support and academic guidance. I am also deeply appreciative of Edoardo M. Acabbi, Marco Francischello Luca Mazzone, and Matteo Paradisi for their help and friendship. My thanks extend to Bruno Cassiman, Edoardo Di Porto, Salvatore Lattanzio, Francesco Serti and Cristina Tealdi for their comments and suggestions. I am also grateful to the participants of internal seminars at IMT Lucca, KU Leuven, CERGE-EI, and the SOLE 2023 conference. As part of the VisitINPS Initiative, data access was provided within an initial project in collaboration with Nicolò Vallarano, to whom I am thankful. Edoardo Di Porto and Paolo Naticchioni helped significantly with data access. The views expressed in this work are solely my own and do not reflect those of INPS. I alone am responsible for any errors or omissions.

young workers are inversely correlated with their employer's degree of centrality in the network, which I interpret as evidence of the described compensating differential. Second, leaving highly-central firms pays off up to 6% more than leaving regular ones upon a job transition—which provides evidence of faster career progressions facilitated by firms' connectivity. Finally, I structurally estimate this framework on Italian data to quantitatively account for this channel in assessing earnings variance.

JEL classification: J31, J24, J62, E24.

Keywords: Search frictions, connectivity, on-the-job search, mobility network.

1.1 Introduction

Firms' non-monetary characteristics greatly affect workers' mobility decisions and their career outcomes (Sorkin, 2018). Many of these characteristics pertain to *current* opportunities, such as alternative payment schemes (Card et al., 2018), variations in training and human capital accumulation rates (Gregory, 2020; Arellano-Bover and Saltiel, 2021), or differences in job security (Jarosch, 2021). A burgeoning body of recent research has sought to understand how these channels offer compensating differentials to workers. Other characteristics are related to *future* opportunities, such as the likelihood of securing higher-paying jobs later in one's career. The ability of workers to select their employers based on the potential for receiving better external offers represents a significant yet under-researched compensating differential channel, which is the primary focus of this paper.

For younger workers, the choice of early-career firms is pivotal as it impacts their current working conditions and future wages and opportunities. It is thus crucial to understand the role of today's workplaces in facilitating better career opportunities in the future, as this understanding can inform assessments of how workers make their occupational choices and how these choices can help to reduce labor market frictions, a recognized contributor to earnings inequality (Hornstein et al., 2011). Exploring how employers offer compensating differentials to alleviate these inefficiencies can provide a more transparent understanding of labor market operations and how frictions contribute to observed lifetime earnings inequality.

This paper investigates, for the very first time, the role of firms' ability to reduce search frictions for their employees and how such a contribution explains earnings dispersion across workers. It provides fresh insights into how workers select their workplaces, focusing on the differential contributions of firms in terms of future opportunities. My key intuition is that firms vary in their ability to enhance the likelihood of their employees receiving attractive outside offers. This perspective shifts the heterogeneity in search behavior from workers to firms. While differences in search behavior typically depend on the varying efforts of workers, in this context, the current employer plays a pivotal role in providing differential opportunities for faster career progression to its employees.

Through the paper, I first build a search model of the labor market with heterogeneous workers and firms where search frictions depend on the present employer, as firms' characteristics matter for the probability of receiving outside offers. Through the paper, I first construct a search model of the labor market that features heterogeneous workers and firms, where the severity of search frictions a worker faces depends on the current employer. Indeed, the characteristics of firms influence the probability of receiving external job offers. Next, I present empirical reduced form evidence supporting two key predictions of the model, leveraging the unique matched employer-employee dataset from the Italian Social Security Institute (INPS). By associating a model's primitive with a measure of degree centrality in the job-to-job network, I initially demonstrate that young workers are willing to accept lower initial salaries in exchange for centrality, thereby indicating the presence of a compensating differential. Furthermore, I show that utilizing this connectivity channel pays. Workers who pass through high-centrality firms earn, on average, approximately 5% more within one year of departure compared to workers leaving regular firms. Following this, I bring the entire model to the data by structurally estimating it, demonstrating that the mechanism I propose provides a good fit for the data. I estimate a skewed distribution of connectivity, with a large proportion of firms exhibiting a low capacity to intermediate workers across a wide range of other firms.

I start by building a search model of the labor market that incorporates heterogeneous workers and firms, which formalizes my core intuition. The essential element of my framework is indeed that firms vary along two dimensions: productivity and connectivity. Productivity dictates the combined output of the worker-firm match, while connectivity influences the probability of receiving external job offers while employed. In this context, this parameter encapsulates how employment at different firms leads to varying rates of offer arrivals. Consequently, workers who search on the job face a trade-off between productivity, which ensures higher immediate wages, and connectivity, which facilitates a quicker ascent of the job ladder. As a result, workers may be willing to accept positions at low-productivity firms, provided these firms significantly increase the likelihood of receiving new job offers. This allows employees to transition to higher-paying jobs more swiftly. In this way, firms can offer differential compensation to their employees by reducing future search frictions. While the model does not explicitly incorporate life-cycle dynamics, the connectivity value is higher for younger workers. This is because the connectivity value diminishes as employees ascend the job ladder (and thus become older,) transitioning into increasingly productive firms. I derive a closed-form wage equation that explicitly links wages to job values across the two dimensions of firm heterogeneity. The impact of a firm's productivity on worker compensation is ambiguous, as it is contingent on the bargaining power that employees hold with respect to the firm. However, connectivity functions as a pure compensating differential, as wages are weakly decreasing in the firm's connectivity.

Connectivity can be conceptualized as the firm's capacity to dispatch and receive employees from a variety of sources and destinations. It is worth discussing how this notion can be interpreted as the result of several plausible underlying mechanisms. For instance, firms may contribute to reducing search frictions through input-output relationships or through connections with past co-workers, as suggested by Caldwell and Harmon (2019). Corporate practices like transition assistance (where former employees may become future clients) can also contribute to a firm's connectivity. Additionally, variations in screening and hiring capabilities can lead to a situation where certain firms consistently hire, and thus supply, superior workers, thus providing their employees a higher probability of being contacted by other firms. At this stage, I will not delve further into the mechanisms underlying the differences among firms in reducing search frictions. The focus of this paper is not to understand the sources of such heterogeneity but rather to explore their implications.

I connect some of the model's predictions to empirical evidence in data by observing that the firm's connectivity parameter aligns with the expected degree centrality of an employer in a job-to-job network. In this network, a node represents a firm with a worker leaving or arriving within the sample period, and a (directed) link signifies a movement either inside or outside a firm. To this end, I utilize a comprehensive administrative matched employer-employee dataset that covers the entirety of the private sector in Italy. This data allows me to construct a job-to-job network of white-collar employees transitioning between relatively large firms from 2008 to 2018. I leverage the structure of this network to derive various measures of a firm's degree centrality. In other words, I account for the relative importance of each firm in controlling the flow of workers between different destinations and sources. To avoid merely capturing the largest firms, which are central due to a mechanical argument, I scale these measures by the firm size.

I then link some model's predictions to reduced form evidence in data by noting that the firm's connectivity parameter maps in the expected degree centrality of an employer in a job-to-job network where a node is a firm with a worker leaving or arriving within the sample period and a (directed) link is a movement either inside or outside a firm. I thus take advantage of a rich administrative matched employer-employee covering the universe of the private sector in Italy to build a job-to-job network of white-collar employees moving between relatively large firms between 2008 and 2018. I exploit the network structure to obtain different measures of a firm's degree centrality, i.e., I account for how relatively important each firm is in controlling the worker flows between different destinations and sources. I keep these measures scaled by firm size to avoid capturing only the largest firms, which are central by a mechanical argument.

First, given the network structure, I use between-firms movements between 2018 and 2020 to show that young workers are willing to accept lower initial earnings in exchange for higher firm centrality, provided the latter is substantial enough to offer credible value. This negative relationship does not apply to older workers, who are more likely to have already sorted into productive matches on average. For these employees, higher chances of receiving new offers are less important compared to workers who have recently entered the market. I interpret this empirical finding as evidence of the compensating differential implied by my model, where firms more richly connected with others can pay young workers less, compensating them with more valuable future opportunities.

Next, I aim to demonstrate that workers transitioning out of highly-central firms typically achieve higher earnings on average. To accomplish this, I divide the sample of firms into two groups using an unsupervised clustering algorithm based on the centrality measures that account for employee inflows and outflows.[2] This approach ensures that the splitting procedure is entirely data-driven. I then use a two-way fixed effects specification to dynamically compare workers leaving high-connectivity firms (those with higher centrality, approximately 12% of the sample of firms) with workers leaving regular firms. On average, the former group earns 6% more than the latter group one and a half years after leaving the firm, controlling for individual- and firm-specific heterogeneity in local labor markets. I interpret these results as evidence of another key mechanism implied by my model, which predicts that workers use firms with higher connectivity to ascend the job ladder more quickly. I interpret these results as evidence of another key mechanism implied by my model that predicts that workers use higher connectivity firms to climb the job ladder faster. Employees who pass through these firms are thus expected to earn more due to the increased likelihood of being matched with a higher-productivity firm, compared to those who transition through less connected employers.

Finally, I estimate the model through indirect inference—a simulated method of moments (SMM) that relies on additional reduced form models in addition to standard moment conditions (McFadden, 1989; Gourieroux et al., 1993)—on the same dataset used in the reduced form. I discuss that identification comes by targeting three different sets of moments or estimated parameters from re-

[2]In particular, I use a k-means algorithm to split firms into the two groups based on different possible measures of degree centrality in the job-to-job network.

duced form specifications for each component of the model that affects wage dynamics: individual ability, the firm's connectivity, and the firm's productivity. I first separate the workers' components from the firms' components by targeting a two-way fixed effect specification in the spirit of Abowd et al. (1999, AKM). I then target a set of moments related to network centrality distribution to inform the parameters governing connectivity. In this case, identification is possible since the model endogenously produces a job-to-job network where higher connectivity maps into higher expected centrality, as already discussed.

I first separate the workers' from the firms' components targeting a two-way fixed effect specification in the spirit of Abowd et al. (1999, AKM). I then target a set of moments related to network centrality distribution to inform the parameters governing the connectivity. Here, identification is possible since the model endogenously produces a job-to-job network where the higher the connectivity, the higher the centrality. Finally, I target several moments related to within-job and between-jobs wage growth to discipline the productivity distribution parameters. The model provides a decent fit of real data. In particular, it very closely matches the set of moments coming from the centrality measures on the job-to-job network, exhibiting an appreciable capacity to simulate firms' role in intermediating workers' flows. Finally, I target several moments related to within-job and between-jobs wage growth to discipline the productivity distribution parameters. The model provides a decent fit of real data. In particular, it very closely matches the set of moments coming from the centrality measures on the job-to-job network, demonstrating a notable ability to simulate firms' role in intermediating workers' flows. The estimated distribution of connectivity is highly skewed, with a large proportion of firms exhibiting low connectivity and a small number of companies intermediating a wide variety of sources and destinations. This result aligns coherently with the unsupervised data-driven split I implemented in the reduced-form exercises.

Related literature This paper intersects and contributes to several distinct areas of literature, specifically those concerning heterogeneity in labor market outcomes, worker dynamics across firms, and optimal search behavior. To the best of my knowledge, this is the first study to explicitly consider the heterogeneity of firms' contributions to future job opportunities, utilizing comprehensive administrative data on private-sector contracts in a large country.

A substantial and well-established body of work has explored how firm-specific characteristics influence workers' labor market outcomes and, reciprocally, how matching with diverse employers impacts wages and career trajectories. Many of these studies, including those by Andersson et al. (2012), Card et al. (2013), Card et al. (2018), and Song et al. (2019) employ the exogenous-mobility approach pioneered by Abowd et al. (1999) on matched employer-employee data. These papers primarily account for characteristics that affect the current worker-firm relationship, while this study underscores the existence of a channel that only impacts the worker's future value.Abowd et al. (2018) and Bonhomme et al. (2019) address employment at heterogeneous firms within dynamic frameworks, thereby connecting past and present employers.

Several papers addressing the long-term labor market effects of firm differences do so by incorporating a combination of search and human capital accumulation, as in the seminal work by Bagger et al. (2014). Among these, Arellano-Bover (2020) use an instrumental variable specification to assess the long-term effects of heterogeneity in size among the first employers in workers' careers. Di Addario et al. (2021) focus on a two-way fixed effect reduced-form model that encapsulates the past firm's effect on present wages, finding no significant contribution from the last employer, on average. Jarosch (2021) demonstrates that heterogeneity in firms' provision of labor security explains a significant amount of wage variance in the long run. Gregory (2020) quantifies the variation in life-cycle earnings profiles that differences in firm-specific human capital accumulation can explain. Arellano-Bover and Saltiel (2021) addresses a similar question through a reduced-form approach relying on firms' clustering in skill-learning groups. Wang (2021) estimates the persistent effect of firms' heterogeneous internal career promotion policies on workers' earnings, showing little benefits for workers in moving to high-promotion firms. The importance of workplace characteristics also aligns with De La Roca and Puga (2017), which explores three channels that could potentially account for the correlation between salaries and city size across regions. However, none of these papers either explicitly or implicitly incorporate a connectivity mechanism in explaining past-to-present firm relationships, nor do they structurally link firms' characteristics in reducing search frictions for their employees via an increased probability of receiving outside offers.

The concept of identifying significant compensating differentials explaining wage cuts upon job-to-job transitions is shared by a considerable number of recent papers, such as Nunn (2012), Sullivan and To (2014), Hall and Mueller (2018), Taber and Vejlin (2020), and Caplin et al. (2022). In particular, Sorkin (2018) adopts a revealed-preferences approach, exploiting a centrality measure on the job-to-job network to assess the importance of compensating differentials in workers' mobility behavior. In addition to this latter work, others like Nimczik (2020) and Huitfeldt et al. (2021) are also built on the workers' mobility network. In this paper, I explicitly connect the firms' centrality in the job-to-job network to their connectivity parameter in the model, thereby investigating a novel channel that firms use to deliver value to workers.

This paper also draws on several studies that address the relevance of labor market frictions through random search models with sequential auctions. Specifically, the bargaining protocol comes from the seminal works of Postel-Vinay and Robin (2002) and Cahuc et al. (2006). Some features of the model also echo Bagger et al. (2014). Similar to Gregory (2020) and Jarosch (2021), firms in my model are heterogeneous along a second dimension other than productivity (in their cases, the quality of the learning environment and job security, respectively). The mechanism through which workers value firms' connectivity in my model pertains to the reduction in search frictions, which provides workers with a higher likelihood of better opportunities, independent of the human capital dynamic.

Lastly, mechanisms closely related to the economic intuition behind the role of firms' connectivity, such as screening and sorting, are investigated by Cai et al. (2021) in a search model with information frictions. However, they focus on the strategic decision of the firm regarding the optimal size of screening pools rather than assessing how employers' screening capacities may explain wage heterogeneity. It is s worth noting that while this version of the paper does not include a microfoundation of the meeting mechanism, I plan to characterize firms with higher connectivity as firms with superior screening technologies.

Paper's structure The remainder of this paper is organized as follows. Section 2.4 presents and discusses a random search model where firms exhibit heterogeneity in productivity and connectivity. Section 1.3 introduces the data sources utilized in this study, explains the process of constructing the job-to-job network and the centrality measures employed, provides evidence of a compensating differential based on firms' centrality, and presents reduced-form results on the relationship between workers' wages and firms' connections in the network. Section 1.4 discusses the identification strategy and estimates the model using Italian administrative data. Section 1.5 concludes the paper.

1.2 A Search model with connectivity

This section introduces an equilibrium model of the labor market that accounts for heterogeneity among both workers and firms. Firms are uniquely characterized by two attributes: productivity and connectivity. This model's novelty lies in the inclusion of the latter, which governs the firm-specific likelihood of employed workers receiving or making job offers during their on-the-job search.

1.2.1 Heterogeneous agents

The market consists of a continuum of workers who are infinitely-lived and differentiated by ability, denoted a. These abilities are distributed exogenously over a continuous set $[\underline{a}, \bar{a}]$, following a cumulative distribution function $A(\cdot)$. [3] Workers have linear preferences for a single good and can either be employed or unemployed.

On the other side of the market, firms are represented by the type $\boldsymbol{\theta} = (\theta_p, \theta_c)$, where θ_p and θ_c represent the firm's productivity and connectivity, respectively. These parameters are distributed exogenously according to cumulative distribution functions $P(\cdot)$ over $[\underline{\theta}_p, \bar{\theta}_p]$ and $T(\cdot)$ over $[\underline{\theta}_c, \bar{\theta}_c]$, respectively. Their joint distribution is denoted as $F(\cdot)$. Workers and firms alike discount future returns at a common rate β.

1.2.2 Meetings and production

Time is discrete. Both workers and firms search on the market for (possibly better) matches, while unemployed workers search for employment. This search process is random, undirected, and incurs costs. Workers and firms meet at each period.

Employed worker. In conventional random search models, workers seeking opportunities on the job receive external offers with a consistent probability that is independent of the types of firms and might be linked to workers' characteristics through individual search effort. When firms are indistinguishable from jobs, one can interpret such a meeting mechanism as a firm-to-firm interaction: the current and potential employer meet, and the worker observes the resulting offer with an exogenous probability. In the proposed model, a meeting between two firms does not guarantee an employed worker a job offer. The formalization of the offer depends on the connectivity of both the current employer (the "incumbent") and the potential employer (the "challenger"). For the worker to be aware of the interaction between the firms—and therefore the offer—the combined connectivity of both firms must exceed an exogenous threshold, denoted

[3] In this preliminary version of the model, human capital accumulation is not considered. Although its inclusion would enhance the model's credibility and data-matching capacity, it would also complicate the model's tractability. A simple human capital dynamic is planned for future iterations.

as I. Consequently, a worker employed at firm B will receive a valid job offer from firm A only if the combined connectivity of firms A and B is larger than I: $\theta_c^A + \theta_c^B \geq I$.

This simple model's characteristic attempts to formalize a widely observed labor market phenomenon: the rate at which workers receive job offers varies depending on their current employment. The higher the connectivity of the incumbent firm, the higher the likelihood of the worker receiving an external offer. Similarly, firms with extensive connectivity are more likely to engage workers employed at incumbent firms. The economic rationale behind this process can be interpreted in several ways. For instance, firms with higher connectivity may systematically provide higher-quality workers due to superior rates of firm-specific human capital accumulation or more efficient screening technologies. Alternatively, companies with extensive connectivity may reduce search frictions for potential future employers because of existing business relationships, such as sales and purchases. Variations in connectivity could also reflect differences in corporate culture. For example, some companies actively assist their former employees in securing new employment. For now, I will abstract from discussing one specific underlying mechanism that can explain why job offer rates vary across employers.

Unemployed worker. In this model, unemployment is defined as a firm characterized by a productivity-connectivity pair $u = (u_p, 0)$, where $u_p < \theta_p$ for any given p. The connectivity parameter is irrelevant in the context of unemployment, which justifies setting it to zero. As a result, the rate at which an unemployed worker receives job offers, denoted as λ, is independent of the firm's characteristics in attempting to hire the worker. This rate is considered exogenous, highlighting that connectivity does not influence the job offer rate for unemployed individuals, as in a standard random search model.

1.2.3 Wage setting protocol

Wages in this model are viewed as fixed contracts that can be renegotiated under specific conditions, particularly when credible threats arise. These threats might occur when workers receive an external job offer substantial enough to be leveraged for renegotiating their current wage with their existing employer or when they transition to a new company. When such a formal offer is made, the incumbent and challenging firms engage in Bertrand competition for the worker, making repeated bids. This sequential auction mechanism was initially proposed by Postel-Vinay and Robin (2002) and later refined by Cahuc et al. (2006). The notation used in what follows is partly borrowed from Jarosch (2021).

Let W, U, and J denote the value of an employed worker, an unemployed worker, and a job for a firm, respectively. $S(a, \theta_p, \theta_c)$ represents the joint surplus generated by a match between a worker of type a and a firm of type (θ_p, θ_c). As in the classic sequential auction setting, both the worker's wage $w(a, \theta_p, \theta_c, \hat{\theta}_p, \hat{\theta}_c)$ and the value $W(a, \theta_p, \theta_c, \hat{\theta}_p, \hat{\theta}_c)$ depend on the worker's ability a, the current employer's type (θ_p, θ_c), and the type of the firm involved in the last wage negotiation, $(\hat{\theta}_p, \hat{\theta}_c)$.

Unemployed worker. If an unemployed worker forms a match with a firm $\boldsymbol{\theta} = (\theta_p, \theta_c)$, the wage should satisfy

$$W\left(a, \theta_p, \theta_c, u_p, 0\right) - U = \sigma S\left(a, \theta_p, \theta_c\right) \tag{1.1}$$

where $\sigma \in [0, 1]$ represents the worker's bargaining power over the match surplus. As $S(a, u_p, 0) = 0$, the set of firms an unemployed worker is willing to work for is represented by $\mathcal{F}_1(u) \equiv (\theta_p, \theta_c) \mid S(a, \theta_p, \theta_c) > 0$.

Employed worker. For an employed worker of type a currently working at the incumbent firm $\boldsymbol{\theta}^1 = (\theta_p^1, \theta_c^1)$, three mutually exclusive cases may arise if challenged by a firm $\boldsymbol{\theta}^2 = (\theta_p^2, \theta_c^2)$.

1. The worker produces a higher joint surplus with the firm (θ_p^2, θ_c^2) than with the firm (θ_p^1, θ_c^1), i.e., $S(a, \theta_p^2, \theta_c^2) > S(a, \theta_p^1, \theta_c^1)$. As a result, the incumbent employer becomes the new negotiation benchmark, and the worker transitions to the challenger firm (θ_p^2, θ_c^2) with a wage such that[4]

$$W\left(a, \theta_p^2, \theta_c^2, \theta_p^1, \theta_c^1\right) - U = S\left(a, \theta_p^1, \theta_c^1\right) + \sigma\left[S\left(a, \theta_p^2, \theta_c^2\right) - S\left(a, \theta_p^1, \theta_c^1\right)\right] \tag{1.2}$$

   The worker, therefore, receives the whole surplus of the incumbent match plus a share σ of the net gains from the movement to θ_2. For a worker employed at θ, the set of firms that allow this first case is $\mathcal{F}_1(\theta_p, \theta_c) \equiv \{(\theta_p', \theta_c') \mid S(a, \theta_p', \theta_c') > S(a, \theta_p, \theta_c)\}$. This set thus includes all firms where the surplus generated with the worker is greater than the surplus generated at the incumbent firm.

2. The worker produces a higher joint surplus with the incumbent after a renegotiation. This would occur if $S(a, \theta_p^2, \theta_c^2) < S(a, \theta_p^1, \theta_c^1)$, but the current negotiation benchmark is still lower than $S(a, \theta_p^2, \theta_c^2)$. This means

[4]As I will extensively discuss, later on, the new wage a worker obtains upon a movement can be *lower* that the wages set with the incumbent, due to a compensating differential mechanism.

the challenging firm could offer a wage that's more attractive than the worker's current wage. The worker could then use this external job offer to negotiate a higher salary while staying with the incumbent firm. The incumbent firm would then have to raise the worker's salary just enough to keep them. The new wage would meet the following indifference condition:

$$W\left(a,\theta_p^1,\theta_c^1,\theta_p^2,\theta_c^2\right)-U=S\left(a,\theta_p^2,\theta_c^2\right)+\sigma\left[S\left(a,\theta_p^1,\theta_c^1\right)-S\left(a,\theta_p^2,\theta_c^2\right)\right] \tag{1.3}$$

I will refer to the set of firms where this second scenario applies as $\mathcal{F}_2(\theta_p,\theta_c,\hat{\theta}_p,\hat{\theta}_c)\equiv\{(\theta_p',\theta_c')\mid S(a,\theta_p,\theta_c)>S(a,\theta_p',\theta_c')>S(a,\hat{\theta}_p,\hat{\theta}_c)\}$.

3. The value generated by the offer is entirely dominated by the current negotiation benchmark, i.e., the previous outside option. In this case, the surplus the worker could generate with the challenging firm is less than what they could generate with the incumbent firm. Moreover, the worker cannot use the external job offer to negotiate a higher wage. As a result, the worker simply dismisses the offer and continues to work for the incumbent firm at the same wage.

The sequential auction wage setting protocol outlined above generates frictional wage dispersion and governs both wage dynamics and job-to-job transitions, depending on the worker's recent employment history (their negotiation benchmark). As long as workers remain employed, they ascend the job ladder by transitioning to firms that offer increasing value. They also utilize external job offers to influence their wage dynamics, taking advantage of these opportunities to negotiate higher wages and secure better positions.

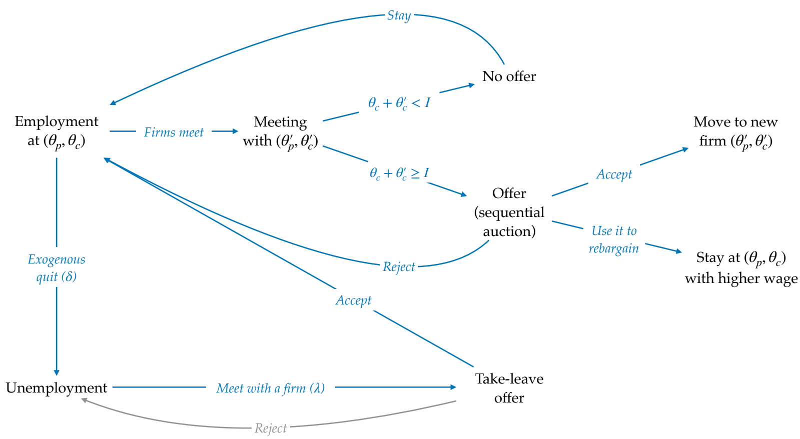
Figure 1.1 provides a graphical representation of the model's dynamics.

1.2.4 Value functions

This paragraph will illustrate the value functions that summarize the model previously outlined. To enhance readability, in what follows the symbol $\boldsymbol{\theta}$ will represent the pair (θ_p,θ_c), and $\hat{\boldsymbol{\theta}}$ will denote the pair $(\hat{\theta}_p,\hat{\theta}_c)$. This approach, while constituting a minor abuse of notation—given that the value functions' arguments change in number depending on context—leads to a cleaner expression of equations.

Employed worker. The value of being employed for a worker of ability a at a firm $\boldsymbol{\theta}$ with a negotiation benchmark $\hat{\boldsymbol{\theta}}$ can be expressed by The employment

FIGURE 1.1: Diagram of the model



Note: A firm is a productivity-connectivity couple (θ_p, θ_c). Employed workers receive offers if the connectivities of the incumbent and challenging firm are sufficiently large. Unemployed workers receive offers with exogenous probability λ and do not reject job offers. Matches are destroyed at the exogenous rate δ. Employed workers who receive an offer decide whether to move to the challenging firm, use the outside offer to rebargain their wage with the incumbent firm or discard it, as explained in Section 1.2.3.

value for a worker of ability a at a firm $\boldsymbol{\theta}$ with negotiation benchmark $\hat{\boldsymbol{\theta}}$ is

$$\begin{aligned} W\left(a, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) = w\left(a, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) + \beta \Bigg\{ (1-\delta) \Bigg[\int_{I-\theta_c}^{\bar{\theta}_c} \Bigg(\int_{x \in \mathcal{F}_1(\theta_p, y)} W\left(a, x, y, \boldsymbol{\theta}\right) \mathrm{d}P(x) \\ + \int_{x \in \mathcal{F}_2(\theta_p, y, \hat{\boldsymbol{\theta}})} W\left(a, \boldsymbol{\theta}, x, y\right) \mathrm{d}P(x) \Bigg) \mathrm{d}T(y) \\ + \left(1 - \int_{I-\theta_c}^{\bar{\theta}_c} \int_{x \in \mathcal{F}_1(\theta_p, y) \cup \mathcal{F}_2(\theta_p, y, \hat{\boldsymbol{\theta}})} \mathrm{d}P(x)\, \mathrm{d}T(y)\right) W\left(a, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) \Bigg] + \delta U(a) \Bigg\} \end{aligned} \tag{1.4}$$

The interpretation of equation (1.4) follows the wage-setting protocol.

This equation reflects the wage-setting protocol; the value of employment is comprised of the current wage $w(a, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$, plus a discounted future value that takes into account the possibility of exogenous job loss, which happens with a probability of δ. If the worker maintains their employment and the combination of incumbent and challenger firms meets the necessary connectivity threshold, three mutually exclusive outcomes can occur: *a)* The worker may receive an offer from a firm in set $\mathcal{F}_1(\boldsymbol{\theta})$, thus opting to move to this new firm and establishing the incumbent firm as the new negotiation benchmark; *b)* The worker may receive an offer from a firm in $\mathcal{F}_2(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})$, allowing them to stay with their current employer but with an updated negotiation benchmark and wage; *c)* The worker may choose to remain in their current position with no changes to the negotiation benchmark or wage. If employment ends, the worker transitions to unemployment and receives a flow income of au_p, which they must relinquish upon gaining new employment.

Unemployed worker. The value for an unemployed worker of ability a is described by

$$
\begin{aligned}
U(a) = au_p + \beta \Bigg[\lambda \iint_{x,y\in\mathcal{F}_1(u)} & W\left(a,x,y,u_p,0\right)\,\mathrm{d}P(x)\,\mathrm{d}T(y) \\
+ \left(1-\lambda \iint_{x,y\in\mathcal{F}_1(\boldsymbol{\theta})} \mathrm{d}P(x)\,\mathrm{d}T(y)\right) U(a) \Bigg]
\end{aligned}
\tag{1.5}
$$

Unemployed workers receive an income flow of au_p. With a probability of λ, they may receive an offer which they will invariably accept. If no offers arrive, their continuation value remains the same as the current value of unemployment. It is crucial to note that the connectivity mechanism does not apply to unemployed workers. This is due to the assumption that unemployment does not carry any connectivity attributes. Furthermore, an unemployed worker will accept any job offer they receive, irrespective of the connectivity or productivity of the offering firm.

Firm. The value for a firm $\boldsymbol{\theta}$ matched with a worker a who has a negotiation benchmark $\hat{\boldsymbol{\theta}}$ is given by

$$
\begin{aligned}
J\left(a,\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\right) = {} & a\theta_p - w\left(a,\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\right) \\
& + \beta\left(1-\delta\right)\Bigg[\int_{I-\theta_c}^{\bar{\theta}_c}\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})} J\left(a,\boldsymbol{\theta},x,y\right)\,\mathrm{d}P(x)\,\mathrm{d}T(y) \\
& + \left(1-\int_{I-\theta_c}^{\bar{\theta}_c}\int_{x\in\mathcal{F}_1(\theta_p,y)\cup\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})} \mathrm{d}P(x)\,\mathrm{d}T(y)\right) J\left(a,\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\right)\Bigg]
\end{aligned}
\tag{1.6}
$$

The value for the firm includes its current profit (the match's production less the wage) and the continuation value of employing the worker. Should workers receive an offer from a challenging firm selected from $\mathcal{F}_2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}})$, they stay with the incumbent employer but with updated wages. Since matches cease with worker departure, the firm does not receive any future value once the worker leaves, which happens both in the case of exogenous separation and the worker moving to a better firm. If no offers are presented or the offer is rejected, the match remains unchanged, and the continuation value for the subsequent period is simply the discounted current value.

Joint surplus. Assuming free entry, the joint surplus generated by a worker with ability a, matched with a firm $\boldsymbol{\theta}$, can be defined as the sum of the worker's and firm's values, minus the unemployment value. Combining the three Bellman equations and applying the bargaining protocol enables us to express it as follows:

$$\begin{aligned} S(a, \boldsymbol{\theta}) = & \max \left\{0, W\left(a, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right) - U(a) + J\left(a, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\right)\right\} \\ = & a(\theta_p - u_p) + \beta \Bigg[(1-\delta) \Bigg(S(a, \boldsymbol{\theta}) \\ & + \sigma \int_{I-\theta_c}^{\bar{\theta}_c} \int_{x \in \mathcal{F}_1(\theta_p, y)} \left[S(a, x, y) - S(a, \boldsymbol{\theta}) \right] \mathrm{d}P(x) \, \mathrm{d}T(y) \Bigg) \\ & - \sigma\lambda \iint_{x, y \in \mathcal{F}_1(u)} S(a, x, y) \, \mathrm{d}P(x) \, \mathrm{d}T(y) \Bigg] \end{aligned} \tag{1.7}$$

The continuation value of the joint surplus accounts for the option value of on-the-job search, which can be delivered through both dimensions of the firm. The continuation value is the sum of the present value of the joint surplus and an additional term, which accounts for the fact that a worker transitioning to another firm not only receives the total surplus of the current match but also a fraction, denoted by σ, of the net surplus gains. It is important to note that since the present component of the value function is already net of the unemployment benefit that would be forfeited, the future value is likewise net of the optional value of search during unemployment that would be foregone. Furthermore, the surplus is independent of the negotiation benchmark $\hat{\boldsymbol{\theta}}$. This is because, under transferable utility, the distribution of the rents within the match does not change its value. Hence, wages, being a pure within-match redistribution, do not enter the equation. Finally, the surplus is strictly increasing in both θ_p and θ_c, ranking jobs across productivity and connectivity according to their appeal to workers.

Equation (1.7) governs all worker transitions, including between employment and unemployment, as well as between different firms. Crucially, these transitions are independent of the distribution of workers across different states, which considerably reduces computational effort when numerically solving the equation.

1.2.5 Wage equation

It is possible to solve the model to derive a convenient closed-form wage equation. This equation pins down the wages showing how they deliver values according to the wage-setting protocol, as delineated in Section 1.2.3 for each incumbent-negotiation benchmark firm pair. The wage equation, along with the surplus value function (1.7), oversees the earnings dynamics for each worker's labor market history, just as the surplus value function regulates worker flows.

I build the wage equation exploiting the wage setting protocol given by equations (1.1)-(1.3) together with the surplus value function given by (1.7) and the employed (1.4) and unemployed (1.5) value functions, as detailed in Appendix F. The equation reads as

$$\begin{aligned} w\left(a,\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\right) = {} & \kappa + \sigma a \theta_p \\ & + \beta(1-\delta)\Bigg(\underbrace{\sigma^2 \int_{I-\theta_c}^{\bar{\theta}_c} \int_{x\in\mathcal{F}_1(\theta_p,y)} \left[S(a,x,y) - S(a,\boldsymbol{\theta})\right] \mathrm{d}P(x)\,\mathrm{d}T(y)}_{\text{Gains from new employer}} - \underbrace{G\left(a,\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\right)}_{\substack{\text{Gains from} \\ \text{otj search}}}\Bigg) \end{aligned} \tag{1.8}$$

where k gathers all the terms that do not depend on $\boldsymbol{\theta}$ or $\hat{\boldsymbol{\theta}}$. The function $G(a,\boldsymbol{\theta},\hat{\boldsymbol{\theta}})$ encapsulates the worker's gains from on-the-job search:

$$\begin{aligned} G\left(a,\boldsymbol{\theta},\hat{\boldsymbol{\theta}}\right) = & \int_{I-\theta_c}^{\bar{\theta}_c} \Bigg(\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})} \overbrace{(1-\sigma)\left[S(a,x,y) - S\left(a,\hat{\boldsymbol{\theta}}\right)\right]}^{\text{Re-bargaining with the incumbent}} \mathrm{d}P(x) \\ & + \int_{x\in\mathcal{F}_1(\theta_p,y)} \underbrace{(1-\sigma)\left[S(a,\boldsymbol{\theta}) - S\left(a,\hat{\boldsymbol{\theta}}\right)\right]}_{\text{New negotiation benchmark}} \\ & + \underbrace{\sigma\left[S(a,x,y) - S(a,\boldsymbol{\theta})\right]}_{\text{New employer}} \mathrm{d}P(x) \Bigg) \mathrm{d}T(y) - U(a) \end{aligned}$$

The on-the-job search component delivers value to the worker through three distinct channels. First, employees can leverage viable outside options to renegotiate the current wage with the incumbent firm. Second, workers who choose to transition to the challenging firm establish a new negotiation benchmark, thereby setting a precedent for the incumbent firm. Finally, the transitioning workers gain rents from the difference between the surplus generated with the new firm and with the incumbent.

The on-the-job-search component reduces the wages as per equation (1.8), as the prospective value of searching from the firm is discounted upon transition. A similar reasoning applies when a new negotiation benchmark is established due to an outside offer from a competing firm. Furthermore, wages exhibit an inverse correlation with the firm's connectivity θ_c. This is attributed to the surplus splitting mechanism that creates compensating differentials. In essence, workers are willing to accept lower present wages in exchange for potential future opportunities arising from increased meeting probabilities, leading to quicker advancement on the job ladder either through re-negotiations utilizing outside options or through job-to-job transitions. This purely-compensating differentials effect echoes the one of Jarosch (2021), where workers trade-off wages for job security, and the one of Gregory (2020), where workers are compensated through faster rates of human capital accumulation, given their age. Still, the mechanism through which the worker improves its future value is entirely different in my model, as it entirely attains the heterogeneity in the firm-specific offers' arrival rates.

On the other hand, the relationship between a firm's productivity θ_p and wages is ambiguous and dependent on the worker's bargaining power σ. Indeed, θ_p influences wages in two significant ways: directly, where more productive firms command higher wages due to increased output; and indirectly, with more productive firms promising greater future wage growth—a compensating differential mechanism similar to the one associated with connectivity. Consider the two extreme scenarios for clarity. In the instance where workers possess no bargaining power ($\sigma = 0$), the hiring wage is set to compensate for the entire surplus from their previous employment upon transitioning to a new firm. Consequently, as the productivity type of the new firm increases, there's a larger scope for future wage growth through on-the-job search gains, which, in turn, lowers the current wage. Essentially, the firm is discounting the future wage growth it offers to the worker. Contrarily, if $\sigma = 1$, indicating that workers have the entire bargaining power, workers get the whole surplus, the on-the-job gains only become significant upon transitions, and the value delivered through wages match the employer's productivity. This signifies that more productive firms yield higher wages. This uncertain relationship between productivity and wages is a well-established outcome of sequential auctions random search models, as first presented in Cahuc et al. (2006).

1.2.6 Equilibrium

Given the exogenous distributions $A(a)$, $P(\theta_p)$ and $T(\theta_c)$, a steady-state equilibrium is:

- a surplus function $S(a, \theta_p, \theta_c)$ satisfying the Bellman equation given in (1.7);
- a worker net surplus function $W(a, \theta_p, \theta_c, \hat{\theta}_p, \hat{\theta}_c) - U(a)$ satisfying the bargaining protocol given by equations (1.1), (1.2) and (1.3);
- a wage equation $w(a, \theta_p, \theta_c, \hat{\theta}_p, \hat{\theta}_c)$ satisfying (11);
- a steady state distribution of workers across employment states such that
  - inflows of workers equate outflows of workers
  - the distribution of workers across employment and unemployment states evolves according to the wage-setting rules and the transitions determined by the surplus value function.

I borrow the convenient notation from Jarosch (2021) in calling $g(\theta, \hat{\theta})$ the density of workers employed in a firm θ with negotiation benchmark $\hat{\theta}$, $g(\theta, u)$ the density of workers in a firm θ with benchmark unemployment, and u the measure of unemployed workers. Then, in equilibrium, one has the following set of flow balances:

$$
\begin{aligned}
g^-(\theta, \hat{\theta}) &= g^-(\theta, \hat{\theta}) \left(\delta + (1-\delta) \int_{I-\theta_c}^{\bar{\theta}_c} \int_{x \in \mathcal{F}_1(\theta_p, y) \cup \mathcal{F}_2(\theta_p, y, \hat{\boldsymbol{\theta}})} \mathrm{d}P(x) \, \mathrm{d}T(y) \right) \\
g^+(\theta, \hat{\theta}) &= f(\theta) \left[\mathbb{1}_{\theta \in \mathcal{F}_1(\hat{\theta})} (1-\delta) \left(\int g(\hat{\theta}, x) \, \mathrm{d}x + g(\hat{\theta}, u) \right) \right] \\
&\quad + f(\hat{\theta}) \left[\iint \mathbb{1}_{\hat{\theta} \in \mathcal{F}_2(\theta, x)} (1-\delta) g(\theta, x) \, \mathrm{d}x \right] \\
g^+(\theta, u) &= \lambda u f(\theta) \\
g^-(\theta, u) &= g(\theta, u) \left(\delta + (1-\delta) \left[\int_{I-\theta_c}^{\bar{\theta}_c} \left(\int_{x \in \mathcal{F}_1(\theta_p, y)} \mathrm{d}P(x) + \int_{x \in \mathcal{F}_2(\theta_p, y, \hat{\boldsymbol{\theta}})} \mathrm{d}P(x) \right) \mathrm{d}T(y) \right] \right) \\
u^+ &= \delta \iint g(x, y) \, \mathrm{d}x \, \mathrm{d}y \\
u^- &= \lambda(1-u)
\end{aligned}
$$

where $\mathbb{1}$ is the indicator function.

1.2.7 Model discussion

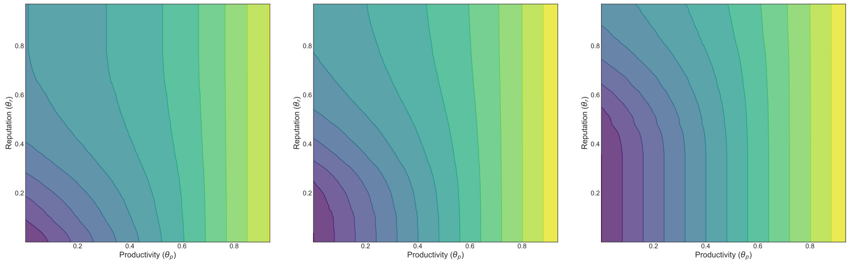
I next discuss some relevant properties of the model.

Productivity/connectivity trade-off. Workers receive value from the firm along its two heterogeneity dimensions. As equation (1.7) shows in its first term, the firm's productivity in a match directly increases the surplus generated, given the worker's ability. The higher the firm's productivity, the higher the output yielded and, therefore, the higher the value for the worker. Conversely, a higher firm's connectivity does not translate into any present worth for the workers, only increasing their future likelihood of receiving more offers. It follows that a worker values a firm's connectivity as long as it cannot convey direct value through its productivity. Thus, a less productive firm can still attract employees thanks to its connectivity, reducing their on-the-job search frictions, thereby increasing their likelihood of meeting a higher-productivity firm later on. Matching with a high-connectivity firm essentially allows the workers to climb the job ladder faster, improving their probability of meeting a 'good firm' sooner.

Figure 1.2 shows this productivity-connectivity trade-off for three different levels of the meeting threshold I. It plots indifference curves for a worker of a given ability as a function of the two firm's attributes. The Figure shows that a worker demands higher compensation in terms of connectivity for less productive firms. Notably, workers disregard the connectivity of productive-enough firms since they are satisfied with staying in an establishment that directly delivers considerable value. This property has intriguing implications for a worker's life cycle, even if the model does not account for it straight away. Indeed, workers tend to sort into higher productive firms as they climb the job ladder, implying that they attach a higher value to a firm's connectivity in the earlier stage of their career. This feature recalls the motivating evidence shown in Figure 1.3, indicating that the compensating differential channel in the entry wage is present for young workers.

Figure 1.2 also shows that the relative importance of a firm's connectivity depends on how easy it is to concrete the meeting between the incumbent and challenging firms, i.e., how easy it is to formalize the offer. The third panel of the Figure gives the intuition behind this result, displaying that the connectivity compensation for low-productivity firms is higher when it is more challenging to observe successful meetings since workers discount that low-connectivity firms will not likely impact future opportunities. Therefore, they are willing to forego more connectivity for productivity than in cases where the threshold is lower.

FIGURE 1.2: The productivity / connectivity trade-off



$I = 0.8$ $I = 1$ $I = 1.3$

Note: The figure presents isoquants (indifference curves) of worker value as a function of a firm's productivity (x-axis) and connectivity (y-axis) for a specific worker type. Each panel sets a different value for the exogenous connectivity threshold necessary for successful meetings. The figure illustrates the trade-off between a firm's productivity and connectivity. A worker values connectivity within a firm unless it directly conveys enough value through productivity. Hence, a less productive firm can attract workers thanks to its connectivity, which decreases search frictions and thereby facilitates quicker movement to more productive firms. Conversely, when workers encounter highly productive firms, they place less value on the firm's connectivity, given the already high wages—thus, the relative importance of the connectivity channel changes with the connectivity threshold. This figure is obtained through the numerical solution of equation (1.7) on an 80×80 grid, given a particular draw in ability. Productivity follows a $\mathcal{B}(5,5)$ distribution, while connectivity follows a $\mathcal{B}(3,2)$ distribution. The integrals are computed using Monte Carlo simulations with 5,000 draws.

Endogenous mobility and the job-to-job network. The meeting mechanism between the firms described in Section 1.2.2 implies that the higher the connectivity of a firm, the higher its degree centrality in the endogenous job-to-job network generated by the model. [5]

Proposition 1. The connectivity parameter c maps into the degree centrality of the network $G = (V, E)$ where vertexes V are firms and edges E are workers transitions across firms.

Proof. The proof can be found in Appendix F. □

Due to the mobility dynamics implied by the meeting mechanism, firms with

[5] Degree centrality, in graph theory, is a measure of a node's importance based on the number of edges it has, i.e., it is the number of its direct connections. I will extensively discuss the concept of degree centrality and how I use it in the context of connectivity in Section 1.3.2.

higher connectivity parameters will eventually trade more workers than firms with lower connectivity—both in hiring and relinquishing. Moreover, firms with higher connectivity will exchange workers more frequently with other similar firms. Thus, after enough iterations, the higher the θ_c parameter, the higher the degree centrality value in the job-to-job network. This is true for both out-degree and in-degree measures. I exploit this close relationship to build reduced-form results that align with predictions of the model, and to inform the parameters related to the connectivity's distribution when estimating the model, as detailed in Section 1.4.1.

Sorting. Since the production function is additively separable in my setup, the model has no predicted sorting. Indeed, there are no complementarities between the workers' ability and the firms' productivity. Given two firm types θ and θ', it never happens that θ is preferred to θ' for some workers and the other way around for others. All matches generate a positive surplus, and there exists a wage always acceptable for every worker-firm couple.

1.3 Two key insights from the job-to-job network

This section provides reduced form evidence of some implications of the model that leverage the link between the connectivity parameter and the firm's centrality in the job-to-job network. First, it presents the data sources used in this exercise. Following this, it outlines the construction of the job-to-job network and explains how it can be used to identify firms that play a relatively more significant role in employee transitions in terms of network centrality. Subsequently, it provides evidence of a compensating differential mechanism associated with firms' centrality, specifically for younger workers. This mechanism suggests that employees are willing to accept lower wages in exchange for other benefits, such as better career prospects, when employed by firms with high centrality. Finally, I demonstrate that this mechanism is beneficial, showing that workers who leave high-centrality firms earn more than those who exit "regular" firms. This suggests that the trade-off workers make when accepting lower wages at high-centrality firms can lead to higher earnings in the long run. Both of these behaviors are coherently predicted by the model.

1.3.1 Matched employer-employee data

This paper relies on confidential administrative datasets provided by the Italian National Social Security Institute (Istituto Nazionale di Previdenza Sociale, INPS). More specifically, it draws on a comprehensive, matched employer-employee dataset comprising monthly-level data for all non-agricultural private firms in Italy that employ at least one salaried worker. Each worker-firm record provides detailed insight into various aspects of the employment match, including contract start and end dates, reasons for commencement or termination, contract type, work schedule, employee's occupational category, earnings, and actual days worked. This employer-employee dataset is supplemented with additional detailed data at both the worker level—like demographic characteristics—and the firm level—such as industry, location, and key dates of the firms' lifespan.

The analysis is restricted to active contracts from 2008 to 2020, specifically those in firms which employed no fewer than 15 workers at least once during this period. I concentrate on large firms, arguing they can better convey connectivity value compared to smaller firms, which often lack the necessary organizational infrastructure for significant connectivity amenities.[6] My sample includes full-time, permanent contracts among employees who held a white-collar position within a sample firm for a minimum of one year during the period under analysis. The goal is to focus on employees who play pivotal roles in their firms' operations and stand to gain most from connectivity channels. The monthly earnings of an employee are unaffected by transitory shocks such as leaves of absence and bonuses.

Appendix B details the data cleaning decisions. This process ultimately yields a panel of 2,742,853 workers across 197,347 firms between 2008 and 2020.

1.3.2 Job-to-job network

I exploit the panel structure of the described dataset to obtain detailed information about worker movements between firms throughout the sample period. The goal of this process is to identify firms with the widest variety of sources and destinations for job-to-job transitions—i.e., firms that attract and dispatch employees to the most diverse range of other employers. To achieve this, I reconstruct the job-to-job transition network from the panel, where firms are nodes and directed links between nodes represent movements of workers between firms.

I exploit the panel structure of the dataset just described obtaining detailed information on the workers' flows between firms during the sample period. The target of this process is to identify those firms that exhibit the widest variety

[6]More precisely, I further restrict the sample to workers that worked exclusively in large firms during the reference period.

of sources and destinations regarding job-to-job transitions—i.e., those firms that receive workers from and send workers to the most extensive assortment of other employers. To do so, I reconstruct the network of the job-to-job transitions from the panel. Here, firms are nodes, and directed links between nodes are movements between firms. More formally, the job-to-job network $N = \{V, E\}$ consists of the set of nodes V—the firms in the sample involved in at least one job-to-job transition in the reference period—and a set of links E—the workers' movements between firms. An adjacency matrix A can represent this network, where $A_{ij} = 1$ if at least one worker moves from firm i to firm j.[7] I define a movement as a worker changing between two firms within no more than two months from quitting the old firm and starting at the new one. The granular information on the reason behind a spell's start or end allows me to identify proper job-to-job movements, distinguishing them from layoffs or changes in the firm's identifier due to internal reorganization.[8] The job-to-job network is a *directed* network since the connections between its nodes are directional and, in general, $A_{ij} \neq A_{ji}$.

Descriptive statistics: panel and network

Descriptive statistics of the job-to-job transitions are detailed in Table A.1 for the entire period, with further breakdown into four-year sub-periods. The demographic composition of workers transitioning between employers remains consistent over time, with a notably low proportion of women involved in movements (Rubolino, 2022). Both average tenure and age at the time of a standard movement increased over time, whereas the age at the first movement decreased by nearly 1.5 years from the 2008-2011 to 2017-2020 periods. Workers showed a decreasing trend in transitioning within the same industry and province, indicative of broadening labor markets. Interestingly, minor average wage cuts (around 3% across the entire period) are associated with movements, possibly reflecting changes in the non-monetary dimension of job value (Caplin et al., 2022). In total, nearly 1.5 million workers transitioned across at least two firms between 2008 and 2020.

Table A.2 provides descriptive statistics for the network, once again segmented

[7] I abstract from the network weights, i.e., the strength of links based on the number of workers flowing from one firm to another. I will relax this when considering centrality measures that account for weights, considering an adjacency matrix A such that $A_{ij} = k$ where $k > 0$ is the number of workers flowing from i to j in the reference period.

[8] For example, a firm changing its business name or tax code for fiscal reasons changes its identifier in the administrative data, potentially leading to an *apparent* job-to-job transition, even though the worker remains in the same firm. Given the information I have on the motivation behind a movement, I avoid this risk.

into four-year sub-periods. The number of between-firm movements declined over time, mirroring the decrease in nodes and links within the sub-networks. The table also outlines the number of connected components in the network. A connected component consists of a network's subset of nodes, such that a path connects each pair. In this paper, my focus is on the largest connected set or the connected component containing the most nodes. Such a restriction is meant to focus on the most significant part of the network, where the centrality analysis I propose is most relevant. Additionally, restricting to this sub-network incurs a small sampling cost, as the largest connected component comprises 91.1% of nodes and 99.2% of links.

Table A.3 presents descriptive statistics of the sample in the entire panel and within the largest connected component. Firms in the latter are, on average, larger but younger. The share of firms operating in the services sector is higher in the largest component than in the full panel—however, more than two-thirds of employers are still in manufacturing. Demographic characteristics, as well as average tenure and experience, remain consistent across both panels. Workers in the largest connected set earn marginally more, reflecting the presence of larger firms. Overall, the descriptive evidence demonstrates significant consistency between the entire panel and the largest connected component in the job-to-job network.

Firms' centrality in the network

I employ the job-to-job network to examine the extent to which firms intermediate worker flows across a variety of sources and destinations. Proposition 1 links a firm's connectivity parameter in the model to the degree centrality of the job-to-job network. Degree centrality (Freeman, 1978) is the number of other nodes each node connects to in the directed network. Formally, for a node i in a network with total nodes N, it reads as

$$D(i) = \sum_{j}^{N} A_{ij}$$

where A is the adjacency matrix and $A_{ij} = 1$ if a link between i and j exists. Conceptually, a firm's relative significance depends on its capacity to link workers with other firms—thus effectively "controlling" the flows between employers.

Degree centrality offers a simple measure of a node's participation in a network, as it relies solely on the local structure surrounding it. In the case of a directed network like the job-to-job one, it is natural to divide degree centrality into in-degree and out-degree—in this context, the number of in- and out-going links a firm has.[9] To further clarify, a firm that sends workers to 10 different firms over the sample period will have an out-degree centrality of 10. Likewise, a firm that receives workers from 8 different firms within the same period will have an in-degree centrality of 8. In the case of an unweighted network—one that does not account for the strength of links between nodes—degree centrality only considers the variety of connections each node possesses, disregarding the *intensity* of worker flows.

The empirical analysis in this paper will primarily focus on out-degree centrality, as it best conceptually aligns with the model's connectivity parameter. I interpret out-degree centrality as an empirical measure of the variety of job offers received by employees of a given firm over time. Still, taking into account in-degree centrality helps distinguish firms that are frequently left for a multitude of other destinations due to their low quality (i.e., workers might end up there due to labor market frictions and wish to leave as soon as possible) from those that workers deliberately choose. However, larger firms often have, on average, greater degree centrality for mechanical reasons unrelated to the economic intuition behind what I have termed 'connectivity': in this paper, I am not examining the relationship between an employer's size and worker wages. Thus, I re-scale each firm's centrality by its average number of employees during the sample period—a proxy for its size. This adjustment ensures that a firm is considered more central in the network if it truly maintains richer connections with other firms, rather than just being larger.

Table A.4 presents summary statistics of the worker-firm panel, categorized by quartiles of normalized out-degree centrality. On average, firms with high centrality employ younger workers, initiate contracts earlier, and offer higher salaries compared to firms at the lower end of the out-degree centrality distribution. Moreover, firms with greater centrality tend to hire more foreign workers and fewer female workers. Notably, firm-specific tenure decreases with centrality, potentially suggesting that central firms serve as "springboards" for workers' future career trajectories. In addition, Figure A.1 displays the mean values of relevant financial measures from the Cerved database, categorized by the ventile of both out- and in-degree centrality. Firms with higher centrality show lower levels of tangible assets and net purchases, indicating a prevalence of intangible, service-related activities. Simultaneously, these firms exhibit higher intangible and financial assets, as well as increased liquidity and profitability indexes.

[9] As discussed by Borgatti (2005), degree centrality measures are particularly suited for walk-based transfer processes along the graph, which applies to job-to-job networks.

While my primary focus is on unweighted in- and out-degree centralities for their tractability and intuitive alignment with the model, I also incorporate other measures considering flow intensity in my supplementary results. Specifically, I evaluate two other degree centrality measures: weighted degree centrality and Opsahl. These additional measures are detailed in Appendix **??**. In Table A.5, I provide different average degree centrality measures organized by industry. The average centrality ranking by industry remains largely stable across these measures. As expected, the most central firms are predominantly in the service sector (information and communications, financial services and insurance, accommodation and food services). Firms with the lowest centrality are typically found in heavy industries (mining and quarrying, water supply and waste management, transports) and education, where connectivity effects are arguably less crucial. Consistency is maintained when modifying the normalization criterion, transitioning from average to maximum employee count.

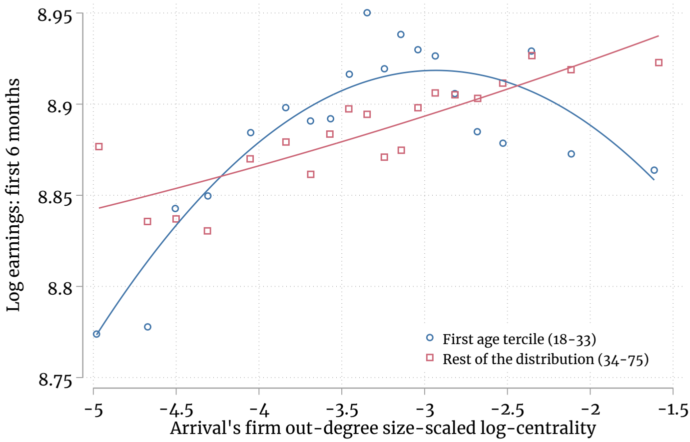
1.3.3 Trading entry wage for connectivity

I now turn to document the presence of substantial heterogeneity in the firms' paying schemes at hiring, depending on their centrality and the age of the hired workers. This exercise aims to build an empirical case showing that firms partly pay young workers with their connectivity, i.e., the higher likelihood of finding a new and better employer in the future. I focus on out-degree centrality as the main degree centrality measure while providing evidence of robustness across different gauges in Appendix E.

Figure 1.3 displays the logarithm of a worker's initial earnings in a new job role in relation to the normalized log out-degree centrality of each firm. For improved graph readability, I've binned the centrality and plotted the average variables per bin. Each line fits a quadratic regression on the 20 bins into which centrality is divided. The direction of the relationship changes along the age distribution of the hired worker. The starting salary of younger workers (blue line, circles) follows a U-shaped relationship. For the lowest part of the centrality distribution, a firm's connectivity doesn't convey particular value, and the relationship with the hiring salary is not dissimilar from the one for older workers. However, as age increases, the relationship inverts and becomes negative: hiring log earnings decrease with an increase in the log centrality of the hiring firms. I interpret this relationship as evidence of the compensating differential described by my model.

The variation in the slopes between age groups suggests that a firm's connectivity in the job-to-job network plays different roles at different stages of a worker's

Figure 1.3: Entry earnings and firms' out-degree centrality, by hiring age



Note: The graph displays the average hiring log earnings, adjusted for worker-firm observables, plotted against the log out-degree centrality of the hiring firm, normalized by its average employer size over the period. Each line signifies a binned quadratic regression for two separate age groups: 18-33 (represented by the blue line and circles) and 34-75 (the red line and squares). There is substantial variation in the hiring earnings of workers depending on the firm's centrality. For younger workers, the hiring wage decreases with the number of unique connections the hiring firm possesses in the job-to-job network, once this measure gains significance. For older workers, this relationship becomes linear and positive. This variability is interpreted as indicative evidence of a compensating differential channel, suggesting that workers may balance immediate earnings against future opportunities. Source: Uniemens data, Istituto Nazionale della Previdenza Sociale (INPS).

career. I argue that this variation serves as evidence of a compensating differential mechanism in which a firm provides value to a worker in terms of future possibilities outside the firm itself. Younger workers, in particular, are more inclined to trade off a portion of their salary for the value delivered by better firm connections. Conversely, older workers, who are in the later stages of their careers, value immediate earnings more than future opportunities.

Of course, the heterogeneity reported in Figure 1.3 might be due to other possible mechanisms. For example, firms' workers turnover or industry characteristics may well explain differences in wages across centrality, which would capture some other firm's dimensions. To exclude these and other possible explanations, both measures are residualized for some observable characteristics that may explain the differences in hiring wages: a second-degree polynomial for age and dummies for the year, nationality, sex, destination province, and industry. Appendix E reports different specifications of the hiring wages-centrality relationship. Qualitatively, results remain strongly consistent when when considering daily wages rather than quarterly earnings (Figure E.1).

1.3.4 Is trading wage for connectivity worth it?

Following the previous discussion, a natural question arises: Is it beneficial for workers to accept lower earnings today in exchange for potentially higher earnings in the future due to increased opportunities? This question is tied to the notion that working for high-connectivity firms may yield higher rewards when leaving them, compared to regular (i.e., non-high-connectivity) firms, as per the opportunities of faster climbing of the job ladder.

A practical approach to answering this question is to compare workers' earnings and wages in the period after they leave these two categories of employers. By doing so, we can gain insights into the potential long-term benefits of accepting lower entry earnings at high-connectivity firms.

A data-driven procedure to identify high-connectivity firms

To precisely delineate the correlation between wages and earnings and the propensity to leave a highly-central firm, the sample of employers is divided into two distinct groups: high-connectivity and regular firms. This segregation is carried out using a k-means clustering algorithm applied to various degree centrality measures ascribed to firms.[10]

Specifically, I address the distance-minimization problem outlined as

$$\arg\min_{k_1,k_2} \sum_{i=1}^{K=2} \sum_{j\in k_i} \|\boldsymbol{C}(j) - \boldsymbol{\mu}_i\|^2$$

Here, k_1 and k_2 represent the $K = 2$ clusters, $\boldsymbol{C}(j)$ denotes the degree centrality vector for firm j, on which the algorithm clusters, and μ_i is the mean vector of centralities within cluster k_i. The distance between the centrality vector and the mean vector is calculated using the L-2 norm. The underlying rationale for this partitioning process is to allocate each firm to a cluster in such a way that

[10]The k-means algorithm is chosen for three primary reasons: a) It is an unsupervised learning algorithm, ensuring the procedure is entirely data-driven; b) Its simplicity and intuitive nature promote clear understanding; c) Its extensive application in social sciences (Steinley, 2006) and particularly in economics (for example, Bonhomme et al., 2019) make it a well-established choice.

minimizes the within-cluster variance of the two centrality measures while maximizing the variance between clusters. Although k-means is an unsupervised algorithm, it presumes the number of partitions (in this case, $k = 2$) as a constant. Further reasoning supporting the choice for this two-group split, along with other insightful details on the process, can be found in Appendix D.

The primary clustering measures are the out-degree centrality alone and both in- and out-degree centrality. The preferred normalization is made by using the average firm size during the sample period. When clustering is based solely on out-degree centrality, the k-means algorithm divides the sample into two significantly different sizes: the high-connectivity firms account for 12.4% of the sample, while the remaining 87.6% fall under the regular category. Similar results are observed in the bi-dimensional clustering scenario, which also includes in-degree centrality, where the high-connectivity cluster comprises 15.2% of the employers in the sample. Table A.6 illustrates the top five representative industries within each group, considering the out-degree centrality. It is noteworthy that this entirely data-driven procedure identifies intuitively appropriate sectors among the high-connectivity firms, primarily consultancy and professional activities. This finding serves as descriptive validation, supporting the general premise underlying the mechanism I am investigating.

How wages change upon leaving a high-connectivity firm

I now turn to show that leaving a high-connectivity firm as the first job-to-job transition pays more than leaving a regular firm. To do so, I confront workers at a generic employment transition in their career in a dynamic differences-in-differences setting in which treated units are employees leaving a high-connectivity firm, and controls are those leaving a regular firm, as previously defined. In particular, I estimate the specification

$$\begin{aligned} y_{it} = & \sum_{\substack{k=-4 \\ k \neq -1}}^{8} \beta_k \times \mathrm{HR}(i, o) \times \mathbb{1}(t - t_i^* = k) \\ & + \alpha_i + \tau_t + \gamma_{s^O(i),t} + \eta_{s^D(i),t} + \psi_{p^O(i),t} + \lambda_{p^D(i),t} + \varepsilon_{it} \end{aligned} \tag{1.9}$$

where the dependent variable y_{it} is either the quarterly log-earnings of individual i at calendar time t, or his/her daily log-wage. The coefficients of interest $(\beta_k)_{-4 \leq k \leq 8}$ measure the impact on earnings of exiting from a high-connectivity firm with respect to exiting from a regular one, k periods away from the movement. HR_i is a dummy equal to one if the origin firm o that individual i leaves is a high-connectivity firm, and $\mathbb{1}(t - t_i^* = k)$ is an indicator function equal to

1 at time t if worker i is k periods away from having left the firm in t_i^*. The error term is captured by ε_{it}. I also include worker (α_i) and time (quarter) (τ_t) fixed effects, and I allow for non-parametric time trends in the origin firm's sector ($\gamma_{s^O(i),t}$) and province ($\psi_{p^O(i),t}$) and destination's firm sector ($\eta_{s^D(i),t}$) and province ($\lambda_{p^D(i),t}$). These latter inclusions allow me to control for heterogeneous slopes in the firm- and province-specific time trends of both departure and arrival firms that might explain a relevant component of the variation in observed wages and earnings due to the job-to-job transition.

Figure E.2 shows the estimation results of the baseline specification given in (1.9). Panel A reports the estimated coefficients for the specification using quarterly log earnings, while Panel B does so with daily log wage. Each Panel reports the bi-dimensional clusterization based on in- and out-degree centrality (blue line, circles) and the one on out-degree only (red line, squares). Standard errors are clustered at the worker-by-quarter level. Overall, results suggest that leaving a high-connectivity firm pays off. Workers that move away from such employers earn, on average, around 5% more than workers leaving regular firms after 1.5 years from the movement. Wages move similarly, showing a gap of around 7% between the two groups of employees. After around 1.5 years, the difference remains stable, if not slightly declining. Moreover, Figure E.2 suggests that out-degree centrality matters more than in-degree: sorting into a firm with a higher degree of connection in exit pays more for future wages than one with stronger entering connections. This is a very natural result, given the economic intuition behind the connectivity mechanism. Appendix E reports different robustness specifications, showing remarkable consistency across different specifications.

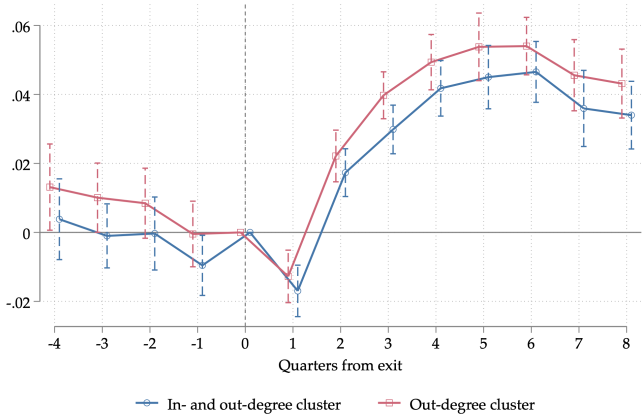
Of course, claiming causal identification in this specification is made hard by selection concerns since it may be the case that workers with higher ability systematically self-select into higher-connectivity firms. Still, results reported in Figures 1.3 and E.2 are notably coherent with the evidence of compensating differentials. They show that young workers accept lower wages when entering central firms in the job-to-job network for an investment of future value and that such investment is rebated when they leave those firms. The remainder of this paper is devoted to investigating further the economic mechanism behind these differences in wage dynamics.

1.4 Estimation

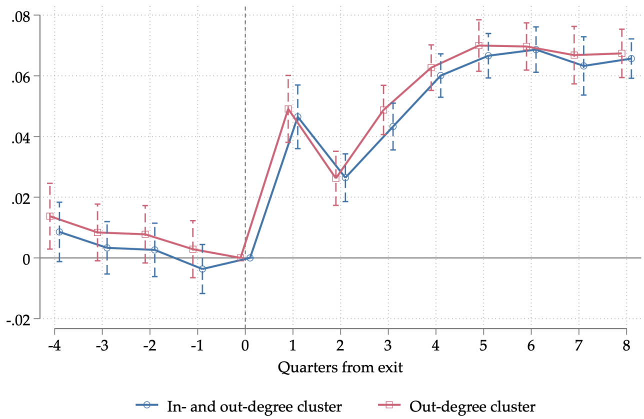
Having discussed the model qualitatively and shown reduced form evidence of some of its properties, I now turn to its quantitative implications on work-

Figure 1.4: Leaving a high-connectivity firm vs. a regular one

(A) Quarterly earnings



(B) Daily wage



Note: These graphs show the relationship between leaving a high-connectivity firm on a worker's quarterly log earnings (A) and daily log wages (B). Blue lines report results for clustering on in- and out-degree centrality; red lines with only out-degree. Coefficients are obtained estimating the specification in (1.9). 95% confidence intervals are obtained by clustering at the individual-by-quarter level. Source: Istituto Nazionale della Previdenza Sociale (INPS).

ers' wage dispersion. For that, I estimate the model at a quarterly frequency on the same administrative dataset provided by INPS already detailed in Section 1.3.1. I do so using the so-called *Indirect Inference* technique. At its core, it comes down to a Simulated Method of Moments (McFadden, 1989) that uses auxiliary reduced-form specifications to provide more refined sets of moments to match, minimizing the distance between the data-generated ones from those that are model-generated.[11] More formally, I solve

$$\hat{\phi} = \arg\min_{\phi} \left\{ \left(m(\phi_0) - \tilde{m}(\phi_0) \right)' W \left(m - \tilde{m}(\phi) \right) \right\}$$

where $\hat{\phi}$ is the $K \times 1$ estimated parameters vector, $m(\phi_0)$ is the $N \times 1$ vector of moments derived from data as a function of the true parameter values ϕ_0, $\tilde{m}$ is its simulated counterpart, and W is an appropriate weighting matrix.[12] K is the number of parameters of interest, while N is the number of targeted moments.

Parametrization The model is fully parametrically estimated under some assumptions. First, as Jarosch (2021), I parametrize the marginal distributions governing firms' heterogeneity as betas: $\theta_p \sim \mathcal{B}(a_p, b_p)$ and $\theta_c \sim \mathcal{B}(a_r, b_r)$.[13] Moreover, I set the ability distribution as a standardized log-normal: $a \sim \log \chi(1, \sigma_a^2)$. For the numerical solution of the model, I approximate the employers' productivity and connectivity distributions on 50 gridpoints each. I similarly approximate workers' ability on 7 gridpoints. Therefore, numerically solving Equation (1.7), I build a multidimensional grid on which I will interpolate surplus value when simulating data. Finally, all along the estimation, I assume the model is in steady state, i.e., workers' inflows and outflows across states are balanced.

1.4.1 Identification

I next discuss how to identify the different parameters of the model that determine its outcomes. Even if the estimation is done jointly, it can be useful to heuristically discuss how different sets of moments inform different parameters.

In particular, in what follows, I assess the problem of informing the three components of wage dynamics to identify them separately: the worker-specific one,

[11] The main reference for the estimation through indirect inference is Gourieroux et al. (1993), but the method can be easily seen as a generalization of the simulated method of moments.

[12] Since I build the vector of moments differences as relative deviations to avoid the risk of heterogeneous moments weighting, I rely on a simple yet consistent $K \times K$ identity matrix.

[13] For the moment, I assume the two are independently distributed. I plan to relax this assumption in future updates of this work.

governed by the ability's distribution $A(a)$, and the two firm-specific components, governed by the connectivity's distribution $T(\theta_c)$ and the productivity's distribution $P(\theta_p)$. I further comment on linking some model features to convenient metrics in the data to disentangle the two firm-specific components. Finally, I present the set of moments that inform the other more standard parameters that do not directly relate to the model's agents.

Worker-specific determinants of wage heterogeneity

First, I want to separate the effects on wages of workers' heterogeneity from those of firms' heterogeneity to account for variation in workers' abilities. To do so, I adopt a two-way fixed effect specification (Abowd et al., 1999, AKM), running the following regression in the data

$$\log w_{it} = \alpha_i + \psi_{j(i,t)} + \gamma_t + X_{it}\beta + \varepsilon_{it} \tag{1.10}$$

where α_i, $\psi_{j(i,t)}$ and γ_t are the worker-, firm-, and time-fixed effects, respectively, and X_{it} is a set of time-varying worker characteristics that comprehend polynomials of degree two for age, experience, and dummies for worker's qualifications. Since I do not explicitly consider these characteristics in the model, I run the exact specification on the simulated data, including experience-related only time-varying controls.

Equation (1.10) is identified only within connected components of the job-to-job network. As I have outlined in section 1.3.2, I have restricted my sample to the largest connected component, which includes 98.5% of the employees' transitions (Table A.2). Therefore, I do not need to operate any further intervention on my data to accommodate the connected-set requirement.

Moreover, as extensively discussed in the most recent literature on these models, variances of the fixed effects estimates tend to be biased upward due to limited mobility in (both real and simulated) data: intuitively, since identification comes from workers switching their job, if too few workers move, fixed effects estimate may result to be noisy—thus, increasing the variance. Papers like Bonhomme et al. (2019) and Kline et al. (2020) have directly tackled this problem suggesting possible corrections that, while substantially different, end up being computationally costly if embedded in an SMM procedure. I indeed adopt the same approach as Gregory (2020) in only reducing the differences between the simulated data (a strongly-balanced panel) and the real one, randomly truncating workers' histories to mimic the average job experience that comes from my sample.

I estimate the reduced-form specification in (1.10) on real data for two consecutive sub-samples: 2006-2012 and 2012-2018. I do the same in the simulated panel, splitting it into two groups of 24 periods, matching the quarterly structure of real data. For each sub-sample, I then target the mean and the variance of the workers' fixed effects. Since my model does not predict sorting across any dimension of heterogeneity, I do not target the correlation between the two.

Firm-specific determinants of wage heterogeneity

connectivity. As discussed among the implications of the theoretical setting, the model predicts a job-to-job network where firm centrality maps into their connectivity. Therefore, I target different moments of the out-degree centrality to precisely inform these parameters. In particular, I match the mean, the variance, and the interquartile range of the observed ratio distribution between the two network centrality measures. Moreover, the job-to-job (EE) rate is another crucial source of identifying the connectivity's distribution parameters since connectivity governs the meeting rate of firms—and thus, the likelihood of a movement to take place. Among the parameters informed by this set of moments, I also include the connectivity threshold I.

Productivity. I discipline the parameters governing the heterogeneity in the productivity of the firms targeting several wage moments, as in Bagger et al. (2014), Gregory (2020), and Jarosch (2021). In particular, I exploit wage changes between and within jobs: for the latter, I use the average wage change upon a job-to-job transition, while for the former, I use the average quarterly change in wages for stayers and the average wage change from the start to the end of a spell. Moreover, I target the interquartile range of the wage distribution. Clearly, these moments also convey information on the bargaining power parameter σ, which governs the magnitude of the wage responses both to employer changes and outside offers that lead to a renegotiation of the current compensation.

Other parameters

I exploit moments related to standard labor market flows to identify the job destruction rate. More specifically, since this parameter is exogenously set in the model, the unemployment-to-employment (EU) rate perfectly informs it. In particular, I calculate the period-specific rate and target its mean over the sample period. Since my dataset does not allow for observing unemployment-to-employment transitions directly—it is impossible to distinguish a worker in unemployment from one self-employed or working, for example, in the pub-

TABLE 1.1: Estimated parameters and targeted moments

PARAMETER	DESCRIPTION	ESTIMATE	TARGETED MOMENT(S)	MODEL	DATA
			Panel A. Externally set or normalized		
β	Discount factor	85%	Herkenhoff et al. (2018)	-	-
λ	Job finding rate in unemployment	0.273	Gregory (2020)	-	-
			Panel B. Internally estimated		
δ	Job destruction rate	1.43%	EU rate	0.0143	0.0143
σ	Workers' bargaining power	47.2%	Wage iq. range	0.6738	0.596
			Between-job wage change	0.0011	-0.01
a_p, b_p	Firm's productivity distribution	7.574, 6.977	Avg. quarterly wage change	0.0003	0.0003
			Avg. wage change within spell	0.0110	0.1072
a_c, b_c	Firm's connectivity distribution	1.299, 3.989	EE rate	0.0042	0.0149
			Avg. degree centrality	0.0496	0.0426
I	Meeting threshold	0.659	Degree centrality var.	0.0031	0.0027
			Degree centrality iq. range	0.06	0.0562
σ_a^2	Worker's ability distribution	0.299	Workers FE var. (1st period)	0.1052	0.0792
			Workers FE var. (2nd period)	0.0906	0.0802

Note: Parameters within each group of moments are always estimated *jointly*. All moments except for the AKM-derived ones are computed for each time cross-section and then averaged over the periods.

lic sector—I externally set the job-finding rate in unemployment. The workers' bargaining power, σ, is estimated together with the productivity distribution's parameters. Finally, as in Engbom (2020) and Gregory (2020), I externally set the discount factor, β, to a 3.75% quarter rate, following Herkenhoff et al. (2018).

1.4.2 Results

Table 1.1 provides a comprehensive summary of the estimation results. It reports the descriptions and estimated values of parameters, accompanied by the corresponding targeted moments in the model and real data. While moments are categorized in accordance with the identification arguments presented in Subsection 1.4.1, all parameters are estimated jointly. On the whole, the model fits the data reasonably well. Specifically, it exhibits an excellent fit with the centrality moments pertaining to firms' centrality in the job-to-job network, closely matching the values observed in the data. This not only affirms the accuracy of network centrality measures mapping into parameters governing firms' connectivity distribution but also signifies that the firm-specific contribution to the reduction in search friction can quantitatively account for observed job-to-job transitions in real data. Additionally, the model demonstrates a satisfactory fit with the wage dynamics moments and predicts average positive between-job wage changes, albeit minuscule. The match for the AKM moments is qualitatively good, suggesting that the model can adequately explain wage dynamics in relation to firm- and worker-specific attributes. Finally, my framework generates markedly lower job-to-job transitions, a prediction that is reflected in the mismatch in the EE rate. One significant limitation of my model is the lack of a dynamic for individual human capital. Incorporating this additional element will likely facilitate an even more precise fit with the data.

1.5 Conclusions

This paper pioneers an exploration into the firms' role in mitigating search frictions for their employees. Within my conceptual framework, I propose that employees may value a trade-off where they accept lower wages from their current employer in return for a higher likelihood of securing a better job match in the future. In this setting, for the first time, the likelihood of receiving an offer while searching on the job attains to the characteristics of the firm rather than the worker's effort.

In the first part of the paper, I draw upon a job-search model where firms are heterogeneous in their productivity—governing the match's output—and connectivity—governing the arrival rate of outside offers for employed workers. The model illustrates how workers balance productivity against connectivity, exploiting the latter for its capacity to accelerate career progression, allowing workers to climb the job ladder faster. I provide a theoretical link between the connectivity parameter in the model and the degree centrality of firms in a mobility network where nodes are employers and links are employees transitions in a given period.

Using a unique Italian administrative dataset, I present reduced-form evidence signifying a negative correlation between young workers' hiring earnings and the job-to-job network centrality of the firms hiring them. Subsequently, I demonstrate that employees transitioning from highly-central firms tend to earn more than those leaving regular companies. These two pieces of reduced-form evidence coherently align with my model's prediction, providing evidence of a relevant compensating differential channel that allows workers to climb the job ladder faster.

I then estimate the model on data, matching a set of moments that allows for joint identification of the relevant parameters. The model nicely fits the data, not only strengthening the validity of the channel I explore but also opening possibilities for further investigation using experiments from simulations. Such an analysis could facilitate insightful breakdowns of life-cycle earnings trajectories and their variance, providing a deeper understanding of the impact of the connectivity channel. I plan to incorporate this enhancement shortly.

Finally, although my theoretical framework sheds fresh light on a previously underexplored mechanism explaining wage disparities among workers, it has two primary constraints. Firstly, it currently overlooks the dynamics of human capital—a recognized determinant of wage fluctuation throughout an individual's career. Secondly, my model treats firms' connectivities as exogenously determined, thereby offering limited insight into the economic mechanisms that drive the heterogeneous likelihood of firms attracting and dispatching workers to various sectors. I aim to address both these concerns in an upcoming iteration of this study.

A Additional figures and tables

Table A.1: Descriptive statistics of the transitions in the panel

	2006-2009	2009-2012	2012-2015	2015-2018	2006-2018 (whole)
Demographics					
share of movements made by females	0.32	0.32	0.32	0.32	0.32
	(0.467)	(0.468)	(0.467)	(0.467)	(0.467)
share of movements made by Italians	0.95	0.95	0.95	0.95	0.95
	(0.209)	(0.218)	(0.217)	(0.213)	(0.213)
Tenure when leaving	1.58	2.73	3.49	3.90	2.63
	(0.994)	(1.858)	(2.912)	(3.639)	(2.519)
Avg. age at					
generic movement	36.0	37.0	37.0	37.4	36.7
	(9.144)	(9.417)	(9.641)	(9.910)	(9.482)
first movement	35.8	35.8	35.0	34.4	35.4
	(9.176)	(9.434)	(9.625)	(9.707)	(9.445)
Share of movements within the same					
2-digits industry	0.44	0.42	0.35	0.33	0.40
	(0.497)	(0.494)	(0.477)	(0.469)	(0.489)
province	0.47	0.43	0.45	0.43	0.45
	(0.499)	(0.496)	(0.497)	(0.495)	(0.497)
Daily wage					
when leaving	123.0	131.8	130.3	131.7	128.1
	(103.9)	(133.1)	(148.2)	(109.3)	(120.9)
when arriving	120.7	124.4	122.6	129.5	124.0
	(88.31)	(99.45)	(103.7)	(108.9)	(98.79)
difference	-2.29	-7.43	-7.70	-2.22	-4.07
	(94.50)	(130.0)	(146.2)	(111.2)	(117.4)
Movers	642,130	453,518	350,350	352,809	1,494,080

Note: the table reports selected descriptive statistics regarding the job-to-job transitions in the panel. Each column reports summaries for a four-years subperiod, while the last one considers the entire time sample. Source: Istituto Nazionale della Previdenza Sociale (INPS).

Table A.2: Nodes and links in the job-to-job network

	2008-2011	2011-2014	2014-2017	2017-2020	2008-2020 (complete)
Number of nodes	120,121	111,555	105,651	90,241	198,036
Number of links	965,195	886,608	870,499	676,010	1,936,383
Number of connected components	8,210	8,732	7,658	6,390	7,709
In the largest connected component					
% of nodes	84.6	82.1	83.5	84.1	91.1
% of links	98.3	97.8	98.0	98.0	99.2

Note: Number of nodes and links in the job-to-job network, between 2006 and 2018, with four sub-periods breakdown. Each node is a firm that has experimented at least one employment transition in the sample period and each link is a transition. The largest connected component is the maximal set of nodes such that each pair of them is connected by a path, which accounts for nearly all the observed movements between firms. Source: Istituto Nazionale della Previdenza Sociale (INPS).

Table A.3: Descriptive statistics of the panel

	Whole panel		**Largest CC**	
	Mean	SD	Mean	SD
Firm size	13.89	333.12	14.33	404.62
Firm age	16.36	13.02	15.83	12.99
Sh. in manufacturing	0.67	0.47	0.64	0.48
Sh. in services	0.29	0.45	0.31	0.46
Sh. of females	0.36	0.48	0.35	0.48
Sh. of italians	0.97	0.18	0.97	0.18
Sh. of under-35	0.28	0.45	0.27	0.45
Sh. between 35 and 55	0.64	0.48	0.64	0.48
Sh. of over-55	0.08	0.28	0.08	0.28
Tenure	4.49	2.29	4.49	2.27
Experience	18.64	10.02	18.67	10.00
Monthly wage	3,503.10	2,160.46	3,581.72	2,169.52
Number of workers	2,742,853		2,577,544	
Number of firms	197,347		179,585	

Note: the table reports selected descriptive statistics for the whole panel (first column) and the largest connected component in the job-to-job network (second column). Source: Istituto Nazionale della Previdenza Sociale (INPS).

TABLE A.4: Descriptive statistics of the worker-firm panel, by quartiles of normalized centrality

Normalized centrality quartile	Age	Age at hiring	Quarterly wage	Firm-specific tenure (quarters)	Females	Italians	Num. of workers	Num. of firms
1	45.41 (9.49)	42.15 (9.48)	10,243 (5,976)	35.50 (15.73)	35.4%	98.0%	933,710	44,897
2	42.04 (9.52)	39.64 (9.68)	10,292 (6,475)	36.11 (15.48)	34.1%	96.3%	1,019,827	44,896
3	40.80 (9.45)	38.43 (9.20)	10,608 (7,036)	35.05 (16.25)	34.7%	96.3%	799,927	44,896
4	39.86 (9.15)	38.04 (8.88)	11,443 (7,884)	33.58 (16.09)	32.1%	96.7%	979,254	44,896
Entire Panel	42.08 (9.64)	39.62 (9.46)	10,654 (6,896)	35.06 (15.89)	33.9%	96.8%	2,742,853	179,585

Note: the table reports firm-level means and standard deviation (in parenthesis) of selected measures by quartiles of out-degree centrality, normalized by average firm's size. The panel comprises full-time, white collars workers in Italian private large firms with at least one job-to-job transition. Source: Istituto Nazionale della Previdenza Sociale (INPS).

Table A.5: Centrality by industry

	Mining & Quarrying	Manufacturing	Electricity & Gas	Water supply / sewerage / waste	Construction	Wholesale / retail trade	Transporting & storage	Accomodation & Food service	Information & Communication	Finance / Insurance	Real estate	Professional / Technical activities	Admin. & Support	PA & Defence	Education	Health & Social work	Arts / Entertainment
Norm. by max size																	
Out unw.	0.030	0.037	0.031	0.031	0.039	0.034	0.033	0.073	0.059	0.057	0.042	0.032	0.028	0.048	0.022	0.033	0.038
	(0.0332)	(0.0416)	(0.0394)	(0.0353)	(0.0462)	(0.0455)	(0.0331)	(0.0865)	(0.0684)	(0.0711)	(0.0583)	(0.0477)	(0.0351)	(0.0567)	(0.0311)	(0.0402)	(0.0477)
Out weighted	0.047	0.055	0.064	0.040	0.059	0.053	0.040	0.15	0.11	0.091	0.063	0.056	0.044	0.076	0.030	0.050	0.059
	(0.0813)	(0.0841)	(0.104)	(0.0622)	(0.0893)	(0.0899)	(0.0557)	(0.203)	(0.158)	(0.136)	(0.105)	(0.106)	(0.0760)	(0.115)	(0.0571)	(0.0798)	(0.100)
Out Opsahl	0.035	0.042	0.041	0.034	0.046	0.040	0.035	0.096	0.074	0.069	0.050	0.039	0.033	0.058	0.024	0.038	0.045
	(0.0408)	(0.0500)	(0.0532)	(0.0414)	(0.0553)	(0.0558)	(0.0373)	(0.110)	(0.0868)	(0.0863)	(0.0706)	(0.0595)	(0.0430)	(0.0717)	(0.0363)	(0.0498)	(0.0587)
In unw.	0.049	0.055	0.056	0.048	0.063	0.051	0.042	0.096	0.091	0.088	0.072	0.050	0.047	0.066	0.030	0.066	0.057
	(0.0491)	(0.0529)	(0.0540)	(0.0502)	(0.0622)	(0.0600)	(0.0622)	(0.0843)	(0.0782)	(0.0801)	(0.0599)	(0.0590)	(0.0533)	(0.0645)	(0.0399)	(0.0606)	(0.0611)
In weighted	0.16	0.18	0.20	0.092	0.20	0.19	0.084	0.56	0.44	0.45	0.18	0.23	0.22	0.23	0.069	0.52	0.21
	(0.239)	(0.234)	(0.247)	(0.161)	(0.260)	(0.287)	(0.327)	(0.477)	(0.458)	(0.472)	(0.229)	(0.372)	(0.365)	(0.305)	(0.158)	(0.625)	(0.317)
In Opsahl	0.074	0.084	0.092	0.050	0.096	0.081	0.044	0.21	0.18	0.18	0.093	0.089	0.086	0.11	0.034	0.16	0.090
	(0.0889)	(0.0937)	(0.102)	(0.0753)	(0.112)	(0.113)	(0.128)	(0.164)	(0.158)	(0.165)	(0.101)	(0.120)	(0.112)	(0.121)	(0.0662)	(0.144)	(0.118)
Norm. by avg. size																	
Out unw.	0.045	0.052	0.043	0.051	0.059	0.052	0.051	0.11	0.092	0.091	0.070	0.050	0.043	0.073	0.033	0.048	0.057
	(0.0543)	(0.0640)	(0.0568)	(0.0616)	(0.0737)	(0.0737)	(0.0533)	(0.140)	(0.115)	(0.119)	(0.0991)	(0.0779)	(0.0578)	(0.0912)	(0.0505)	(0.0613)	(0.0768)
Out weighted	0.068	0.078	0.080	0.066	0.087	0.082	0.062	0.22	0.17	0.14	0.10	0.085	0.066	0.11	0.046	0.073	0.088
	(0.114)	(0.127)	(0.117)	(0.103)	(0.134)	(0.146)	(0.0903)	(0.296)	(0.275)	(0.218)	(0.172)	(0.166)	(0.118)	(0.191)	(0.0815)	(0.126)	(0.154)
Out Opsahl	0.052	0.060	0.054	0.056	0.068	0.062	0.055	0.14	0.12	0.11	0.082	0.060	0.050	0.087	0.038	0.056	0.067
	(0.0650)	(0.0774)	(0.0684)	(0.0720)	(0.0874)	(0.0911)	(0.0597)	(0.176)	(0.148)	(0.144)	(0.119)	(0.0956)	(0.0697)	(0.116)	(0.0574)	(0.0804)	(0.0939)
In unw.	0.072	0.078	0.086	0.080	0.096	0.079	0.067	0.15	0.14	0.15	0.12	0.078	0.071	0.10	0.047	0.089	0.087
	(0.0770)	(0.0795)	(0.0851)	(0.0871)	(0.0970)	(0.0936)	(0.104)	(0.137)	(0.127)	(0.132)	(0.102)	(0.0967)	(0.0878)	(0.102)	(0.0624)	(0.0865)	(0.0974)
In weighted	0.23	0.25	0.28	0.15	0.30	0.28	0.13	0.86	0.68	0.73	0.29	0.35	0.31	0.35	0.10	0.72	0.30
	(0.354)	(0.341)	(0.335)	(0.257)	(0.379)	(0.415)	(0.559)	(0.760)	(0.737)	(0.771)	(0.357)	(0.554)	(0.512)	(0.496)	(0.238)	(1.197)	(0.487)
In Opsahl	0.10	0.12	0.14	0.082	0.14	0.12	0.068	0.33	0.27	0.29	0.15	0.14	0.12	0.16	0.052	0.21	0.13
	(0.129)	(0.136)	(0.147)	(0.124)	(0.166)	(0.166)	(0.216)	(0.264)	(0.251)	(0.267)	(0.163)	(0.187)	(0.167)	(0.189)	(0.101)	(0.204)	(0.183)

Note: the table reports average centrality measures by 2-digit industries. The top panel normalizes centrality by firm's maximum number of employees over the period; the bottom panel does so by its average. Each panel reports the following measures, in order: out-degree unweighted centrality, out-degree weighted centrality, out-degree Opsahl centrality, in-degree unweighted centrality, in-degree weighted centrality, in-degree Opsahl centrality. Opsahl centrality assumes $\alpha = 0.5$. Source: Istituto Nazionale della Previdenza Sociale (INPS).

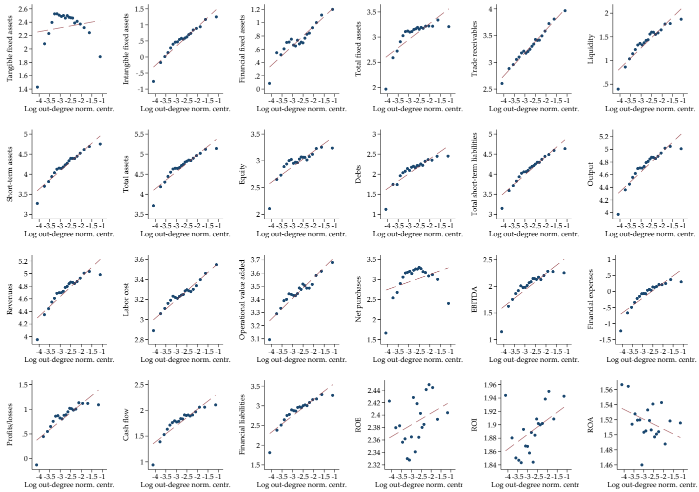
Table A.6: Top 5 representative industries among high-connectivity and regular firms

High-rep firms	**Regular firms**
ICT consultancy	Constructions
Fiscal / law / commercial consultancy	Retail trade
Other professional activities	Metallurgical
Retail trade	Other professional activities
Wholesale	Plant engineering

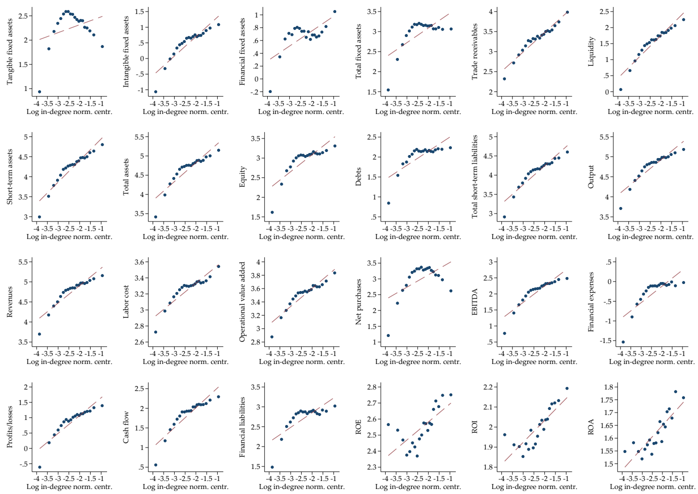
Note: This table reports the top-five most common industries among high-connectivity firms and regular firms, as split by the k-means algorithm run on normalized out-degree and in-degree centrality of the firms in my sample. Source: Istituto Nazionale della Previdenza Sociale (INPS).

FIGURE A.1: Firm-level financial measures by degree centrality

(A) By out-degree centrality



(B) By in-degree centrality



Note. This figure plots means of relevant financial variable by ventiles of out-degree centrality (Panel A) and in-degree centrality (Panel B). Source: Istituto Nazionale della Previdenza Sociale (INPS) and Cerved.

B Additional information on data cleaning

Here, I detail the data cleaning operation undertaken on the firm panel and the matched employer-employee dataset.

Firm panel I start from the *Uniemens* database provided by INPS. The primary cleaning required by this firm-by-year panel is to assign a single province and four-digit industry for each observation—the same firm might indeed operate in multiple sectors or geographical areas within the same period. For both, I do so by imputing the observation with the highest number of employees in the year. Then, I restrict the sample to firms that have employed at least fifteen workers at least once over the 2006-2018 period.

Matched employer-employee I build the matched employer-employee dataset at the monthly level starting by appending separate yearly files. First, I restrict the sample of workers and firms. I keep only workers that have been employed in a white-collar position at least once over the reference period. Among these employees, I further restricted to those that have moved between large firms only—as previously defined. Within this subsample, I drop contracts that lasted less than nine weeks in a given year and contracts with zero wage, and I winsorize the wage outliers at the 0.45 and 99.5 percentiles. I also restrict the contract sample to full-time jobs. Then, I assign each worker to *one* firm with *one* contract each year. To do so, I need to solve for the occurrence of multiple spells, both *within* and *between* worker-firm pairs. When facing multiple spells in the same month within the same employer—i.e., two contemporaneous contracts within the same firm in a given period—I keep the one that pays more. Then, I resolve multiple spells across different employers within the same month through a nested criterion: I keep the one that involves more worked days and, subordinately, the one that pays more. Finally, I perform minor cleanings related to unreliable measures, such as dropping workers that have been paid more than 365 days per year and workers that entered the job market when younger than 18 or older than 50.

C Additional centrality measures

For complementary results and robustness checks, I rely on two additional centrality measures.

The first is the weighted degree centrality. Unlike its unweighted counterpart, which simply counts the number of links to a node, the weighted variant sums the weights—i.e., the intensity of the worker flow entering or leaving the firm. Formally, in an undirected scenario, it's expressed as

$$D^w(i) = \sum_{j}^{N} W_{ij}$$

where W represents the weighted adjacency matrix and W_{ij} the flow between i and j over a given period. This measure only accounts for a node's total involvement in the network, ignoring the number of other nodes it's connected to.

Therefore, I rely on Opsahl et al. (2010) to create a hybrid measure considering both a node's degree and strength. Specifically, the Opsahl centrality equals the number of connected nodes times the adjusted average weight to these nodes:

$$D^\alpha(i) = D(i)\left(\frac{D^w(i)}{D(i)}\right)^\alpha = D(i)^{1-\alpha} \cdot D^w(i)^\alpha$$

where $\alpha > 0$ is a tuning parameter fixing the importance of links quantity relative to their weight. Here, α is set to 0.5.

D K-means

The k-means algorithm is an unsupervised clustering method that, *ex-ante*, only asks for the number of partitions to split the sample in. In this appendix, I discuss the choice of dividing the firms in two groups: high- and low-connectivity. Following Makles (2012), I involve four measures as an optimality criterion to infer the optimal number of clusters: the within sum of squares (WSS), its logarithm, the η^2 coefficient defined as

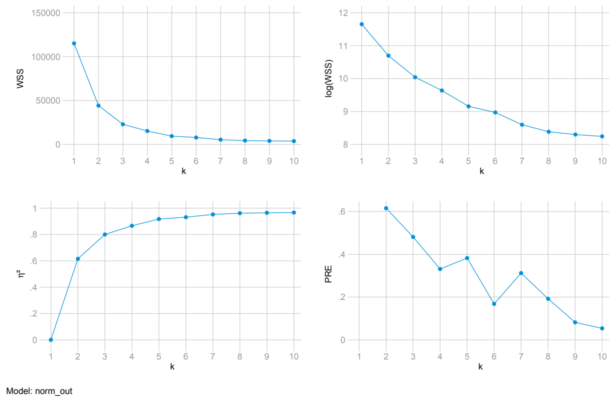
$$\eta_k^2 = 1 - \frac{\text{WSS}(k)}{\text{WSS}(1)} = 1 - \frac{\text{WSS}(k)}{\text{TSS}} \quad \forall k$$

where $\text{WSS}(k)$ is the WSS for a clustering with k partitions; and the proportional reduction error (PRE) given by

$$\text{PRE}_k = \frac{\text{WSS}(k-1) - \text{WSS}(k)}{\text{WSS}(k-1)} \quad \forall k \geq 2$$

Basically, the η_k^2 accounts for the proportional reduction of the WSS for each clustering with k partitions, compared with the total sum of squares (TSS). PRE_k measures the proportional reduction of the WSS for each added cluster.

FIGURE D.1: Optimal-splitting criteria for 1-to-10 clusters



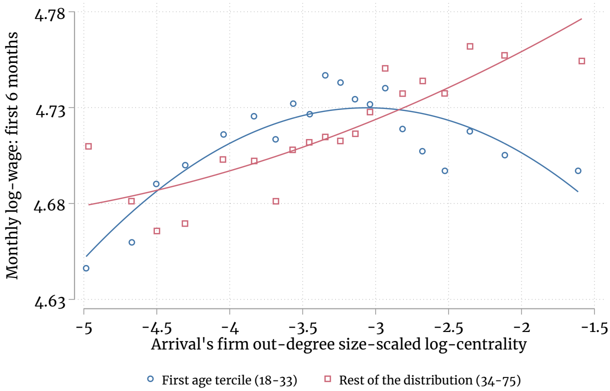
Note: The figure shows the optimal splitting criteria from 1 to 10 clusters. The highest gains in both relative and absolute terms is achieved for $k = 2$.

Figure D.1 plots these four indicators computed on the firms' sample split on the normalized out-degree centrality. Splitting the sample in two is the best choice in terms of WSS and $\log(\text{WSS})$ reduction, as the two top panels in the figure show the deepest kink for $k = 2$. Moreover, η_2^2 records a 60% reduction in the WSS in absolute terms, while PRE_2 gives the highest gain in terms of proportional WSS decrease.

E Alternative specifications and robustness checks

E.1 Centrality and earnings

FIGURE E.1: Entry daily wages and firms' out-degree centrality, by hiring age

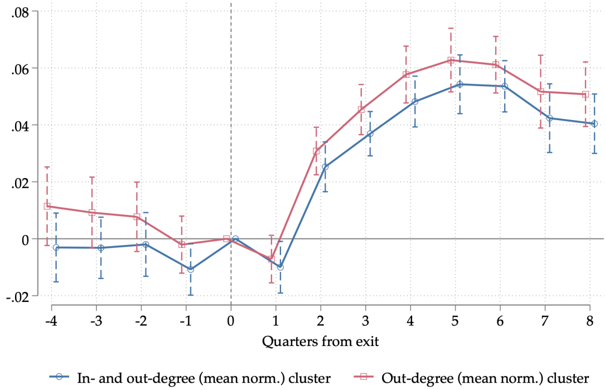


Note: The graph displays the average hiring log daily wages, residualized for worker-firm observables, plotted against the log out-degree centrality of the hiring firm, normalized by its average employer size over the period. Each line signifies a binned quadratic regression for two separate age groups: 18-33 (represented by the blue line and circles) and 34-75 (the red line and squares). There is substantial variation in the hiring earnings of workers depending on the firm's centrality. For younger workers, the hiring wage decreases with the number of unique connections the hiring firm possesses in the job-to-job network, once this measure gains significance. For older workers, this relationship becomes linear and positive. This variability is interpreted as indicative evidence of a compensating differential channel, suggesting that workers may balance immediate earnings against future opportunities. Source: Uniemens data, Istituto Nazionale della Previdenza Sociale (INPS).

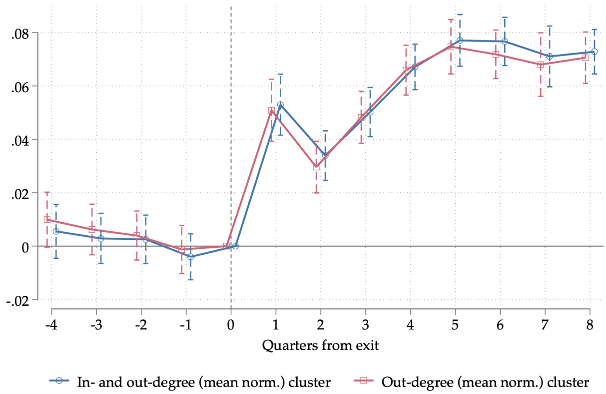
E.2 Leaving high-connectivity firms

FIGURE E.2: Leaving a high-connectivity firm vs. a regular one (max size normalization)

(A) Quarterly earnings



(B) Daily wage



Note: These graphs show the relationship between leaving a high-connectivity firm on a worker's quarterly log earnings (A) and daily log wages (B). Blue lines report results for clustering on in- and out-degree centrality; red lines with only out-degree. Coefficients are obtained estimating the specification in (1.9). 95% confidence intervals are obtained by clustering at the individual-by-quarter level. Source: Istituto Nazionale della Previdenza Sociale (INPS).

F Theory

F.1 Derivation of the wage equation

On the lines of Jarosch (2021), I derive the wage equation using the employed worker value function (1.4) and the unemployed worker value function (1.5), applying the bargaining protocol to obtain the joint surplus for all $(\theta, \hat{\theta})$.

I start by writing the net-of-unemployment value for a worker

$$
\begin{aligned}
W\left(a,\theta,\hat{\theta}\right) - U(a) = {} & w\left(a,\theta,\hat{\theta}\right) \\
& + \beta\Bigg\{(1-\delta)\Bigg[\int_{I-\theta_r}^{\bar{\theta}_r}\Bigg(\int_{x\in\mathcal{F}_1(\theta_p,y)} W(a,x,y,\theta)\,\mathrm{d}P(x) - U(a) \\
& + \int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})} W(a,\theta,x,y)\,\mathrm{d}P(x) - U(a)\Bigg)\,\mathrm{d}T(y) \\
& + \left(1 - \int_{I-\theta_r}^{\bar{\theta}_r}\int_{x\in\mathcal{F}_1\cup\mathcal{F}_2}\mathrm{d}P(x)\,\mathrm{d}T(y)\right)\left(W\left(a,\theta,\hat{\theta}\right) - U(a)\right)\Bigg] \\
& + \delta U(a)\Bigg\} - U(a)
\end{aligned}
\tag{11}
$$

for which I apply the bargaining protocol, obtaining

$$
\begin{aligned}
W\left(a,\theta,\hat{\theta}\right) - U(a) = {} & w\left(a,\theta,\hat{\theta}\right) + \beta\Bigg\{(1 \\
& -\delta)\Bigg[\int_{I-\theta_r}^{\bar{\theta}_r}\Bigg(\int_{x\in\mathcal{F}_1(\theta_p,y)} \left[(1-\sigma)S(a,\theta)+\sigma S(a,x,y)\right]\mathrm{d}P(x) \\
& + \int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})} \left[(1-\sigma)S(a,x,y)+\sigma S(a,\theta)\right]\mathrm{d}P(x)\Bigg)\,\mathrm{d}T(y) \\
& + \left(1 - \int_{I-\theta_r}^{\bar{\theta}_r}\int_{x\in\mathcal{F}_1\cup\mathcal{F}_2}\mathrm{d}P(x)\,\mathrm{d}T(y)\right)\left(W\left(a,\theta,\hat{\theta}\right) - U(a)\right)\Bigg] \\
& + \delta U(a)\Bigg\} - U(a)
\end{aligned}
$$

Then, I substitute the value of unemployment from (1.5) to obtain

$$
\begin{aligned}
W\left(a,\theta,\hat{\theta}\right) - U\left(a\right) = w\left(a,\theta,\hat{\theta}\right) - au_p + \beta\Bigg\{ (1 \\
-\delta)\Bigg[\int_{I-\theta_r}^{\bar{\theta}_r}\Bigg(\int_{x\in\mathcal{F}_1(\theta_p,y)}\left[(1-\sigma)\,S\left(a,\theta\right)+\sigma S\left(a,x,y\right)\right]\,\mathrm{d}P(x) \\
+\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})}\left[(1-\sigma)\,S\left(a,x,y\right)+\sigma S\left(a,\theta\right)\right]\,\mathrm{d}P(x)\Bigg)\,\mathrm{d}T(y) \\
+\left(1-\int_{I-\theta_r}^{\bar{\theta}_r}\int_{x\in\mathcal{F}_1\cup\mathcal{F}_2}\mathrm{d}P(x)\,\mathrm{d}T(y)\right)\left(W\left(a,\theta,\hat{\theta}\right)-U\left(a\right)\right)\Bigg] \\
+\delta U(a) - \lambda\iint_{1(u)} W\left(a,x,y,u_p,0\right)\,\mathrm{d}P(x)\,\mathrm{d}T(y) \\
-\left(1-\lambda\iint_{1(u)}\mathrm{d}P(x)\,\mathrm{d}T(y)\right)U(a)\Bigg\}
\end{aligned}
$$

which, after some simple manipulation, reads as

$$
\begin{aligned}
W\left(a,\theta,\hat{\theta}\right) - U\left(a\right) = w\left(a,\theta,\hat{\theta}\right) - au_p + \beta\Bigg\{ (1 \\
-\delta)\Bigg[\int_{I-\theta_r}^{\bar{\theta}_r}\Bigg(\int_{x\in\mathcal{F}_1(\theta_p,y)}\left[(1-\sigma)\,S\left(a,\theta\right)+\sigma S\left(a,x,y\right)\right]\,\mathrm{d}P(x) \\
+\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})}\left[(1-\sigma)\,S\left(a,x,y\right)+\sigma S\left(a,\theta\right)\right]\,\mathrm{d}P(x)\Bigg)\,\mathrm{d}T(y) \\
+\left(1-\int_{I-\theta_r}^{\bar{\theta}_r}\int_{x\in\mathcal{F}_1\cup\mathcal{F}_2}\mathrm{d}P(x)\,\mathrm{d}T(y)\right)\left(W\left(a,\theta,\hat{\theta}\right)-U\left(a\right)\right) \\
-U(a)\Bigg] - \lambda\iint_{1(u)}\sigma S\left(a,x,y\right)\Bigg\}
\end{aligned}
$$

Defining

$$\eta \equiv au_p + \beta\lambda\iint_{1(u)}\sigma S\left(a,x,y\right)$$

and collecting terms on the lhs one has

$$
\begin{aligned}
&[1-\beta(1-\delta)]\left(W\left(a,\theta,\hat{\theta}\right)-U(a)\right)\\
&\quad = w\left(a,\theta,\hat{\theta}\right)-\eta\\
&\qquad +\beta(1-\delta)\left[\int_{I-\theta_r}^{\bar{\theta}_r}\left(\int_{x\in\mathcal{F}_1(\theta_p,y)}[(1-\sigma)S(a,\theta)+\sigma S(a,x,y)]\,\mathrm{d}P(x)\right.\right.\\
&\qquad\qquad +\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})}[(1-\sigma)S(a,x,y)+\sigma S(a,\theta)]\,\mathrm{d}P(x)\\
&\qquad\qquad \left.\left.-\int_{x\in\mathcal{F}_1\cup\mathcal{F}_2}\left[(1-\sigma)S\left(a,\hat{\theta}\right)+\sigma S(a,\theta)\right]\mathrm{d}P(x)\right)\mathrm{d}T(y)-U(a)\right]
\end{aligned}
$$

and since the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ are mutually disjoint:

$$
\begin{aligned}
&[1-\beta(1-\delta)]\left(W\left(a,\theta,\hat{\theta}\right)-U(a)\right)=w\left(a,\theta,\hat{\theta}\right)-\eta\\
&\quad+\beta(1-\delta)\left[\int_{I-\theta_r}^{\bar{\theta}_r}\left(\int_{x\in\mathcal{F}_1(\theta_p,y)}\left[(1-\sigma)S(a,\theta)+\sigma S(a,x,y)-(1-\sigma)S\left(a,\hat{\theta}\right)-\sigma S(a,\theta)\right]\mathrm{d}P(x)\right.\right.\\
&\quad+\left.\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})}\left[(1-\sigma)S(a,x,y)+\sigma S(a,\theta)-(1-\sigma)S\left(a,\hat{\theta}\right)-\sigma S(a,\theta)\right]\mathrm{d}P(x)\right)\mathrm{d}T(y)\\
&\qquad\left.-U(a)\right]
\end{aligned}
$$

Now it is possible to simplify the previous expression as

$$
[1-\beta(1-\delta)]\left(W\left(a,\theta,\hat{\theta}\right)-U(a)\right)=w\left(a,\theta,\hat{\theta}\right)-\eta+\beta(1-\delta)\,G\left(a,\theta,\hat{\theta}\right)
$$

where $G(a,\theta,\hat{\theta})$ collects the gains from on-the-job search for the worker:

$$
\begin{aligned}
G\left(a,\theta,\hat{\theta}\right)=&\int_{I-\theta_r}^{\bar{\theta}_r}\left(\int_{x\in\mathcal{F}_1(\theta_p,y)}\left[(1-\sigma)\left[S(a,\theta)-S\left(a,\hat{\theta}\right)\right]\right.\right.\\
&\qquad\left.+\sigma\left[S(a,x,y)-S(a,\theta)\right]\right]\mathrm{d}P(x)\\
&\left.+\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})}\left[(1-\sigma)S(a,x,y)-(1-\sigma)S\left(a,\hat{\theta}\right)\right]\mathrm{d}P(x)\right)\mathrm{d}T(y)\\
&-U(a)
\end{aligned}
$$

Thus, the net-of-unemployment worker value can be expressed as

$$W\left(a,\theta,\hat{\theta}\right)-U(a)=\frac{w\left(a,\theta,\hat{\theta}\right)+\beta(1-\delta)G\left(a,\theta,\hat{\theta}\right)-\eta}{1-\beta(1-\delta)}$$

and, recalling once again the bargaining protocol, one has

$$w\left(a,\theta,\hat{\theta}\right)+\beta(1-\delta)G\left(a,\theta,\hat{\theta}\right)-\eta=[1-\beta(1-\delta)]\left[(1-\sigma)S\left(a,\hat{\theta}\right)+\sigma S(a,\theta)\right] \tag{12}$$

Finally, it is possible to use the surplus value function (1.7) to write

$$[1-\beta(1-\delta)]S(a,\theta)=a\theta_p-\eta+\beta(1-\delta)\sigma\int_{I-\theta_r}^{\bar{\theta}_r}\int_{x\in\mathcal{F}_1(\theta_p,y)}[S(a,x,y)-S(a,\theta)]\,\mathrm{d}P(x)\,\mathrm{d}T(y)$$

and exploit this latter expression to solve equation (12) for the wages:

$$\begin{aligned}w\left(a,\theta,\hat{\theta}\right)=&\underbrace{(1-\sigma)[1+\eta-\beta(1-\delta)]S\left(a,\hat{\theta}\right)}_{\kappa}+\sigma a\theta_p\\&+\beta(1-\delta)\left(\sigma^2\int_{I-\theta_r}^{\bar{\theta}_r}\int_{x\in\mathcal{F}_1(\theta_p,y)}[S(a,x,y)-S(a,\theta)]\,\mathrm{d}P(x)\,\mathrm{d}T(y)\right.\\&\left.-G\left(a,\theta,\hat{\theta}\right)\right)\end{aligned}$$

where κ collects all the terms that do not depend on θ.

F.2 Comparative statics for the wage equation

With the wage equation, it is possible to operate a comparative statics exercise on wages with respect to the two dimensions of firms' heterogeneity. First, rewrite $\Pi(\theta,\hat{\theta})$ as

$$\begin{aligned}\Pi(\theta,\hat{\theta})=&\int_{I-\theta_c}^{\bar{\theta}_c}\left(\int_{x\in\mathcal{F}_1(\theta_p,y)}\overbrace{(\sigma^2-\sigma)\left[S(x,y)-S(\theta)\right]}^{<0}\mathrm{d}P(x)\right.\\&\left.-\underbrace{(1-\sigma)\left[S(\theta)-S(\hat{\theta})\right]}_{>0}-\int_{x\in\mathcal{F}_2(\theta_p,y,\hat{\boldsymbol{\theta}})}\underbrace{(1-\sigma)\left[S(x,y)-S(\hat{\theta})\right]}_{>0}\mathrm{d}P(x)\right)\mathrm{d}T(y)<0\end{aligned}$$

Then, one has

$$\frac{\partial w(a,\theta,\hat{\theta})}{\partial \theta_c} = \frac{\partial \Pi(\theta,\hat{\theta})}{\partial \theta_c} < 0$$

since the surplus is strictly increasing in θ_c.

When considering θ_p, one has

$$\frac{\partial w(a,\theta,\hat{\theta})}{\partial \theta_p} = \sigma a + \frac{\partial \Pi(\theta,\hat{\theta})}{\partial \theta_p}$$

Therefore, the sign of the derivative depends on α. In particular, when $\alpha = 0$, $\partial w/\partial \theta_p < 0$; when $\alpha = 1$, the opposite.

F.3 Proof of Proposition 1

Proof. Let A represent a node in the network G—specifically, a firm—with a connectivity parameter c_A. The search process in the model can be divided into two stages at any given time. The first concerns meetings between firms, and the second involves formalizing the offer to the worker and potentially establishing a link. As the first step occurs with a constant, uniform probability independent of firms' connectivity, after a sufficient number of iterations, each firm (node) will eventually connect with the mass of all possible nodes to which it can link, i.e., $1 - T(I - c_A)$. This constitutes the expected relative degree centrality for an infinite number of iterations and is a function that increases with c_A. Assuming a large enough number of iterations for convergence completes the proof. □

Chapter 2

The Heterogeneous Consequences of Reduced Labor Cost on Firm Productivity[1]

Abstract. We explore the relationship between labor market institutions and firm productivity by examining the heterogeneous effects of a labor cost change indirectly induced by a reform that lowered employment protection legislation in Italy. The decrease in labor cost—isolated due to unique features of the Italian collective bargaining institutions—resulted in a reduction in average total factor productivity TFP among less productive firms in manufacturing, while increasing productivity at the top of the distribution. We pair these findings

[1]This chapter is joint with Paolo Zacchia. We express our gratitude to Raffaele Saggio for generously sharing his data on the renewal of collective agreements. We are also thankful to Edoardo M. Acabbi, Bruno Cassiman, Luca Citino, Edoardo Di Porto, Kenan Huremovic, Mattia Nardotto, Matteo Paradisi, Santiago Pereda Fernández, Francesco Serti, and Cristina Tealdi for their valuable comments and suggestions. We extend our thanks to the participants of the seminars at VisitINPS, KU Leuven, SIE's 63rd RSA, XX Brucchi Luchino workshop, Econometric Society Winter Meeting 2022, and SOLE 2023. We would like to acknowledge the support of the "VisitINPS Scholars" program, which made this project possible. We are particularly grateful to Edoardo Di Porto and Paolo Naticchioni for their assistance with the data and helpful feedback. We also thank the entire INPS' *Direzione Studi e Ricerche* for their support. Any errors are entirely ours.

with increased entry and exit dynamics among low-productivity firms to suggest the presence of an adverse selection mechanism at the bottom of the TFP distribution. We formalize this concept through a model that links equilibrium productivity to labor market frictions, and we use this model to assess the welfare impact of the reform.

JEL classification: D21, D22, D24, E24, J08, O14.

Keywords: Productivity, TFP, labor flexibility, EPL, labor cost.

2.1 Introduction

What is the relationship between labor market institutions and firms' productivity? This complex, broad question involves many faces of how a labor market works. Employment protection legislation (EPL) plays a crucial role. By enabling firms to more easily adjust their workforce, lower EPLs can minimize production choice distortions and enhance efficiency (Autor et al., 2007). Conversely, higher protection can suppress job creation and destruction rates, theoretically diminishing aggregate productivity (Lazear, 1990; Lagos, 2006). Institutional mechanisms such as minimum wages (Dustmann et al., 2021), unionization (Haucap and Wey, 2004), and the structure of collective bargaining (Jäger et al., 2022) can also influence productivity, both directly and by interplaying with changes in EPLs. Shading light on this complex mechanism is central in understanding whether more flexible labor markets genuinely function as a *rising tide that lifts all boats* towards productivity.

This paper presents new evidence suggesting that the relationship between labor market rigidity and firm productivity is more nuanced than often thought, with significant variation across the productivity distribution. We demonstrate that increased flexibility—achieved through a reform that reduced firms' labor costs by decreasing workers' bargaining power—adversely affects total factor productivity (TFP) among less productive firms in the manufacturing sector. This negative effect diminishes as we move up the pre-intervention productivity distribution and eventually reverses, becoming positive for the most productive firms. To interpret these results, we construct a comprehensive general equilibrium model that links the equilibrium TFP distribution across sectors with labor and capital frictions. We use this model to theoretically decompose the cost reduction effect along the productivity distribution, attributing it to a combination of a selection mechanism specific to the left tail and an incentive for productivity-enhancing investments due to downward pressure on labor costs. Highlighting these heterogeneous effects underscores the need for nuanced policy approaches, such as addressing capital misallocation based on the model's insights, to effectively address the diverse productivity levels exhibited by firms.

To empirically investigate the heterogeneous effects of flexibilization on productivity, we leverage a 2001 Italian reform designed to lower the barriers for firms to initiate new temporary contracts. We conduct a series of event studies to examine the staggered adoption of the intervention across collective bargaining agreements (Contratti Collettivi Nazionali del Lavoro, CCNL). As in Daruich et al. (2023),oour identification strategy relies on the dynamic comparison of firms that implemented the reform early—based on the prevalent contract they used prior to the reform—with those that adopted it later.

Contrary to Daruich et al. (2023), our empirical analysis primarily focuses on the firm level, examining the effect of the reform on firms' TFP. In the manufacturing sector, this intervention led to a decrease in labor costs triggered by a reduction in workers' bargaining power, indirectly caused by the reform's original mechanism—a change in the use of temporary jobs—which we document did not occur within our sample. This setting enables us to more accurately isolate a causal relationship between pure labor cost reduction and productivity.

Interestingly, the adoption of temporary contracts has remained consistently low in manufacturing over time despite the reform—a contrast to what occurs in other industries, as documented by Daruich et al. (2023). However, we still observe a significant reduction in labor costs. This paradox is explained by the loss in workers' bargaining power that followed the reform, which propagated through collective contracts across industries. We present three key facts to support this claim. First, the vast majority of Italian firms use a single collective contract for nearly their entire workforce. Second, collective agreements are employed across sectors, regardless of the specific industry they were initially designed for. Third, there exists a relationship between workforce composition and salaries across macro-sectors. In other words, changes in the share of temporary workers used in the service industries are associated with changes in salaries in manufacturing.

Our empirical analysis utilizes a comprehensive, matched employer-employee administrative dataset from the Italian Social Security Institute (Istituto Nazionale di Previdenza Sociale, INPS), further enriched by incorporating firm-level financial data from balance sheets.[2] We provide evidence of a significant heterogeneous response across the ex-ante industry-specific TFP distribution for firms within the Italian manufacturing sector. The reform reduced average TFP among already unproductive firms, and increased it for the most productive ones. We further complement these linear specifications with a quantile treatment effect approach, enabling a more nuanced examination of heterogeneity in productivity outcomes, and we provide evidence of reduced exit rates and increased entry rates for the least productive firms. Overall, our empirical results support a dual mechanism where a reduction in labor cost aids the survival and

[2]This data allows us to construct three distinct TFP measures using various estimation techniques—namely, those proposed by Levinsohn and Petrin (2003), Ackerberg et al. (2015), and Gandhi et al. (2020)—and to incorporate worker, firm, and province-by-sector level outcomes into our analysis.

new entrance of unproductive firms, thereby impeding allocative efficiency and overall productivity, while productive firms benefit from enhanced investment incentives and potential TFP growth due to a more flexible labor market.

In more detail, we first document that the reform led to a substantial decrease in labor costs. On average, firms experienced a reduction of up to 6% in per-worker earnings within two years of adopting the new framework. Importantly, this outcome is not attributed to a shift in firms' reliance on flexible work arrangements in our sample. Indeed, the adoption of temporary contracts has remained unaltered in manufacturing following the reform. As we have argued, the combination of this evidence, the distinctive patterns in the utilization of CCNLs both across and within firms, and the lack of a response to the reform in terms of temporary work arrangements usage, allow us to interpret the reform itself as a pure labor cost shifter.

Turning to productivity, we estimate the effect of the reform on a variety of alternative measures of TFP. Plain staggered difference-in-differences around the firm-specific year of the reform adoption reveal a marginally negative average impact on firm-level TFP, which remains consistent across the three measures employed. However, this overall effect conceals significant heterogeneity, as a large portion of the impact is attributable to firms in the lowest quartile of the ex-ante productivity distribution. Post-reform, the average effect within this bottom quartile is negative, resulting in a productivity loss of up to 14% three years after adoption. Simultaneously, we observe a positive and nearly symmetrical average effect within the highest quartile of the distribution, albeit estimated with slightly less precision than the impact on the lower end. Notably, these findings remain qualitatively consistent across all three productivity measures. The evidence we present stems from linear specifications that measure the average effect of the EPL reduction within given quartiles of the ex-ante TFP distribution. To refine our heterogeneity analysis, we employ a non-parametric specification for quantile treatment effects, comparing post-reform TFP distribution to a counterfactual distribution, had the reform never been implemented. Our findings reveal a monotonic influence on the TFP distribution, with negative effects on the lowest deciles and positive effects on the most productive firms.

Additionally, we demonstrate that the labor cost reduction had heterogeneous impacts on firms' turnover. While there is no effect at all on the top quartile of the distribution regarding entry and exit rates at the province-by-subsector level, there is strong evidence of a substantial reduction (approximately 7% after two years) in exit events and an increase (up to 10% after two years) in entries for less productive firms. In essence, firms seemed to exploit the reduced labor costs to bolster their chances of survival, compared to a scenario without the reform. Simultaneously, this decrease in labor costs attracted more unproductive firms to enter the market.

We propose a new interpretation of these heterogeneous results on productivity, suggesting the presence of a mixed mechanism at work, which can be understood through a straightforward intuition. A reduction in EPL leads to a change in workers' outside options that in turn decreases both the absolute and relative prices of labor faced by firms. Consequently, this change offers survival opportunities to unproductive firms that would have had a higher probability of exiting the market if the reform had not occurred. As a result, an adverse selection effect arises on the left side of the TFP distribution, where firms that should not have continued production manage to survive, entry barriers are lowered, allowing low-productivity firms to enter the market, and the reduced labor cost discourages capital deepening and investments. This combination, in turn, impairs allocative efficiency and suppresses overall productivity. In contrast, firms on the right side of the distribution, which are already productive, face increased incentives to continue investing due to the relative price change of production factors. In this case, no negative selection occurs, allowing these firms to potentially experience TFP growth as a result of efficiency gains in labor force adjustments driven by a more flexible labor market.

To lend greater rigor to the arguments supporting our findings, we develop a model that incorporates the proposed mechanisms and builds upon the foundation of general equilibrium models featuring monopolistic competition (Dixit and Stiglitz, 1977) and heterogeneous firms (Melitz, 2003). Our analysis integrates two key elements previously introduced in the literature, with novel approaches in each case: financial frictions (see, for example, Manova, 2013) and endogenous productivity (Bustos, 2011; Zhelobodko et al., 2012). We address financial frictions as an asymmetric information issue: firms require financial intermediaries (FIs) to provide credit for market entry, while FIs only have access to a noisy signal of the firms' true productivity. We model endogenous productivity in a similar manner to Bustos (2011), but treat the cost of productivity-enhancing investments (PEIs) as a continuous variable, rather than binary. Our model predicts that stronger EPLs result in reduced entry of low-productivity firms and hinder PEIs, particularly on the right tail. Our empirical findings offer robust evidence supporting the former mechanism and mixed evidence for the latter. Excluding the consideration of EPL's utility value for workers, the net welfare effect of these two mechanisms remains ambiguous, dependent on the relative impact at the tails.

Related literature Many papers have addressed the link between EPL and productivity. Three works closely related to our research include Autor et al. (2007), Cappellari et al. (2012), and Dolado et al. (2016). Autor et al. (2007) examined US plant-level data to study the effects of state courts adopting wrongful-discharge protection provisions, finding a decrease in job flows, entries, and TFP. Cappellari et al. (2012) used the same Italian reform as our study to demonstrate productivity losses resulting from the substitution of temporary employees for external staff and a decrease in capital intensity. Dolado et al. (2016) associated the cost gap between permanent and temporary jobs with firms' TFP, arguing that a larger gap lowers the temp-to-perm conversion rate, which in turn reduces worker effort and firm-level paid-for training—thus decreasing productivity.

While our results are consistent with many of the findings from these three works, we diverge from them in several aspects. First, we emphasize the heterogeneous effects of increased labor flexibility on TFP based on ex-ante productivity, which highlights the potential harm to already unproductive firms. Second, our study relies on an institutional framework that enables us to claim a causal interpretation of our results. In comparison to Cappellari et al. (2012), we access an exceptionally rich administrative dataset, allowing us to observe the entire worker-firm match universe and the specific collective bargaining adopted by each worker, as seen in Daruich et al. (2023) and Acabbi and Alati (2021). Third, our setting allows us to interpret the EPL reduction as a *ceteris paribus* decrease in labor cost that is not directly contingent upon the use of temporary work arrangements, due to the distinctive use of the Italian CCNLs, which permit spillover effects. Fourth, we introduce a new mechanism to explain the heterogeneous impact of easy access to temporary jobs on TFP, which combines a selection effect with altered investment incentives.

Other research has examined various firm-level dimensions, associating them with reforms that affect EPL more generally. Kugler and Pica (2008) find that increased dismissal costs for small firms lead to lower worker accessions and separations, diminished employment adjustments on the internal margin, reduced entry rates, and increased exit rates. Bassanini et al. (2009) report a negative impact of dismissal regulation in the OECD on productivity growth, concentrated in industries where layoff restrictions are more likely to be binding—although they do not find evidence of productivity effects from temporary contract regulation. Cingano et al. (2016) show that introducing unjust-dismissal costs for Italian firms with fewer than 15 employees causes an increase in the capital-labor ratio and a decline in TFP for small firms relative to larger firms. Acabbi and Alati (2021) leverage the same reform as our study to reveal how firms use contract composition to manage labor-induced operating leverage risk: for firms with an ex-ante rigid labor cost structure, a more flexible workforce composition results in a profit margin increase. Other studies have also focused on the economic impact of EPL changes using cross-country analyses with aggregate data, primarily assessing the effects on unemployment and wages (Lazear, 1990; Bertola, 1990; Bertola and Rogerson, 1997; Garibaldi and Violante, 2005).[3]

Gnocato et al. (2020) examine a different channel of the heterogeneous effect of easing temporary contracts by considering the size-productivity covariance as a measure of allocative efficiency, following the approach of Hsieh and Klenow (2009). In their study, heterogeneity is driven by geographical differences in the length of labor court disputes. Our results remain consistent with their proposed mechanism of heterogeneous gains in labor productivity, which suggests that more productive firms tend to gain market shares due to longer tenures at the workplace for fixed-term workers.

Different partial equilibrium approaches have discussed the ambiguous effect of easing access to fixed-term contracts on unemployment and wages from a theoretical standpoint. Bentolila and Bertola (1990) show that higher EPL increases average employment as the reduction in lay-offs dominates the adverse effect coming from lower hiring. Blanchard and Landier (2002) and Cahuc and Postel-Vinay (2002) argue that temporary contracts may result in higher turnover in entry-level jobs, leading to increased unemployment, slower job ladder climbing, and lower match productivity. General equilibrium effects remain ambiguous, with a relevant dependence on the specific model considered (Ljungqvist, 2002).

Several mechanisms can explain the negative correlations between EPL and labor productivity. High EPL may hinder the efficient allocation of resources within the economy (Hopenhayn and Rogerson, 1993); depress workers' effort (Ichino and Riphahn, 2005; Engellandt and Riphahn, 2005; Dolado et al., 2016); diminish the incentive to acquire general skills rather than specific ones, thereby hindering worker reallocations across firms (Wasmer, 2006); and reduce LP through increased substitution between permanent and temporary contracts (Cahuc et al., 2016). Similarly, a reduction in EPL (especially when facilitating the use of temporary contracts) might be employed as a screening device by firms—serving as stepping stones into permanent contracts that improve match quality (Ichino et al., 2008; Faccini, 2014). Furthermore, an increase in EPL could depress TFP by reducing job creation and job destruction rates in an aggregate model that takes individual search behavior into account (Lagos, 2006). On the other hand,

[3]For a review of the different cost margins associated with structural reforms that improved labor flexibility, see Boeri et al. (2015).

there are two main theoretical arguments that might link increases in EPL to increased productivity. First, employment protection might encourage workers to invest in match-specific human capital, which can benefit LP, especially when other labor market rigidities exist (Belot et al., 2007). Second, increasing EPL induces a selection of the most productive firms that can accommodate the rise in labor costs (Poschke, 2009). Although our model is built on a different mechanic, it relies on an intuition that is related to the latter.

Outline of the paper The paper is organized as follows. Section 2.2 presents the reform we study, discusses some peculiar facts about collective bargaining institutions in Italy, and illustrates the data and the sample. Section 2.3 outlines the empirical approach; then, it presents and discusses our findings. Section 2.4 illustrates the model. Section 1.5 concludes.

2.2 Institutional setting and data

This section is composed of three parts. Initially, we present and analyze the reform in question, emphasizing how our identification strategy capitalizes on its staggered implementation across different Italian national collective contracts. Subsequently, we explore the distinctive characteristics of Italian collective bargaining institutions and their implications for our empirical analyses. Lastly, we delineate the data sources utilized in our study and provide information regarding our selected sample.

2.2.1 The Italian 368/2001 decree

Labor legislation in Italy distinguishes between regulations for permanent and temporary employment contracts. Permanent contracts, lacking a predefined termination date, necessitate substantial severance packages if an employer decides to terminate an employee. These costs depend on variables such as the size of the company and the employee's length of service. In contrast, temporary contracts, agreements with a set termination date, allow employers to dismiss employees post-contract without additional costs. Prior to the enactment of Decree 368, compliant with the EU directive 1999/70/CE on September 6, 2001, Italian companies could only employ temporary contracts under certain conditions, which needed explicit reporting to the Italian social security institute (INPS). The reform removed numerous restrictions associated with temporary contracts, leaving the permanent ones unaffected. Consequently, this led to a easier use of fixed-term employment.[4]

Though the reform was officially enacted on a specific date, it only took effect in different occupations upon the renewal of the corresponding Collective Bargaining Agreements (CCNLs). Each Italian union or union group negotiates its own CCNL, hence expiration dates differ and are known well in advance. This resulted in a staggered implementation of the new regulations on temporary contracts across CCNLs, without disrupting their usual renewal timelines.

Given this framework, we utilize the staggered renewal of 181 Italian CCNLs, exploiting the staggered adoption of more liberal temporary employment policies across these contracts. The timing of this setup allows us to leverage a potentially exogenous shift in the reform's application, yielding a quasi-experimental variation in labor market flexibility across collective agreements, as seen in Acabbi and Alati (2021) and Daruich et al. (2023). This approach enables us to identify the causal impact of the reform on temporary employment on productivity and other crucial firm-level outcomes. However, due to the unique characteristics of Italy's collective contract negotiation system, we refrain from attributing the changes prompted by the reform solely to the modified conditions for using temporary labor arrangements. The following section delves into the reasons for this stance.

2.2.2 The collective bargaining agreements in Italy

Institutional Overview Collective bargaining in Italy is a well-structured process, characterized by the existence of hundreds of national sector-wide collective national labor contracts. These contracts (*Contratti Collettivi Nazionali del Lavoro*, CCNLs), negotiated by trade unions and employers' associations, primarily aim to establish minimum pay levels at the national, industry-wide level within the private sector. These compensation floors, known as *contractual wages*, are set for each job title, typically encompassing between five and ten occupations. They are immune to reductions at the local level and apply to all employees within the contract, regardless of their union membership. Contractual wages are not just seen as a wage floor, but also as a fixed component of the wage (Fanfani, 2022).

The number and nature of collective agreements within an industry are not uniform, due to both historical and organizational factors, and because the activi-

[4] It is worth noting that the reform did not amend the existing employment protection measures for ongoing and permanent contracts, increasing the difference in worker protection levels across contract types. Moreover, even post-reform, there were restrictions on the duration a firm could employ a worker under temporary contracts.

TABLE 2.1: Share of firms applying a single CCNL to 80%+ of workforce

	All panel		Within year	
	All industries	Manufacturing	All industries	Manufacturing
All	0.83	0.80	0.95	0.96
>15 empl.	0.81	0.83	0.94	0.96
>50 empl.	0.82	0.86	0.92	0.96

Note. The table presents the proportion of companies in the sample that apply a single CCNL code to at least 80% of their workforce, broken down by size. The left panel reports the share for the entire 1996-2016 panel, while the right panel displays the within-year averages. Source: Istituto Nazionale della Previdenza Sociale (INPS).

ties defined and regulated by each collective agreement do not map to a standard sector classification. As a result, as we will document in this section, it is common to observe multiple collective contracts coexisting within a single sector, with multi-sector contracts also being a frequent occurrence. The activities governed by these agreements are delineated by the bargaining parties and explicitly articulated within each contract. Employers are formally obligated to apply the contract that is most relevant to the activities performed by each employee, and this contract must bear the signatures of the most representative unions and employers' associations at the national level.

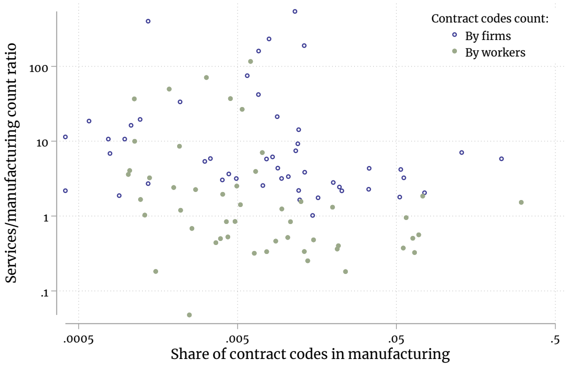
We now document three main facts regarding the use of collective bargaining in Italy. This evidence is noteworthy not only *per se*, as it contributes to a better understanding of how firms utilize CCNLs, but it will also be useful for later discussions of the mechanisms behind the effect of the reform on firm-level productivity.

Fact 1. Most firms apply a single collective contract to the vast majority of their workforce.

The evidence of the presence of a dominant contract within a firm is presented in Table 2.1. This table shows the proportion of companies in the sample that apply a single CCNL code to at least 80% of their workforce, broken down by size and sector. The vast majority of Italian companies apply a single collective contract to almost all of their workforce, regardless of their size or sector. Furthermore, such a contract remains stable over time.

Fact 2. Collective contract types are not segregated by macro-sector (manufacturing vs. services). Instead, they are typically used across these sectors.

FIGURE 2.1: CCNLs' overlap across macro-industries



Note. This figure illustrates the overlap in the use of CCNLs across manufacturing and service industries. It plots the proportion of a collective contract's use in manufacturing against the count ratio of observations in services vs manufacturing. The count is carried out both at the worker level (blue empty circles) and at the firm level (grey filled circles), i.e., by assigning a modal collective contract to each firm. Axes are in log-scale. Source: Istituto Nazionale della Previdenza Sociale (INPS).

Despite their names, which often refer to industry-specific occupations, the same CCNLs are widely used across sectors. Evidence of this fact is presented in Figure 2.1, which shows, for each collective contract, the proportion of that contract's use in manufacturing against the count ratio of observations in services vs manufacturing. The count is carried out both at the worker level, i.e., by counting the single contracts, and at the firm level, i.e., by assigning a modal collective contract to each firm. The high dispersion of the points indicates an overlap of contracts across industries: if this were not the case, we would observe points clustered rather than spread across the area. This dispersion remains consistent regardless of the count, a result that aligns with the previously established evidence of within-firm contracts' homogeneity. Fact 2 can be explained as a by-product of low enforcement (Garnero, 2018) and legal ambiguity regarding the sectoral specificity of the contracts.[5]

[5]The debate on whether a firm can choose to apply a Collective Bargaining Agreement (CCNL) that differs from its sector's has been settled by the Italian Supreme Court, favoring the view that supports a company's freedom to choose its CCNL. This principle hinges on union freedom and the contract's effectiveness between the parties involved. Despite this, choosing a different CCNL may indirectly affect aspects like contractual wage, fiscal benefits of social burdens, and other legal facilitations. Labor inspectors, though unable to compel a firm to change its CCNL, can address wage

TABLE 2.2: Within-CCNL temporary-to-earnings transmission

Dep. variable	*Earnings*		*Log-earnings*	
	Manufacturing	Services	Manufacturing	Services
Temporary share				
in services	-1.37**	-6.04*	-0.040**	-0.31**
	(0.54)	(3.20)	(0.020)	(0.14)
in manufacturing	-8.13**	-0.45	-0.34**	-0.029
	(3.81)	(0.54)	(0.13)	(0.023)
South share				
in services	-0.98	-5.84***	-0.051*	-0.23**
	(0.65)	(2.06)	(0.030)	(0.11)
in manufacturing	-4.03***	0.066	-0.20***	0.014
	(1.38)	(0.48)	(0.057)	(0.022)
Observations	5,245	5,245	5,245	5,245
Adj, R-squared	0.996	0.987	0.995	0.988
CCNL + Year FE	✓	✓	✓	✓
CCNL-specific time trends	✓	✓	✓	✓

Note. The table presents the estimates of an OLS regression of earnings on the share of temporary workers and on the share of workers in the South of Italy as a control, broken down by industry. Regressions are weighted for workers numerosity in each macro-industry. Standard errors are provided in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Source: Istituto Nazionale della Previdenza Sociale (INPS) and CNEL.

Fact 3. Wages in one macro-sector display significant residual correlation with the share of temporary workers in the other macro-sector.

The evidence supporting this last fact is presented in Table 2.2. It reports the estimated coefficient of the OLS regression

$$\mathrm{E}^{I}_{ct} = \beta_1 \mathrm{TempSh}^{M}_{ct} + \beta_2 \mathrm{TempSh}^{S}_{ct} + \beta_3 \mathrm{SouthSh}^{M}_{ct} + \beta_4 \mathrm{SouthSh}^{S}_{ct} + \alpha_c + \tau_t + \gamma_c \mathrm{Year} + \varepsilon_{ct}$$

where the dependent variable is the CCNL-year average earnings for macro-industry $I \in \{S, M\}$, regressed on the shares of workers under a temporary workers and workers in the South of Italy (as a control) for both the manufacturing and services macro-sectors. We include CCNL and year fixed effects (α_c and τ_t, respectively), and we control for CCNL-specific linear time trends (γ_c).

differences, recalculate minimum contributions, and reclaim economic benefits derived during the period the firm used a non-corresponding CCNL.

We are mainly interested in estimating β_2 when $I = M$, to assess the correlation between the share of temporary workers in the service macro-industry and the wages in manufacturing. This coefficient is negative and statistically significant in both the levels and logs specifications. A 10 percentage points increase in the number of temporary workers in the service macro-industry is estimated to be associated with a 0.4% decrease in earnings in manufacturing, on average. Such an association remains qualitatively consistent (although statistically not significant) when considering the opposite relationship, i.e., the one between the share of temporary contracts in manufacturing and earnings in services. We interpret these results as evidence of interdependence between macro-industries within a CCNL. An increase in the use of temporary contracts can affect salaries even outside the sector where this shift occurs due to the reduction in bargaining power that follows the change in the industrial relationship. The presence and relevance of such a transmission channel are confirmed by Facts 1 and 2.

2.2.3 Data, sample, and summary statistics

Our paper relies on a diverse set of administrative data sources to create a comprehensive panel that connects workers and firms. This panel is enhanced by extensive details on *i)* national labor contracts (CCNLs), and their renewal dates and *ii)* data from firms' balance sheets, which we leverage to construct various total factor productivity measures. In the following section, we delve deeper into these data sources, elaborate on the TFP measures we use, and present some descriptive statistics of our sample.

Data sources

The empirical analysis in this paper is based on three distinct data sources:

1. The firm-level balance sheet panel data for incorporated firms in Italy from 1996 to 2016 (*Cerved* dataset).

2. The matched employer-employee panel data that covers the entirety of employment relationships in the Italian non-agricultural private sector over the same time period (*Uniemens* database).

3. Information on national collective bargaining agreements (CCNLs) detailing their renewal dates.

The first two datasets are sourced from the Italian Social Security Institute (*Istituto Nazionale di Previdenza Sociale*, INPS), and were exclusively accessible to us through the VisitINPS Scholars program.

Firms' Financial Data (Cerved) We use proprietary firm-level data on balance sheets from the Cerved database to construct three distinct TFP measures, which are discussed in detail in Section 2.2.3. The sample encompasses a period of twenty years from 1996 to 2016, incorporating standard account variables such as revenues, value-added, labor costs, tangible and intangible assets, and the cost of materials. These measures are deflated using three indices: monetary value, industry prices, and industry costs at the three-digit sector level. The deflation primarily targets the manufacturing industry, as the sample is limited to sectors for which reliable deflators are available from the Italian National Institute of Statistics (*Istituto Nazionale di Statistica*, ISTAT). The procedure for deflation and data cleaning is further detailed in Appendix B.

MEE Data (INPS' Uniemens) The National Institute for Social Security (INPS) provided us with detailed matched employer-employee records for all non-agricultural firms in the Italian private sector employing at least one worker. This unique panel comprises monthly employment histories of workers, yielding comprehensive employee-level information on demographic characteristics, labor earnings, contract type (temporary, permanent, apprenticeship), and working time arrangement (part-time or full-time). It also includes data on the collective contract applied to each worker. On the firm side, we observe company demographics—including establishment and cessation, suspension periods—the industry they operate in, and their workforce composition, size, and the total labor cost. To clean this dataset, we first select the primary employment relationship for each worker-year pair based on duration and, secondarily, earnings. We limit our sample to establishments employing at least five workers for a minimum of one year within our sample period to exclude very small firms for which a reliable TFP measure is unattainable. To match this information with firm-level data, we further restrict our *Uniemens* sample to firms present in the *Cerved* database. Appendix B provides additional details on our data cleaning choices for this dataset.

CCNL Data Our panel is supplemented with data on the renewal dates of each CCNL, provided by the National Center for Economy and Labor (*Centro Nazionale dell'Economia e del Lavoro*, CNEL).[6] This data allows us to capitalize on the staggered implementation of the employment legislation reform, forming the cornerstone of our reduced-form analysis, as elaborated in Section 2.3.1.

[6] We express our gratitude to Raffaele Saggio and his co-authors for sharing this dataset, which they initially gathered and used in their study Daruich et al. (2023).

TFP measures

Throughout the analysis, we focus on total factor productivity as the main measure of a firm's efficiency in production. TFP allows us to assess the effect of changes induced by the reform on a margin resulting from mechanisms that differ from (or complement) the simple adjustment of production factors that follow a change in their relative prices.

We leverage the panel structure of the data to derive a parametric estimate of the residual, ω_{it}, from a firm-specific Cobb-Douglas production function, represented as:

$$Y_{it} = K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M} \exp \omega_{it}$$

where Y_{it} are deflated sales, K_{it} is capital (assets), L_{it} is the labor force, and M_{it} is the deflated cost of materials for firm i at time t.

We have employed three distinct methodologies to measure TFP, each applied as firm-specific estimations within the industry:

1. The semi-parametric estimation using the control function approach proposed by Levinsohn and Petrin (2003)—referred to as LP.

2. The output-based proxy variable approach, which originated from Olley and Pakes (1996) and further refined by Ackerberg et al. (2015)—referred to as ACF.[7]

3. The nonparametric approach introduced by Gandhi et al. (2020)—referred to as GNR.

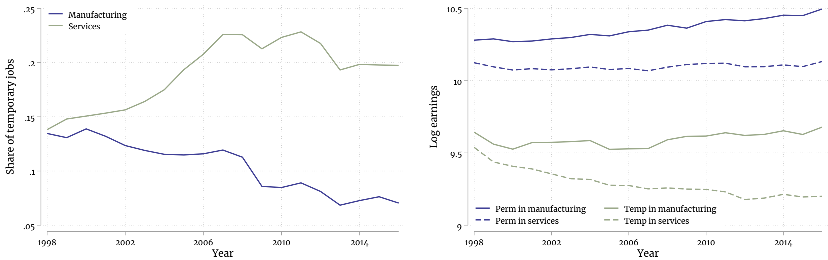
Incorporating these three measures is designed to reduce the dependence of our results on a specific model, given the potential for TFP variation across the numerous estimation methods outlined in the literature. Moreover, our use of output-based measures permits a larger sample size in our analysis. This contrasts with value-added measures that restrict the sample to only those firm-year pairs where the value remains non-negative after logarithmic transformation.

[7]Ackerberg et al. (2015) discuss that a purely Cobb-Douglas production function, similar to the one presented above, cannot be identified via a control function approach if one of its inputs is used as a proxy variable, *unless* there are frictions that impact firms' input timing decisions. EPLs, the focus of our paper, serve as a perfect example of such frictions influencing labor input. Consequently, we find the ACF method appropriate for estimating the entire production function in this setting.

FIGURE 2.2: Permanent vs. temporary contracts by macro-industry

(A) Share of temporary contracts

(B) Temp / perm earnings



Note.
Source: Istituto Nazionale della Previdenza Sociale (INPS).

2.2.4 Descriptive Statistics of the Sample

Following the data cleaning process detailed in Appendix B, we are left with a selected sample comprising 50,000 to 70,000 firms per year across approximately 65 three-digit sectors in manufacturing. The variability in the sample (Figure A.1) is primarily due to changes in the coverage of the Cerved database, which included an increasing number of firms over the years. Moreover, the sample size is influenced by two main factors: firstly, we retain firms for which we can observe balance sheet data from Cerved; secondly, we limit our analysis to sectors for which we can use industry-specific prices and cost indexes as meaningful deflators.

Figure 2.2 presents two notable descriptive statistics. First, the proportion in the use of temporary contracts has demonstrated different patterns across the manufacturing and service sectors. In the manufacturing sector, the share of temporary contracts remained largely stable over time, even showing a slight decrease. Conversely, in the service sector, this proportion rose from about 13% in 1998 to 23% in 2007, as depicted in Panel A. Concurrently, Panel B reveals that the real earnings of workers in the manufacturing sector remained largely constant over time, regardless of contract type, with a minor increase for permanent contracts. In contrast, temporary contracts in the service sector experienced a steady decline in real earnings, negatively mirroring the increase in their usage.[8]

[8]The slight decrease in temporary contract usage within the manufacturing industry may be attributed to two factors. First, the manufacturing macro-sector is characterized by unique technological and institutional attributes. Technologically, the less volatile nature of manufacturing pro-

Figure A.2 presents the time trend of selected statistics for the sample: mean, variance, interquartile range (p75-p25), and inter-extreme range (p90-p10) for the three considered TFP measures. The TFP growth trend is notably negative across all three measures. The overall variance shows an increase in the early years of the sample and a subsequent reduction that continued until 2016, the last year we consider. This trend remains consistent across the TFP measures. The same observation applies to the interquartile and the inter-extreme ranges, which display a much more stable evolution over time.

Figure A.3 depicts the dynamics of firm entry and exit in our sample, taken at the province-by-3-digit-sector level each year.[9] The number of exit events, considered as potentially temporary suspensions, has increased over time. The same trend is observed when looking at the rates. This may reflect the stagnation in growth that the country has experienced over the last 30 years. Consistently, the number of entry events, defined as possible reactivations, began to decline after the 2008 financial crisis when viewed in absolute numbers, and even earlier when considering the rates. However, both panels show that new firms continue to be established and enter the market at a higher rate than the rate at which they cease their activity. In addition to these macro-trends, this paper will examine how changes in EPLs have impacted firm dynamics and through which channels.

2.3 Empirical strategy and results

This section outlines the empirical approach utilized in our reduced-form analysis for assessing the impact of the labor cost shifts, induced by the reform, on firm productivity and other outcomes. Moreover, here we present and discuss our findings.

We first establish that the reform led to a decrease in average labor cost without altering the use of temporary contracts in the manufacturing industry. Next,

duction limits the necessity for rapid workforce adjustments in response to unexpected demand fluctuations. Institutionally, manufacturing firms frequently use *cassa integrazione*—an Italian wage support scheme during downturns—maintaining workforce flexibility when needed. Second, manufacturing firms might prefer to outsource labor through temporary employment agencies. This approach potentially minimizes administrative and bureaucratic burdens and mitigates litigation risk, which is highly relevant in the union-dense manufacturing sector. However, the dataset that we are currently using does not allow us to verify this hypothesis or to delve into the reasons behind the differential use of outsourced labor between the service and manufacturing sectors. As such, a comprehensive quantitative assessment of these issues is left for future research.

[9] For details on the construction of the two groups of entry and exit measures, please refer to Appendix B.

we provide robust causal evidence showing this decrease negatively impacted average firm productivity. Specifically, we demonstrate the reform's uneven distributional effects, which negatively impacted firms on the lower end of the ex-ante TFP distribution and moderately benefited those at the upper end. This finding indicates the average negative effect conceals substantial heterogeneity, mainly driven by numerous already underperforming firms. Lastly, we illustrate how the reform impacted entry and exit rates variably, enabling previously unproductive companies to endure.

2.3.1 Overview of the empirical strategy

Our empirical strategy, following Daruich et al. (2023) and Acabbi and Alati (2021), leverages the quasi-experimental variation introduced by the 368/2001 decree, as detailed in Section 2.2.1.

The reform allowed firms to apply a collective contract with eased use of temporary contracts right after its predetermined renewal date. To leverage this staggered implementation, we initially assign each firm a contract-specific renewal date by computing the modal collective contract code applied by a firm in 2001—the last year before the reform. In essence, we regard a firm as treated in the year its 2001 modal collective contract was renewed. This assignment procedure is reinforced by the evidence of a single contract's high representativeness within a firm, as indicated in Table 2.1 and discussed in Section 2.2.2. As a result, we can examine within-firm changes in productivity or other pertinent firm-level outcomes before and after the reform. This is achieved by dynamically comparing firms primarily operating under the new rules with a control group of firms that, within the same year, are still adhering to pre-reform requirements. Therefore, causal identification arises from comparing as-good-as-randomly early-treated firms to later-treated ones, given the covariates.

Our empirical analysis begins by using this identification strategy to establish two facts: the reform had no uptake within our manufacturing industry sample, and it led to an average, reduction in labor cost. We then estimate the impact on firm-level TFP. Initially, we run our specification on the whole sample to gauge the average effect of the reform-induced changes on overall productivity. We then broaden our analysis to evaluate the heterogeneity of effects based on firms' positions along the sector-specific TFP distribution *before* the reform. Considering the heterogeneity in pre-reform TFP allows us to determine if labor market changes, induced by the intervention, impacted the productivity of already less productive firms differently than more productive ones. We perform this heterogeneity analysis in two ways: first, we assign a time-invariant pre-reform TFP quartile to each firm in our sample, and then run separate specifications for the top and bottom quartile of pre-reform TFP. Second, we shift from comparing within-quartile average effect estimates to the distributional one by conducting a quantile treatment effect analysis. This exercise compares the observed TFP distribution after the reform to an imputed counterfactual distribution if the reform had not occurred.

Lastly, we shift our focus to other outcomes that we deem significant for explaining the mechanisms we believe underpin the productivity results. Specifically, we transition from a firm-level specification to a cell (province-by-industry) one to evaluate the reform's effect on firms' entry and exit events., again dividing the sample by ex-ante (for exits) and ex-post (for entries) TFP quartiles.

2.3.2 Specifications and results

In what follows, we discuss the empirical specification we run for each different outcome of interest, and we present the associated results.

Baseline event study (average effects)

We start our empirical analysis with a baseline event study to assess the average effects of the reform on different firm level outcomes.

Specification We quantify the effect of the reform's adoption on the use of temporary contracts estimating the average treatment on the treated firms following Callaway and Sant'Anna (2021).[10] More in detail, we estimate the following specification:

$$\text{ATT}(g,t;X) = \mathbb{E}\left[Y_{f,t} - Y_{f,g1}|G_f = g\right] - \mathbb{E}\left[Y_{f,t} - Y_{f,g-1}|G_f \in \mathcal{G}\right] \tag{2.1}$$

where $\text{ATT}(g,t;X)$ gives the average treatment effect at time t for the cohort of firms treated in time g: for example, $\text{ATT}(2003, 2005)$ measures the effect of the reform in 2005 on the group of firms that adopted the reform in 2003. For each firm f, we have that $g = c(f, 2001)$ where c is a function assigning the modal collective contract employed by firm f in 2001. Thus, g is the treatment year for

[10]The methodology proposed by Callaway and Sant'Anna (2021) addresses two main issues associated with standard dynamic TWFE specifications. Firstly, it offers accurate estimations even when dealing with variable treatment effects, as in our case. Specifically, this method not only avoids negative weighting but also provides control over how effects across cohorts are weighted. Secondly, this approach explicitly defines the units used as a control group to infer unobserved potential outcomes. This is in contrast to traditional TWFE models, which can result in perplexing comparisons during staggered implementations.

all firms that in 2001 used to employ a modal CCNL that was renewed in year g. As discussed, identification comes from comparing the expected change in the outcome of interest for cohort g between periods $g-1$ and t to that of a control group of not-yet treated firms in t. This set of dynamic controls is represented by $\mathcal{G}$, i.e., $\mathcal{G} \equiv \{g' : g' > t\}$. $Y_{f,t}$ is the outcome of interest for firm f at time t—in this section, the ratio of temporary contracts use by the firm in each year, the firm's per-worker labor cost and the TFP measures. More specifically, we run specification 2.1 on the residuals obtained from the intermediate specification

$$\tilde{Y}_{ft} = \psi_f + \lambda_{p(f),t} + \eta_{s(f),t} + \varepsilon_{ft}$$

where ψ_f is a firm fixed effect, $\lambda_{p(f),t}$ and $\eta_{s(f),t}$ are province- and sector-by-time fixed effects, respectively, and ε_{ft} is an error term.

To visualize the estimated effect in time deviation from the reform, we are interested in an *event study* parameter that represents the weighted average of the treatment effect k periods away from the adoption across the cohorts:

$$\text{ATT}_k = \sum_g w_g \text{ATT}(g, g+k) \tag{2.2}$$

where weights w weight cohorts for their relative frequencies in the treated population. Specifically, we consider the integers $k \in [-3, 3]$, thereby concentrating on a symmetric three-year window around the modal CCNL's renewal. Unless otherwise stated, we cluster the standard errors at the firm level.

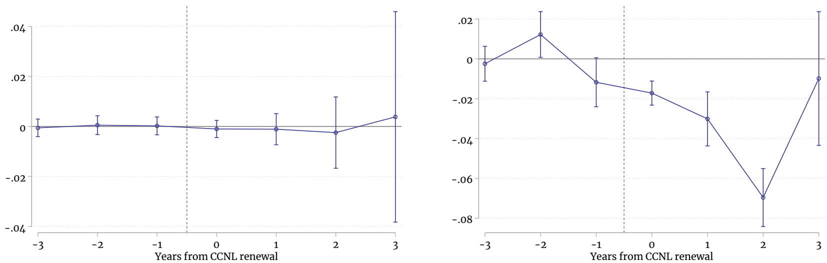
Results on the use of temporary contracts and labor cost On average, the reform did not result in an increased use of temporary contracts in the manufacturing industry. As shown in Figure 2.3, which illustrates the event study coefficients computed as in (2.2), the estimated effect is precisely null, with a perfectly sharp trend (Panel A). Concurrently, labor cost underwent a significant decrease. Panel B depicts the estimated coefficients on per-worker labor earnings, demonstrating a reduction of up to 7% after two years.

These results can be explained by focusing on the specificities of national collective bargaining institutions discussed in Section 2.2.2. Firstly, other studies that exploit the same reform across samples including non-manufacturing firms demonstrate that the intervention led to an average increase in the use of temporary contracts (Daruich et al., 2023). This suggests that the effect on labor cost we observe may be attributed to a transmission mechanism across sectors through collective bargaining. Specifically, we suggest that the increased use of temporary contracts and lower conversion rate from temporary to permanent in the

FIGURE 2.3: Avg. effect of the reform on use of temporary contracts and labor cost

(A) Use of temporary contracts

(B) Labor cost: per-worker earnings

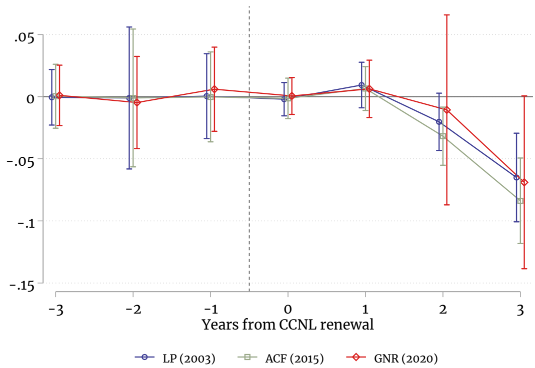


Note. This figure presents the event study coefficients calculated in accordance with equation (2.2) for the share of temporary contracts in the workforce (Panel A) and the labor cost as per-worker earnings (Panel B). Confidence intervals at the 95 percent level are obtained from firm-level cluster-robust standard errors. Source: Istituto Nazionale della Previdenza Sociale (INPS) and Cerved.

service sector documented by Daruich et al. (2023) decreased the overall workers' bargaining power *within* a collective bargaining agreement. As these agreements are extensively shared *across* industries (Figure 2.1) and exhibit minimal within-firm heterogeneity (Table 2.1), the change in the workforce composition induced in the service sector by the intervention deploys effects in the manufacturing sector, even without direct uptake of temporary work arrangements. This mechanism is particularly intriguing as it enables us to interpret the effect within manufacturing as a plausibly exogenous shift in labor cost—a change not endogenous to alterations in firms' choices regarding work arrangements.

Results on the average TFP The average overall impact on TFP due to the labor cost shift induced by the reform becomes noticeable after two years, and it's negative when statistically different from zero. Figure 2.4 shows that between two and three years after the modal 2001 CCNL has been renewed, a firm experiences, on average, a TFP reduction between 2 (two years) and 6 (three years) percent. These results remain qualitatively consistent across the three TFP estimation methods we use. Moreover, it's worth noting that our negative average result on productivity aligns with the findings of Cappellari et al. (2012), who also leverage the same reform but use a much smaller CCNL sample and survey data on firms' sectors.

FIGURE 2.4: Avg. effect of the reform on TFP



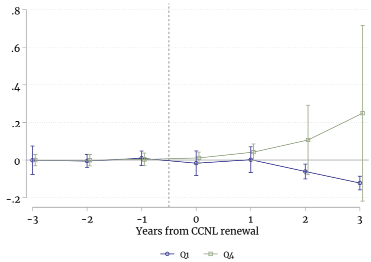
Note. This figure presents the event study coefficients calculated in accordance with equation (2.2) for the three different measures of TFP detailed in Section 2.2.3. Confidence intervals at the 95 percent level are obtained from firm-level cluster-robust standard errors. Source: Istituto Nazionale della Previdenza Sociale (INPS) and Cerved.

Results on the heterogeneous ex-ante TFP quartiles A natural question arises within our conceptual framework concerning the driver of this average effect. Specifically, we aim to understand whether the reform's impact varies across the heterogeneity margin defined by ex-ante productivity. In other words, we seek to answer: did the reform differentially impact firms that were initially more or less productive? To this end, we assign each firm to a time-invariant quartile of TFP, following a process similar to the one proposed by Devicienti and Fanfani (2021). More precisely, we first calculate the firm's position in the TFP distribution within a given sector-year pair. Subsequently, we assign each firm the modal quartile in which it was classified in the five years preceding the reform's enactment.

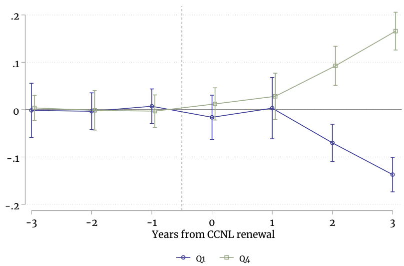
Figure 2.5 showcases the outcomes of this exercise, focusing on the top and bottom quartiles of the ex-ante TFP distribution. The labor cost shift caused by the reform had symmetrical effects on these quartiles. For firms that were already significantly unproductive, the impact is detrimental, with TFP declining between 4 and 8% after two years, depending on the estimation method applied. Conversely, among the most productive firms prior to the reform, the average effect is positive, though slightly less precisely estimated, with a 10% increase in TFP two years post-reform when productivity is calculated using the ACF method (Panel B). Despite minor discrepancies in the statistical significance of

Figure 2.5: Avg. effect of the reform on TFP, by pre-reform TFP quartiles

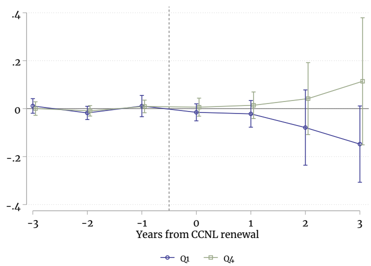
(A) Levinsohn and Petrin (2003)



(B) Ackerberg et al. (2015)



(C) Gandhi et al. (2020)



Note. This figure presents the event study coefficients calculated in accordance with equation (2.2) for firms within the bottom and the top time-invariant quartile of the pre-reform TFP distribution. Each panel shows the results using a different measure of TFP as the dependent variable: Levinsohn and Petrin (2003) (A); Ackerberg et al. (2015) (B); Gandhi et al. (2020) (C). Confidence intervals at 95 percent are obtained from firm-by-year level cluster-robust standard errors. Source: Istituto Nazionale della Previdenza Sociale (INPS) and Cerved.

the estimates, the symmetrical nature of the effect remains qualitatively consistent across all three methods. Furthermore, all three panels reveal no differences between treated and yet-to-be-treated firms in the pre-reform period, suggesting that these firms followed parallel trends prior to the reform. These findings substantiate considerable heterogeneity behind the results presented in Figure 2.4. Our estimates indicate that the reform negatively impacted the lower end of the distribution, with these underperforming firms primarily driving the overall negative effect.

Quantile Treatment Effects (heterogeneous effects on the TFP distribution)

In this section, we examine how the change in labor market flexibility induced by the reform has differently impacted total factor productivity based on firms' pre-reform productivity levels. So far, we've explored the average effects of the intervention *within* the quartiles of the ex-ante TFP distribution. However, we have yet to assess the direct distributional impact of the policy. We aim to compute the Quantile Treatment Effects (QTE) of the reform on the TFP distribution for firms. Specifically, we want to evaluate how the changes in labor market conditions has differentially affected the TFP distribution itself. Essentially, our aim is to estimate the TFP distribution across firms following the reform compared to the distribution had the liberalization not occurred.

Specification To estimate the Quantile Treatment Effect on the Treated (QTT), we use the following equation:

$$\mathrm{QTT}(\tau) = F^{-1}_{\mathrm{TFP}_{1,t}\,|\,D=1}(\tau) - F^{-1}_{\mathrm{TFP}_{0,t}\,|\,D=1}(\tau) \tag{2.3}$$

Here, τ is a quantile of the TFP distribution, and $F_{\mathrm{TFP}_{1,t}\,|\,D=1}$ and $F_{\mathrm{TFP}_{0,t}\,|\,D=1}$ represent, respectively, the distribution of a firm's potential productivity $\mathrm{TFP}_{1,t}$ and $\mathrm{TFP}_{0,t}$, conditional on running under the new rules. To accurately estimate the QTT, we need to identify the marginal distributions of potential productivity, which requires us to make two empirical assumptions.

Empirical Assumption 1 (Distributional Parallel Trends)**.** Define $\Delta\mathrm{TFP}_{0,t} = \mathrm{TFP}_{0,t} - \mathrm{TFP}_{0,t-1}$. Then,

$$\Delta\mathrm{TFP}_{0,t} \perp\!\!\!\perp D$$

In words, the distribution of the change in the untreated potential TFP must not depend on the treatment status. This is a generalization of the standard difference-in-differences parallel trends assumption applied to a non-linear context. Essentially, conditioned on covariates, the TFP trajectory observed after the reform should have been the same had the temporary contracts not been liberalized. Despite the usefulness of the Distributional Parallel Trends assumption, it isn't sufficient to fully identify the counterfactual distribution of the outcome on its own, as demonstrated by Fan and Yu (2012). To point identify the counterfactual distribution, an additional assumption, as suggested by Callaway and Li (2019), is needed.

Empirical Assumption 2 (Copula Stability)**.** Let $C\left(\Delta \text{TFP}_{0,t}, \text{TFP}_{0,t-1} \,|\, X, D = d\right)$ be the copula between the change in untreated potential TFP and its starting level, conditional on covariates X and being treated. Then,

$$C\left(\Delta \text{TFP}_{0,t}, \text{TFP}_{0,t-1} \,|\, X, D = 1\right) = C\left(\Delta \text{TFP}_{0,t-1}, \text{TFP}_{0,t-2} \,|\, X, D = 1\right)$$

The Copula Stability Assumption presumes that the copula, which describes the statistical dependence between the change in untreated potential TFP and its baseline level, remains constant over time for treated firms. [11] This is to say, if firms with higher TFP have historically exhibited greater increases in TFP, this trend will continue in the present, assuming no treatment is applied.

Given these two assumptions, we can identify the counterfactual marginal distribution in (2.3), thereby estimating the QTT.[12] Conceptually, the Copula Stability Assumption aids in identifying the joint distribution of $\left(\Delta \text{TFP}_{0,t}, \text{TFP}_{0,t-1} \,|\, D = 1\right)$, from which one can derive $F_{\text{TFP}0,t \,|\, D=1}$.[13]

Results Figure 2.6 presents the estimated distributional effects of the reform on firm-level TFP for ten deciles.[14] These results represent the short-term impact of the reform on the productivity distribution. Specifically, the specification (2.3) does not account for a dynamic effect k periods post-event; instead, it offers a straightforward pre-vs-post comparison. Therefore, we separately estimate the QTT for two specific cohorts, which comprise approximately 90% of the treated

[11]It is important to note that this assumption does not necessitate a specific parametric copula or a particular form of dependence, provided that a form of dependence exists and is consistent over time.

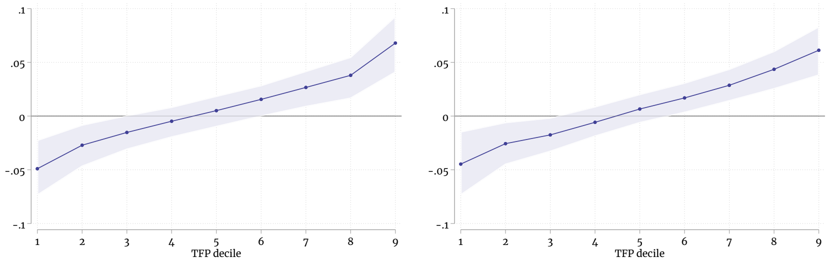
[12]To estimate the first term of equation (2.3), we simply need to invert the observed empirical distribution of the TFP for firms that adopt the reform within a specific year.

[13]It is worth noting that while the marginal distributions of $\Delta \text{TFP}_{0,t}$ and $\text{TFP}_{0,t-1}$ are identified by the Distributional Parallel Trend Assumption and data respectively, this does not inherently allow for the identification of the joint distribution. As proposed by Callaway and Li (2019), we utilize the observed dependence (the past copula) to construct the necessary information to identify $F_{\Delta \text{TFP}_{0,t}, \text{TFP}_{0,t-1} \,|\, D=1}$, leveraging the connection between the joint distribution and the copula function as established by Sklar's Theorem (Sklar, 1959).

[14]In this case, we employ the measure based on Ackerberg et al. (2015). The results remain consistent across other TFP specifications.

FIGURE 2.6: Quantile Treatment Effect for selected cohorts

(A) Reform adopted in 2002 (B) Reform adopted in 2003



Note. This figure presents the estimate of the QTT(τ) as specified in equation (2.3) for $\tau = (.1, \ldots, .9)$ on the residualized TFP. The results reveal considerable heterogeneity across the TFP distribution: the effect of the reform strictly increases with the quantiles, leading to a sign reversal in the coefficients. Confidence intervals at the 95 percent level are derived through bootstrapping with 1000 iterations. Source: Istituto Nazionale della Previdenza Sociale (INPS) and Cerved.

firms in our sample (Figure A.4): 2002 (Panel A) and 2003 (Panel B). Each panel thus illustrates the impact of the reform on the ten deciles of the TFP distribution for the two firm cohorts, two years subsequent to the intervention.

Our findings not only corroborate the significant heterogeneity observed in the event-study estimates, but also reveal a monotonic effect across the TFP distribution for both treated cohorts. In both instances, the impact of the reform varies from a 5% decrease for the bottom decile to a positive effect of the same magnitude on the top decile, relative to a counterfactual scenario where the reform did not occur. The negative effect at the lower end of the distribution diminishes as we ascend the distribution, eventually reversing. This result imparts additional insights beyond the heterogeneity analysis conducted through linear models in preceding sections. Through the event studies, we have quantified the average marginal effect of belonging to a specific segment of the pre-reform TFP distribution on the TFP. The QTT enables us to examine the direct effect of the reform *on* the quantiles of the TFP distribution, offering evidence of the heterogeneous distribution shifts that ensued from the increase in labor flexibility.

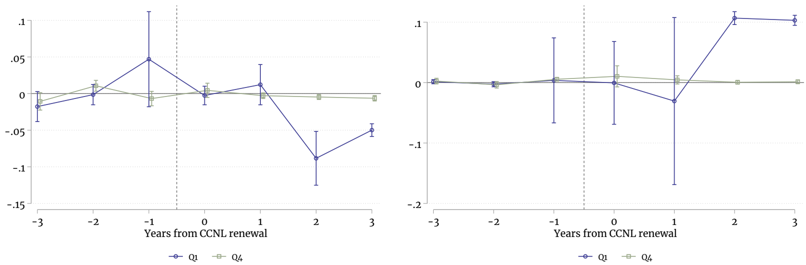
Cell-level event studies (entry and exit dynamics)

This section evaluates the impact of the reform on entry and exit dynamics at the cell level. A cell is defined as a specific combination of province and industry

FIGURE 2.7: Reform's effect on entry and exit, by TFP quartile

(A) Firms' exits

(B) Firms' entry



Note. This figure presents the event study coefficients calculated in accordance with equation (2.2) for entry (Panel A) and exit (Panel B) dynamics at the province-sector level, for the two extreme quartiles of ex-ante TFP. Source: Istituto Nazionale della Previdenza Sociale (INPS) and Cerved.

(classified to the 3-digit level). We apply the same specification detailed in equation (2.1) to this revised unit of observation. The appearance of a firm within the panel dataset is considered an entry, while its disappearance indicates an exit.[15] For this analysis, the dependent variable is the ratio of the number of entries or exits to the total number of firms in the cell in 2001. This ratio is residualized for year and industry fixed effects, in addition to industry-specific linear time trends. Each cell is allocated an event year, during which it is considered treated based on the modal CCNL implemented in that particular cell in 2001. Similar to the firm-level event studies, we execute quartile-specific estimations by assigning each firm to a time-invariant TFP quartile prior to the renewal of the CCNL (for exits) or following it (for entries).[16] Again, our aim is to determine whether the reform had differential effects on firm dynamics across the top and bottom of the productivity distribution within each cell.

Results The estimated ATT computed following (2.2) is reported in Figure 2.7. On average, the reform caused an increase in entry and a derease in exit, up to 10% and 8%, respectively, within two years from the CCNL renewal. This happened only within the bottom quartile of the ex-ante TFP distribution. On the contrary, the reform led completely unaltered the entry/exit dynamics for

[15]Firms persist in the MEE records as long as they maintain at least one employee in the given period.

[16]For this analysis, we utilize the TFP measure based on Ackerberg et al. (2015), but the results remain qualitatively consistent across the other measures.

firm the where already productive.

We interpret these results as a sharp evidence of the use of the decrease in labor cost low-productivity firms made to survive—or to enter the market. Thus, by lowering labor cost, the reform gave the opportunity to firms that would have probably left the market otherwise to continue operating, and allowed low-productivity firms that wouldn't have make it to the market, to enter it. The next Section discusses the intuition behind this mechanism and its links with results on productivity, thus introducing the model presented in Section 2.4.

2.3.3 A brief discussion on the mechanism at play

Here, we briefly discuss the intuition behind our empirical findings thus far. This discussion wraps up our empirical observations, links them to specific characteristics of the Italian labor market, and introduces some key ideas for the model presented in the next section.

First, we demonstrated that for manufacturing industries, the reform did not lead to increased use of temporary contracts. Yet, labor costs consistently decreased despite the absence of change in the composition of work arrangements. We explain the coexistence of these two phenomena with the unique characteristics of Italian labor bargaining, as discussed in Section 2.2.2. Briefly, the observed surge in the use of temporary jobs in the service sector, both in raw descriptive statistics (as shown in Figure 2.2) and following the reform (as shown in Daruich et al., 2023), weakened the bargaining power of workers *within* a collective contract. Since collective contracts are highly homogeneous within a firm and the same contract is applied across sectors, such propagation explains the observed decrease in labor costs. Thus, we consider the shift induced by the reform as a shift in firm-level labor costs, and interpret our results based on this assumption.

Next, we showed that the reform adversely affected the productivity of already unproductive firms, while improving it for already productive ones. Notably, this relationship holds not only within each quartile of ex-ante productivity, but also along the entire distribution in a monotonic way. Furthermore, the reform resulted in less market exit and more entries for unproductive firms.

We hypothesize that, thanks to the decrease in labor costs, a double negative selection mechanism was triggered at the bottom of the productivity distribution. This mechanism allowed unproductive firms to survive or enter the market. In equilibrium, the TFP composition was altered, demonstrating the productivity effects we documented. We formalize this intuition in the following section through a general equilibrium model that links labor wedges to the productivity distribution in the market.

2.4 Model

In this section, we present a comprehensive theoretical framework to help interpret our empirical findings. This framework connects the equilibrium productivity distributions across economic sectors with frictions in both labor and capital markets. Besides offering predictions that largely align with our primary empirical results, the model underscores the potential for labor wedges to have heterogeneous effects and an ambiguous net impact, as they may mitigate misallocation effects stemming from other types of distortions.

Our model builds on the well-established closed-economy monopolistic competition framework featuring heterogeneous firms, as developed by Melitz (2003). This approach allows us to separate various aspects of firm behavior—such as entry, exit, and investment—from other elements of the economy like labor supply. We incorporate financial frictions (FFs) due to asymmetric information and post-entry productivity-enhancing investments (PEIs) into our model. While both FFs and PEIs have been explored in prior work, our combined treatment is both innovative and tractable, as we will elaborate. Throughout our discussion, we maintain notation consistent with the original Melitz (2003) model.

We examine a standard Melitz (2003) economy where preferences for individual goods are described by a Constant Elasticity of Substitution (CES) with $\sigma > 1$. A representative consumer has a utility function $U^{\frac{\sigma-1}{\sigma}} = \int_{\omega \in \Omega} q(\omega)^{\frac{\sigma-1}{\sigma}} d\omega$, where Ω represents the set of varieties available in equilibrium, and $q(\omega)$ denotes the quantity of product $\omega \in \Omega$ consumed. Firms, which provide these varieties, are heterogeneous in productivity $\varphi(\omega) > 0$ and are characterized by a linear cost function with increasing returns due to fixed costs $f > 0$ incurred during each period of operation. The labor demand function, also linear in quantity, is expressed as $l(q) = f + q/\varphi$. In this economy, labor is inelastically supplied by a mass of workers L. Each unit of labor receives a wage w, which we normalize to unity ($w = 1$).

This economy inherits all the standard properties of the monopolistic competition model by Dixit and Stiglitz (1977) as developed later by Melitz (2003). In particular, for each firm optimal quantity is a power function of productivity with exponent σ, whereas both revenue and profit scale with exponent $\sigma - 1$. Hence, for any two firms with productivity φ_1 and φ_2, the ratio of their equilibrium revenues is $r(\varphi_1)/r(\varphi_2) = (\varphi_1/\varphi_2)^{\sigma-1}$. As in Melitz (2003), the probability distribution of productivity in this model is endogenous, represented by a density function $\mu(\varphi)$. However, the determination of this distribution in equilibrium differs in our model. Alongside a modified entry decision that firms undertake, they also need to obtain setup financing from financial intermediaries. In what follows, we refer to financial intermediaries more simply as "banks."[17]

The set of varieties Ω with associated productivity $\varphi(\omega)$ is determined through the interaction between *entrepreneurs* and *banks*. We define an entrepreneur as a pair (φ, θ), where $\theta > 0$ is an individual signal about the productivity φ. These two random variables are drawn from a joint probability distribution $G(\varphi, \theta)$, which is common knowledge; however, they are initially unobserved by all agents involved. Banks are instead described as a mass B of risk-neutral workers endowed with the ability to convert any unit of labor into a unit of "capital," a unique good used solely to set up firms.[18] The process of creating firms goes as follows:

1. A given mass of entrepreneurs decides whether to attempt setting up a firm. To do so, they must incur a one-time, sunk experimentation cost f_n. This provides information about the signal θ, which both entrepreneurs and banks observe.

2. Next, firms must secure capital financing equal to f_b units of labor, which only banks can provide. The true productivity φ is revealed only after both f_n and f_b are paid. In return for paying f_b, banks demand a permanent claim over a share $b(\omega) \in (0, 1]$ of *all* future profits $\pi(\omega)$ of a firm supplying variety ω. The capital market is perfectly competitive: entrepreneurs can purchase capital from any bank without frictions.

3. Lastly, all extant firms set their prices and quantities and may even choose to exit and supply zero output if the optimal profits conditional on producing are negative (due to fixed costs). Firms then operate in the economy until an event occurring with exogenous probability δ forces them to exit.

This enhanced entry-stage model incorporates financial frictions (FFs) due to informational asymmetries. At the financing stage, banks are unable to see or verify entrepreneurs' true productivity (irrespective of whether the entrepreneurs themselves can). Existing versions of the Melitz model that incorporate FFs (see Manova, 2013 and Chaney, 2016) typically introduce liquidity constraints that

[17]This is fitting in the Italian setting, where commercial banks dominate capital markets.

[18]This is a simplification and normalization, as a more elaborate production function for the capital good would not significantly alter the analysis.

firms face only when they encounter costs for entering foreign markets. In our model, FFs also impact entry into the domestic market, as our primary objective is to elucidate potential sources of misallocation while abstracting from trade considerations.[19]

Our choice to model FFs through informational frictions enhances the tractability of the model. It helps isolate one specific channel: the firm selection on the extensive margin in the left tail of the productivity distribution.[20] For the sake of clarity, we will momentarily abstract from post-entry investments (PEIs). Following our analysis of the (closed) economy with financial frictions at the entry stage, we will discuss the implications of post-entry choices about PEIs made by firms.

2.4.1 Analysis

Once the set of firms that have paid both entry fixed costs f_n and f_b is determined, firm behavior proceeds as in the Melitz model. To understand how financial frictions affect firms' selection, we solve the entry stage recursively, starting from the financing stage. To clarify the trade-offs that banks face, we introduce an innocuous assumption.

Assumption 1. *Signal informativeness*: if $\theta_1 > \theta_2$ are two different realizations of the signal θ, then $G(\varphi|\,\theta_1) \leq G(\varphi|\,\theta_2)$ for any $\varphi > 0$.

This assumption states that signals are ordered in a way that higher values lead to conditional distributions of productivity that first-order stochastically dominate those from lower values.[21]

There are two key implications of Assumptions 1: first, lower signals imply a higher risk for banks; second, as θ is the only information that banks receive about firms, set shares $b(\theta)$, that is the fraction of total equity they demand to entrepreneurs in exchange for f_b, which is only a function of the signal. Therefore, when financing an entrepreneur with signal θ, a bank's expected profit is

[19]Our distinction between the two fixed costs f_n and f_b can be viewed as a form of liquidity constraint: of the full Melitz entry cost f_e, entrepreneurs are only able to pay $f_n < f_e$ upfront, with $f_b < f_e - f_n$ to be financed by banks.

[20]In a recent contribution, Unger (2021) introduced an augmented Melitz model where firms face financial frictions in the post-entry stage, as they need to anticipate part of both variable and fixed production costs in every period before realizing revenues. Contrary to our model, his framework (which features moral hazard) predicts that financial frictions lead to a more intense selection effect, as the least productive firms face tighter access to credit.

[21]This comes without any loss of generality: as signals are abstract, they can always be transformed in such a way that Assumption 1 holds by construction.

$\widetilde{\pi}(\theta)\, b(\theta)/\delta - f_b$, where $\widetilde{\pi}(\theta)$ is the *unconditional* per-period profit (which incorporates the probability that a firm exits after observing φ) that one can expect from setting up a firm under signal θ.

Perfect competition in capital markets leads to an equilibrium where banks make zero profits in expectation. The reason is straightforward: there cannot be an equilibrium where $\widetilde{\pi}(\theta)\, b(\theta)/\delta > f_b$ for any value of $\theta > 0$, or else any subsets of banks with mass $B' < B$ would find it profitable to set a strictly lower share $b'(\theta) < b(\theta)$ and capture all the profits from firms generated by signal θ. Banks would not make negative profits either, as they would simply deny financing to all entrepreneurs with signal values such that $\widetilde{\pi}(\theta)\, b(\theta)/\delta < f_b$. If such a strict inequality is theoretically possible in the support of θ for some fixed primitives of the model, Assumption 1 implies the existence of a threshold signal that makes banks indifferent towards financing an entrepreneur under the assumption that they would capture all the profits of the resulting firm, i.e., the smallest positive number θ^* such that

$$\frac{\widetilde{\pi}(\theta^*)}{\delta} - f_b = 0. \tag{2.4}$$

We guess that a suitable value of θ^* exists; we verify *ex post* whether this is true.

This analysis implies that in equilibrium, only those firms with signal $\theta \geq \theta^*$ receive financing, $b(\theta^*) = 1$, and for any two signals $\theta_1 \geq \theta^*$ and $\theta_2 \geq \theta^*$, banks set shares that yield zero profits in expectation with the property that $b(\theta_1)/b(\theta_2) = \widetilde{\pi}(\theta_2)/\widetilde{\pi}(\theta_1)$, and $\widetilde{\pi}(\theta)\, b(\theta) = \widetilde{\pi}(\theta^*)$ for any $\theta \geq \theta^*$. [22] Since (2.4) completely summarizes the trade-off faced by banks and the equilibrium in the capital markets, we call it (with some abuse of terminology) the Arbitrage Condition (AC), as it subsumes the fact that banks demand higher shares in exchange for riskier signals.

The initial entry decision by entrepreneurs is conceptually simpler. The expected value of generating a business idea is $v_n = \delta^{-1} \int_{\theta^*}^{\infty} \widetilde{\pi}(\theta) \left[1 - b(\theta)\right] dC(\theta)$, where $C(\theta)$ is the marginal cumulative distribution of the signal θ. Since entrepreneurs are free to attempt entering the economy and generate new signals, they would only refrain from doing so if the value of entry v_n falls shorter of the experimentation cost f_n. Thus, incorporating the equilibrium in the subsequent financing subgame and the value of the bank share $b(\theta)$ implies the following Free Entry (FE) condition in the economy:

$$\int_{\theta^*}^{\infty} \frac{\widetilde{\pi}(\theta)}{\delta} dC(\theta) - \left[1 - C(\theta^*)\right] f_b - f_n = 0. \tag{2.5}$$

[22] This result can be formulated formally as a Bayes-Nash equilibrium.

Together with the Arbitrage Condition (2.4), this equation characterizes the economy's equilibrium. As (2.5) shows, entrepreneurs anticipate the probability of bearing the financing cost f_b, which they only bear if they receive a signal $\theta \geq \theta^*$.

To complete the analysis, it is necessary to characterize the function $\widetilde{\pi}(\theta)$. Following the analysis of the post-entry phase of the Melitz model, given a value of θ one has:

$$\widetilde{\pi}(\theta) = \mathbb{E}_{\varphi|\theta}\left[\pi(\varphi)\middle|\theta\right] = f\left\{\int_{\varphi^*}^{\infty}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} g(\varphi|\theta)\,d\varphi - \left[1 - G(\varphi^*|\theta)\right]\right\}, \quad (2.6)$$

where $g(\varphi|\theta)$ is a conditional density function derived from $G(\varphi|\theta)$, and φ^* is the threshold value of productivity below which, in equilibrium, firms find production unprofitable and exit. Note that (2.6) implicitly embeds a "Zero Profit Condition" *à la* Melitz, which is specific to θ. A pair of thresholds (θ^*, φ^*), one for the signal and one for productivity, completely determines the equilibrium—if one exists.

Here, we demonstrate the existence and uniqueness of equilibrium in a specific case outlined in the following Assumption.

Assumption 2. *Log-normality*: $G(\varphi, \theta)$ is a cumulative bivariate (joint) log-normal distribution with standard log-normals as its marginals. Let $\rho \equiv \mathbb{C}\text{orr}(\log\theta, \log\varphi)$.

This assumption of normality simplifies the analysis without compromising the model's realism, given that a (truncated) normal distribution is a well-known approximation for log-productivity distributions. The assumption that the marginals are standard is a normalization that does not limit the generality of the model. It should be noted that Assumptions 1 and 2 imply that $\rho \geq 0$: the signal and the true productivity are non-negatively correlated.

Proposition 2. An equilibrium pair (θ^*, φ^*) always exists, is unique, and is identified by the intersection of the curve of points satisfying the AC, given by $\varphi^* = A(\theta^*)^\rho$ for a suitable constant $A > 0$, and a globally concave curve tracing the points that satisfy the FE condition. The intersection always occurs at the global maximum of the implicit function of φ^* for θ^*, as traced out by the FE curve.

Proof. The proof is provided in Appendix C.1. □

Figure 2.8 illustrates the equilibrium as the intersection between the two solid lines. The AC curve is consistently increasing because higher threshold values set by banks result in higher average productivity due to the selection of better firms, and vice-versa. The FE curve, on the other hand, is concave due to the interaction of two mechanisms. As in the Melitz model, the higher the productivity threshold, the higher the profits required to motivate entry. The same mechanism applies to the signal threshold; hence for low values of θ^*, the latter increases alongside φ^* on the FE curve. However, a higher signal threshold implies a lower probability that firms repay the financing cost f_b, thereby increasing the relative entry value. This latter effect dominates at high values of θ^* and causes the FE curve to decrease in that section. The equilibrium is located at the maximum of the FE curve due to perfect competition among banks: they lend the financing cost f_b as long as the benefits outweigh the costs.

Proposition 3. Introducing a wedge $\tau > 0$ to firms' labor costs (but not to either entry cost f_n or f_b) such that the effective wage increases from $w = 1$ to $w_{(\tau)} = 1 + \tau$, results in an equilibrium $\left(\theta^*_{(\tau)}, \varphi^*_{(\tau)}\right) \gg (\theta^*, \varphi^*)$ where both thresholds are higher with the wedge.

Proof. The proof is provided in Appendix C.2. □

The intuition behind this is straightforward: higher labor costs make it more challenging for firms to repay their fixed costs and remain in the market, leading to a more stringent selection. The Melitz model translates this effect in a downward rotation of the Zero Profit Condition curve. In our model, the wedge induces a leftward rotation of the AC curve and a rightward shift of the FE curve. As the two curves must still intersect at the maximum of the implicit function for θ^* as traced out by the FE curve, both equilibrium thresholds inevitably increase. This is graphically represented by the two dashed curves in Figure 2.8.

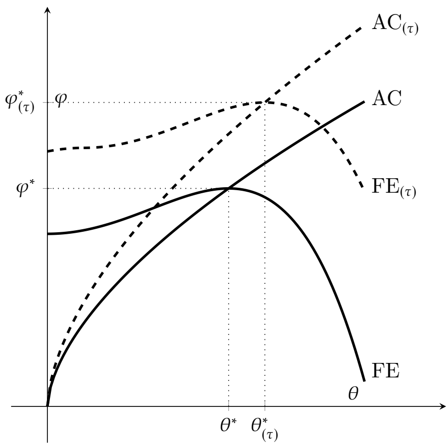
2.4.2 Welfare Implications

This model presents intriguing and non-trivial implications for welfare. There are two primary distinctions from the Melitz model. Firstly, there are three types of labor to be compensated: entrepreneurial (L_n), banking (L_b), and production (L_p) labor, with the total labor force being $L = L_n + L_b + L_p$. Secondly, the presence of FFs leads to the optimality result by Dhingra and Morrow (2019), which posits that the Melitz economy provides the optimal product diversity and firm size distribution, to collapse. This is summarized in the following proposition.

Proposition 4. Social welfare increases with ρ, and reaches its maximum at $\rho = 1$.

Proof. The proof is provided in Appendix C.3. □

FIGURE 2.8: Equilibrium of the model and comparative statics



Note. This figure illustrates the equilibrium pair (θ^*, ϕ^*) as the intersection between the solid lines representing the AC condition (2.4) and the FE condition (2.5). The AC curve slopes upward because a higher signal threshold leads to higher productivity due to selection, and vice versa. For $\theta \leq \theta^*$, increasing productivity and signal thresholds necessitate higher profits to incentivize entry, hence they increase alongside the two. Conversely, if $\theta \geq \theta^*$, a higher signal threshold reduces the probability of repaying the financing cost f_b, thereby increasing the relative entry value. In this region, the FE curve slopes downward. The dashed line illustrates the shift in the FE curve and the leftward rotation of the AC curve resulting from the introduction of a wedge $\tau > 0$ in the labor cost.

The intuition behind this is straightforward: more informative signals result in a more efficient allocation of the financing cost f_b. This is best exemplified by two extreme, degenerate cases, also discussed in the Appendix, where $\rho = 0$ and $\rho = 1$ respectively, which can be viewed as two Melitz economies with different primitives. If $\rho = 0$, all entrepreneurs are financed in equilibrium, even if they exit the economy after the revelation of their true productivity. Conversely, if $\rho = 1$, no financial resources are wasted as all financed entrepreneurs remain in the economy.

A significant implication is that price distortions may lead to second-best outcomes.

Proposition 5. Introducing a wedge $\tau > 0$ to firms' labor costs (but not to either entry cost f_n or f_b), which does not factor into workers' utility, increases average productivity in equilibrium and has ambiguous effects on social welfare.

Proof. The proof is provided in Appendix C.4. □

This concept is illustrated in the Appendix, also referring to the two extreme cases mentioned above. The intuition is as follows: in the presence of FFs, higher labor costs make entry less profitable (thus reducing social welfare), but they also raise the signal threshold θ^* as per Proposition 3. This in turn leads to higher equilibrium productivity (due to a pure *selection* effect) and less "waste" on the financing cost f_b. Both of these mechanisms enhance welfare, counteracting the negative impact on welfare from reduced entry. This holds under the assumption that τ does not affect workers' compensation and social welfare *per se*. If τ is due to EPLs, it is likely that workers derive utility from it.

2.5 Conclusions

This paper introduces new evidence concerning the relationship between labor market institutions and productivity within the manufacturing sector. We analyze an Italian labor reform intended to ease restrictions on the use of temporary contracts to evaluate its impact on total factor productivity (TFP), following an indirect reduction in labor cost. Fundamental characteristics of Italian collective bargaining institutions are documented, enabling us to leverage the within-collective contract relationship between changes in contract composition and labor cost across macro-industries.

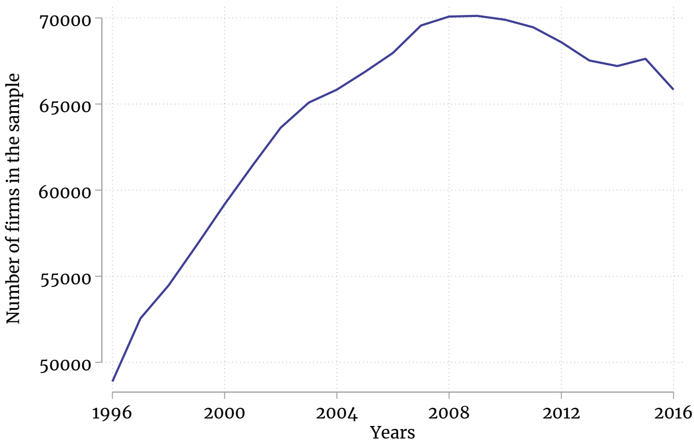
Indeed, we demonstrate that while the reform did not affect the use of temporary contracts within the manufacturing sector, it did reduce labor cost. We then illustrate how the reform decreased the TFP within the lowest quartile of the ex-ante productivity distribution, while inversely affecting the top of the distribution. We also note that these distributional effects grow progressively within the distribution, suggesting the more productive the firm, the greater the reform's impact. Combining this evidence with the observations of reduced exits and increased entries solely among the lowest quartile of ex-ante productivity, we propose that the reform induced a negative selection mechanism at the bottom of the TFP distribution.

We rationalize our findings through a steady-state general equilibrium model featuring monopolistic competition and heterogeneous firms. Financial frictions necessitate an upfront investment for firms to enter the market, supplied by financial intermediaries. An asymmetric information problem arises, as financial intermediaries receive only a noisy signal about firms' productivity. Our model indicates that increased labor costs contribute to selection at the lower end of the productivity distribution, leading to fewer entries. This mechanism aligns with our empirical evidence, as lower labor costs and decreased exits among unproductive firms jointly explain the observed heterogeneous effect on TFP. Our model also accounts for productivity gains on the right tail due to incentives to invest arising from labor cost savings.

Overall, this work shows that large-scale policy interventions aimed at improving labor market flexibility can have ambiguous interpretations. A variety of labor market mechanisms contribute to determining observable outcomes, with some them warranting further investigation. A full-fledged welfare analysis of our model's implications is left for future refinement.

A Additional figures and tables

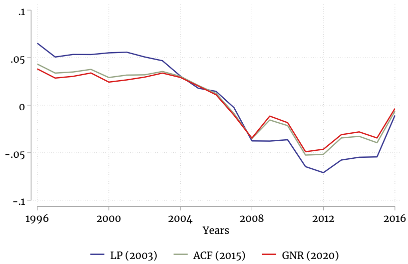
FIGURE A.1: Sample size evolution



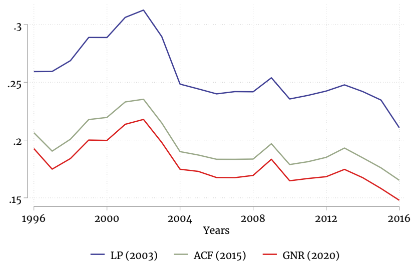
Note. This figure illustrates the evolution of the sample size over the sample period from 2006 to 2016. The observed change is primarily due to the expansion of firms' balance sheets recorded in the Cerved database. Source: Cerved.

FIGURE A.2: Descriptive statistics of the sample

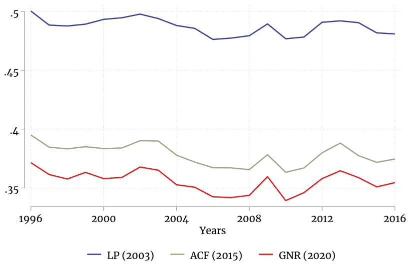
(A) Mean



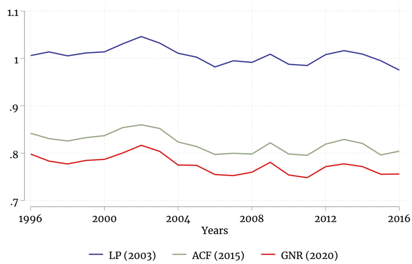
(B) Variance



(C) Inter-quartile range (p75-p25)



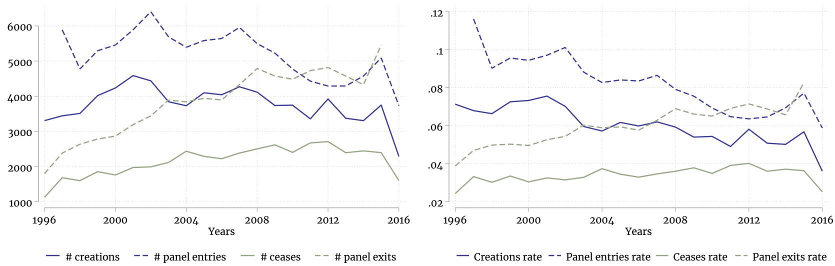
(D) Inter-extreme range (p90-p10)



Note. This figure reports time series of different descriptive statistics of the sample—mean (A), variance (B), inter-quartile range (C), inter-extreme range (D)—for the three TFP measures employed in the paper. Source: Istituto Nazionale della Previdenza Sociale (INPS) and Cerved.

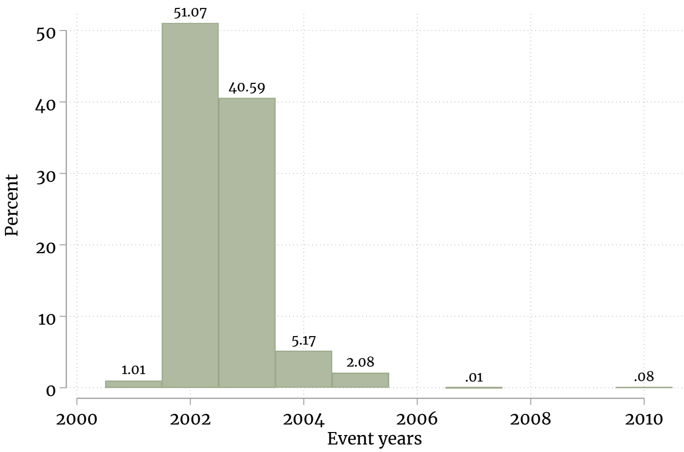
Figure A.3: Firm dynamics in the sample

(A) Exit / entry events　　　　(B) Exit / entry rates



Note. This figure reports the entry and exit events (Panel A) and the entry and exit rates (defined as the ratio between the events out of the firms population in each year; Panel B) for the sample. Source: Istituto Nazionale della Previdenza Sociale (INPS).

Figure A.4: Event years distribution



Note. This figure presents the relative percentages of event years at the firm level. An event year is defined as the year of renewal for the most commonly used National Collective Labor Contract (CCNL) among a firm's workforce in 2001. Source: Istituto Nazionale della Previdenza Sociale (INPS) and Consiglio Nazionale dell'Economia del Lavoro (CNEL).

B Additional Information on Data Cleaning and Deflation

B.1 Deflation of Cerved Measures

We adjust the firm-level balance sheet measures for inflation using three different indexes from ISTAT, the Italian National Institute of Statistics. Firstly, we adjust capital measures—fixed assets and liquidity—using a purchasing power index. Secondly, we adjust revenues using an industry-specific price index, matched at the finest available digit each year. Lastly, we use an imputed cost index at the industry level to adjust production inputs—net purchases and labor costs, matching these indexes at the best possible digit.

In more detail, we construct this latter measure through the following steps.

1. we normalize the input-output table so that each matrix element represents the relative weight an input has in the output costs in a given year;
2. for each output sector-year pair, we construct a cost index through a weighted sum of the cost indexes of the input sectors;
3. each cost index is assigned to the best available industry-specific price index in ISTAT.

We set the base year of all three indexes in 2015. When the industry price indexes are unavailable from 1996 to 1999, we extrapolate the available points of the series to predict these observations. Specifically, we use an ARIMA$(0, 1, 0)$ with a subset of external predictors, primarily the series of the lagged salary index.

B.2 Details on Data Cleaning

Here, we detail all the data-cleaning operations performed on the different datasets.

We construct a firm-year panel dataset starting from the Uniemens database. In this dataset, we assign a unique province and industry for each observation—since a single firm might operate in more than one sector or geographical region with some branch—keeping the observation with the highest number of employees.

For the matched employer-employee dataset, we first discard the contracts that lasted less than nine weeks in a year. Then, we assign to each worker-year firm only one establishment. Firstly, we resolve multiple spells *within* the same employer in a year by keeping the one that pays more. Then, we resolve multiple spells in different employers within the same year by adopting a double criterion: we keep the one that pays more and, subordinately, the one that involves more worked months. Finally, we discard contracts with no wage, and we winsorize the wage outliers on the right at the 99.7 percentile.

We clean the Cerved firm-level panel dataset by winsorizing all the relevant balance sheet variables at the 0.1 and 99.9 percentiles to remove outliers and by replacing negative values among variables that should not contain one (costs, revenues, purchases, and assets) as missing values: this way, we can still use the observation in its valid information. Besides the labor cost measure that this dataset contains, we also add the one obtained by the matched employer-employees, where we can collapse individual wages at the firm-year level. We then remove industries with less than forty firms in the entire period and industry-province-year cells that do not contain at least three establishments. Finally, after each TFP estimation, we discard the estimates that report at least one negative coefficient, and we further restrict the sample to industries for which we have non-missing estimates of all three productivity measures.

B.3 Details on Exit and Entry Events

We use the record of a registered cessation of a firm in a specific year from the INPS dataset to identify permanent exit events, i.e., those cases in which a firm permanently exits the markets, communicating this fact to the Social Security Institute. Similarly, we use the registered firms' creations to identify entries of newborn employers.

Moreover, since firms appear in the INPS panel as long as they employ at least one worker (either full- or part-time; either with a temporary or permanent contract), we interpret disappearances from the dataset in specific years as a signal of firms' inactivity in those periods. Specifically, we flag as a *disappearance* event a period of at least two consecutive years in which the firm does not appear in the panel. Conversely, we consider a reappearance in the panel as an indication of a firm's reactivation—with the same reasoning applied for the disappearances from the panel. It is important to note that in both cases, the second definition of an exit or entry event encompasses the first one.

C Additional analysis of the model

C.1 Analysis of Proposition 2

It is useful to establish some auxiliary notation first. Let:

$$\begin{aligned} t &= \log\theta \\ p &= \log\varphi \\ u &= -\log\theta \\ u' &= -\log\theta + \rho(\sigma-1) \\ z &= \frac{\log\varphi - \rho\log\theta}{\sqrt{1-\rho^2}} \end{aligned}$$

and use asterisks to denote the values of these transformations evaluated at the corresponding threshold value of their argument(s): thus, $t^* = \log\theta^*$, $p = \log\varphi^*$, *et cetera* (but $z^* = (p^* - \rho t)/\sqrt{1-\rho^2}$ is a function of $t = \log\theta$). In addition, let $\phi(x)$ be the probability density function of the standard normal distribution and $\Phi(x)$ the corresponding cumulative distribution – both evaluated at a given point x—and $\Phi_\varrho(x, y)$ be the cumulative bivariate normal distribution with standard normal marginals and correlation parameter ϱ—evaluated at point (x, y).

We start by elaborating expression (2.6) under the model's assumptions:

$$\begin{aligned} \frac{\widetilde{\pi}\left(e^t\right)}{f} &= \int_{p^*}^{\infty} \frac{e^{(\sigma-1)(p-p^*)}}{\sqrt{1-\rho^2}} \phi\left(\frac{p-\rho t}{\sqrt{1-\rho^2}}\right) dp - \left[1 - \Phi\left(\frac{p^*-\rho t}{\sqrt{1-\rho^2}}\right)\right] \\ &= \int_{z^*}^{\infty} e^{(\sigma-1)\left(\sqrt{1-\rho^2}z+\rho t-p^*\right)} \phi(z)\, dz - \left[1-\Phi(z^*)\right] \\ &= e^{(\sigma-1)(\rho t-p^*)+\frac{1}{2}(\sigma-1)^2(1-\rho^2)} \int_{z^*}^{\infty} \phi\left(z-(\sigma-1)\sqrt{1-\rho^2}\right) dz - \Phi(-z^*) \\ &= e^{(\sigma-1)(\rho t-p^*)+\frac{1}{2}(\sigma-1)^2(1-\rho^2)} \Phi\left((\sigma-1)\sqrt{1-\rho^2}-z^*\right) - \Phi(-z^*) \\ &= e^{(\sigma-1)(\rho t-p^*)+\frac{1}{2}(\sigma-1)^2(1-\rho^2)} \Phi\left(\frac{\rho t-p^*+(\sigma-1)\left(1-\rho^2\right)}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\rho t-p^*}{\sqrt{1-\rho^2}}\right). \end{aligned}$$

Therefore, the Arbitrage Condition (2.4) reads:

$$e^{(\sigma-1)(\rho t^*-p^*)+\frac{1}{2}(\sigma-1)^2(1-\rho^2)} \Phi\left(\frac{\rho t^*-p^*+(\sigma-1)\left(1-\rho^2\right)}{\sqrt{1-\rho^2}}\right) - \Phi\left(\frac{\rho t^*-p^*}{\sqrt{1-\rho^2}}\right) - \frac{\delta f_b}{f} = 0,$$

with an associated implicit function $p^* = \rho t^* + a$ where $a = \log A$—as one can verify by setting the total differential at zero. It is also possible to verify that plugging this implicit function back into the right-hand side of the above AC delivers a decreasing monotone function of a which cuts the x-axis so long as $\delta f_b / f > 0$. Therefore, a (and hence A) is unique, and it is both decreasing in f_b and increasing in f.

To analyze the Free Entry Condition, it is helpful to define $\widetilde{v} \equiv \int_{\theta^*}^{\infty} \widetilde{\pi}(\theta)\, dC(\theta)$ as the expected joint profits that accrue to both the entrepreneur and the bank following the experimentation stage. This quantity can be expressed as the following function of the two threshold values (t^*, p^*):

$$
\begin{aligned}
\widetilde{v}(t^*, p^*) &= f \int_{t^*}^{\infty} \pi\left(e^t\right) \phi(t)\, dt \\
&= f \int_{t^*}^{\infty} e^{(\sigma-1)(\rho t - p^*) + \frac{1}{2}(\sigma-1)^2\left(1-\rho^2\right)} \Phi\left(\frac{\rho t - p^* + (\sigma-1)\left(1-\rho^2\right)}{\sqrt{1-\rho^2}}\right) \phi(t)\, dt \\
&\qquad - f \int_{t^*}^{\infty} \Phi\left(\frac{\rho t - p^*}{\sqrt{1-\rho^2}}\right) \phi(t)\, dt \\
&= f e^{\frac{1}{2}(\sigma-1)^2 - (\sigma-1)p^*} \int_{-\infty}^{-t^* + \rho(\sigma-1)} \Phi\left(\frac{-\rho u' - p^* + (\sigma-1)}{\sqrt{1-\rho^2}}\right) \phi(u')\, du' \\
&\qquad - f \int_{-\infty}^{-t^*} \Phi\left(\frac{-\rho u - p^*}{\sqrt{1-\rho^2}}\right) \phi(u)\, du \\
&= f \left[e^{\frac{1}{2}(\sigma-1)^2 - (\sigma-1)p^*} \Phi_\rho\left(-p^* + \sigma - 1, -t^* + \rho(\sigma-1)\right) - \Phi_\rho\left(-p^*, -t^*\right) \right],
\end{aligned}
$$

where the last line follows from the analysis of the moments of the standard normal cumulative distribution as in **?**. Write the Free Entry condition as follows:

$$
\begin{aligned}
\mathcal{H}(p^*, t^*) = e^{\frac{1}{2}(\sigma-1)^2 - (\sigma-1)p^*} \Phi_\rho\left(-p^* + \sigma - 1, -t^* + \rho(\sigma-1)\right) - \\
- \Phi_\rho\left(-p^*, -t^*\right) - \frac{\delta f_b}{f} \Phi(-t^*) - \frac{\delta f_n}{f} = 0.
\end{aligned}
$$

The derivative of the above with the respect to the log-productivity threshold p^* is, following some manipulation, shown to be always negative:

$$
\frac{\partial \mathcal{H}(p^*, t^*)}{\partial p^*} = -(\sigma-1)\, e^{\frac{1}{2}(\sigma-1)^2 - (\sigma-1)p^*} \Phi_\rho\left(-p^* + \sigma - 1, -t^* + \rho(\sigma-1)\right) < 0.
$$

Instead, the derivative with respect to the log-signal threshold t^* is shown to be:

$$\frac{\partial \mathcal{H}(p^*, t^*)}{\partial t^*} = -\left[e^{(\sigma-1)(\rho t^* - p^*) + \frac{1}{2}(\sigma-1)^2(1-\rho^2)} \Phi\left(\frac{\rho t^* - p^* + (\sigma-1)(1-\rho^2)}{\sqrt{1-\rho^2}} \right) - \right.$$
$$\left. - \Phi\left(\frac{\rho t^* - p^*}{\sqrt{1-\rho^2}} \right) - \frac{\delta f_b}{f} \right] \phi(t^*)$$

which is not a monotone function of t^*. However, an analysis of this derivative shows that, for a fixed p^*, it is $\lim_{t^* \to -\infty} \partial \mathcal{H}(p^*, t^*) / \partial t^* = \lim_{t^* \to \infty} \partial \mathcal{H}(p^*, t^*) / \partial t^* = 0$; that the derivative equals exactly 0 whenever $t^* = (p^* - a)/\rho$ (observe that the expression in brackets matches the Arbitrage Condition); and that to the left of this value, the derivative is positive, while on the right, it is negative. These results give rise to the pattern depicted in Figure 2.8, with the interpretation given in the text. Also, observe that the line $p^* = \rho t^* + a$ can only intersect the implicit function of p^* with respect to t^* based on the Free Entry condition at a stationary point of the implicit function because a is unique. Since there is only one such stationary point, there is only one intersection point and, therefore, only one equilibrium of the model.

C.2 Analysis of Proposition 3

This is straightforward: as already mentioned a (and thus A) is increasing in f, while $\partial \mathcal{H}(p^*, t^*; f) / \partial f = \delta [f_b \Phi(-t^*) + f_n] f^{-2} > 0$. Hence, the AC and FE curves shift, following an increase of the fixed cost of production from f to $f(1+\tau)$—with f_n and f_b staying unchanged—according to the pattern depicted in Figure 2.8. Since the two curves must always meet at the maximum of the implicit function of p^* over t^*, both threshold values are higher in the new equilibrium.

C.3 Analysis of Proposition 4

This is a particular instance where informational frictions lead to a deadweight welfare loss, which is larger the more marked frictions are. As in the original Melitz model, we analyze the welfare implications of the model's steady state. First, define:

$$\mathcal{P}^*_\theta \equiv \mathbb{P}\mathrm{r}(\theta \geq \theta^*)$$
$$\mathcal{P}^*_\varphi \equiv \mathbb{P}\mathrm{r}(\varphi \geq \varphi^*)$$

as the two unconditional probabilities that in equilibrium, before the draw of (θ, φ) pair, a firm-entrepreneur passes either threshold. Also define:

$$\widetilde{\pi} \equiv \mathbb{E}\left[\pi | \theta \geq \theta^*\right]$$

that is the expected market profits (including the share to be paid out to banks) that firms expect in equilibrium conditional upon passing the signal threshold. Thus, in steady state the mass of entering firms M_e and that of active firms M must comply to $\delta M = \mathcal{P}^*_\varphi M_e$; the total remuneration of entrepreneurial labor is $L_n = M_e f_n$; and bank labor amounts in equilibrium to $L_b = M_e \mathcal{P}^*_\theta f_b$. Moreover, free entry implies the following relationship in steady state:

$$\mathcal{P}^*_\theta \left(\widetilde{\pi} - \delta f_b\right) - \delta f_n = 0.$$

Lastly, recall that $\bar{r}$ and $\bar{\pi}$, in the original Melitz model, indicate average equilibrium revenues and profits conditional on successful entry, respectively.

Combining everything, it is:

$$\begin{aligned} L = L_p + L_b + L_n &= M\left(\bar{r} - \bar{\pi}\right) + M_e\left(\mathcal{P}^*_\theta f_b + f_n\right) \\ &= M\left[\bar{r} - \bar{\pi} + \frac{\delta}{\mathcal{P}^*_\varphi}\left(\mathcal{P}^*_\theta f_b + f_n\right)\right] \\ &= M\left(\bar{r} - \bar{\pi} + \widetilde{\pi}\frac{\mathcal{P}^*_\theta}{\mathcal{P}^*_\varphi}\right) \end{aligned}$$

where the first line exploits $L_p = M\left(\bar{r} - \bar{\pi}\right)$ as in Melitz; the second line leverages stationarity, and the third lines makes use of the Free Entry condition in steady state. Therefore, welfare per worker $\mathcal{W}$ equals the inverse of the price level, that is:

$$\mathcal{W} = \frac{\sigma - 1}{\sigma} L^{\frac{1}{\sigma-1}} \left(\bar{r} - \bar{\pi} + \widetilde{\pi}\frac{\mathcal{P}^*_\theta}{\mathcal{P}^*_\varphi}\right)^{-\frac{1}{\sigma-1}} \widetilde{\varphi}$$

where $\widetilde{\varphi}$, using the same notation as in Melitz, is the productivity of the representative firm. Note that $\widetilde{\varphi}$ is increasing in ρ: an argument akin to that of Proposition 3 would show that a higher ρ leads to higher equilibrium thresholds (θ^*, φ^*), and hence to a higher average productivity (thanks to a sharper selection by banks).

Further observe (although this is tedious to show), that for $\rho \geq 0$ it is:

$$\frac{\mathcal{P}^*_\theta}{\mathcal{P}^*_\varphi} \geq \frac{\bar{\pi}}{\widetilde{\pi}} \geq 1,$$

with both relationships becoming equalities if and only if $\rho = 1$. In addition, the two inequalities widen the closer ρ gets to zero. To conclude, social welfare is maximized under the perfect information case $\rho = 1$ when the economy reduces to Melitz's, and hence the optimality result by **?** is restored. A deviation of ρ from the optimal benchmark leads to two sources of inefficiency: first, representative productivity $\widetilde{\varphi}$ falls due to a selection effect; second, the number of available varieties decreases by a factor $\bar{r}/\left(\bar{r} - \bar{\pi} + \widetilde{\pi}\mathcal{P}^*_\theta/\mathcal{P}^*_\varphi\right) < 1$, as some resources in the economy are wasted to finance entrepreneurs-firms that pass the signal threshold θ^* but fail to meet the productivity threshold φ^*.

C.4 Analysis of Proposition 5

Adding a wedge τ to firms labor costs, holding everything else equal, raises the two equilibrium threshold are raised (per Proposition 3), hence $\widetilde{\varphi}$ increases while the gap between $\bar{\pi}$ and $\widetilde{\pi}\mathcal{P}^*_\theta/\mathcal{P}^*_\varphi$ also narrows. At the same time, fewer firms can repay production costs and survive in the economy, leading to higher average revenues $\bar{r}$ and fewer product varieties. This makes the overall welfare effects of the wedge ambiguous and dependent on the specific parametrization of the model.

C.5 Extensions

We next sketch a version of the model that features PEIs. In the analysis developed so far, the equilibrium productivity distribution of the model obtains as a truncated version of the distribution firms draw their productivity from, as in the Melitz model: $\mu(\varphi) = [1 - G_0(\varphi^*)]^{-1} g_0(\varphi)$, where $g_0(\varphi)$ and $G_0(\varphi)$ are the marginal p.d.f. and c.d.f. for φ, respectively. We now allow firms to adjust their productivity after entry. Specifically, we add a further, final stage of the model where firms are allowed to *set* a productivity level $\breve{\varphi}$ subject to a decreasing cost in their original draw φ.

We specify the firm optimization problem as the difference between the additional profits obtained by raising productivity from φ to $\breve{\varphi}$ and the cost of the raise:

$$\max_{\breve{\varphi}} B\left(\breve{\varphi}^{\sigma-1} - \varphi^{\sigma-1}\right) - \kappa\left(\frac{\breve{\varphi}}{\varphi}\right)^{\alpha}, \tag{7}$$

where B is a constant that comes from the Dixit-Stiglitz analysis of monopolistic competition, while κ and α are two technological constants. We assume $\alpha > \sigma - 1$ to ensure that the cost of the raise scales faster than the benefit, thereby making the problem salient. The problem is globally concave, and the solution is straightforward:

$$\breve{\varphi} = \left[\frac{B(\sigma - 1)\varphi^{\alpha}}{\alpha\kappa} \right]^{\frac{1}{\alpha+1-\sigma}}. \tag{8}$$

This delivers a monotone increasing mapping $\breve{\varphi}(\varphi)$ and an equilibrium productivity distribution expressed by $\mu(\varphi) = [1 - G_0(\breve{\varphi}^*)]^{-1} g_0(\breve{\varphi}^{-1}(\varphi)) \frac{d}{d\varphi}\breve{\varphi}^{-1}(\varphi)$, where $\breve{\varphi}^*$ is the new productivity threshold that obtains in the new equilibrium where firms have enhanced their productivity.

The implications of adding PEIs to our analysis of labor market distortions differ slightly depending on the interpretation one gives to the cost side of (7). On the one hand, wedges to labor costs definitely reduce the benefit side of (7), as they depress equilibrium profits. On the other hand, they may also raise the cost of PEIs, if the latter depends, at least in part, on human labor. We summarize these considerations with the following statement.

Proposition 6. When firms can perform PEIs as in (7), adding a wedge $\tau > 0$ to firms' labor costs (but not to either entry cost f_n or f_b), has ambiguous effects on average productivity: the positive effect due to a higher threshold (per Proposition 5) is mitigated by a negative effect due to lower PEIs. This negative effect is larger if the wedge also leads to a multiplicative increase in the cost side of PEIs.

Proof. The proof can be found in Appendix C.6. □

Adding PEIs helps make sense of our empirical results at the distribution level. Under our specification of PEIs (7), firms that are *ex ante* highly productive (on the right tail of the distribution) benefit from labor market reforms that decrease effective labor costs. Conversely, on the left tail, the selection effect dominates, which contributes to depressing average productivity. Note that our analysis is silent on the overall welfare effects of the reform: even if the net impact on average productivity is lower, consumers may still benefit from lower product varieties. In future work, we plan to provide structural estimates of the model that would let us make preliminary conclusions about the overall welfare effects.

The analysis so far was confined to a closed economy and neglected considerations about trade, as this is not the key concern of this paper. In this regard, we plan to develop a suitable extension in future work, which is natural for an extension of the Melitz framework like ours. We expect to formalize the intuition according to which adding (removing) labor market distortions harms (helps) those firms in the right tail of the productivity distribution that are more likely to engage in foreign markets.

C.6 Analysis of Proposition 6

Use the (τ) subscript to denote the values of the constants featured in (7) following the addition of a wedge τ to labor costs. From the Dixit-Stiglitz analysis of monopolistic competition one has:

$$B_{(\tau)} = \frac{B}{(1+\tau)^{\sigma-1}},$$

as both revenues and profits decrease because of higher labor costs. Let the wedge τ also cause the cost side of (7) to increase, say because part of the cost of enhancing productivity involves human resources, as follows:

$$\kappa_{(\tau)} = \kappa\,(1+\tau)^{\zeta},$$

for some $\zeta \geq 0$. Therefore, by (8) the wedge leads to a multiplicative transformation of the equilibrium productivity distribution, which is expressed as follows:

$$\breve{\varphi}_{(\tau)} = (1+\tau)^{\frac{\sigma-1+\zeta}{\sigma-1-\alpha}}\,\breve{\varphi} < \breve{\varphi},$$

where $\breve{\varphi}$ is as in (8), while $\breve{\varphi}_{(\tau)}$ is the updated value of post-investment productivity following the addition of the wedge.

C.7 Analysis of the two extreme cases

The critical properties of the model are perhaps best appreciated by looking at two "extreme" cases about the statistical relationship between the signal θ and productivity φ. In one case, signals are not informative at all, and the two random variables are fully independent. In the other case, the signal is fully informative, and the two random variables are perfectly correlated. The analysis of these two cases can be conducted without maintaining either Assumptions 1 or 2. Under these assumptions, however, the cases in question correspond to those where $\rho = 0$ and $\rho = 1$, respectively.

Signals are not informative If signals deliver no information about productivity, the two random variables are independent: $G(\varphi|\,\theta) = G_0(\varphi)$ for all pairs (φ, θ), and $\widetilde{\pi}(\theta) = (1 - G_0(\varphi^*))\,\bar{\pi}$, where $\bar{\pi}$ are the expected profits conditional upon successful entry as in Melitz, for all values of θ. At stage 2. of the model, banks set their share uniformly for all firms: hence $\theta^* = 0$ and $b(\theta) = \delta f_b/(1 - G_0(\varphi^*))\,\bar{\pi}$. Back in the firm entry stage (stage 1.) it is $C(\theta^*) = 0$ and free entry reduces to:

$$\frac{(1 - G_0(\varphi^*))}{\delta}\bar{\pi} - f_b - f_n = \frac{f}{\delta}(1 - G_0(\varphi^*))\,k\,(\widetilde{\varphi}(\varphi^*)) - f_e = 0$$

for $f_e = f_b + f_n$ and where $\bar{\pi} = f k(\widetilde{\varphi}(\varphi^*))$ is the Zero Profit Condition (ZPC) as in Melitz. This is precisely the equilibrium condition of the original Melitz model.

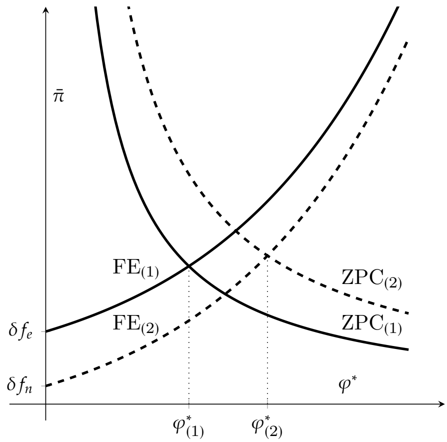
Signals are fully informative If signals predict productivity with probability one, θ^* and φ^* are jointly determined, hence one can safely focus on productivity φ only while disregarding the signal θ. To solve the model, observe that at stage 2. banks will finance only firms that are able to repay f_b in present value. This translates, relative to the Melitz benchmark, into an actual *per-period* post-entry fixed cost of $f + \delta f_b$: therefore the Zero Profit Condition becomes as $\bar{\pi} = (f + \delta f_b) k(\widetilde{\varphi}(\varphi^*))$. In the firm entry stage (stage 1.) entrepreneurs only need to bear their own entry cost f_n, and Free Entry implies $\bar{\pi} = \delta f_n / (1 - G_0(\varphi^*))$. Combining everything, the equilibrium solution is given as follows, and it shown to be unique by Appendix B in Melitz.

$$\frac{(1 - G_0(\varphi^*))}{\delta}\bar{\pi} - f_n = \frac{f + \delta f_b}{\delta}(1 - G_0(\varphi^*)) k(\widetilde{\varphi}(\varphi^*)) - f_n = 0$$

One can show analytically that the second scenario leads to a higher threshold productivity value φ^*. An easier way to appreciate this is by comparing Melitz's ZCP and FE curves between the two cases: when moving from the first scenario (no information) to the second (full information), both curves shift outward in such a way that leads to a higher value of φ^*, as it is shown in Figure C.5. As in Proposition 5, the second scenario is more efficient for two reasons: entering firms are, on average, more productive, and no intermediary-specific fixed entry cost f_b is wasted on firms that eventually fail to pass the final threshold and produce.

It is interesting to analyze the effect of adding a wedge τ to labor costs in the first of the two extreme scenarios, where signals provide no information. Conditional on expected post-entry profits $\bar{\pi}$ staying constant, the wedge does not affect the cost side of firm entry decisions. However, it obviously affects the benefits side, in a way that is summarized by the ZPC, which becomes $\bar{\pi} = (1 + \tau) f k(\widetilde{\varphi}(\varphi^*))$. Hence, graphically the ZPC curve shifts outward thus leading to a higher productivity threshold (from φ^* to $\varphi^*_{(\tau)}$ in the representation given by Figure C.6). Thus, the wedge τ can in principle be tailored to make the resulting productivity threshold equal to that of the "full information," efficient outcome shown in Figure C.5. Observe that this would not, however, restore the full efficiency properties of the model! The intuition is that by introducing the wedge, only the ZCP curve shifts, but the FE curve does not. In equilibrium, the lower expected profits dissuade some firms from entering, thus decreasing

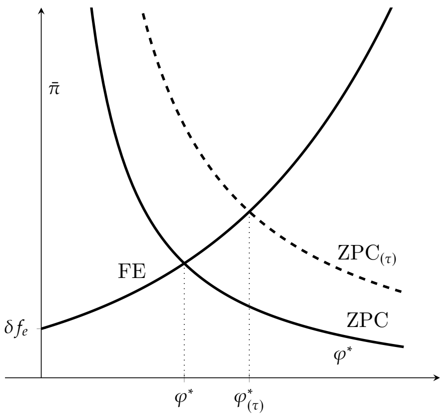
FIGURE C.5: Analysis of the two extreme cases



Note. Numbers apposed to curves or variables denote one of the two "extreme" scenarios, as described above.

the number of varieties and increasing average profits. Therefore, the overall welfare effect is ambiguous, as expressed by Proposition 5 for the general case.

FIGURE C.6: Introduction of labor frictions in the Melitz model



Note. τ included in a curve's or variable's subscript represents the implications of introducing labor price wedges equal to τ on it.

Chapter 3

Optimal Ramsey Taxation with Social Security[1]

Abstract. We develop an OLG model with heterogeneous agents and aggregate uncertainty to study optimal Ramsey taxation when the government can use a credible set of social security instruments. Social security mitigates the income effect in optimal labor tax smoothing and, together with heterogeneity, adds new redistributive motives to both labor and capital taxes while crowding out others. We calibrate the model on three different economies: the US, Netherlands, and Italy. We argue that the three countries would experience heterogeneous gains, in redistributive and efficiency terms, by moving from the status-quo allocations to those prescribed by a utilitarian Ramsey planner. Our simulations show that retirement benefits in the current economies are higher than their Ramsey-optimal level while we argue that the use of funded social security schemes, neglected in current actual policies, could be welfare improving.

JEL classification: H21, H23, H24, H31, H55, E62.

[1]This chapter is joint with Marco Francischello and Matteo Paradisi. We especially thank Cristiano Ricci for the research assistance with the numerical simulations. We also thank the participants to the Netspar IPW 2022 for various comments and suggestions. The realization of this article was possible thanks to the financial support of the Netspar comparative research grant 2021.

Keywords: Optimal taxation, Social Security, tax smoothing, efficiency, redistribution.

3.1 Introduction

Three well-established trends have recently characterized advanced economies. First, labor productivity growth has halved since the 1980s and particularly after the 2007-2009 global financial crisis (Dieppe, 2021). Second, both wealth and income inequalities have been increasing, most notably in the US (Chancel et al., 2021). Third, the population has dramatically aged: over the last forty years, the number of over-65 per 100 people of working age in OECD countries increased from 20 to 31 (OECD, 2019). These trends have heterogeneous impacts across generations and worker groups, pushing toward policy reassessments. For example, a well-designed pension system can help to reduce inequalities among retirees, while the generational policy imbalance observed in many western countries will cause the future generation to pay significantly higher taxes in the years ahead (Kotlikoff, 2009). This paper provides a new normative framework for studying how the optimal policy mix of taxes and social security instruments can help deal with these trends.

In our analysis, we rely on a neoclassical, general equilibrium Ramsey (1927) taxation model augmented along three directions: *i)* we substitute the infinite-horizon households with a series of overlapping generations (OLG); *ii)* we add aggregate and idiosyncratic risks; *iii)* we include, among the policies, a rich set of social security instruments. The OLG structure breaks the Ricardian equivalence and lets us capture two relevant factors: first, the policies' incidence on a single generation and, therefore, the reduced scope for their perfect smoothing; second, the across-generations redistributive and risk-sharing concerns. Aggregate uncertainty allows us to study how the government responds to temporary shocks that impact asymmetrically on different generations, while through idiosyncratic shocks, we can investigate the need for redistribution within a single generation—in a similar way to equity concerns in static models. Finally, we add social security to the standard policy mix usually studied in Ramsey problems because it improves the credibility of our framework, being a relevant-yet-understudied government's leverage to impact all the trends we described earlier. This is why we find it crucial to model social security to mimic how many pension schemes work closely. In our model, workers receive a retirement transfer from the government made of two components: one is fixed (i.e., a minimum pension benefit), while the other is work-related (i.e., fees paid by workers during their working period, reevaluated at a specific replacement rate). The government finances this social security scheme by mixing two methods: one is taxing current payrolls (PAYGO or unfunded approach); the other is capitalizing taxes levied on the current generation of retirees while they were working (funded approach).

Through this framework, we theoretically characterize optimal distortionary labor and capital taxation, and we comment on their interactions with social security. The key question here is to understand how social security can be optimally used *together with* classic Ramsey policy instruments at the optimum.

First, we provide conditions for optimal labor taxation smoothing over time. In particular, we show that the government has preferences for smoothing across generations the deadweight loss of labor taxation, i.e., the net loss in welfare efficiency that follows the collection of an additional dollar of revenue through labor taxes. We formalize this intuition under a simplifying quasi-linear assumption,when such smoothing is independent of the households' types: the government fixes the same distortion pattern over time for each type so that such a pattern is decided only through preferences for intergenerational redistribution. We then introduce concavity in preferences and decompose the marginal effects of taxation in three elements: a mechanical, a distortionary, and an income effect. In doing this, we extend the smoothing formula proposed by Hipsman (2018) in three directions. First, we adjust the income effects to accommodate the mitigation brought in by the social security instruments, which reduce the response in labor supply that would follow a marginal increase in labor taxes. Second, since agents are heterogeneous in our setting, we derive formulas describing the optimal smoothing pattern of the marginal contribution to the deadweight loss of a given type. Third, we obtain additional motives for labor income taxation that stem from redistributional objectives. These new motives mark the distance between our formula and a perfect smoothing result.

We then turn to capital taxation, deriving an optimal capital taxes formula that we decompose in different terms, each representing an additional motive for taxing capital. We start showing that in the baseline case—homogeneous agents without social security—optimal taxes can be broken down into a budget, an aggregate resources, an intertemporal, and a hedging component, on the lines of Farhi (2010). Each accounts for a specific taxation motive that affects social welfare through one particular channel. In particular, the budget and the intertemporal components refer to individual contributions to capital taxation, while the other two relate to aggregate motives. Allowing for household heterogeneity then introduces an additional redistributional component that adds a taxation motive to impose distortions to the more productive type for equity purposes. Moreover, we show that when the Ramsey planner has access to social security instruments and households are homogeneous, the aggregate terms of the decomposition disappear. The funded component of social security is indeed financed through a specific tax levied on income and capitalized on the market, thus acting as a mandatory saving from the point of view of the households. Therefore, this new instrument crowds out both the aggregate resources and hedging components. The very same does not hold when allowing for agents' heterogeneity. Actually, we argue that, compared to the baseline decomposition in the heterogeneous setting, all the terms remain unchanged, but a newly funded social security term substitutes the hedging one. This new component shows that, at the Ramsey optimum, the change in the income tax base that follows a variation in the capital taxes must be taken into account by the planner, who now levies specific taxes to fund social security instruments. Thus, the planner uses this new income-related channel to smooth capital taxes when expecting adverse shocks to the economy.

Finally, we briefly discuss the insightful relationship between the pension replacement rate, i.e., the conversion rate from income to social security benefit the government fixes for the work-related component of the retirement transfer, and the debt issued by the government for financing expenses. We show that, in equilibrium, the optimal replacement rate needs to balance the present labor distortions with future risk-free capitalization. Intuitively, a higher replacement rate calls for a larger labor supply in the present period while changing the bond's return in the next one. This effect on returns manifests both because the change in the social security instrument changes the aggregate income and because the replacement rate acts, conceptually, as another risk-free asset. Thus, the planner needs to consistently coordinate the two risk-free instruments at the optimum.

With this rich theoretical setting at our disposal, we then turn to the quantitative implication of our model. To allow for a richer and more credible discussion of the policy implications we derive in our numerical exercise, we calibrate our model on three different economies: the United States, Italy, and the Netherlands. These countries exhibit different mixes of current policy and macroeconomic parameters, thus enriching our calibration's power to credibly inform policy. For computational feasibility, we focus on the deterministic steady state of the model for each country.

First, we compare the allocations achieved in the benchmark economy, i.e., those stemming from calibrating the model with the status-quo parameters, with the social optimum and Ramsey allocations, i.e., the first-best allocations that maximize social welfare with no use of the policy instruments, and the feasible allocations a Ramsey planner reach at the optimum. Overall, the social optimum implements the highest consumption for every household type and a more efficient labor supply. Still, the status-quo economy and the Ramsey optimum achieve allocations mutually more similar than those of the social optimum since they can use the same set of policy instruments. In particular, the consumption of young agents increases across types for all three equilibria. We can show that marked differences in the consumption inequalities across the three countries persist and that the three face different gains in moving from the benchmark economy to the Ramsey optimum. For example, the US shows a consumption pattern for young households, which is remarkably close to the optimal Ramsey allocation, while Italy could obtain much larger benefits from a policy shift. Moreover, the Ramsey optimal allocation of old consumption is steeper for all the countries except the US, where this quantity is suboptimally high for the wealthiest. Instead, old households consume the same independently of their type at the social optimum. Finally, labor supply is flat at the Ramsey optimum because of the separability of consumption and labor, and all three countries exhibit too much labor effort exerted by low types. The social planner, on the contrary, maximizes efficiency by reaching a corner solution where the lowest types do not work at all. Overall, these numerical results provide valuable insights into the direction of policy-adjusting intervention in the three countries from both a redistributional and an efficiency perspective.

To further investigate the gains from a policy shift from the status quo to a Ramsey optimal allocation, we analyze the consumption-equivalent variation by household type in each country. Intuitively, this measure quantifies the percentage variation in consumption each type would need to be indifferent between the two scenarios, keeping labor constant: the larger it is, the higher the gains from the shift for that type. This exercise informs in particular how (un)equal such compensation should be—something insightful on the distribution of the distance from the Ramsey optimality across types in the status quo. For example, in the Netherlands, the gains from the shift are much more equal than in the US, where the less productive types demand substantial compensation to be indifferent after the change.

We then turn to social security. We find that retirement benefits in the benchmark economies are higher compared to those the Ramsey optimum prescribes—with a much wider gap for Italy and the Netherlands than for the US. We treat this more as a qualitative rather than quantitative result since our two-period OLG structure mechanically calls for a low rate of work-to-retirement years, thus artificially depressing the optimal replacement rate. Remarkably, our simulations prescribe a positively funded social security component, i.e., a mandatory saving the government imposes on workers to finance their pensions. This instrument is currently unused in all three countries we consider.

3.1.1 Related literature

Our contribution lies at the intersection between the literature on Ramsey optimal taxation and the one addressing retirement policies. Classic findings in the former include results on income tax smoothing (Lucas and Stokey, 1983; Werning, 2007) and no capital taxes (Judd, 1985; Chamley, 1986). The optimal taxes have been characterized under further assumptions such as incomplete markets (Aiyagari, 1995; Aiyagari et al., 2002; Farhi, 2010) or heterogeneous households (Werning, 2007). Economies in these works feature a standard infinitely-lived cohort of households, where intergenerational redistribution is not a concern. Other papers have used the OLG setting to discuss optimal Ramsey taxation both theoretically (Atkinson and Sandmo, 1980; Erosa and Gervais, 2002; Garriga, 2019) and quantitatively (Conesa et al., 2009). These papers do not include aggregate uncertainty, and so they do not inform optimal policies for responding to stochastic shocks. Hipsman (2018) adds aggregate stochasticity to a rich OLG setting similar to ours, but with no households' heterogeneity—thus, not discussing within-generation redistribution—and no social security. Saitto (2020), on the contrary, builds an OLG optimal taxation model with households' heterogeneity but no aggregate uncertainty and, again, no social security. More precisely, none of these papers exploit social security instruments as Ramsey policies, while we characterize their interaction with classical income and capital taxes.

The literature on retirement policies has separately addressed some relevant aspects we discuss in this work, such as financial sustainability—while much less is available on intergenerational inequality. Still, no study adopts a comprehensive setting such as ours. While İmrohoroğlu et al. (1998) and Krueger and Kubler (2006) have investigated the welfare effect of introducing a PAYGO social security system, they respectively allow for either idiosyncratic or aggregate risk only, while we pool the two together. Harenberg and Ludwig (2019) address the same question considering both risks, but under benefit schemes that are extremely simplified, thus different from those that are currently present in advanced economies. Ciurila (2017) explores the relationship between different benefit schemes (defined benefit vs. notional defined contribution) on long-run macroeconomic variables and welfare without considering differences in the system funding. Bonenkamp et al. (2017) analyze how pension reforms may be an instrument to respond to demographic and financial aggregate risk only. Hosseini and Shourideh (2019) look for Pareto-efficient reforms in social security systems, proposing a test on earning and asset taxes.

The remainder of the paper is organized as follows. Section 3.2 lays out the model's primitives, the agents' problems, and the social welfare function. Section 3.3 discusses the social planner's problem, while Section 3.4 presents and comments the Ramsey problem. Sections 3.5 and 3.6 illustrate our results on optimal labor and capital taxes, respectively, and their interactions with social security instruments. Section 3.7 discusses the relationship between the replacement rate and the risk-free bonds. In Section 3.8 we calibrate the model and in Section 3.9 we discuss our quantitative results in steady state. Section 3.10 concludes.

3.2 Model

We study a closed, neoclassical overlapping generation economy with two generations, nesting an infinite horizon model, in the spirit of Atkinson and Sandmo (1980). Agents are heterogeneous in their productivity and work when young in the period they are born, whereas they retire in the following period, once old. Firms are homogeneous and maximize their profit statically. The economy faces aggregate shocks, and markets are incomplete.

3.2.1 Uncertainty and heterogeneity

We model aggregate risk as a discrete set of states $s_t \in \mathcal{S}$ and histories of those states $s^t = (s_0, s_1, \ldots, s_{t-1}, s_t)$. The state s_t is characterized by a shock to the economy γ_t, and evolves according to a Markov process described by the transition matrix $\boldsymbol{M}$. Government consumption is exogenous and takes the form of a government spending shock $G(s_t)$ that only depends on the present state and not on the history. For each history s^t, there exists a one-period risk-free bond that pays a premium $R^b(s^t) = 1 + r^b(s^t)$ at all histories $s^{t+1} \succeq s^t$.[2] Throughout the paper, we suppress the argument of all the terms depending on the stories, replacing it with according subscripts: for each symbol x, the reader can think of x_t as equivalent to $x(s^t)$, unless differently specified.

Households have heterogeneous labor productivity, characterized by the discrete type θ, exogenously taken from a time-varying probability density Θ_t on the support $\mathbb{S}_t$. We denote with $\Theta_t(\theta)$ the probability that a worker is of type θ at time t.

[2]Since the bond is risk-free, its return at $t+1$ only depends on history s^t.

3.2.2 Demographics and timing

Young. Time is discrete. At each time t, a new generation of size n_t^y is born.[3] In this period, agents observe their type θ and the realized state of the world s_t, and choose their labor supply $l_{\theta,t}$, the investment in risky capital $q_{\theta,t}$, the investment in public debt $b_{\theta,t}$, and their consumption $c_{\theta,t}^y$.

Old. A share ψ_{t+1} of the young generation at time t survives in the following period so that $\psi_{t+1} n_{t+1}^o$ old workers retire at $t+1$. Old households consume the returns from the previous-period investment choices and receive a social security benefit from the government. The aggregate capital income not enjoyed by old agents due to mortality risk becomes unintended bequests and is added to the government resources for the sake of the model's tractability.[4]

3.2.3 Government policies

We let the government choose an optimal policy mix of public debt, capital and labor taxes. In addition, we allow for an additional instrument to redistribute across generations by including a social security system that encompasses two benefit schemes and two financing mechanisms. Retirees obtain a mix of defined benefits and work-related contributions. The government finances these benefits through a funded and unfunded component.

Formally, the government sets capital taxes τ_t^K one period in advance so that returns at time $t+1$ are taxed at a rate that depends on history s^t (Farhi, 2010). These taxes are levied on the returns on risk-free bonds and on capital returns net of the exogenous depreciation rate δ. Thus, the after-tax return is $R_t^{b,\tau} = 1 + (1-\tau_t^K)\, r_t^b$ for the risk-free bond with return r_t^b, and $R_{t+1}^{K,\tau} = 1 + (1-\tau_t^K)\, r_{t+1}^K$ for the risky capital, where r_{t+1}^K is the risky rate in $t+1$, net of capital depreciation. Moreover, we call $R_{t+1}^K = 1 + r_{t+1}^K$ the pre-tax return on risky capital and with $R_t^b = 1 + r_t^b$ the pre-tax return on the bond.

Moreover, the government levies the following total income taxes on a worker with productivity θ earning wage $w_{\theta,t}$:

$$T\left(w_{\theta,t} l_{\theta,t}\right) = \left[\tau_t^l + \tau_t^{\mathrm{SS}}\right] w_{\theta,t} l_{\theta,t}$$

where $\tau^l \in (-\infty, 1)$ is the linear labor income tax rate, and $\tau^{\mathrm{SS}} = \tau^{\mathrm{SS,U}} + \tau^{\mathrm{SS,F}}$ is the income tax financing the social security, consisting of the unfunded compo-

[3]These agents are then *young* at t and *old* at $t+1$. Therefore, $n_t^o = n_{t-1}^y$.

[4]Optimal bequest taxation with heterogeneous households has been extensively studied in papers as Farhi and Werning (2010) and Piketty and Saez (2013).

nent $\tau^{\mathrm{SS,U}}$, which bankrolls current benefits, and the funded one, $\tau^{\mathrm{SS,F}}$, invested in the capital market to finance future benefits. Moreover, $w_\theta(s^{t-1})$ is the wage paid to a worker with productivity θ. Notice that, from the point of view of the household, such a tax structure implies that it is impossible to distinguish the unfunded component of social security from the labor income tax. Thus, we will treat the former as an implicit component of the latter in the whole paper.

Finally, the government transfers a social security benefit $y^{\mathrm{SS}}_{\theta,t}$ to a retired worker of type θ in each period. The benefit consists of a work-related and a fixed component (i.e., a minimum pension benefit):

$$y^{\mathrm{SS}}_{\theta,t} = \kappa^{\mathrm{SS}}_{t-1} w_{\theta,t-1} l_{\theta,t-1} + \bar{y}^{\mathrm{SS}}_{t-1}$$

κ^{SS} is the replacement rate, regulating the amount of labor income converted to pension, and $\bar{y}^{\mathrm{SS}}_{t-1}$ is the fixed component of the benefit—both decided in $t-1$.

3.2.4 The agents' problems

The economy is populated by three types of agents: homogeneous firms, heterogeneous households, and the government.

Firms A representative firm has constant returns to scale technology $F(K, L)$ using capital and efficient labor and statically maximizes its profit given a collection of prices $\boldsymbol{Q} \equiv \{w_{\theta,t}, r^K_t\}$. Thus, it solves the following problem

$$\begin{aligned} \underset{K_{t-1},\{l_{\theta,t}\}_{\theta\in\mathbb{S}}}{maximize} \quad & \gamma_t F(K_{t-1}, L_t) - n^y_t \sum_\theta w_{\theta,t} l_{\theta,t} \Theta_t(\theta) - r^K_t K_{t-1} \\ \text{subject to} \quad & L_t = n^y_t \sum_\theta \theta l_{\theta,t} \Theta_t(\theta) \end{aligned}$$

Optimality conditions equate capital and labor prices to their marginal product[5]:

$$\theta \gamma_t F_{L,t} = w_{\theta,t} \quad \text{and} \quad \gamma_t F_{K,t} = r^K_t$$

To stress the dependence of prices from θs and the aggregate shock, we denote $\bar{w}(s^t) = F_{L,t}$ and $\bar{r}^K(s^t) = F_{K,t}$ so that a worker earns $w_{\theta,t} = \gamma_t \bar{w}(s^t)\theta$ per unit of labor and receives $r^K_t = \gamma_t \bar{r}^K(s^t)$ per unit of capital invested.

[5]For better readability, we omit everywhere the arguments of first and second derivatives of the production function.

Government At time t, the government sets a collection of policies $\mathcal{P} \equiv \{\tau_t^l, \tau_t^{\text{ss,U}}, \tau_t^{\text{ss,F}}, \kappa_t^{\text{ss}}, \tau_{t-1}^K, r_{t-1}^b\}$, given the exogenous spending G_t, that satisfies the following budget constraint

$$
\begin{aligned}
& n_t^o R_t^{b,\tau} \sum_\theta b_{\theta,t-1}\Theta_t(\theta) + G_t + n_t^o \psi_t \sum_\theta y_{\theta,t}^{\text{ss}}\Theta_t(\theta) \\
& \leq n_t^y \sum_\theta (\tau_t^l + \tau_t^{\text{ss,U}}) w_{\theta,t} l_{\theta,t}\Theta_t(\theta) + n_t^o \tau_{t-1}^K r_t^K \sum_\theta q_{\theta,t-1}\Theta_t(\theta) \\
& + n_t^y \sum_\theta b_{\theta,t}\Theta_t(\theta) + n_t^o \tau_{t-1}^{\text{ss,F}} R_t^K \sum_\theta w_{\theta,t-1} l_{\theta,t-1}\Theta_t(\theta) + n_t^o(1-\psi_t) R_t^K \sum_\theta q_{\theta,t-1}\Theta_t(\theta)
\end{aligned}
$$

i.e., the government uses debt to balance the difference between expenditures and revenues from income, capital, and social security taxes.

Households A worker of type θ from a given generation who is young at time t chooses the allocations $\mathcal{H} \equiv \{c_{\theta,t}^y, c_{\theta,t+1}^o, l_{\theta,t}, q_{\theta,t}, b_{\theta,t}\}$ to maximize life-time utility discounted at rate $\beta \in (0,1)$

$$
\begin{aligned}
\underset{H}{maximize} \quad & U\left(c_{\theta,t}^y, c_{\theta,t+1}^o, l_{\theta,t}\right) = u\left(c_{\theta,t}^y, l_{\theta,t}\right) + \beta\psi_{t+1}\,\mathrm{E}_t\left[u\left(c_{\theta,t+1}^o, 0\right)\right] \\
\text{subject to} \quad & c_{\theta,t}^y + q_{\theta,t} + b_{\theta,t} \leq \left[1 - \tau_t^l - \tau_t^{\text{ss,F}} - \tau_t^{\text{ss,U}}\right] w_{\theta,t} l_{\theta,t}, \qquad (3.2.1) \\
& c_{\theta,t+1}^o \leq R_{t+1}^{K,\tau} q_{\theta,t} + R_t^{b,\tau} b_{\theta,t} + \kappa_t^{\text{ss}} w_{\theta,t} l_{\theta,t} + \bar{y}_t^{\text{ss}} \qquad (3.2.2)
\end{aligned}
$$

Optimal allocations for a household must satisfy the no-arbitrage condition[6]

$$
\mathrm{E}_t\left[R_{t+1}^{K,\tau}\right] - R_t^{b,\tau} = \frac{-\mathrm{Cov}\left(R_{t+1}^{K,\tau}, u_{c,\theta,t+1}^o\right)}{\mathrm{E}_t\left[u_{c,\theta,t+1}^o\right]} \qquad (3.2.3)
$$

and the labor-leisure condition

$$
-\frac{u_{l,\theta,t}^y}{u_{c,\theta,t}^y} = \left[1 - \tau_t^l - \tau_t^{\text{ss}} + \frac{\kappa_t^{\text{ss}}}{R_t^{b,\tau}}\right] w_{\theta,t} \qquad (3.2.4)
$$

where the labor wedge depends on labor income taxes and social security taxes and is decreased by the discounted expected return on social security, which depends on the replacement rate. Notice that an increase in the replacement rate reduces the marginal disutility of labor since it gives households incentives to work today to enjoy higher consumption tomorrow. Indeed, an instrument

[6]For readability, we omit the arguments of the marginal utilities in all the text.

that links present earnings to future consumption acts as a risk-free mandatory saving program. This explains why the return on the government's debt scales it in (3.2.4): the government must set the two hand in hand to avoid arbitrages between the two risk-free assets. We will discuss this implication extensively later on when talking about optimal taxes.

3.2.5 Feasibility, market clearing, and equilibrium

Three equations regulate the flow of resources in the economy. Total investment in risky assets must equate to the aggregate level of capital. Hence, the market clearing condition in the capital market requires

$$K_t = n_t^y \sum_\theta q_{\theta,t} \Theta_t(\theta) + n_t^y \tau_t^{\text{ss,F}} \sum_\theta w_{\theta,t} l_{\theta,t} \Theta_t(\theta) \tag{3.2.5}$$

The efficient amount of labor employed by the firm must be consistent with the households' labor supply:

$$L_t = n_t^y \sum_\theta \theta l_{\theta,t} \Theta_t(\theta) \tag{3.2.6}$$

Finally, the aggregate resource constraint implies the following feasibility condition

$$n_t^y \sum_\theta c_{\theta,t}^y \Theta_t(\theta) + n_t^o \psi_t \sum_\theta c_{\theta,t}^o \Theta_t(\theta) + G_t + K_t \leq \gamma_t F\left(K_{t-1}, L_t\right) \tag{3.2.7}$$

Using the setup outlined above, we can define a competitive equilibrium as follows:

Definition 1 (Competitive equilibrium)**.** Given the initial stocks $b_\theta(s^{-1})$ and $q_\theta(s^{-1})$ for each $\theta \in \Theta_{-1}$, and the initial prices and policies $r^K(s^{-1})$, $r^b(s^{-1})$, $\tau^K(s^{-1})$ and $\kappa^{\text{SS}}(s^{-1})$, a competitive equilibrium is a set of prices $\mathcal{Q}$, policies $\mathcal{P}$ and allocations $\mathcal{H}$ such that households and firms maximize their utilities under the budget constraints, the feasibility constraint holds, and markets clear.

3.2.6 Social welfare function

We measure social welfare as the discounted sum of households' utilities weighted by two sets of welfare weights. Preferences for redistribution across generations are captured by the set of generational weights $\{\phi_t\}_{t\geq 0}$, while weights $\{g(\theta)\}_{\theta\in\mathbb{S}}$

measure the social desire for redistribution within a given generation. Therefore, we define the ex-ante social welfare function as

$$W \equiv \phi_0 \beta \psi_0 n_0^o \sum_{\theta} g(\theta)\, u\left(c_\theta^0, 0\right) \Theta_t(\theta) + \mathrm{E}_0 \left[\sum_{t \geq 0} \phi_t n_t^y \sum_{\theta} g(\theta)\, U\left(c_{\theta,t}^y, l_{\theta,t}, c_{\theta,t+1}^o\right) \Theta_t(\theta)\right] \tag{3.2.8}$$

The concavity of the utility function already provides motives for redistribution even when the government is utilitarian and $g(\theta) = \bar{g}$ for any θ. Indeed, a decreasing marginal utility implies that a government prefers allocations with low dispersion in consumption. Therefore, welfare weights $g(\theta)$ capture preferences for redistribution on top of the desire for redistribution implied by the agents' preferences.

3.3 Social planner

We introduce a social planner optimum to characterize the first-best allocation that will serve as a benchmark to evaluate welfare in the Ramsey optimum. Under the feasibility constraint and the market clearing condition for labor, a social planner maximizes the discounted sum of the generations' average utilities by directly choosing among implementable allocations. Therefore, we state the planner's problem as

Definition 2 (Planner's problem)**.** The planner's problem is a maximization of the ex-ante social welfare function (3.2.8) subject the following constraints:

- the feasibility constraint (3.2.7);
- the market clearing condition for labor (3.2.6);
- the initial condition on capital $K_0 = n_0^o \sum_\theta q_\theta(s_0)\Theta_t(\theta)$.

Since the planner has direct access to allocations, the problem does not account for the instruments detailed in Section 3.2.3. The only relevant constraints are the feasibility and efficient labor ones, plus an initial condition on capital. We refer to the solution of the social planner problem as a *social optimum*.

Proposition 7 (Social optimum)**.** The social optimum allocations satisfy the following conditions:

$$\frac{u_{c,\theta,t}^y}{u_{c,\theta',t}^y} = \frac{g(\theta')}{g(\theta)} \text{ and } \frac{u_{c,\theta,t}^o}{u_{c,\theta',t}^o} = \frac{g(\theta')}{g(\theta)}, \quad \forall \theta, \theta' \tag{3.3.1}$$

$$-\frac{u^y_{l,\theta,t}}{u^y_{c,\theta,t}} = \theta\gamma_t F_{L,t} \tag{3.3.2}$$

$$\frac{u^y_{c,\theta,t}}{\beta u^o_{c,\theta,t}} = \frac{\phi_{t-1}}{\phi_t} \tag{3.3.3}$$

$$\frac{\phi_t}{\phi_{t+1}} \frac{u^y_{c,\theta,t}}{\mathrm{E}_t\left[u^y_{c,\theta,t+1}\left(1+\gamma_{t+1}F_{K,t+1}-\delta\right)\right]} = 1 \tag{3.3.4}$$

Equations (3.3.1)-(3.3.4) are derived from the first-order conditions on the social planner's problem as illustrated in Appendix A.2.

First, equation (3.3.1) shows how the planner trades off the marginal utilities of different types. In the corner case with constant welfare weights, i.e., $g(\theta) = \bar{g}$ for each θ, the planner is utilitarian and equalizes the marginal utility of consumption across types for both young and older workers. Hence, the planner allocates the same consumption to all the old types within a given generation since their marginal utility does not depend on labor. This equality result holds for the young generation if and only if the preferences are separable in labor and consumption. Outside the utilitarian benchmark, given any two types within the same generation, the social planner trades off their marginal utilities of consumption depending on the ratio of their social welfare weights. In particular, the marginal utility of consumption across two generic types θ and θ' grows proportionally to the inverse ratio of their welfare weights, guaranteeing that types with lower $g(\cdot)$ get lower consumption in equilibrium.

Equation (3.3.2) is the standard labor-leisure condition that equates the marginal disutility of working to the marginal product of labor. It shows that the social planner requires more productive individuals to work more for efficiency reasons. More hours of work are compensated as described by equation (3.3.3). Because the marginal utility of consumption decreases in labor when preferences are not separable, and the consumption of old agents is constant across households, the planner must promise larger consumption when young to high-ability individuals to compensate them for their higher labor supply. This compensation disappears if preferences are separable and the marginal utility of consumption is independent of labor supply.

Finally, equation (3.3.4) pins down the optimal savings path, which in turn determines how aggregated capital is transferred across generations. Capital is set in equilibrium such that the marginal effect of transferring one unit of consumption across young agents of two different generations equates to the ratio of their generational welfare weights.

The social planner problem presumes that the government observes individual types and that there are unconstrained instruments. In the next Section, we set up a Ramsey problem where the government chooses optimal policies from a given set of instruments to maximize social welfare, keeping agents' optimal responses into account.

3.4 Ramsey planner

A Ramsey planner is a benevolent government that maximizes social welfare by choosing an optimal mix of policies $\mathcal{P}$. Hence, the planner chooses the Competitive Equilibrium that maximizes the social welfare function. We define the Ramsey Planner's problem as follows.

Definition 3 (Ramsey Problem)**.** The Ramsey problem is a maximization of the ex-ante social welfare function (3.2.8) subject to the following constraints:

- the young (3.2.1) and old (3.2.2) generations budget constraints;
- the feasibility constraint (3.2.7);
- the market clearing conditions for capital (3.2.5) and labor (3.2.6);
- the no-arbitrage (3.2.3) and labor-leisure (3.2.4) conditions.

We rewrite the problem using a variation on the so-called *primal approach* (Lucas and Stokey, 1983) since we have a rich set of policies for the government's use. Therefore, we let the government search for the optimal allocations *given* the social security instruments $\tau^{\mathrm{SS,F}}$ and κ^{SS}, and then we derive the other Ramsey optimal supporting policies and prices. Then, we ensure that the obtained set of policies, prices, and allocations satisfy the conditions for a competitive equilibrium. Since we will focus our analysis, in the following section, on the distortion caused by the policies enacted by the government to reduce the dimensionality of the problem, we assume $\bar{y}^{\mathrm{SS}} = 0$. The following Definition and Proposition formalize this approach.

Definition 4 (Implementable allocation)**.** Given a set of cross-periods optimal policies $\{\tau_t^{\mathrm{SS,F}}, \kappa_t^{\mathrm{SS}}\}_{t\geq 0}$, we say that an allocation $\left\{c_{\theta,t}^{y}, c_{\theta,t}^{o}, q_{\theta,t}, b_{\theta,t}, l_{\theta,t}\right\}_{t\geq 0}$ is implementable if the following implementability conditions (IC) hold for each $s^t, t \geq 0$ and for each $\theta \in \mathbb{S}$:

IC on the young generation:

$$u^y_{l,\theta,t} l_{\theta,t} + u^y_{c,\theta,t} \left[c^y_{\theta,t} + q_{\theta,t} + b_{\theta,t} \right] = -\beta \psi_{t+1} \kappa^{\text{ss}}_t \theta \gamma_t F_{L,t} l_{\theta,t} \, \mathrm{E}_t \left[u^o_{c,\theta,t+1} \right] \qquad \forall \theta \in \mathbb{S} \tag{3.4.1}$$

IC on the old generation:

$$\begin{aligned} c^o_{\theta,t+1} = & \frac{u^y_{c,\theta,t}}{\beta \psi_{t+1} \, \mathrm{E}_t \left[u^o_{c,\theta,t+1} \right]} b_{\theta,t} + \kappa^{\text{ss}}_t \theta \gamma_t F_{L,t} + \\ & q_{\theta,t} \left(1 + \frac{u^y_{c,\theta,t} - \beta \psi_{t+1} \, \mathrm{E}_t \left[u^o_{c,\theta,t+1} \right]}{\beta \psi_{t+1} \, \mathrm{E}_t \left[\hat{F}_{K,t+1} u^o_{c,\theta,t+1} \right]} \hat{F}_{K,t+1} \right) \qquad \forall \theta \in \mathbb{S} \end{aligned} \tag{3.4.2}$$

IC on the feasibility constraint:

$$\begin{aligned} \gamma_t \, F\left(K_{t-1}, L_t\right) = & n^y_t \sum_\theta c^y_{\theta,t} \Theta_t(\theta) + n^o_t \sum_\theta c^o_{\theta,t} \Theta_t(\theta) + G_t \\ & + n^y_t \sum_\theta q_{\theta,t} \Theta_t(\theta) + n^y_t \tau^{\text{ss,F}}_t \gamma_t F_{L,t} \sum_\theta \theta l_{\theta,t} \Theta_t(\theta) \\ & - (1-\delta) \left[n^y_{t-1} \sum_\theta q_{\theta,t-1} \Theta_{t-1}(\theta) + n^y_{t-1} \tau^{\text{ss,F}}_{t-1} \gamma_{t-1} F_{L,t-1} \sum_\theta \theta l_{\theta,t-1} \Theta_{t-1}(\theta) \right. \end{aligned} \tag{3.4.3}$$

IC on the income taxes:

$$\left[1 - \tau^l_t - \tau^{\text{ss,F}}_t - \tau^{\text{ss,U}}_t \right] = - \frac{u^y_{l,\theta,t} + \beta \psi_{t+1} \kappa^{\text{ss}}_t F_{L,t} \, \mathrm{E}_t \left[u^o_{c,\theta,t+1} \right]}{\theta \gamma_t F_{L,t} u^y_{c,\theta,t}} \qquad \forall \theta \in \mathbb{S} \tag{3.4.4}$$

IC on the capital taxes:

$$1 - \tau^K_t = \frac{u^y_{c,\theta,t} - \beta \psi_{t+1} \, \mathrm{E}_t \left[u^o_{c,\theta,t+1} \right]}{\beta \psi_{t+1} \, \mathrm{E}_t \left[\hat{F}_{K,t+1} u^o_{c,\theta,t+1} \right]} \qquad \forall \theta \in \mathbb{S} \tag{3.4.5}$$

where $\hat{F}_{K,t+1} = \gamma_{t+1} F_{K,t+1} - \delta$.

Proposition 8 (Implementability conditions)**.** If a set of cross-periods optimal policies $\tau^{\text{ss,F}}_t$, κ^{ss}_t and an allocation $c^y_{\theta,t}, c^o_{\theta,t}, q_{\theta,t}, b_{\theta,t}, l_{\theta,t}$ solves the Ramsey problem, then the allocation must be implementable, i.e., the ICs (3.4.1), (3.4.2), (3.4.3), (3.4.4), (3.4.5) must hold.

Proof. The proof can be found in Appendix A.3. □

To solve the Ramsey problem, we attach an agent-specific—except for the feasibility constraint—Lagrange multiplier for every $t \geq 0$ to each constraint in Proposition 8. In particular, we define the multipliers $\lambda^{y}_{\theta,t}$ for (3.4.1), $\lambda^{o}_{\theta,t}$ for (3.4.2), and λ^{f}_{t} for (3.4.3). Moreover, since the policies are not θ-specific, ICs on income (3.4.4) and capital (3.4.5) taxes ensure that the agents' choices are consistent across all workers' types. Thus, the differences between the rhs of the two constraints must be zero for every couple θ, θ', and we attach the multipliers $\lambda^{L}_{\theta,\theta',t}$ and $\lambda^{K}_{\theta,\theta',t}$ to these differences. This approach is similar to the one adopted in Chari et al. (1994).

3.5 Deadweight loss from labor taxation and social security

In this section, we study the optimal smoothing of labor taxes and their interaction with social security instruments. Since the structure of the model does not allow us to derive explicit formulas for marginal tax rates (or the labor wedge), we focus on the marginal deadweight loss (DWL) from labor taxation as a measure of labor distortions. The DWL is the net loss in efficiency that arises from collecting an additional dollar of revenue through the labor tax.

We link the DWL to our formulas with the following argument. At a Ramsey optimum, the DWL from labor taxes must equate the monetary social value of providing a lump-sum transfer to all individuals. The Lemma below formalizes this intuition.

Lemma 1. At the optimum, the marginal deadweight loss associated with an increase in labor taxes at the story s^t, $\mathcal{D}^{l}_{t}$, is equal to the total government value of collecting a dollar through a lump-sum transfer from young agents:

$$\mathcal{D}^{l}_{t} = -\sum_{\theta} \frac{\lambda^{y}_{\theta,t}}{\mathcal{W}_{\theta,t}} \tag{3.5.1}$$

where $\mathcal{W}_{\theta,t} = n^{y}_{t}\phi_{t} g(\theta)\,\Theta_{t}(\theta)$ is the social marginal value of increasing utils for a young worker of type θ at time t.

Proof. The proof can be found in Appendix A.4. □

The multipliers $\lambda^y_{\theta,t}$ in the numerator of (3.5.1) capture the value of transferring a dollar to young individuals, while the denominator represents the government value of increasing by one unit the utility of young agents and converts utils into money.

Due to the richness of our model, it is not always possible to make $\mathcal{D}^l$ explicit in our formulas. Thus, we also discuss the individual contribution to the DWL, *i.e.*, the value of relaxing the implementability condition for a young agent of type θ, normalized by the marginal social utility of increasing their consumption. Formally:

Definition 5 (Individual contribution to the DWL from labor taxation)**.** The individual contribution of a type-θ individual born at s^t to the DWL caused by labor taxation is defined as $d^y_{\theta,t} = -\frac{\lambda^y_{\theta,t}}{\mathcal{W}_{\theta,t}}$.

3.5.1 The deadweight loss from labor taxation in the quasi-linear case

For illustrative purposes, we start our discussion on the optimal DWL from a model with quasi-linear preferences, which delivers simple formulas. We assume that $u\left(c^y_{\theta,t}, l_{\theta,t}\right) = c^y_{\theta,t} + h(l_{\theta,t})$, where h is a decreasing concave function. The marginal utility of consumption is, therefore, constant across θs and equal to 1.

Proposition 9. At the Ramsey optimum, the DWL from labor taxation under quasi-linear preferences evolves as follows

$$\phi_t n^y_t \left(1 + \bar{\mathcal{D}}^l_t\right) = \frac{\phi_{t+1}}{\beta} n^y_{t+1} \, \mathrm{E}_t \left[1 + \bar{\mathcal{D}}^l_{t+1}\right] \tag{3.5.2}$$

where $\bar{\mathcal{D}}^l_t$ is the weighted average DWL across θs. Similarly, the individual contribution follows

$$\phi_t n^y_t \, \Theta_t(\theta) \left(1 + d^y_{\theta,t}\right) = \frac{\phi_{t+1}}{\beta} n^y_{t+1} \, \Theta_{t+1}(\theta) \, \mathrm{E}_t \left[1 + d^y_{\theta,t+1}\right] \tag{3.5.3}$$

Proof. Proof can be found in Appendix A.4. □

Equations (3.5.2) and (3.5.3) show the government's preferences for smoothing the DWL from labor taxation over time. The former establishes that the evolution of the aggregate DWL follows a martingale. Such a structure is a byproduct of market incompleteness, and it echoes the tax smoothing results in incomplete markets established in Barro (1979) and in Farhi (2010) who obtains a similar martingale structure for the government needs for funds and the marginal cost of taxation.

Equation (3.5.3) extends this result to the smoothing of individual DWL contributions. It shows that the DWL smoothing is independent of θ. In other words, the ratio between the contribution of young agents' of a certain θ to the DWL for two contiguous generations only depends on the proportion of the generational welfare weights. As an implication, the government chooses the same distortion pattern over time for every young agent's type so that the growth in individual distortions is only pinned down by preferences for redistribution across generations. This result is a byproduct of the quasi-linearity: since the marginal utility of consumption is constant, the government has no incentive to compress consumption heterogeneously across types. Concavity in consumption utility breaks down this result, as we illustrate in the next Subsection.

3.5.2 The deadweight loss from labor taxation in the general case

We now discuss the general case where preferences are concave, and income effects exist. To make formulas more intuitive, we preserve the assumption of separability between consumption and labor, but we report the generic formula in Appendix A.4. We start with optimal DWL smoothing in the homogeneous agents' case to distinguish between taxation motives based on the intertemporal balancing of income and substitution effects and those based on the preferences for redistribution across different θs.

Proposition 10. (DWL with homogeneous agents). Suppose agents' preferences are separable in consumption and labor and that there is a representative agent for each generation, at the Ramsey optimum, the DWL evolves following

$$\tilde{\mathcal{W}}_t \left[1 + \mathcal{D}_t^l \left(1 + MB_t^L\right)\right] = R_t^{b,\tau} \psi_{t+1} \,\mathrm{E}_t \left[\tilde{\mathcal{W}}_{t+1} \left(1 + \mathcal{D}_{t+1}^l \left(1 + MB_{t+1}^L\right)\right)\right] \tag{3.5.4}$$

where

$$MB_t^L = \varepsilon_{b,t}^{R_t^{b,\tau}} \left(1 - \frac{\kappa_t^{\mathrm{SS}}}{R_t^{b,\tau}} \frac{w_t l_t}{b_t} \Pr\left(s^{t+1} | s^t\right)\right) + \frac{\mathrm{d}\tau_t^K}{\mathrm{d}b_t} \frac{q_t \,\mathrm{E}_t \left[\hat{F}_{K,t+1} \lambda_{t+1}^o\right]}{R_t^{b,\tau} \,\mathrm{E}_t \left[\lambda_{t+1}^o\right]} \tag{3.5.5}$$

and

$$MB^{L}_{t+1} = -\sigma^{y}_{t+1}\left[1 - \frac{q_{t+1}}{c^{y}_{t+1}}\left(\frac{\mathrm{E}_{t+1}\left[\lambda^{o}_{t+2}\hat{F}_{K,t+2}\right]}{r^{b}_{t+1}\,\mathrm{E}_{t+1}\left[\lambda^{o}_{t+2}\right]} - 1\right)\right] \tag{3.5.6}$$

Proof. Proof can be found in Appendix A.4. □

Equation (3.5.4) mimics the optimal smoothing in (3.5.2) up to the use of social marginal utilities $\tilde{\mathcal{W}}_t$ that arise from concave preferences, and the presence of the intertemporal marginal benefits of labor taxation MB^L. The formula balances (in expectation) the marginal costs of labor taxation at two consecutive histories. An increase in the labor tax has a mechanical effect of 1 on consumption, captured by the first terms on the left- and right-hand sides. Moreover, the tax increase generates mechanical distortions in both histories that are measured by the second terms on both sides. Finally, intertemporal marginal benefits of distortions that depend on income effects arise and are measured by MB^L. These terms arise because income effects are not considered pure distortions and, therefore, should not be incorporated in the marginal cost of taxation.[7]

The intertemporal marginal benefit at s^t (i.e., MB^L_t) has two components that partly resemble those in Hipsman (2018). First, an increase in labor taxes reduces the number of issued bonds, whose return will decrease proportionally to the elasticity of risk-free returns to issued bonds $\varepsilon^{R^{b,\tau}_t}_{b,t}$. This change in $R^{b,\tau}_t$ will make the young poorer. To offset it, they will work more because of the income effect and will thus diminish the distortionary cost of taxation. The social security benefit partially mitigates this effect since lower interest rates make the replacement rate κ^{SS} more appealing for the household reducing the incentives to supply labor in response to a change in $R^{b,\tau}_t$. Second, a change in the risk-free interest rate implies that the government must increase taxes by $\mathrm{d}\tau^K_t / \mathrm{d}b_t$ to reduce the after-tax return on capital (by a no-arbitrage argument). Here, $\mathrm{d}\tau^K_t / \mathrm{d}b_t$ is the increase in capital taxes needed to keep a young agent's consumption constant after a change in issued bonds. Therefore the second term in MB^L_t captures how this change in capital taxes will mechanically affect the old generation at time $t+1$ (i.e., the young at time t).

The intertemporal marginal benefit at s^{t+1} (i.e., MB^L_{t+1}) also consists of two components. First, the direct intertemporal distortionary cost generated from the consumption drop, which is proportional to the intertemporal elasticity of

[7]Notice that the MB terms are equal to zero in the case of quasi-linear preferences since they only depend on income effects that are absent in the quasi-linear case.

substitution $\sigma^{y}_{\theta,t+1}$. Second, because the tax hike induced by the no-arbitrage condition will be phased out after s^{t+1}, the last term balances the capital tax term in MB^{L}_{t+1}.

Introducing heterogeneity and the need for redistribution

So far, we have ignored the within-generation heterogeneity and all the motives to tax labor that arise from this model's feature. In this Section, we add this ingredient and discuss how it augments the formula in (3.5.4). Let us define with $MB^{L}_{\theta,t}$ and $MB^{L}_{\theta,t+1}$ the versions of equations (3.5.5) and (3.5.6) where we employ the θ-specific quantities or prices for all quantities and prices that can be θ-specific. The following Proposition provides a DWL formula for the heterogeneous agents model.

Proposition 11. (DWL with heterogeneous agents). Suppose agents' preferences are separable in consumption and labor and that there are two types $\theta > \theta'$ for each generation. At the Ramsey optimum, the DWL evolves as follows

$$\tilde{\mathcal{W}}_{\theta,t}\left[1 + d^{y}_{\theta,t}\left(1 + \widetilde{MB}^{L}_{\theta,t}\right)\right] = R^{b,\tau}_{t}\psi_{t+1}\,\mathrm{E}_{t}\left[\tilde{\mathcal{W}}_{\theta,t+1}\left(1 + d^{y}_{\theta,t+1}\left(1 + \widetilde{MB}^{L}_{\theta,t+1}\right)\right)\right] \tag{3.5.7}$$

where

$$\widetilde{MB}^{L}_{\theta,t} = MB^{L}_{\theta,t} - \frac{1}{R^{b,\tau}_{t}\,\mathrm{E}_{t}\left[\lambda^{o}_{\theta,t+1}\right]}\left(\varepsilon^{R^{b,\tau}_{t}}_{b_{\theta,t}}\frac{\lambda^{L}_{\theta,\theta',t}}{b_{\theta,t}}\frac{\kappa^{\mathrm{SS}}_{t}}{R^{b,\tau}_{t}}\psi_{t+1} + \frac{\mathrm{d}\tau^{K}_{t}}{\mathrm{d}b_{\theta,t}}\frac{\lambda^{K}_{\theta,\theta',t}}{\beta}\right) \tag{3.5.8}$$

and

$$\widetilde{MB}^{L}_{\theta,t+1} = MB^{L}_{\theta,t+1} - \frac{\sigma^{y}_{\theta,t+1}}{c^{y}_{\theta,t+1}}\frac{\lambda^{K}_{\theta,\theta',t+1} - \left(1-\tau^{l}_{t+1}\right)\lambda^{L}_{\theta,\theta',t+1}}{r^{b}_{t+1}\beta\,\mathrm{E}_{t+1}\left[\lambda^{o}_{\theta,t+2}\right]} \tag{3.5.9}$$

Proof. Proof can be found in Appendix A.4. □

The smoothing of DWL contributions closely follows the structure of (3.5.4), where aggregate DWL is replaced by individual contributions to the DWL, and the intertemporal marginal benefits of taxation are augmented by two factors presented in equations (3.5.8) and (3.5.9) that capture the motives for redistribution. These extra terms depend on the income effects discussed in (3.5.5) and (3.5.6). Indeed, while income effects do not affect efficiency, they do have an impact on the redistribution across agents within a generation and should therefore be taken into account.

The second term in (3.5.8) measures how the changes in bond returns and in capital taxes discussed in the discussion to Proposition 10 affect the margins of choice of an agent θ. Under the assumption that $\theta > \theta'$, we obtain that $\lambda^L_{\theta,\theta',t}, \lambda^K_{\theta,\theta',t} > 0$. It follows that larger labor supply reductions on richer individuals (from the fact that κ^{SS} becomes more attractive relative to bonds) decrease the intertemporal marginal benefit and generate a motive to increase distortions on type θ. At the same time, the increase in the capital tax $d\tau^K_t / db_{\theta,t}$ decreases the intertemporal marginal benefit if it affects richer individuals more, and therefore induces the government to set a larger DWL contribution for θ. The term would have the opposite sign if the agent was of type θ'. Similarly to what we discussed for the marginal benefit in (3.5.6), the second term in (3.5.9) balances the latter term in (3.5.8) and measures the benefits from reducing capital taxes between period $t+1$ and $t+2$. If the benefit is larger for richer individuals, it generates motives to reduce the DWL contributions of these individuals.

3.6 Capital taxes and social security

This section presents the results regarding optimal capital taxes and their interactions with social security through an incremental approach, starting from a benchmark case with homogeneous agents and no social security instruments, then building the intuition of the marginal effects of relaxing these initial assumptions.

We express τ^K as a sum of different components, each of which represents an additional motive to tax capital. Some of these terms are related to the benchmark result on optimal capital taxes of Farhi (2010), but with notable differences that stem from the relevant additions in our model: the OLG structure, the households heterogeneity, and the social security instruments.

In particular, we show that the funded component of social security taxes crowds out the capital taxation motives related to aggregate quantities in the homogeneous workers' case and adds new motives when assuming heterogeneity among households.

3.6.1 Capital taxes with homogeneous agents and no social security

Proposition 12 (Optimal capital taxes with homogeneous agents and no social security)**.** If agents are homogeneous and the government does not have access to social security instruments, optimal capital taxes at the story s^t with $t \geq 1$ are composed of three terms: aggregate resources, intertemporal, and hedging—each scaled by the mechanical budget component.

$$\tau_t^K = \frac{1}{B_t(\lambda_{t+1}^o)} \left[T_t^A \left(\lambda_t^f, \lambda_{t+1}^f \right) + T_t^I \left(\lambda_t^y, \lambda_{t+1}^o \right) + T_t^H \left(\lambda_{t+1}^o \right) \right] \tag{3.6.1}$$

Proof. Proof can be found in Appendix A.5. □

Through the decomposition given in Proposition 12, we provide a distinct interpretation of each motive of the optimal capital tax. In the remainder of this section, we separately discuss the contribution of each term of the decomposition.

Budget component. The budget component is given by

$$B_t(\lambda_{t+1}^o) = \frac{1}{n_t^y} \beta \, \mathrm{E}_t \left[\lambda_{t+1}^o \hat{F}_{K,t+1} \right] \tag{3.6.2}$$

An increase in τ_t^K mechanically strengthens the budget constraint on the old generation, proportionally to the capital tax base $\hat{F}_{K,t+1} = \gamma_{t+1} F_{K,t+1} - \delta$. Since capital taxes are fixed one period in advance, this mechanical effect is appropriately discounted by β. Multiplying by the marginal social utility converts the quantity in welfare utils while dividing by n_t^y delivers an average for each agent subject to the tax.

Aggregate resources component. The term representing the motive for capital taxes linked to aggregate resources is

$$T_t^A \left(\lambda_t^f, \lambda_{t+1}^f \right) = \mathrm{E}_t \left[\lambda_{t+1}^f R_{t+1}^K - \lambda_t^f \right]$$

This component accounts for the mechanical effect on the economy's resources at large that follows a rise in capital taxes at story s^t. Capital taxes affect capital accumulation, influencing the amount of available resources in the whole economy. Therefore, an increase in the tax affects the feasibility constraints of two consequent periods. First, when the tax variation is decided since agents incorporate the information responding to the change in the tax when choosing optimal consumption and saving. Then, it affects the feasibility constraint of the subsequent period to an extent proportional to the return on risky capital. The difference between the two terms controls part of the optimal smoothing behavior of capital taxes. Indeed, a negative shock at the story s^t increases the multiplier at that time (the social welfare effect of a marginal relaxation on the feasibility constraint), reducing the motives of taxation. Intuitively, the government expects the shock to be transitory and the economy to grow in the next period. Therefore it decides to shift the tax burden accordingly, moving resources from one period to another. This two-period dynamic also informs the optimal capital taxes behavior in the secular trend. A government expecting positive growth in the future prefers to shift the tax burden away from the present and progressively increase taxes over time. By the same argument, when expecting a slowdown in the economy, such an increasing path becomes more concave, increasing the smoothness of the tax burden.

Intertemporal component. The component of the optimal capital tax that accounts for intertemporal taxation motives is given by

$$T_t^I\left(\lambda_t^y, \lambda_{t+1}^o\right) = \frac{\mathrm{E}_t\left[\beta\lambda_{t+1}^o R_{t+1}^K\right] - \lambda_t^y u_{c,t}^y}{n_t^y} \tag{3.6.3}$$

This term highlights motives for capital taxation that stem from the agents' intertemporal consumption tradeoff. It captures the incidence of an increase in capital taxes on the taxed generation and its distortions on the lifetime consumption-saving path. Its structure recalls the one of the aggregate resource component, although it focuses on a single generation. Adding an OLG structure to the model requires the government to take into account the effect of a capital tax not only on aggregate resources but also on the agents that directly pay for it. Indeed, increasing capital taxes shifts consumption from old to young agents of the taxed generation. Waiving a consumption unit today to accumulate capital relaxes the young budget constraint proportionally to the marginal utility of consumption while mechanically impacting the old budget constraint proportionally to the risky capital return. Such a relationship pinpoints the social value of this intertemporal shifting as an accounting association between the Lagrange multipliers on the two households' constraints.

Hedging component. The hedging term reads as

$$T_t^H\left(\lambda_{t+1}^o\right) = -\beta\,\mathrm{E}_t\left[(1-\tau_t^K)q_t\gamma_{t+1}F_{KK,t+1}u_{c,t+1}^o\right]C^H\left(\lambda_{t+1}^o\right)$$

where

$$C^H\left(\lambda_{t+1}^o\right) = \frac{\mathrm{Cov}_t\left[\lambda_{t+1}^o, q_t\hat{F}_{K,t+1}\right]}{\mathrm{E}_t\left[q_t\hat{F}_{K,t+1}u_{c,t+1}^o\right]} - \frac{\mathrm{Cov}_t\left[\lambda_{t+1}^o, q_t\gamma_{t+1}F_{KK,t+1}\right]}{\mathrm{E}_t\left[q_t\hat{F}_{K,t+1}u_{c,t+1}^o\right]}$$

This component echoes the one derived by Farhi (2010). The term in parenthesis is proportional to the inverse of the elasticity of capital to capital taxes. The higher this elasticity, the lower will be the tax rate's absolute value.

The term C^H characterizes a motive for capital taxation that attains to smooth the effects of a tax increase on the agents' marginal consumption across states (hence, hedging). This term balances the opposite direct and indirect impact of the tax increase, consisting of the difference between two covariances. The first measures the relationship between the direct effect of increased capital taxes on investments and the multiplier on the old budget constraint. A larger correlation implies a lower optimal τ_t^K because of the depressive effects on retirees' consumption. The other covariance accounts for an indirect effect of an increase in the tax, which distorts labor and investment allocations to a magnitude pinned down by the covariance between the size of the tax base adjustment (i.e., the derivative of the marginal product of the tax base F_{KK}) and the multiplier on the old budget constraint.

3.6.2 Capital taxes with heterogeneous agents and no social security

We now introduce agents' heterogeneity while keeping the government from using social security instruments. We characterize the composition of the optimal capital taxes at the Ramsey equilibrium in the following Proposition.

Proposition 13 (Optimal capital taxes with heterogeneous agents and no social security)**.** Suppose, without loss of generality, there exists two agents' types $\theta > \theta'$ for each generation and that the government does not have access to social security instruments. Then, optimal capital taxes at the story s^t, with $t \geq 1$, read as

$$\tau_t^K = \sum_\theta \frac{1}{B_{\theta,t}(\lambda^o_{\theta,t+1})} \Big[T^A_{\theta,t}\left(\lambda^f_t, \lambda^f_{t+1}\right) + T^I_{\theta,t}\left(\lambda^y_{\theta,t}, \lambda^o_{\theta,t+1}\right) + T^H_{\theta,t}\left(\lambda^o_{\theta,t+1}\right) + T^R_{\theta,t}\left(\lambda^K_{\theta,\theta',t}\right)\Big] \Theta_t(\theta) \qquad (3.6.4)$$

where $B_{\theta,t}(\cdot)$, $T^A_{\theta,t}(\cdot)$, $T^I_{\theta,t}(\cdot)$ and $T^H_{\theta,t}(\cdot)$ are the same terms of the Ramsey optimal tax in the homogeneous case (3.6.1), where we use the θ-specific quantities or prices for all quantities and prices that can be θ-specific; and $T^R_{\theta,t}(\lambda^K_{\theta,\theta',t})$ is a redistributional component.

Proof. The proof can be found in Appendix A.5. □

Redistributional component The redistributional component in (3.6.1) is given by

$$T^R_{\theta,t}(\lambda^K_{\theta,\theta',t}) = \frac{(1-\tau^K_t)}{r^b_t}\lambda^K_{\theta,\theta',t}\Theta_t(\theta)\,C(\theta,\theta') \tag{3.6.5}$$

where

$$C(\theta,\theta') = \frac{\mathrm{Cov}_t\left[\gamma_{t+1}F_{KK,t+1}, u^o_{c,\theta,t+1}\right]}{u^o_{c,\theta,t+1}} - \frac{\mathrm{Cov}_t\left[\gamma_{t+1}F_{KK,t+1}, u^o_{c,\theta',t+1}\right]}{u^o_{c,\theta',t+1}}$$

As with labor taxes, by allowing for agents' heterogeneity, we introduce an additional motive of capital taxation, i.e., one for redistribution. The first term in equation 3.6.5 illustrates how the government increases distortions on the θ-type agent. If $\theta > \theta'$, the concavity in preferences implies $\lambda^K_{\theta,\theta',t} > 0$. This, in turn, generates taxation motives to impose distortions on the more productive type. The magnitude of such a distortion is pinned down by a covariance term $C(\theta,\theta')$ which measures how differently the two types respond to the indirect adjustments of the aggregate tax base that follows an increase in capital taxes. In other words, each covariance term sizes how the θ-specific marginal utility of consumption is sensitive to changes in the return on risky capital caused by changes in the tax rate. Therefore, a positive covariance means that the change in the yield is positive in the states for which it generates positive marginal utility. How differently the two types respond to such a change pinpoints the extent of the redistributional intervention through capital taxes. If the covariance for the lower type θ' is higher than the one for θ, motives for reducing capital taxes stem since their marginal increase would harm the former more than the latter. Notice that the covariance term would be zero under agents' homogeneity, thus shutting down the entire component in that case.

3.6.3 Capital taxes with homogeneous agents and social security

In this subsection, we revert to the homogeneous case while letting the government use the social security instruments. The following Proposition shows how some of the motives for capital taxation are crowded out by social security instruments.

Proposition 14. If agents are homogeneous and the government has access to the social security instruments, capital taxes at the Ramsey optimum are given by:

$$\tau_t^K = \frac{T_t^I\left(\lambda_t^y, \lambda_{t+1}^o\right)}{B_t(\lambda_{t+1}^o)} \tag{3.6.6}$$

where $T_t^I(\cdot)$ is given by (3.6.3) and $B_t(\cdot)$ by (3.6.2).

Proof. The proof can be found in Appendix A.5. □

Here, we show that the motives to tax capital boil down to the intertemporal component scaled by the welfare one when the government can levy taxes to fund the social security system. This is because $\tau^{\text{ss,F}}$ takes now care of the terms in (3.6.1) that do not appear in (3.6.6), i.e., the aggregate resources and the hedging components. Notice that these two components reflect aggregate motives, which the Ramsey planner can thus shift on social security taxes.

3.6.4 Capital taxes with heterogeneous agents and social security

Finally, we discuss here the case in which the government can impose social security taxes in a setting with heterogeneous agents. Under these assumptions, a new motive for capital taxes arises, substituting the hedging component of (3.6.4), as formalized in the following Proposition.

Proposition 15. Suppose, without loss of generality, there exists two agents' types $\theta > \theta'$ for each generation and that the government has access to social security instruments. Then, optimal capital taxes at the story s^t, with $t \geq 1$, read as

$$\tau_t^K = \sum_\theta \frac{1}{B_{\theta,t}(\lambda_{\theta,t+1}^o)} \left[T_{\theta,t}^A\left(\lambda_t^f, \lambda_{t+1}^f\right) + T_{\theta,t}^I\left(\lambda_{\theta,t}^y, \lambda_{\theta,t+1}^o\right) + T_{\theta,t}^R\left(\lambda_{\theta,\theta',t}^K\right) + T_{\theta,t}^{\text{ss}}\left(\lambda_{t+1}^f, \lambda_{t+2}^f\right)\right]\Theta_t(\theta)$$

where $B_{\theta,t}$, $T_{\theta,t}^A(\cdot)$, $T_{\theta,t}^I(\cdot)$, $T_{\theta,t}^R(\cdot)$ are the same terms of (3.6.4), and $T_{\theta,t}^{\text{ss}}(\cdot)$ is a funded social security component.

Proof. The proof can be found in Appendix A.5. □

Funded social security component The funded social security component is

$$T^{\text{ss}}_{\theta,t}\left(\lambda^f_{t+1},\lambda^f_{t+2}\right)=\mathrm{E}_t\Bigg[\Big((1-\delta)\lambda^f_{t+2} - \lambda^f_{t+1}\Big)\, n^y_{t+1}\tau^{\text{ss,F}}_{t+1}\gamma_{t+1}F_{LK,t+1}\left(\sum_\theta \theta l_{\theta,t+1}\Theta_{t+1}(\theta)\right)\Bigg]$$

Since the funded component of social security acts as a mandatory saving in risky capital for the households, it distorts their labor choices with effects on the amount of produced resources in the economy. This term shows that even considering this new channel, the government looks for tax smoothing when it expects adverse shocks. In fact, an increase in the capital tax rate at the story s^t has a mechanical effect on the aggregate income tax base of young agents at s^{t+1}. The magnitude of such an effect is pinned down by the overall wage change and the aggregate labor supply's measure. This shrink in the social security tax base affects the feasibility constraint at s^{t+1} since it reduces the funded component of social security in the economy and the feasibility constraint at s^{t+2} by an amount proportional to the net-of-depreciation rate. Intuitively, the Ramsey planner looks to smooth taxes across two periods when it expects a negative transitory shock by an argument similar to the one discussed for the aggregate resources component. With respect to that case, here, the timing is translated by one period due to the channel through which taxes are linked to the resource constraint, i.e., labor income. The expectation of a negative transitory shock at the story s^{t+1} reduces the motives for taxation at s^t to avoid a harmful distortion of the aggregate income tax base at the time of the shock. At the same time, the smaller the depreciation rate of capital, the higher the optimal capital taxes since the erosion effect on aggregate capital— which would call for higher tax smoothing—decreases.

3.7 Social security replacement rate

We now briefly turn the discussion to the optimal social security replacement rate and, in particular, to its relationship with the government's debt.

The first-order condition of the Ramsey problem on κ^{SS}_t reads as

$$\sum_\theta w_{\theta,t}l_{\theta,t}\lambda^y_{\theta,t}u^y_{c,\theta,t}=\beta R^{b,\tau}_t\sum_\theta w_{\theta,t}l_{\theta,t}\,\mathrm{E}_t\left[\lambda^o_{\theta,t+1}\right] \tag{3.7.1}$$

The optimal replacement rate is chosen to balance, in aggregate, labor distortion today with risk-free capitalization tomorrow. Indeed, an increase in κ_t^{SS} improves the long-term returns to labor and creates incentives for all θs to increase their labor supply. The size of this incentive is proportional to the current income $w_{\theta,t} l_{\theta,t}$ earned by each type and is rescaled by $\lambda_{\theta,t}^{y}$ to capture the amount of the distortion. The right-hand side instead measures the future returns of increasing κ^{SS}, which depend on the current income and the value of transferring money to older workers as measured by $\lambda_{\theta,t+1}^{o}$. Importantly, returns are determined by the risk-free premium on bonds since the replacement rate acts in practice as a risk-free asset. This has to do with the timing structure of the social security system. In particular, the replacement rate is set one period in advance: for any amount of current income, agents know with certainty how much they will get in social security benefits when they are old. An immediate consequence of this observation is that the replacement rate and debt management must show some consistency at the optimum.

We show this equilibrium coordination between replacement rate and debt management by comparing the optimality conditions of the Ramsey problem for the two instruments. In particular, we find that equation (3.7.1) is implied by the first-order condition on the individual investment in bonds, which reads as

$$\lambda_{\theta,t}^{y} = \frac{\mathrm{E}_t\left[\lambda_{\theta,t+1}^{o}\right]}{\psi_{t+1}\,\mathrm{E}_t\left[u_{c,\theta,t+1}^{o}\right]}.$$

This suggests that at the optimum, the government sets the two rates following a non-arbitrage logic between the two risk-free assets to which agents have access.

3.8 Calibration

This section discusses how we calibrate the model outlined in Section 3.2, choosing the parameters' values and the parametric specification of the households' distribution. We let the model match targets from three relevant western economies: the United States, Italy, and the Netherlands. We chose these countries because we argue they represent an insightful example of different possible policy mixes. In particular, the USA has shown a recent trend of increasing debt-to-GDP ratio, with constant relatively low average income and capital taxes and a low social security replacement rate. On the contrary, Italy has one of the highest public debts in the world, supported by considerable tax rates. Moreover, it shows a high replacement rate. The Netherlands is characterized instead by a small stock of public debt, significant income taxes, average capital tax rates, and a large replacement rate. Moreover, the USA exhibits a sharply lower survival probability rate for the male population with respect to the other two countries.

In what follows, we illustrate how we calibrate every parameter in the model, breaking down our description by group. Table 3.1 presents the summary of the calibration exercise on the three benchmark economies.

Period length and discount rate We calibrate the model so that a period corresponds to 40 calendar years, and accordingly, we set the discount factor β to 0.50, matching a yearly discount of 0.98—common in the literature.

Households' heterogeneity To calibrate the distribution of the discrete household types, we non-parametrically fit different quantiles of the per-hour labor income that we compute from the Luxembourg Income Study (LIS) database. In particular, we focus on the 2004 wave as the first pre-crisis year common to the three countries. Notice that, in our model, individual productivity is the only driver of wage heterogeneity. Thus, this information is sufficient for our calibration purposes.

Demographics We assume that population growth is constantly equal to zero since we focus on the steady state. We also assume a constant survival rate calibrated on the males' survival probabilities to 65 years old from the World Development Indicators of the World Bank. We obtain two very close values for Italy and the Netherlands—0.90 and 0.89, respectively—while the fraction of American males that reach 65 years old is, on average, only 0.80. Given the time structure of our model, this number well informs the probability for a given generation to reach the retirement period.

Government policies We calibrate average capital taxes paid by households on their saving using a dedicated institutional source for each country.[8] In particular, we set a 15% capital tax rate for the US given the information offered by the Internal Revenue Service (IRS) of the US Government; a 26% rate for Italy as provided by the parliamentary documentation on the financial revenues taxation[9]; and a 31% in the Netherlands as discussed in Klemm et al. (2021). All

[8]Despite some works proposed estimated measures of savings taxes (see, e.g., the review in Sørensen and Sørensen 2004), there exists a strong model dependence in those estimates, as discussed by Hosseini and Shourideh (2019). Therefore, we opted for a parametrization that relies more on institutional documents than other academic works.

[9]In particular, we refer to the March 31st, 2021 "focus" accessible here.

three tax rates are taken from 2021 data.

For income taxes, we start from the tax wedge decomposition provided in the OECD's *Taxing wages* database (OECD, 2021c).[10] This is because, in our model, the total labor cost a firm incurs corresponds to the labor income of a household. Firms indeed do not make profits, nor do they pay taxes. Conceptually, they transfer all of their labor costs to workers' wages, who then pay income taxes. From the OECD data, we are able to observe the breakdown of the labor wedge in four relevant components: employer social security contributions, employee social security contributions, income taxes, and cash benefits. Given the structure of our model, the sum of all these four components constitutes the income tax rate we calibrate. Our income taxes calibration closely meets the results from Erosa et al. (2012) for all three countries.

To calibrate the replacement rate, we use data from OECD's *Pensions at a Glance* 2021 (OECD, 2021b), which reports the gross pension replacement rate by country as the ratio between the gross pension entitlement and the gross pre-retirement earnings. This number needs to be re-scaled as the percentage of our income tax base net of social security contributions and income taxes. Therefore, we compute our effective country-specific replacement rate as

$$\kappa^{\mathrm{SS}} = \kappa^{\mathrm{SS}}_{\mathrm{G}} \underbrace{(1 - EC)(1 - \tau^{l}_{\mathrm{E}})}_{\text{Net income}}$$

where $\kappa^{\mathrm{SS}}_{\mathrm{G}}$ is the gross replacement rate as reported in the data, EC is the employer's social security contribution as a fraction of the total labor cost, and τ^{l}_{E} is the effective labor tax levied on the average worker's income net of social security contributions. This way, we assign to each country an average replacement rate value that is fully coherent with our model and reads as 0.32 for the USA, 0.46 for Italy, and 0.51 for the Netherlands.

Finally, we set the funded component of social security to zero for all three countries. Indeed, none of the countries we examine has a mandatory saving program to finance social security. In particular, there is no mandatory program for a funded pension plan in the USA and Italy. The Netherlands commits its employees and employers to pay contributions into pension funds, but these contributions are agreed upon in collective employment agreements—thus, not chosen by the government. This is why we opt to set $\tau^{\mathrm{SS,F}} = 0$ even in this case.

[10]Notice that *tax wedge* here refers to "the ratio between the amount of taxes paid by an average single worker (a single person at 100% of average earnings) without children and the corresponding total labor cost for the employer. The average tax wedge measures the extent to which tax on labor income discourages employment," as the OECD documentation reports.

TABLE 3.1: Model's parameters

Parameter	Description	Value USA	Value ITA	Value NED	Source(s)
Demographics					
ψ	Survival probability (65 yo)	0.80	0.90	0.89	WB World Development Indicator (2019)
β	Discount factor	0.50	0.50	0.50	Common in literature
Government policies					
τ^K	Capital taxes	0.15	0.26	0.31	IRS Topic No. 409 (US); Parliamentary docs (ITA); Klemm et al. (2021) (NED)
τ^l	Labor taxes	0.28	0.46	0.36	OECD (2021c)
k^{ss}	Replacement rate	0.32	0.46	0.51	See text
$\tau^{ss,F}$	Funded component of social security	0	0	0	See text
Macroeconomic parameters					
B/Y	Debt to GDP ratio	1.61	1.83	0.66	OECD (2021a)
G/Y	Expenditure to GDP ratio	0.48	0.57	0.48	OECD (2021a)
Technology					
α	Capital share	0.36	0.36	0.36	Common in literature
δ	Capital depreciation rate	0.88	0.88	0.88	Hosseini and Shourideh (2019)

Macroeconomic parameters We target two macroeconomic parameters: debt-to-GDP ratio and expenditure-to-GDP ratio. For both, we use data from the last available year in OECD's *Government at a Glance* 2021 (OECD, 2021a).

Preferences and technology For the numerical simulations, we use the same functional form as the household utility function adopted in Conesa et al. (2009):

$$u(c,l) = \frac{\left(c^{\eta}(1-l)^{1-\eta}\right)^{1-\sigma}}{1-\sigma} \tag{3.8.1}$$

where η is a share parameter that tunes the relative importance of consumption to labor, and σ determines the household's risk aversion. We take $\eta = 0.18$ and $\sigma = 1.5$. Notice that, for $\sigma = 1$, this parametric form collapses on the standard log-log specification used in Chari et al. (1994), and later in Farhi (2010), and Hipsman (2018). We also assume a Cobb-Douglas production function with a constant capital share of 0.36 across countries, which is standard in the literature. Moreover, we set the capital depreciation rate as equal to 0.88 for our period to meet an annual rate of approximately 0.05 as in Hosseini and Shourideh (2019).

3.9 Quantitative results for the steady state

In this section, we present the results of our numerical simulation. We focus on an economy in a steady state without aggregate uncertainty. In this environment, the scope of debt is smaller since it cannot satisfy the households' need for safe assets, and it can be replicated by a combination of income and funded social security taxes. Since these simplifying assumptions reduce the policy space, we comment on our results, focusing on allocations and welfare metrics rather than policy parameters. Moreover, we assume that the planner is utilitarian so that welfare weights are constant across all θs. We start by defining the main welfare metrics that we employ to quantify the improvements of the Ramsey optimum on the benchmark economies.

3.9.1 Welfare metrics

We quantify the welfare gains and losses of moving across two different policy scenarios with a measure of equivalent variation that keeps labor constant at the reference one. We call this measure a *consumption-equivalent variation*, and we quantify it as the percentage increase in consumption that each type would need to experience to be indifferent between the benchmark economy and the Ramsey optimum, given the constant labor. In a steady state, the levels of consumption and labor along the life cycle are constant across generations, so the definition of this measure simplifies to the following:

Definition 6 (Consumption-equivalent variation)**.** Denote with $(c^{y,B}_{\theta,t}, c^{o,B}_{\theta,t+1}, l^{y,B}_{\theta,t})$ the optimal allocation in a benchmark economy for the θ-type agents of a generation born at s^t, and $(c^{y,R}_{\theta,t}, c^{o,R}_{\theta,t+1}, l^{y,R}_{\theta,t})$ the optimal allocations for the same agents at the Ramsey optimum. Moreover, define $\tilde{c}^{y,B}_{\theta,t} = c^{y,B}_{\theta,t}(1 + \Delta_\theta)$ and $\tilde{c}^{o,B}_{\theta,t+1} = c^{o,b}_{\theta,t+1}(1 + \Delta_\theta)$ the Δ-augmented consumption for a given type in the benchmark scenario. Then, the consumption-equivalent variation for a type θ is a value of Δ_θ that satisfies

$$u\left(\tilde{c}^{y,B}_{\theta,t}, l^{y,B}_{\theta,t}\right) + \beta\psi_{t+1} u\left(\tilde{c}^{o,B}_{\theta,t+1}, 0\right) = u\left(c^{y,R}_{\theta,t}, l^{y,R}_{\theta,t}\right) + \beta\psi_{t+1} u\left(c^{o,R}_{\theta,t+1}, 0\right) \quad (3.9.1)$$

The consumption-equivalent variation measures the willingness to pay for each type to avoid moving from the benchmark economy to the Ramsey optimum. Thus, $\Delta_\theta > 0$ if the θ-type agent would be better off at the Ramsey optimum relative to the benchmark. Therefore, we can compare Δs across the types to understand the redistributive effects of moving between the two economies.

3.9.2 Numerical results: allocations and welfare

We start the discussion on our numerical simulations by comparing allocations in three regimes: *i)* the benchmark economy, *ii)* the Ramsey optimum, and *iii)* the social optimum. In particular, we study the components that determine agents' utility at the optimum, i.e., young and old consumption and labor. Moreover, our three different calibrations allow us to compare the allocations across countries to discuss differences and similarities, which provide more insights about the model.

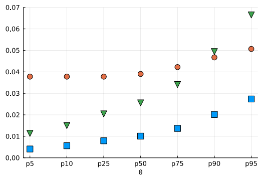
Figure 3.1 reports the allocations for the three countries in the three regimes. Overall, the benchmark economy and the Ramsey optimum tend to behave more similarly since they rely on the same set of instruments. On the other hand, the social optimum dominates the other two through two channels. First, it lets households enjoy more significant consumption when young and old for almost every type. Second, it manages to improve efficiency through an increasing pattern of labor supply over the type space so that more productive households work significantly more than less productive ones. In this calibration, the social optimum achieves a corner solution where no labor supply is required from the least productive types, making the labor-type profile very steep. On the contrary, the Ramsey optimum and the benchmark economies have flatter or even decreasing labor patterns due to limited instruments.

Young consumption increases across types in all countries in the benchmark economy so that wealthier households enjoy more significant consumption levels. This pattern is particularly accentuated due to low labor taxes in the US economy. At the same time, Italy is the country with the lowest inequality in consumption in the benchmark economy since it is the one where labor and capital taxes are larger. Moreover, at the Ramsey optimum young-age consumption is higher for all types in all countries, and the profile on the types' dimension is steeper. Italy has more considerable consumption gains in moving from the benchmark economy to the Ramsey optimum, especially for high-productivity households who enjoy sub-optimally low consumption levels in the status quo. The consumption pattern at the young age in the United States is instead remarkably close to the Ramsey optimum. In contrast with the limited-instrument scenarios, the social optimum displays a non-linear and convex consumption pattern due to the need to compensate high productivity types for their labor effort as prescribed by Equation (3.3.3).

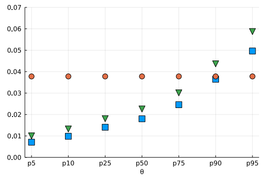
Consumption in old age also increases in the agents' types and tends to be higher than young-age consumption for most of θs. The Ramsey optimum increases consumption for everyone making the pattern steeper except for the US,

FIGURE 3.1: Consumption allocations and labor: benchmark economy, Ramsey optimum, social optimum

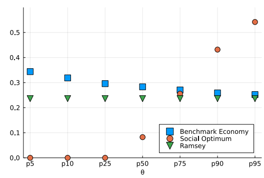
(A) Consumption young (ITA)



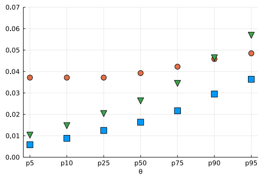
(B) Consumption old (ITA)



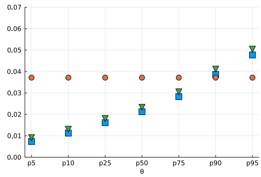
(C) Labor (ITA)



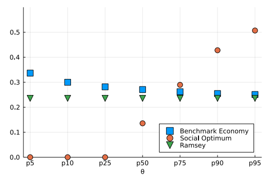
(D) Consumption young (NED)



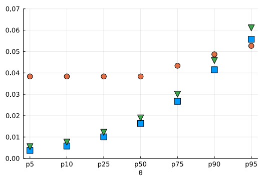
(E) Consumption old (NED)



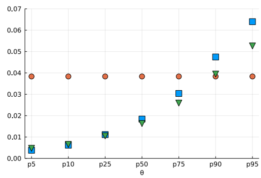
(F) Labor (NED)



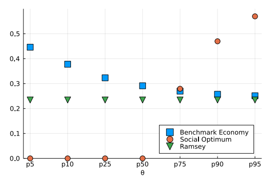
(G) Consumption young (USA)



(H) Consumption old (USA)



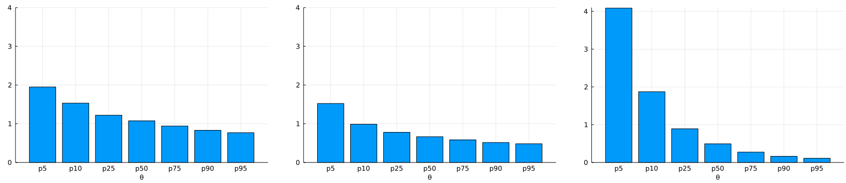
(I) Labor (USA)



Note: This panel shows the allocations of consumption of young agents (first column), old agents (second column), and labor supply (third column). Each row refers to a different country: Italy (ITA), the Netherlands (NED), and the United States (USA), from top to bottom. Each figure plots the benchmark economy allocation (blue square), the Ramsey optimum one (green triangle), and the social optimum one (orange circle).

FIGURE 3.2: Consumption-equivalent variation between benchmark economy and Ramsey optimum

(A) ITA (B) NED (C) USA



Note: This panel shows the consumption-equivalent variation, as stated in Definition 6, for three countries: Italy, the Netherlands and the US. Each figure shows the share of consumption a type should receive to be indifferent between moving to the Ramsey optimum of staying in the benchmark economy while keeping labor constant.

where old-age consumption is suboptimally high in the benchmark economy for the wealthiest types. This is due to the low capital tax rate that US households enjoy compared to the other two countries. At the social optimum, old consumption is flat, as suggested by equation (3.3.1).

Labor effort is flat across types at the Ramsey optimum. This is a well-known result in the case of log-log separable preferences in consumption and labor—a case we are not far from given our utility function's parametrization in (3.8.1). At the same time, the decreasing pattern in labor supply observed in the benchmark economies is explained by significant replacement rates that incentivize labor efforts from low-productivity agents who strive to increase their consumption in retirement. As already discussed, the social optimum would require a steep pattern of labor supply to maximize efficiency.

We conclude our discussion by investigating the gains of a Ramsey optimal policy relative to the benchmark economy. Figure 3.2 reports the consumption-equivalent variations by type in the three countries, as described in the Definition 6. Overall, every household type benefits from the Ramsey optimum, as the variations' positivity suggests. In all countries, we observe a decreasing pattern in the willingness to pay, which implies that moving from the status quo to the Ramsey optimum would benefit poorer households more than richer ones. This result implicitly suggests that the current policy mix in the three countries is less fair to low-productivity households than a Ramsey optimum would prescribe. Notably, the redistributive gain for poorer households would be even larger if we solved the Ramsey optimum with decreasing welfare weights instead of focusing on the utilitarian case. The Netherlands is the country where gains seem to be more equal across productivity types, which implies that the benchmark policy in the country is closer to the optimal level of redistribution that can be achieved in a utilitarian Ramsey equilibrium. The United States, on the other hand, penalizes the low end of the productivity distribution in the benchmark calibration and emerges as the most unequal tax system. Italy seems to be placed between the other two countries.

3.9.3 Numerical results: social security

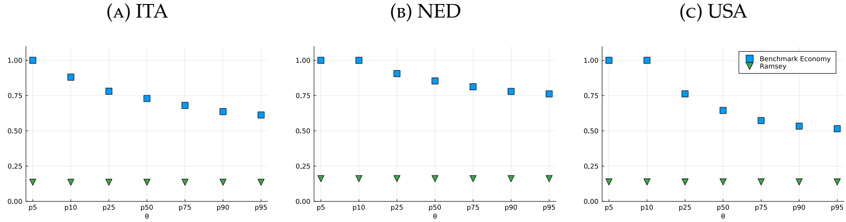
We now turn the discussion to the results on social security, focusing on two aspects. First, we assess the generosity of the benefits system by quantifying the share of old-age consumption that relies on social security transfers. Second, we quantify the importance of the funded component by looking at what share of consumption young households save on funded social security.

Figure 3.3 shows that benefits in the benchmark economies exceed those at the Ramsey optimum in all three countries, suggesting that the status-quo social security systems tend to be too generous. The gap between the two scenarios is much larger in Netherlands and Italy than in the United States, which has a lower replacement rate in the benchmark economy. At the optimum, the Ramsey planner would indeed rely more on private savings rather than on social security. Moreover, the old-age consumption share funded by social security benefits is flat across household types in the Ramsey optimum, while it decreases in the benchmark economies. This is the byproduct of the labor supply patterns discussed in the previous section.

While the qualitative result that the systems seem to be more generous than what is prescribed by the Ramsey optimum is robust to alternative calibrations, the exact extent to which it is generous should be taken with some grain of salt. Indeed, our calibration is limited by the two-period structure of the OLG model, which implements an artificially low rate of work-to-retirement years that depresses the optimal replacement rate in the optimum. Our simulation allows agents to enjoy retirement for the same number of years they work, while a more realistic calibration with multiple periods per generation and a lower share of the life-cycle in retirement would certainly deliver a greater replacement rate in the Ramsey optimum.

Figure 3.4 displays our simulations for the funded component of social security, shown as the share of consumption during the working period. As discussed early, this pillar of the social security system is absent in all three countries in our benchmark economies calibration. Instead, the Ramsey optimum prescribes

FIGURE 3.3: Social security benefit as a share of old consumption: benchmark economy and Ramsey optimum



Note: This panel shows the amount of social security benefit each household type receives, as a share of old-age consumption, under the benchmark economy (blue square) and at the Ramsey optimum (green triangle). The panel reports the exercise for three countries: Italy, the Netherlands, and the US.

a positively funded social security tax and induces the agents to save a positive and reasonably sized share of their consumption when young. This share is wider in the US compared to the Netherlands and Italy because the US exhibits a lower survival rate to retirement age. Thus, young agents face incentives to invest less, which in turn causes a suboptimal capital level in aggregate. Since the funded social security contribution acts as a mandatory saving plan, the Ramsey planner implements positive levels of this instrument when disincentives to investment are more significant. At the same time, the planner aims to subsidize capital through negative capital taxes, as in Saitto (2020). The share of funded contribution is also flat across household types since both consumption and income are proportional to the agents' productivity. Taken together, these results suggest that a funded component in the social security system is desirable for all three countries.

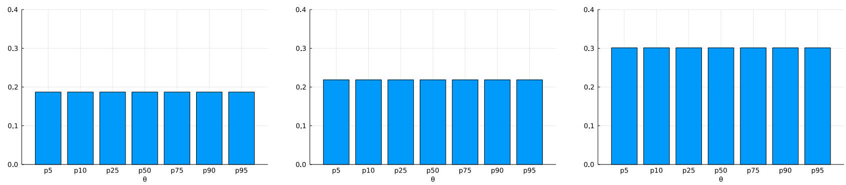
3.10 Conclusions

This paper has provided a theoretical and quantitative analysis of optimal Ramsey taxation when the households are heterogeneously productive, the economy faces aggregate shocks, and the government has access to social security instruments. To model these instruments insightfully, we allow them to account for both a defined benefit and a defined contribution scheme and to be financed through a mix of funded and unfunded systems.

Our theoretical results show that introducing such a pension scheme changes optimal labor taxation smoothing across periods in two directions. First, it erodes

FIGURE 3.4: Social security funded contribution as a share of young consumption at the Ramsey optimum

(A) ITA (B) NL (C) US



Note: This panel shows the amount of funded social security distribution, as a share of young-age consumption, implied by the Ramsey optimum for each household type. The panel reports the exercise for three countries: Italy, the Netherlands, and the US.

the income effects of a tax rate increase since the replacement rate reduces the incentives for a labor supply adjustment. Second, it adds new taxation motives that come from redistributional objectives. At the same time, social security impacts capital taxes crowding out the taxation motives related to aggregate distortions and bringing in a new motive linked to the change in labor supply caused by a change in the capital tax rate. Moreover, we argue that the structure of our risk-free capital assets calls for equilibrium coordination between the social security replacement rate and the public debt.

To keep our numerical analysis tractable, we have focused on the deterministic steady state of three economies: the benchmark one, the Ramsey optimum, and the social optimum. Calibrating our model on three countries (the US, Netherlands, and Italy) shows space for redistributive and efficiency gains by moving from the status-quo allocations to the Ramsey optimal ones. In particular, we show that the Ramsey-optimal social security benefits are lower than the actual policies in all three economies. At the same time, optimal taxes include a non-zero social security-funded component that the government uses to increase aggregate capital.

Two limitations of our work are the low number of generations in our model and the linear structure of the policies we consider. Our OLG model features two periods, thus unnaturally weighting the working period as much as the retirement one—a modeling choice that artificially depresses the optimal replacement rates. Moreover, linear income taxes allow for higher tractability at the expense of lower efficacy in dealing with redistributive motives. Thus, we plan to include non-linear income taxes and add multiple periods for each generation to refine the policy prescriptions on social security in a future version of this work.

Appendix A: Proofs and derivations

A.1 Households

The first-order conditions for a θ-type household problem with respect to savings, borrowings, and labor at time t, respectively, read as

$$\beta\psi_{t+1}\,\mathrm{E}_t\left[R^{K,\tau}_{t+1}u^o_{c,\theta,t+1}\right] = u^y_{c,\theta,t} \tag{A.1a}$$

$$\beta\psi_{t+1}R^{b,\tau}_t\,\mathrm{E}_t\left[u^o_{c,\theta,t+1}\right] = u^y_{c,\theta,t} \tag{A.1b}$$

$$\begin{aligned} &u^y_{c,\theta,t}\left[1-\tau^l_t-\tau^{\mathrm{ss,F}}_t-\tau^{\mathrm{ss,U}}_t\right]w_{\theta,t} \\ &+u^y_{l,\theta,t}+\beta\psi_{t+1}\kappa^{\mathrm{ss}}_t w_{\theta,t}\,\mathrm{E}_t\left[u^o_{c,\theta,t+1}\right]=0 \end{aligned} \tag{A.1c}$$

First, we notice that

$$R^{b,\tau}_t = \frac{u^y_{c,\theta,t}}{\beta\psi_{t+1}\,\mathrm{E}_t\left[u^o_{c,\theta,t+1}\right]} \qquad \text{and} \qquad \frac{u^y_{c,\theta,t}}{\beta\psi_{t+1}\,\mathrm{E}_t\left[R^{K,\tau}_{t+1}u^o_{c,\theta,t+1}\right]} = 1$$

Substituting the expression for $R^{K,\tau}_{t+1}$ at the optimum and rearranging the terms, we can express capital taxes as

$$\beta\psi_{t+1}\left(1-\tau^K(s^t)\right) = \frac{u^y_{c,\theta,t}-\beta\psi_{t+1}u^o_{c,\theta,t+1}}{\mathrm{E}_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^o_{c,\theta,t+1}\right]}$$

Moreover, the first order condition with respect to labor gives an expression for income taxes:

$$\left[1-\tau^l_t-\tau^{\mathrm{ss,F}}_t-\tau^{\mathrm{ss,U}}_t\right]w_{\theta,t} = -\frac{u^y_{l,\theta,t}+\beta\psi_{t+1}\kappa^{\mathrm{ss}}_t w_{\theta,t}\,\mathrm{E}_t\left[u^o_{c,\theta,t+1}\right]}{u^y_{c,\theta,t}}$$

A.2 Social planner

The first order conditions of the social planner problem with respect to young's consumption, old's consumption, capital and labor, respectively, read as:

$$\begin{aligned} g(\theta)\,u^y_{c,\theta,t} &= \frac{\eta_t}{\phi_t} \\ \beta g(\theta)\,u^o_{c,\theta,t} &= \frac{\eta_t}{\phi_{t-1}} \\ \mathrm{E}_t\left[\eta_{t+1}\left(\gamma_{t+1}F_{K,t+1}+1-\delta\right)\right] &= \eta_t \\ \frac{g(\theta)\,u^y_{l,\theta,t}}{\theta\gamma_t F_{L,t}} &= -\frac{\eta_t}{\phi_t} \end{aligned}$$

where η_t is the Lagrange multiplier of the feasibility constraint at time t.

Combining the first two, we obtain

$$\frac{u^y_{c,\theta,t}}{\beta u^o_{c,\theta,t}} = \frac{\phi_{t-1}}{\phi_t}$$

From the first and the third, we have

$$\frac{\phi_t}{\phi_{t+1}} \frac{u^y_{c,\theta,t}}{\mathrm{E}_t\left[u^y_{c,\theta,t+1}\left(1+\gamma_{t+1}F_{K,t+1}-\delta\right)\right]} = 1$$

While the first and the fourth give the standard labor-leisure condition

$$-\frac{u^y_{l,\theta,t}}{u^y_{c,\theta,t}} = \theta\gamma_t F_{L,t}$$

A.3 Ramsey planner

Proof of Proposition 8

Proof. We start by using the first-order conditions for the firm to write wages and capital returns as

$$\theta\gamma_t F_{L,t} = w_{\theta,t} \quad \text{and} \quad \gamma_t F_{K,t} = r^K_t.$$

Then we note that equations (3.4.5) and (3.4.4) are equivalent to the first order conditions (3.2.3) and (3.2.4) for the household problem. Substituting them into the budget conditions (3.2.1) and (3.2.2) we can easily obtain (3.4.1) and (3.4.2). Finally, (3.4.3) is equivalent to (3.2.7) after we substitute (3.2.5). Proceeding as in

Section 3.4, we write the Lagrangian of the Ramsey problem as:

$$
\begin{aligned}
\mathcal{L} = {} & \phi_0 \beta \psi_0 n_0^o \sum_\theta g(\theta)\, u\left(c_\theta^0, 0\right) \Theta_t(\theta) + \mathrm{E}_0 \left[\sum_{t\geq 0} \phi_t n_t^y \sum_\theta g(\theta)\, U\left(c_{\theta,t}^y, l_{\theta,t}, c_{\theta,t+1}^o\right) \Theta_t(\theta) \right] \\
& - \mathrm{E}_0 \left[\sum_{t\geq 0} \sum_\theta \lambda_{\theta,t}^y \left[u_{l,\theta,t}^y l_{\theta,t} + u_{c,\theta,t}^y \left[c_{\theta,t}^y + q_{\theta,t} + b_{\theta,t} \right] + \beta \psi_{t+1} \kappa_t^{\mathrm{ss}} \theta \gamma_t F_{L,t} l_{\theta,t}\, \mathrm{E}_t \left[u_{c,\theta,t+1}^o \right] \right] \right] \\
& - \mathrm{E}_0 \left[\sum_{t\geq 0} \sum_\theta \beta \lambda_{\theta,t+1}^o \left[c_{\theta,t+1}^o - \frac{u_{c,\theta,t}^y}{\beta \psi_{t+1} \mathrm{E}_t \left[u_{c,\theta,t+1}^o \right]} b_{\theta,t} - \kappa_t^{\mathrm{ss}} \theta \gamma_t F_{L,t} \right.\right. \\
& \left.\left. - \ q_{\theta,t} \left(1 + \frac{u_{c,\theta,t}^y - \beta \psi_{t+1} \mathrm{E}_t \left[u_{c,\theta,t+1}^o \right]}{\beta \psi_{t+1} \mathrm{E}_t \left[\hat{F}_{K,t+1} u_{c,\theta,t+1}^o \right]} \hat{F}_{K,t+1} \right) \right] \right] \\
& - \mathrm{E}_0 \left[\sum_{t\geq 0} \lambda_t^f \left[n_t^y \sum_\theta c_{\theta,t}^y \Theta_t(\theta) + n_t^o \sum_\theta c_{\theta,t}^o \Theta_t(\theta) + G_t - \gamma_t F\left(K_{t-1}, L_t\right) + n_t^y \sum_\theta q_{\theta,t} \Theta_t(\theta) \right.\right. \\
& + n_t^y \tau_t^{\mathrm{ss,F}} \gamma_t F_{L,t} \sum_\theta \theta l_{\theta,t} \Theta_t(\theta) \\
& \left.\left. - \ (1-\delta) \left[n_{t-1}^y \sum_\theta q_{\theta,t-1} \Theta_{t-1}(\theta) + n_{t-1}^y \tau_{t-1}^{\mathrm{ss,F}} \gamma_{t-1} F_{L,t-1} \sum_\theta \theta l_{\theta,t-1} \Theta_{t-1}(\theta) \right] \right] \right] \\
& - \mathrm{E}_0 \left[\sum_{t\geq 0} \sum_\theta \sum_{\theta' \neq \theta} \lambda_{\theta,\theta',t}^K \left[\frac{u_{c,\theta,t}^y - \beta \psi_{t+1} \mathrm{E}_t \left[u_{c,\theta,t+1}^o \right]}{\beta \psi_{t+1} \mathrm{E}_t \left[\hat{F}_{K,t+1} u_{c,\theta,t+1}^o \right]} - \frac{u_{c,\theta',t}^y - \beta \psi_{t+1} \mathrm{E}_t \left[u_{c,\theta',t+1}^o \right]}{\beta \psi_{t+1} \mathrm{E}_t \left[\hat{F}_{K,t+1} u_{c,\theta',t+1}^o \right]} \right] \right] \\
& - \mathrm{E}_0 \left[\sum_{t\geq 0} \sum_\theta \sum_{\theta' \neq \theta} \lambda_{\theta,\theta',t}^L \left[\frac{u_{l,\theta,t}^y + \beta \psi_{t+1} \kappa_t^{\mathrm{ss}} F_{L,t} \mathrm{E}_t \left[u_{c,\theta,t+1}^o \right]}{\theta \gamma_t F_{L,t} u_{c,\theta,t}^y} - \frac{u_{l,\theta',t}^y + \beta \psi_{t+1} \kappa_t^{\mathrm{ss}} F_{L,t} \mathrm{E}_t \left[u_{c,\theta',t+1}^o \right]}{\theta \gamma_t F_{L,t} u_{c,\theta',t}^y} \right] \right]
\end{aligned}
\tag{A.2}
$$

□

A.4 The deadweight loss from labor taxation

Proof of Lemma 1

Proof. An individual θ solves

$$
\begin{aligned}
\underset{c_{\theta,t}^y, c_{\theta,t+1}^o}{\mathit{maximize}} \quad & u\left(c_{\theta,t}^y, l_{\theta,t}\right) + \beta\, \mathrm{E}_t \left[u\left(c_{\theta,t+1}^o, 0\right) \right] \\
\text{subject to} \quad & c_{\theta,t}^y + q_{\theta,t} + b_{\theta,t} \leq \left[1 - \tau_t^l - \tau_t^{\mathrm{ss,F}} - \tau_t^{\mathrm{ss,U}} \right] w_{\theta,t} l_{\theta,t} + Tr_t, \\
& c_{\theta,t+1}^o \leq R_{t+1}^{K,\tau} q_{\theta,t} + R_t^{b,\tau} b_{\theta,t} + \kappa_t^{\mathrm{ss}} w_{\theta,t} l_{\theta,t}
\end{aligned}
$$

where Tr_t are lump-sum transfers used by the government to compensate young agents' utility. Now, define $V\left(c^y_{\theta,t}, l_{\theta,t}, c^o_{\theta,t}\right)$ the indirect utility of the individual θ. From the envelope theorem, $V_{Tr} \equiv \partial V/\partial Tr = \mu^y_\theta$, where $\mu^y_\theta = u^y_{c,\theta,t}$ is the Lagrange multiplier on the young budget constraint. It follows that $\partial V/\partial \tau^l_t = -w_{\theta,t} l_{\theta,t} \mu^y_\theta = -w_{\theta,t} l_{\theta,t} \partial V/\partial Tr_t$.

Consider the planner's problem

$$\begin{aligned}
maximize \quad & \phi_0 + \beta\psi_0 n^o_0 \sum_\theta g(\theta) u\left(c^o_{\theta,0}, 0\right)\Theta_t(\theta) \\
& + \mathrm{E}_0\left[\sum_{t\geq 0}\phi_t n^y_t \sum_\theta g(\theta) V\left(c^y_{\theta,t}, l_{\theta,t}, c^o_{\theta,t}\right)\Theta_t(\theta)\right] \\
\text{subject to} & \\
n^o_t R^{b,\tau}_{t-1} \sum_\theta b_{\theta,t-1}\Theta_t(\theta) + G_t \leq & \; n^y_t \sum_\theta \left[1-\tau^l_t - \tau^{\mathrm{ss,F}}_t - \tau^{\mathrm{ss,U}}_t\right] w_{\theta,t} l_{\theta,t}\Theta_t(\theta) \\
& - n^y_t \sum_\theta b_{\theta,t}\Theta_t(\theta) - n^o_t \tau^K_{t-1} r^K_t \sum_\theta q_{\theta,t-1}\Theta_t(\theta) \\
& - n^o_t \tau^{\mathrm{ss,F}}_{t-1} R^{K,\tau}_t) \sum_\theta w_{\theta,t-1} l_{\theta,t-1}\Theta_t(\theta)
\end{aligned}$$

with the associated Lagrangian

$$\begin{aligned}
\mathcal{L} = & \; \phi_0 + \beta\psi_0 n^o_0 \sum_\theta g(\theta) u(c^o_{\theta,0})\Theta_t(\theta) + \mathrm{E}_t\left[\sum_{t\geq 0}\phi_t n^y_t \sum_\theta g(\theta) V\left(c^y_{\theta,t}, l_{\theta,t}, c^o_{\theta,t}\right)\Theta_t(\theta)\right] \\
& - E_0\Bigg[\sum_{t\geq 0}\lambda_t\Big[n^o_t R^{b,\tau}_{t-1}\sum_\theta b_{\theta,t-1}\Theta_t(\theta) + G_t \\
& - n^y_t \sum_\theta \left[1-\tau^l_t - \tau^{\mathrm{ss,F}}_t - \tau^{\mathrm{ss,U}}_t\right] w_{\theta,t} l_{\theta,t}\Theta_t(\theta) - n^y_t \sum_\theta b_{\theta,t-1}\Theta_t(\theta) \\
& n^o_t \tau^K_{t-1} r^K_t \sum_\theta q_{\theta,t-1}\Theta_t(\theta) - n^o_t \tau^{\mathrm{ss,F}}_{t-1} R^{K,\tau}_t) \sum_\theta w_{\theta,t-1} l_{\theta,t-1}\Theta_t(\theta)\Big]\Bigg]
\end{aligned}$$

Naming the different tax basis as $Y_{\theta_t} = w_{\theta,t} l_{\theta,t}$, $Y^{ss}_{\theta_t} = R^{K,\tau}_{t+1} w_{\theta,t} l_{\theta,t}$, $B_{\theta,t} =$

$R_t^{b,\tau} b_{\theta,t}$, $Q_{\theta,t} = r_t^K q_{\theta,t-1}$, we can write the FOC for τ_t^l as

$$\phi_t n_t^y \sum_\theta g(\theta) \frac{\partial V\left(c_{\theta,t}^y, l_{\theta,t}, c_{\theta,t}^o\right)}{\partial \tau_t^l} \Theta_t(\theta) - E_t\Bigg[\overbrace{-\lambda n_t^y \sum_\theta Y_{\theta_t} \Theta_t(\theta)}^{\text{Mechanical revenue}}$$
$$+ \sum_{s\geq t} \lambda_s \sum_\theta \Bigg(n_s^o \frac{\partial B_{\theta,s-1}}{\partial \tau_t^l} - n_s^y \left[1 - \tau_t^l - \tau_t^{\text{ss,F}} - \tau_t^{\text{ss,U}}\right] \frac{\partial Y_{\theta_s}}{\partial \tau_t^l} + n_s^y \frac{\partial b_{\theta,t}}{\partial \tau_t^l}$$
$$- n_s^o \tau_{t-1}^K \frac{\partial Q_{\theta,s-1}}{\partial \tau_t^l} + n_s^o \tau_{t-1}^{\text{ss,F}} \frac{\partial Y_{\theta,s-1}^{ss}}{\partial \tau_t^l} \Bigg) \Theta_t(\theta) = 0$$

We can use the Slutsky equation and the envelope theorem on the individual's problem to obtain the following relationship

$$n_t^y \sum_\theta Y_{\theta_t} \underbrace{\left[-\frac{\phi_t g(\theta)}{\lambda_t} \frac{\partial V\left(c_{\theta,t}^y, l_{\theta,t}, c_{\theta,t}^o\right)}{\partial Tr_t} + 1 - \frac{\partial FE_{s\geq t}}{\partial Tr_t} \right]}_{\text{Govt. value for decreasing an extra transfer to all } \theta\text{s}} \Theta_t(\theta) = \underbrace{\frac{\partial FE_{s\geq t}^c}{\partial \tau_t^l}}_{\substack{\text{Revenue loss from} \\ \text{compensated tax} \\ \text{base change}}} \tag{A.3}$$

where the compensated fiscal externalities are given by

$$\frac{\partial FE_{s\geq t}}{\partial Tr_t} = E_t\Bigg[\sum_{s\geq t} \frac{\lambda_s}{\lambda_t} \Bigg(n_s^o \sum_\theta \frac{\partial B_{\theta,s-1}}{\partial Tr_t} - n_t^y \sum_\theta \left[1 - \tau_t^l - \tau_t^{\text{ss,F}} - \tau_t^{\text{ss,U}}\right] \frac{\partial Y_{\theta_s}}{\partial Tr_t}$$
$$- n_s^y \sum_\theta \frac{\partial b_{\theta,t}}{\partial Tr_t} - n_s^o \tau_{t-1}^K \sum_\theta \frac{\partial Q_{\theta,s-1}}{\partial Tr_t} - n_s^o \tau_{t-1}^{\text{ss,F}} \sum_\theta \frac{\partial Y_{\theta,s-1}^{ss}}{\partial Tr_t} \Bigg) \Theta_t(\theta) \Bigg]$$

and

$$\frac{\partial FE_{s\geq t}^c}{\partial \tau_t^l} = E_t\Bigg[\sum_{s\geq t} \frac{\lambda_s}{\lambda_t} \Bigg(n_s^o \sum_\theta \frac{\partial B_{\theta,s-1}}{\partial \tau_t^l} - n_t^y \sum_\theta \left[1 - \tau_t^l - \tau_t^{\text{ss,F}} - \tau_t^{\text{ss,U}}\right] \frac{\partial Y_{\theta_s}}{\partial \tau_t^l}$$
$$- n_s^y \sum_\theta \frac{\partial b_{\theta,t}}{\partial \tau_t^l} - n_s^o \tau_{t-1}^K \sum_\theta \frac{\partial Q_{\theta,s-1}}{\partial \tau_t^l} - n_s^o \tau_{t-1}^{\text{ss,F}} \sum_\theta \frac{\partial Y_{\theta,s-1}^{ss}}{\partial \tau_t^l} \Bigg) \Theta_t(\theta) \Bigg]$$

Equation (A.3) equates the value of reducing the transfer of an extra unit of consumption to every young individual in t to the marginal excess burden of the tax. The latter is equal to the fiscal externality computed using the compensated responses of different tax bases, including general equilibrium price changes caused by the increase in τ^l. Since everything is normalized by λ_t, quantities are in terms of government revenues.

In our main Ramsey setup, the value of decreasing transfers to a young individual of type θ in t is captured by the multiplier $\lambda^y_{\theta,t}$, i.e., the value of relaxing the individual's implementability condition. To convert this value into government revenues, we normalize it by the marginal social value of increasing utility for a young individual of type θ in t:

$$\tilde{\mathcal{W}}_{\theta,t} = n^y_t \phi_t g(\theta)\, \Theta_t(\theta)$$

Finally, summing across the individuals, we obtain the total value of a reduction of one unit of transfer to all young agents:

$$\lambda^L(s^t) = -\sum_\theta \frac{\lambda^y_{\theta,t}}{\mathcal{W}_{\theta,t}}$$

which, in equilibrium, equates to the deadweight loss and is thus equivalent to the labor wedge. □

Proof of Proposition 9

Proof. This proposition follows from Equation (A.10) noting that in the quasi-linear case $u^y_{c,\theta,t} = u^o_{c,\theta,t+1} = 1$ and $u^y_{cc,\theta,t} = u^o_{cc,\theta,t+1} = u^y_{lc,\theta,t} = 0$. Hence in this case $\varepsilon^{R^{\tau,b}}_{b_\theta,t} = 0$ and $\frac{d\tau^K_t}{db_{\theta,t}} = 0$. □

Proof of Proposition 10

Proof. This proposition follows from Equation (A.10) noting that the terms in $\lambda^K_{\theta,\theta',t}$ and $\lambda^L_{\theta,\theta',t}$ will not appear since, with homogeneous agents, their terms vanish from the Lagrangian (A.2). □

Proof of Proposition 11

Proof. We prove the most general case (for two types θ and θ') in which the utility is not necessarily separable, and then all the other cases will follow from this. To start, we compute the first-order conditions of the Ramsey planner with respect to $c^y_{\theta,t}$ and $c^o_{\theta,t+1}$ by taking the appropriate derivatives of the Lagrangian

(A.2). The first order condition with respect to $c^y_{\theta,t}$ reads:

$$\begin{aligned}
\lambda^f_t n^y_t \Theta_t(\theta) = {} & \phi_t n^y_t g(\theta)\Theta'_t(\theta) u^y_{c,\theta,t} - \lambda^y_{\theta,t}\left[\left(c^y_{\theta,t} + q^y_{\theta,t} + b^y_{\theta,t}\right) u^y_{cc,\theta,t} + u^y_{c,\theta,t} + u^y_{lc,\theta,t} l_{\theta,t}\right] \\
& + E_t\left[\lambda^o_{\theta,t+1} u^y_{cc,\theta,t}\left(\frac{(\gamma_{t+1}F_{K,t+1}-\delta)\, q^y_{\theta,t}}{\psi_{t+1}\,\mathrm{E}_t\left[(\gamma_{t+1}F_{K,t+1}-\delta)\, u^o_{c,\theta,t+1}\right]} + \frac{b^y_{\theta,t}}{\psi_{t+1}\,\mathrm{E}_t\left[u^o_{c,\theta,t+1}\right]}\right)\right] \\
& - \lambda^K_{\theta,\theta',t}\frac{u^y_{cc,\theta,t}}{\beta\psi_{t+1}\,\mathrm{E}_t\left[(\gamma_{t+1}F_{K,t+1}-\delta)\, u^o_{c,\theta,t+1}\right]} \\
& - \lambda^L_{\theta,\theta',t}\left[\frac{u^y_{lc,\theta,t}}{\gamma_t\theta F_{\tilde{L},t} u^y_{c,\theta,t}} - u^y_{cc,\theta,t}\frac{u^y_{l,\theta,t} + \beta\psi_{t+1}\kappa^{ss}_t\gamma_t\theta F_{\tilde{L},t} E_t\left[u^o_{c,\theta,t+1}\right]}{\gamma_t\theta F_{\tilde{L},t}\left(u^y_{c,\theta,t}\right)^2}\right]
\end{aligned} \tag{A.4}$$

while the one with respect to $c^o_{\theta,t}$ is:

$$\begin{aligned}
\lambda^f_t n^o_t \psi_t \Theta_t(\theta) = {} & \phi_{t-1} n^y_{t-1} g(\theta)\,\Theta'_{t-1}(\theta)\,\beta\psi_t u^o_{c,\theta,t} - \lambda^y_{\theta,t-1}\left(\beta\psi_t\kappa^{ss}_{t-1}\theta\gamma_{t-1}F_{\tilde{L},t-1} l_{\theta,t-1} p_{t|t-1} u^o_{cc,\theta,t}\right) \\
& - \beta\lambda^o_{\theta,t} + E_{t-1}\left[\lambda^o_{\theta,t}\right]\frac{u^y_{c,\theta,t-1} u^o_{cc,\theta,t}}{\psi_t E_{t-1}\left[u^o_{c,\theta,t}\right]^2} b^y_{\theta,t-1} \\
& - E_{t-1}\left[(\gamma_t F_{K,t}-\delta)\,\lambda^o_{\theta,t}\right]\left(\frac{\left(u^y_{c,\theta,t-1} - \beta\psi_t E_{t-1}\left[u^o_{c,\theta,t}\right]\right) u^o_{cc,\theta,t}}{\psi_t E_{t-1}\left[(\gamma_t F_{K,t}-\delta)\, u^o_{c,\theta,t}\right]^2}\right)(\gamma_t F_{K,t}-\delta)\, q^y_{\theta,t-1} \\
& - E_{t-1}\left[(\gamma_t F_{K,t}-\delta)\,\lambda^o_{\theta,t}\right]\beta\frac{u^o_{cc,\theta,t}}{E_{t-1}\left[(\gamma_t F_{K,t}-\delta)\, u^o_{c,\theta,t}\right]} q^y_{\theta,t-1} \\
& - \lambda^K_{\theta,\theta',t-1} E_{t-1}\left[\left[-\frac{u^o_{cc,\theta,t}}{E_{t-1}\left[\gamma_t(F_{K,t}-\delta)\, u^o_{c,\theta,t}\right]}\right.\right. \\
& \left.\left. - \frac{u^y_{c,\theta,t-1} - \beta\psi_t E_{t-1}\left[u^o_{c,\theta,t}\right]}{\beta\psi_t\left(E_{t-1}\left[(\gamma_t F_{K,t}-\delta)\, u^o_{c,\theta,t}\right]\right)^2}(\gamma_t F_{K,t}-\delta)\, u^o_{cc,\theta,t}\right]\right] \\
& - \lambda^L_{\theta,\theta',t-1} E_{t-1}\left[\frac{\beta\psi_t\kappa^{ss}_{t-1}\gamma_{t-1}\theta F_{\tilde{L},t-1} u^o_{cc,\theta,t}}{\gamma_{t-1}\theta F_{\tilde{L},t-1} u^y_{c,\theta,t-1}}\right]
\end{aligned} \tag{A.5}$$

In the following it will be useful to use the first order condition of the Ramsey planner with respect to $b_{\theta,t}$,

$$\lambda^{y}_{\theta,t} = \frac{E_t\left[\lambda^{o}_{\theta,t+1}\right]}{\psi_{t+1}E_t[u^{o}_{c,\theta,t+1}]} \tag{A.6}$$

and a similar identity that follows from it by using (A.1b):

$$\frac{\lambda^{y}_{\theta,t}u^{y}_{c,\theta,t}}{\beta R^{\tau,b}_{t}E_t\left[\lambda^{o}_{\theta,t+1}\right]} = 1 \tag{A.7}$$

Furthermore we recall the definition of the marginal social utility of an increase in consumption for a young worker of type θ:

$$\tilde{\mathcal{W}}_{\theta,t} = n^{y}_{t}\phi_t g\left(\theta\right)\Theta_t\left(\theta\right)u^{y}_{c,\theta,t} \tag{A.8}$$

From Equation (A.4) using $\tilde{\lambda}^{y}_{\theta,t} = \frac{\lambda^{y}_{\theta,t}u^{y}_{c,\theta,t}}{\tilde{\mathcal{W}}_{\theta,t}}$ we get:

$$\begin{aligned}
\lambda^{f}_{t}n^{y}_{t}\Theta_t\left(\theta\right) = {} & \tilde{\mathcal{W}}_{\theta,t}\left[1-\tilde{\lambda}^{y}_{\theta,t}\left[-\frac{q^{y}_{\theta,t}+b^{y}_{\theta,t}}{c^{y}_{\theta,t}}\sigma^{y}_{\theta,t}+\eta^{y}_{\theta,t}\right]\right] \\
& + \tilde{\mathcal{W}}_{\theta,t}E_t\left[\frac{\lambda^{o}_{\theta,t+1}}{\mathcal{W}_{\theta,t}}\frac{u^{y}_{cc,\theta,t}}{u^{y}_{c,\theta,t}}\left(\frac{\left(\gamma_{t+1}F_{K,t+1}-\delta\right)q^{y}_{\theta,t}}{\psi_{t+1}\,\mathrm{E}_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^{o}_{c,\theta,t+1}\right]}+\frac{b^{y}_{\theta,t}}{\psi_{t+1}\,\mathrm{E}_t\left[u^{o}_{c,\theta,t+1}\right]}\right.\right. \\
& - \tilde{\mathcal{W}}_{\theta,t}\tilde{\lambda}^{K}_{t}\frac{u^{y}_{cc,\theta,t}}{u^{y}_{c,\theta,t}\beta\psi_{t+1}\,\mathrm{E}_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^{o}_{c,\theta,t+1}\right]} \\
& - \tilde{\mathcal{W}}_{\theta,t}\frac{\tilde{\lambda}^{L}_{t}}{u^{y}_{c,\theta,t}}\left[\frac{u^{y}_{lc,\theta,t}}{\gamma_t\theta F_{\tilde{L},t}u^{y}_{c,\theta,t}}-\sigma_{\theta,t}\frac{\left(1-\tau^{L}_{t}\right)}{c^{y}_{\theta,t}}\right]
\end{aligned}$$

Where $\sigma^{y}_{\theta,t} = -\frac{u^{y}_{cc,\theta,t}}{u^{y}_{c,\theta,t}}c^{y}_{\theta,t}$ represents the inverse of the inter-temporal elasticity of substitution and we define $\eta^{y}_{\theta,t} = 1+\frac{u_{lc,\theta,t}l_{\theta,t}}{u_{c,\theta,t}}-\sigma^{y}_{\theta,t}$. Using then Equations (A.6) and the households first order conditions (A.1a), and (A.1b) we get:

$$\lambda_t^f n_t^y \Theta_t(\theta) = \tilde{\mathcal{W}}_{\theta,t}\left[1 - \tilde{\lambda}_{\theta,t}^y\left[-\frac{q_{\theta,t}^y + b_{\theta,t}^y}{c_{\theta,t}^y}\sigma_{\theta,t}^y + \eta_{\theta,t}^y\right] - \sigma_{\theta,t}^y \frac{b_{\theta,t}^y}{c_{\theta,t}^y}\frac{\beta R_t^{\tau,b}\psi_{t+1}E_t\left[u_{c,\theta,t+1}^o\right]}{u_{c,\theta,t}^y}\frac{\lambda_{\theta,t}^y}{\mathcal{W}_{\theta,t}}\right.$$

$$- \sigma_{\theta,t}^y \frac{q_{\theta,t}^y}{c_{\theta,t}^y}\frac{1}{\psi_{t+1}r_t^b E_t\left[u_{c,\theta,t+1}^o\right]}E_t\left[\frac{\lambda_{\theta,t+1}^o}{\mathcal{W}_{\theta,t}}\left(\gamma_{t+1}F_{K,t+1} - \delta\right)\right]$$

$$\left.+ \tilde{\lambda}_t^K \frac{\sigma_{\theta,t}^y}{c_{\theta,t}^y}\frac{1}{\beta\psi_{t+1}r_t^b E_t\left[u_{c,\theta,t+1}^o\right]} - \frac{\tilde{\lambda}_t^L}{u_{c,\theta,t}^y}\left[\frac{u_{lc,\theta,t}^y}{\gamma_t\theta F_{\tilde{L},t}u_{c,\theta,t}^y} - \sigma_{\theta,t}\frac{\left(1-\tau_t^L\right)}{c_{\theta,t}^y}\right]\right]$$

where $\tilde{\lambda}_t^K = \frac{\lambda_{\theta,\theta',t}^K u_{c,\theta,t}^y}{\tilde{\mathcal{W}}_{\theta,t}}$, and $\tilde{\lambda}_t^L = \frac{\lambda_t^L u_{c,\theta,t}^y}{\tilde{\mathcal{W}}_{\theta,t}}$. After using (A.1a), and (A.6) we proceed collect $\tilde{\lambda}_{\theta,t}^y$ and we finally get:

$$\lambda_t^f n_t^y \Theta_t(\theta) = \tilde{\mathcal{W}}_{\theta,t}\left[1 - \tilde{\lambda}_{\theta,t}^y\left[\eta_{\theta,t}^y + \sigma_{\theta,t}^y\frac{q_{\theta,t}^y}{c_{\theta,t}^y}E_t\left[\frac{\lambda_{\theta,t+1}^o\left(\gamma_{t+1}F_{K,t+1} - \delta\right)}{r_t^b E_t\left[\lambda_{\theta,t+1}^o\right]} - 1\right]\right.\right.$$

$$\left.\left.-\sigma_{\theta,t}^y\frac{\lambda_{\theta,\theta',t}^K}{c_{\theta,t}^y}\frac{1}{\beta r_t^b E_t\left[\lambda_{\theta,t+1}^o\right]} - \frac{\lambda_t^L}{R_t^{b,\tau}\beta E_t\left[\lambda_{\theta,t+1}^o\right]}\left[\frac{u_{lc,\theta,t}^y}{\gamma_t\theta F_{\tilde{L},t}u_{c,\theta,t}^y} - \sigma_{\theta,t}\frac{\left(1-\tau_t^L\right)}{c_{\theta,t}^y}\right]\right]\right]$$

We now turn our attention to Equation (A.5). In order to facilitate the interpretation of this term we need the expressions for the elasticity of risk-free returns to issued bonds and the sensitivity of capital taxes to issued bonds:

$$\begin{aligned}\varepsilon_{b_\theta,t}^{R^{\tau,b}} &= -R_t^{\tau,b} b_{\theta,t}\frac{E_t\left[u_{cc,\theta,t+1}^o\right]}{E_t\left[u_{c,\theta,t+1}^o\right]},\\ \frac{d\tau_t^K}{db_t} &= R_t^{\tau,b}\frac{E_t\left[u_{cc,\theta,t+1}^o\right] + \left(1-\tau_t^K\right)E_t\left[\left(\gamma_{t+1}F_{K,t+1} - \delta\right)u_{cc,\theta,t+1}^o\right]}{E_t\left[\left(\gamma_{t+1}F_{K,t+1} - \delta\right)u_{c,\theta,t+1}^o\right]}\end{aligned} \tag{A.9}$$

The formulas above are obtained by computing the total derivative of (A.1a) and (A.1b) by keeping the young agents consumption constant and assuming that the change in the old agents consumption is given just by a change in $R_t^{\tau,b}$.

After taking the expectation at time $t-1$, and rearranging some terms using the households first order conditions, Equation (A.5) becomes:

$$
\begin{aligned}
E_{t-1}\left[\lambda_t^f\right] n_t^o \psi_t \Theta_t(\theta) = {} & \phi_{t-1} n_{t-1}^y g(\theta)\, \Theta_{t-1}(\theta)\, \beta \psi_t E_{t-1}\left[u_{c,\theta,t}^o\right] \\
& - \lambda_{\theta,t-1}^y \left(\beta \psi_t \kappa_{t-1}^{ss} \theta \gamma_{t-1} F_{\tilde{L},t-1} l_{\theta,t-1} p_{t|t-1} E_{t-1}\left[u_{cc,\theta,t}^o\right]\right) \\
& - \beta E_{t-1}\left[\lambda_{\theta,t}^o\right] \\
& - E_{t-1}\left[\lambda_{\theta,t}^o\right] \beta \varepsilon_{b_\theta,t-1}^{R^{\tau,b}} - E_{t-1}\left[\left(\gamma_t F_{K,t} - \delta\right) \lambda_{\theta,t}^o\right] \beta \frac{d\tau_{t-1}^K}{db_{\theta,t-1}} \frac{q_{\theta,t-1}^y}{R_{t-1}^{\tau,b}} \\
& + \frac{\lambda_{\theta,\theta',t-1}^K}{R_{t-1}^{\tau,b}} \frac{d\tau_{t-1}^K}{db_{\theta,t-1}} + \lambda_{\theta,\theta',t-1}^L \beta \psi_t \kappa_{t-1}^{ss} \frac{\varepsilon_{b_\theta,t-1}^{R^{\tau,b}}}{R_{t-1}^{\tau,b} b_{\theta,t-1}^y}
\end{aligned}
$$

Using Equations (A.1b), (A.6), (A.7), and (A.8) we get:

$$
\begin{aligned}
E_{t-1}\left[\lambda_t^f\right] n_t^o \psi_t \Theta_t(\theta) = {} & \frac{\tilde{\mathcal{W}}_{\theta,t-1}}{R_{t-1}^{b,\tau}} \left[1 - \tilde{\lambda}_{\theta,t-1}^y \left[1 + \varepsilon_{b_\theta,t-1}^{R^{\tau,b}} \left(1 - \frac{1}{R_{t-1}^{b,\tau}} \frac{\kappa_{t-1}^{ss}}{b_{\theta,t-1}^y} \left(z_{\theta,t-1} p_{t|t-1} \right.\right.\right.\right. \\
& + \left.\left.\left. \psi_t \frac{\lambda_{\theta,\theta',t-1}^L}{R_{t-1}^{\tau,b} E_{t-1}\left[\lambda_{\theta,t}^o\right]} \right)\right) \\
& + \left.\left. \frac{d\tau_{t-1}^K}{db_{\theta,t-1}} \frac{q_{\theta,t-1}^y \beta E_{t-1}\left[\left(\gamma_t F_{K,t} - \delta\right) \lambda_{\theta,t}^o\right] - \lambda_{\theta,\theta',t-1}^K}{\beta R_{t-1}^{\tau,b} E_{t-1}\left[\lambda_{\theta,t}^o\right]} \right]\right]
\end{aligned}
$$

where we used the notation $z_{\theta,t-1} = \theta \gamma_{t-1} F_{\tilde{L},t-1} l_{\theta,t-1}$. Finally we can put together the two expressions we found and get:

$$\frac{\tilde{\mathcal{W}}_{\theta,t}}{R_t^{b,\tau}}\left[1-\tilde{\lambda}_{\theta,t}^{y}\left[1+\varepsilon_{b_{\theta},t}^{R^{\tau,b}}\left(1-\frac{1}{R_t^{b,\tau}}\frac{\kappa_t^{ss}}{b_{\theta,t}^{y}}\left(z_{\theta,t}p_{t+1|t}+\psi_{t+1}\frac{\lambda_t^{L}}{R_t^{\tau,b}E_t\left[\lambda_{\theta,t+1}^{o}\right]}\right)\right)\right.\right.$$
$$\left.\left.+\ \frac{d\tau_t^K}{db_{\theta,t}}\frac{q_{\theta,t}^{y}\beta E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)\lambda_{\theta,t+1}^{o}\right]-\lambda_{\theta,\theta',t}^{K}}{\beta R_t^{\tau,b}E_t\left[\lambda_{\theta,t+1}^{o}\right]}\right]\right]$$
$$=\frac{n_{t+1}^{o}\psi_{t+1}\Theta_{t+1}\left(\theta\right)}{n_{t+1}^{y}\Theta_{t+1}\left(\theta\right)}E_t\left[\widetilde{\mathcal{W}}_{\theta,t+1}\left[1-\tilde{\lambda}_{\theta,t+1}^{y}\left[\eta_{\theta,t+1}^{y}\right.\right.\right.$$
$$+\ \sigma_{\theta,t+1}^{y}\frac{q_{\theta,t+1}^{y}}{c_{\theta,t+1}^{y}}E_{t+1}\left[\frac{\lambda_{\theta,t+2}^{o}\left(\gamma_{t+2}F_{K,t+2}-\delta\right)}{r_{t+1}^{b}E_{t+1}\left[\lambda_{\theta,t+2}^{o}\right]}-1\right]$$
$$-\sigma_{\theta,t+1}^{y}\frac{\lambda_{t+1}^{K}}{c_{\theta,t+1}^{y}}\frac{1}{\beta r_{t+1}^{b}E_{t+1}\left[\lambda_{\theta,t+2}^{o}\right]}-\frac{\lambda_{\theta,\theta',t+1}^{L}}{R_{t+1}^{b,\tau}\beta E_{t+1}\left[\lambda_{\theta,t+2}^{o}\right]}\left[\frac{u_{lc,\theta,t+1}^{y}}{\gamma_{t+1}\theta F_{\tilde{L},t+1}u_{c,\theta,t+1}^{y}}\right.$$
$$\left.\left.\left.\left.-\ \sigma_{\theta,t+1}\frac{\left(1-\tau_{t+1}^{L}\right)}{c_{\theta,t+1}^{y}}\right]\right]\right]\right] \tag{A.10}$$

This general expression can be then reduced to the case of a separable utility function noting that in that case

$$\eta_{\theta,t}^{y}=1-\sigma_{\theta,t}^{y}$$

□

A.5 Capital taxes and social security

Proof of Proposition 12

Proof. This proposition follows from A.5 once we set $\tau_t^{\mathrm{ss,F}}$ and κ_t^{ss} to zero, and noting that in the case of homogeneous agents we do not have terms in $\lambda_{\theta,\theta',t}^{K}$. □

Proof of Proposition 13

Proof. This proposition follows from A.5 once we set $\tau_t^{\mathrm{ss,F}}$ and κ_t^{ss} to zero. □

Proof of Proposition 14

Proof. The proof follows form A.5. In particular, we can use (A.12) and note that with homogeneous agents there are no terms in $\lambda^K_{\theta,\theta',t}$. Then we can substitute in (A.14) and realize that almost all terms cancel out and we are left with

$$\tau^K_t = \frac{E_t\left[\beta\tilde{\lambda}^o_{t+1}u^o_{c,t+1}R_{t+1}\right] - \tilde{\lambda}^y_t u^y_{c,\theta,t}}{\beta E_t\left[\tilde{\lambda}^o_{t+1}u^o_{c,t+1}\left(\gamma_{t+1}F_{K,t+1} - \delta\right)\right]}$$

□

Proof of Proposition 15

Proof. We prove the most general case (for two types θ and θ') in which the utility is not necessarily separable and then all the other cases will follow from this. To start, we compute the first order conditions of the Ramsey planner with respect to $q_{\theta,t}$ by taking the appropriate derivative of the Lagrangian (A.2). This first order condition reads:

$$
\begin{aligned}
0 = &- \lambda^y_{\theta,t} u^y_{c,\theta,t} + E_t\left[\left(\lambda^o_{\theta,t+2} - \psi_{t+2}\lambda^y_{\theta,t+1} E_{t+1}\left[u^o_{c,\theta,t+2}\right]\right)\beta\kappa^{ss}_{t+1}\theta\gamma_{t+1}F_{\tilde{L}K,t+1}n^y_t\Theta_t(\theta)\, l_{\theta,t+1}\right] \\
&+ E_t\left[\beta\lambda^o_{\theta,t+1}\left(1 + \left(u^y_{c,\theta,t} - \beta\psi_{t+1}E_t\left[u^o_{c,\theta,t+1}\right]\right)\frac{\left(\gamma_{t+1}F_{K,t+1} - \delta\right)}{\beta\psi_{t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1} - \delta\right)u^o_{c,\theta,t+1}\right]}\right)\right] \\
&+ E_t\left[\beta\lambda^o_{\theta,t+1}q^y_{\theta,t}\left(u^y_{c,\theta,t} - \beta\psi_{t+1}E_t\left[u^o_{c,\theta,t+1}\right]\right)\frac{n^y_t\Theta_t(\theta)}{\beta\psi_{t+1}}\right. \\
&\left.\left(\frac{\gamma_{t+1}F_{KK,t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1} - \delta\right)u^o_{c,\theta,t+1}\right] - \left(\gamma_{t+1}F_{K,t+1} - \delta\right)E_t\left[\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta,t+1}\right]}{E_t\left[\left(\gamma_{t+1}F_{K,t+1} - \delta\right)u^o_{c,\theta,t+1}\right]^2}\right)\right] \\
&- \lambda^f_t n^y_t\Theta_t(\theta) + E_t\left[\lambda^f_{t+1}\Theta_t(\theta)\left(\gamma_{t+1}F_{K,t+1}n^y_t + (1-\delta)n^y_t\right)\right] \\
&+ E_t\left[\left(\lambda^f_{t+2}(1-\delta) - \lambda^f_{t+1}\right)n^y_{t+1}\tau^{ss,F}_{t+1}\gamma_{t+1}F_{\tilde{L}K,t+1}n^y_t\Theta_t(\theta)\sum_\theta \theta l_{\theta,t+1}\Theta_{t+1}(\theta)\right] \\
&+ \lambda^K_{\theta,\theta',t}\left[\frac{u^y_{c,\theta,t} - \beta\psi_{t+1}E_t\left[u^o_{c,\theta,t+1}\right]}{\beta\psi_{t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1} - \delta\right)u^o_{c,\theta,t+1}\right]^2}E_t\left[\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta,t+1}\right]n^y_t\Theta_t(\theta)\right. \\
&\left.-\frac{u^y_{c,\theta',t} - \beta\psi_{t+1}E_t\left[u^o_{c,\theta',t+1}\right]}{\beta\psi_{t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1} - \delta\right)u^o_{c,\theta',t+1}\right]^2}E_t\left[\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta',t+1}\right]n^y_t\Theta_t(\theta')\right] \\
&- E_t\left[\lambda^L_{\theta,\theta',t+1}\frac{\kappa^{ss}_{t+1}\gamma_{t+1}\theta F_{\tilde{L}K,t+1}E_{t+1}\left[u^o_{c,\theta,t+2}\right]}{\gamma_{t+1}\theta F_{\tilde{L},t+1}u^y_{c,\theta,t+1}}n^y_t\Theta_t(\theta)\right. \\
&-\lambda^L_{\theta,\theta',t+1}\frac{\kappa^{ss}_{t+1}\gamma_{t+1}\theta' F_{\tilde{L}K,t+1}E_{t+1}\left[u^o_{c,\theta',t+2}\right]}{\gamma_{t+1}\theta' F_{\tilde{L},t+1}u^y_{c,\theta',t+1}}n^y_t\Theta_t(\theta) \\
&- \lambda^L_{\theta,\theta',t+1}\frac{u^y_{l,\theta,t} + \beta\psi_{t+2}\kappa^{ss}_{t+1}\gamma_{t+1}\theta F_{\tilde{L},t+1}E_{t+1}\left[u_{c,\theta,t+2}\right]}{\gamma_{t+1}\theta u^y_{c,\theta,t+1}F^2_{\tilde{L},t+1}}F_{\tilde{L}K,t+1}n^y_t\Theta_t(\theta) \\
&\left.+ \lambda^L_{\theta,\theta',t+1}\frac{u^y_{l,\theta',t} + \beta\psi_{t+2}\kappa^{ss}_{t+1}\gamma_{t+1}\theta' F_{\tilde{L},t+1}E_{t+1}\left[u_{c,\theta',t+2}\right]}{\gamma_{t+1}\theta' u^y_{c,\theta',t+1}F^2_{\tilde{L},t+1}}F_{\tilde{L}K,t+1}n^y_t\Theta_t(\theta)\right]
\end{aligned}
$$

Using (3.4.4), (3.4.5), and (A.7) we get:

$$
\begin{aligned}
0 = & -\lambda^{y}_{\theta,t} u^{y}_{c,\theta,t} + E_t\left[\left(\lambda^{o}_{\theta,t+2} - \psi_{t+2}\lambda^{y}_{\theta,t+1} E_{t+1}\left[u^{o}_{c,\theta,t+2}\right]\right)\beta\kappa^{ss}_{t+1}\theta\gamma_{t+1}F_{\tilde{L}K,t+1}n^{y}_{t}\Theta_t(\theta)\, l_{\theta,t+1}\right] \\
& + E_t\left[\beta\lambda^{o}_{\theta,t+1}R^{K,\tau}_{t+1}\right] \\
& + E_t\left[\beta\lambda^{o}_{\theta,t+1}q^{y}_{\theta,t}\left(1-\tau^{K}_{t}\right)\frac{n^{y}_{t}\Theta_t(\theta)}{\beta\psi_{t+1}}\right. \\
& \left.\left(\frac{\gamma_{t+1}F_{KK,t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^{o}_{c,\theta,t+1}\right] - \left(\gamma_{t+1}F_{K,t+1}-\delta\right)E_t\left[\gamma_{t+1}F_{KK,t+1}u^{o}_{c,\theta,t+1}\right]}{E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^{o}_{c,\theta,t+1}\right]}\right)\right] \\
& -\lambda^{f}_{t}n^{y}_{t}\Theta_t(\theta) + E_t\left[\lambda^{f}_{t+1}\Theta_t(\theta)\left(\gamma_{t+1}F_{K,t+1}n^{y}_{t} + (1-\delta)n^{y}_{t}\right)\right] \\
& + E_t\left[\left(\lambda^{f}_{t+2}(1-\delta) - \lambda^{f}_{t+1}\right)n^{y}_{t+1}\tau^{ss,F}_{t+1}\gamma_{t+1}F_{\tilde{L}K,t+1}n^{y}_{t}\Theta_t(\theta)\sum_{\theta}\theta l_{\theta,t+1}\Theta_{t+1}(\theta)\right] \\
& + \lambda^{K}_{\theta,\theta',t}\left(1-\tau^{K}_{t}\right)n^{y}_{t}\left(\frac{E_t\left[\gamma_{t+1}F_{KK,t+1}u^{o}_{c,\theta,t+1}\right]}{E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^{o}_{c,\theta,t+1}\right]}\Theta_t(\theta) - \frac{E_t\left[\gamma_{t+1}F_{KK,t+1}u^{o}_{c,\theta',t+1}\right]}{E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^{o}_{c,\theta',t+1}\right]}\Theta_t(\theta)\right) \\
& - E_t\left[\lambda^{L}_{\theta,\theta',t+1}n^{y}_{t}\kappa^{ss}_{t+1}\frac{F_{\tilde{L}K,t+1}}{F_{\tilde{L},t+1}}\left(\frac{E_{t+1}\left[u^{o}_{c,\theta,t+2}\right]}{u^{y}_{c,\theta,t+1}}\Theta_t(\theta) - \frac{E_{t+1}\left[u^{o}_{c,\theta',t+2}\right]}{u^{y}_{c,\theta',t+1}}\Theta_t(\theta)\right)\right. \\
& \left. -\lambda^{L}_{\theta,\theta',t+1}\left(1-\tau^{L}_{t}\right)\frac{F_{\tilde{L}K,t+1}}{F_{\tilde{L},t+1}}n^{y}_{t}\left(\Theta_t(\theta) - \Theta_t(\theta)\right)\right]
\end{aligned}
\tag{A.11}
$$

Rearranging some terms and using (A.1b) we get:

$$
\begin{aligned}
\tau_t^K = & \frac{E_t\left[\left(\lambda^f_{t+1}R_{t+1}-\lambda^f_t\right)\right]n^y_t\Theta_t\left(\theta\right)}{\mathcal{W}_{\theta,t}\beta E_t\left[\tilde{\lambda}^o_{\theta,t+1}u^o_{c,\theta,t+1}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)\right]}+\frac{E_t\left[\beta\tilde{\lambda}^o_{\theta,t+1}u^o_{c,\theta,t+1}R_{t+1}\right]-\tilde{\lambda}^y_{\theta,t}u^y_{c,\theta,t}}{\beta E_t\left[\tilde{\lambda}^o_{\theta,t+1}u^o_{c,\theta,t+1}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)\right]} \\
& +\frac{\beta n^y_t\Theta_t\left(\theta\right)E_t\left[-\left(1-\tau_t^K\right)q^y_{\theta,t}\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta,t+1}\right]}{\beta E_t\left[\tilde{\lambda}^o_{\theta,t+1}u^o_{c,\theta,t+1}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)\right]} \\
& \left(\frac{Cov_t\left[\tilde{\lambda}^o_{\theta,t+1},q^y_{\theta,t}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^o_{c,\theta,t+1}\right]}{E_t\left[q^y_{\theta,t}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^o_{c,\theta,t+1}\right]}-\frac{Cov_t\left[\tilde{\lambda}^o_{\theta,t+1},q^y_{\theta,t}\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta,t+1}\right]}{E_t\left[q^y_{\theta,t}\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta,t+1}\right]}\right) \\
& +\frac{E_t\left[\left(\lambda^f_{t+2}(1-\delta)-\lambda^f_{t+1}\right)n^y_{t+1}\tau^{ss,F}_{t+1}\gamma_{t+1}F_{\tilde{L}K,t+1}n^y_t\Theta_t\left(\theta\right)\left(\sum_\theta\theta l_{\theta,t+1}\Theta_{t+1}(\theta)\right)\right]}{\mathcal{W}_{\theta,t}\beta E_t\left[\tilde{\lambda}^o_{\theta,t+1}u^o_{c,\theta,t+1}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)\right]} \\
& +\frac{E_t\left[\mathcal{W}_{\theta,t+1}\left(\tilde{\lambda}^o_{\theta,t+2}u^o_{c,\theta,t+2}-E_{t+1}\left[\tilde{\lambda}^o_{\theta,t+2}u^o_{c,\theta,t+2}\right]\right)\beta\kappa^{ss}_{t+1}\theta\gamma_{t+1}F_{\tilde{L}K,t+1}n^y_t\Theta_t\left(\theta\right)l_{\theta,t+1}\right]}{\mathcal{W}_{\theta,t}\beta E_t\left[\tilde{\lambda}^o_{\theta,t+1}u^o_{c,\theta,t+1}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)\right]} \\
& +\frac{\lambda^K_{\theta,\theta',t}n^y_t\Theta_t\left(\theta\right)}{\mathcal{W}_{\theta,t}\beta E_t\left[\tilde{\lambda}^o_{\theta,t+1}u^o_{c,\theta,t+1}\left(\gamma_{t+1}F_{K,t+1}-\delta\right)\right]}\frac{\left(1-\tau_t^K\right)}{r^b_t} \\
& \left(\frac{Cov_t\left[\gamma_{t+1}F_{KK,t+1},u^o_{c,\theta,t+1}\right]}{E_t\left[u^o_{c,\theta,t+1}\right]}-\frac{Cov_t\left[\gamma_{t+1}F_{KK,t+1},u^o_{c,\theta',t+1}\right]}{E_t\left[u^o_{c,\theta',t+1}\right]}\right)
\end{aligned}
\tag{A.12}
$$

where $\tilde{\lambda}^y_{\theta,t}=\frac{\lambda^y_{\theta,t}}{\mathcal{W}_{\theta,t}}$, $\tilde{\lambda}^o_{\theta,t+1}=\frac{\lambda^o_{\theta,t+1}}{\mathcal{W}_{\theta,t}u^o_{c,\theta,t+1}}$ and $\mathcal{W}_{\theta,t}=\frac{\tilde{\mathcal{W}}_{\theta,t}}{u^y_{c,\theta,t}}$. Now we can simplify this expression by using the first order condition of the Ramsey planner with respect to $\tau_t^{\text{ss,F}}$:

$$
\begin{aligned}
0 = & -E_t\left[\sum_\theta\left[\lambda^y_{\theta,t+1}\beta\psi_{t+2}\kappa^{ss}_{t+1}E_{t+1}\left[u^o_{c,\theta,t+2}\right]\theta\gamma_{t+1}F_{K\tilde{L},t+1}l_{\theta,t+1}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right]\right] \\
& + E_t\left[\beta\sum_\theta\left[\lambda^o_{\theta,t+1}\left(\frac{u^y_{c,\theta,t}-\beta\psi_{t+1}E_t\left[u^o_{c,\theta,t+1}\right]}{\beta\psi_{t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^o_{c,\theta,t+1}\right]}\right)\gamma_{t+1}F_{KK,t+1}q^y_{\theta,t}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right]\right] \\
& - E_t\left[\beta\sum_\theta\left[\lambda^o_{\theta,t+1}\left(\frac{u^y_{c,\theta,t}-\beta\psi_{t+1}E_t\left[u^o_{c,\theta,t+1}\right]}{\beta\psi_{t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^o_{c,\theta,t+1}\right]^2}\right)\right.\right. \\
& \left.\left. E_t\left[\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta,t+1}\right]\left(\gamma_{t+1}F_{K,t+1}-\delta\right)q^y_{\theta,t}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right]\right] \\
& + E_t\left[\beta\sum_\theta\left[\lambda^o_{\theta,t+2}\kappa^{ss}_{t+1}\theta\gamma_{t+1}F_{\tilde{L}K,t+1}l_{\theta,t+1}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right]\right] \\
& - \lambda^f_t n^y_t\gamma\left(s^t\right)F_{\tilde{L},t}\left(\sum_\theta \theta l_{\theta,t}\Theta_t(\theta)\right) \\
& - n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)E_t\left[\lambda^f_{t+1}n^y_{t+1}\tau^{ss,F}_{t+1}\gamma_{t+1}F_{\tilde{L}K,t+1}\left(\sum_\theta \theta l_{\theta,t+1}\Theta_{t+1}(\theta)\right)\right] \\
& + n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)E_t\left[\lambda^f_{t+1}\gamma_{t+1}F_{K,t+1}\right]+(1-\delta)n^y_t\gamma_t F_{\tilde{L},t}\left(\sum_\theta \theta l_{\theta,t}\Theta_t(\theta)\right)E_t\left[\lambda^f_{t+1}\right] \\
& + (1-\delta)n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)E_t\left[\lambda^f_{t+2}n^y_{t+1}\tau^{ss,F}_{t+1}\gamma_{t+1}F_{\tilde{L}K,t+1}\left(\sum_\theta \theta l_{\theta,t+1}\Theta_{t+1}(\theta)\right)\right] \\
& + \lambda^K_{\theta,\theta',t}\left[\frac{u^y_{c,\theta,t}-\beta\psi_{t+1}E_t\left[u^o_{c,\theta,t+1}\right]}{\beta\psi_{t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^o_{c,\theta,t+1}\right]^2}E_t\left[\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta,t+1}\right]n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right. \\
& \left. -\frac{u^y_{c,\theta',t}-\beta\psi_{t+1}E_t\left[u^o_{c,\theta',t+1}\right]}{\beta\psi_{t+1}E_t\left[\left(\gamma_{t+1}F_{K,t+1}-\delta\right)u^o_{c,\theta',t+1}\right]^2}E_t\left[\gamma_{t+1}F_{KK,t+1}u^o_{c,\theta',t+1}\right]n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right] \\
& - E_t\left[\lambda^L_{\theta,\theta',t+1}\frac{\kappa^{ss}_{t+1}\gamma_{t+1}\theta F_{\tilde{L}K,t+1}E_{t+1}\left[u^o_{c,\theta,t+2}\right]}{\gamma_{t+1}\theta F_{\tilde{L},t+1}u^y_{c,\theta,t+1}}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right. \\
& -\lambda^L_{\theta,\theta',t+1}\frac{u^y_{l,\theta,t}+\beta\psi_{t+2}\kappa^{ss}_{t+1}\gamma_{t+1}\theta F_{\tilde{L},t+1}E_{t+1}\left[u_{c,\theta,t+2}\right]}{\gamma_{t+1}\theta u^y_{c,\theta,t+1}F^2_{\tilde{L},t+1}}F_{\tilde{L}K,t+1}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right) \\
& -\lambda^L_{\theta,\theta',t+1}\frac{\kappa^{ss}_{t+1}\gamma_{t+1}\theta' F_{\tilde{L}K,t+1}E_{t+1}\left[u^o_{c,\theta',t+2}\right]}{\gamma_{t+1}\theta' F_{\tilde{L},t+1}u^y_{c,\theta',t+1}}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right) \\
& \left.+\lambda^L_{\theta,\theta',t+1}\frac{u^y_{l,\theta',t}+\beta\psi_{t+2}\kappa^{ss}_{t+1}\gamma_{t+1}\theta' F_{\tilde{L},t+1}E_{t+1}\left[u_{c,\theta',t+2}\right]}{\gamma_{t+1}\theta' u^y_{c,\theta',t+1}F^2_{\tilde{L},t+1}}F_{\tilde{L}K,t+1}n^y_t\left(\sum_\theta w_{\theta,t}l_{\theta,t}\Theta_t(\theta)\right)\right]
\end{aligned}
\tag{A.13}
$$

Using similar steps to the ones that let us obtain (A.12), we get:

$$\beta E_t \left[\sum_\theta \left(\lambda^o_{\theta,t+2} - E_{t+1} \left[\lambda^o_{\theta,t+2} \right] \right) \kappa^{ss}_{t+1} \gamma_{t+1} F_{\tilde{L}K,t+1} \theta l_{\theta,t+1} \right]$$
$$+ \left[\beta \left(1 - \tau^K_t \right) \sum_\theta q^y_{\theta,t} \mathcal{W}_{\theta,t} E_t \left[\gamma_{t+1} F_{KK,t+1} u^o_{c,\theta,t+1} \right] \right.$$
$$\left. \left(\frac{Cov\left(\tilde{\lambda}^o_{\theta,t+1}, u^o_{c,\theta,t+1} \gamma_{t+1} F_{KK,t+1} \right)}{E_t \left[\gamma_{t+1} F_{KK,t+1} u^o_{c,\theta,t+1} \right]} - \frac{Cov\left(\tilde{\lambda}^o_{\theta,t+1}, u^o_{c,\theta,t+1} \left(\gamma_{t+1} F_{K,t+1} - \delta \right) \right)}{E_t \left[\left(\gamma_{t+1} F_{K,t+1} - \delta \right) u^o_{c,\theta,t+1} \right]} \right) \right]$$
$$+ E_t \left[\left(\lambda^f_{t+2} (1-\delta) - \lambda^f_{t+1} \right) \tau^{ss,F}_{t+1} z_{t+1} \frac{F_{\tilde{L}K,t+1}}{F_{\tilde{L},t+1}} \right] + E_t \left[\lambda^f_{t+1} \left(\gamma_{t+1} F_{K,t+1} + (1-\delta) \right) - \lambda^f_t \right]$$
$$+ \lambda^K_{\theta,\theta',t} \left[\frac{(1-\tau^K_t)}{r^b_t} \left(\frac{Cov_t \left[\gamma_{t+1} F_{KK,t+1}, u^o_{c,\theta,t+1} \right]}{E_t \left[u^o_{c,\theta,t+1} \right]} - \frac{Cov_t \left[\gamma_{t+1} F_{KK,t+1}, u^o_{c,\theta',t+1} \right]}{E_t \left[u^o_{c,\theta',t+1} \right]} \right) \right] = 0 \tag{A.14}$$

Now, we can sum (A.12) across all θ and substituting in (A.14) we arrive to:

$$\tau^K_t \sum_\theta \frac{\mathcal{W}_{\theta,t} \beta E_t \left[\tilde{\lambda}^o_{\theta,t+1} u^o_{c,\theta,t+1} \left(\gamma_{t+1} F_{K,t+1} - \delta \right) \right]}{n^y_t \Theta_t(\theta)} = \left(N^\Theta - 1 \right) E_t \left[\left(\lambda^f_{t+1} R_{t+1} - \lambda^f_t \right) \right]$$
$$+ \sum_\theta \frac{\mathcal{W}_{\theta,t} \left[E_t \left[\beta \tilde{\lambda}^o_{\theta,t+1} u^o_{c,\theta,t+1} R_{t+1} \right] - \tilde{\lambda}^y_{\theta,t} u^y_{c,\theta,t} \right]}{n^y_t \Theta_t(\theta)}$$
$$+ \left(N^\Theta - 1 \right) E_t \left[\left(\lambda^f_{t+2} (1-\delta) - \lambda^f_{t+1} \right) n^y_{t+1} \tau^{ss,F}_{t+1} \gamma_{t+1} F_{\tilde{L}K,t+1} \left(\sum_\theta \theta l_{\theta,t+1} \Theta_{t+1}(\theta) \right) \right]$$
$$+ \lambda^K_{\theta,\theta',t} \left(N^\Theta - 1 \right) \frac{(1-\tau^K_t)}{r^b_t} \left(\frac{Cov_t \left[\gamma_{t+1} F_{KK,t+1}, u^o_{c,\theta,t+1} \right]}{E_t \left[u^o_{c,\theta,t+1} \right]} - \frac{Cov_t \left[\gamma_{t+1} F_{KK,t+1}, u^o_{c,\theta',t+1} \right]}{E_t \left[u^o_{c,\theta',t+1} \right]} \right) \tag{A.15}$$

□

Bibliography

Abowd, John M, Francis Kramarz, and David N Margolis (1999) "High wage workers and high wage firms," *Econometrica*, 67 (2), 251–333. (Cited on pages 9, 10, and 36).

Abowd, John M, Kevin L McKinney, and Nellie L Zhao (2018) "Earnings inequality and mobility trends in the United States: Nationally representative estimates from longitudinally linked employer-employee data," *Journal of Labor Economics*, 36 (S1), S183–S300. (Cited on page 10).

Acabbi, Edoardo Maria and Andrea Alati (2021) "Defusing Leverage: Liquidity Management and Labor Contracts," *SSRN Electronic Journal*, 10.2139/ssrn.3768825. (Cited on pages 63, 66, and 75).

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer (2015) "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83 (6), 2411–2451, 10.3982/ECTA13408. (Cited on pages 60, 72, 80, 82, and 84).

Aiyagari, S. Rao (1995) "Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting," *Journal of political Economy*, 103 (6), 1158–1175. (Cited on page 117).

Aiyagari, S. Rao, Albert Marcet, Thomas J. Sargent, and Juha Seppälä (2002) "Optimal taxation without state-contingent debt," *Journal of Political Economy*, 110 (6), 1220–1254. (Cited on page 117).

Andersson, Fredrik, Elizabeth E Davis, Matthew L Freedman, Julia I Lane, Brian P McCall, and Kristin Sandusky (2012) "Decomposing the sources of earnings inequality: Assessing the role of reallocation," *Industrial Relations: A Journal of Economy and Society*, 51 (4), 779–810. (Cited on page 10).

Arellano-Bover, Jaime (2020) "Career consequences of firm heterogeneity for young workers: First job and firm size." (Cited on page 10).

Arellano-Bover, Jaime and Fernando Saltiel (2021) "Differences in on-the-job learning across firms." (Cited on pages 5 and 10).

Atkinson, Anthony B. and Agnar Sandmo (1980) "Welfare implications of the taxation of savings," *The Economic Journal*, 90 (359), 529–549. (Cited on pages 117 and 118).

Autor, David H., William R. Kerr, and Adriana D. Kugler (2007) "Does Employment Protection Reduce Productivity? Evidence from US States," *The Economic Journal*, 117 (521), F189–F217, 10.1111/j.1468-0297.2007.02055.x. (Cited on pages 59 and 63).

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin (2014) "Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics," *American Economic Review*, 104 (6), 1551–96. (Cited on pages 10, 11, and 37).

Barro, Robert J. (1979) "On the Determination of the Public Debt," *Journal of Political Economy*, 87 (5, Part 1), 940–971. (Cited on page 129).

Bassanini, Andrea, Luca Nunziata, and Danielle Venn (2009) "Job Protection Legislation and Productivity Growth in OECD Countries," *Economic Policy*, 24 (58), 349–402, 10.1111/j.1468-0327.2009.00221.x. (Cited on page 63).

Belot, Michèle, Jan Boone, and Jan Van Ours (2007) "Welfare-Improving Employment Protection," *Economica*, 74 (295), 381–396, 10.1111/j.1468-0335.2006.00576.x. (Cited on page 65).

Bentolila, Samuel and Giuseppe Bertola (1990) "Firing Costs and Labour Demand: How Bad Is Eurosclerosis?" *The Review of Economic Studies*, 57 (3), 381, 10.2307/2298020. (Cited on page 64).

Bertola, Giuseppe (1990) "Job Security, Employment and Wages," *European Economic Review*, 34 (4), 851–879, 10.1016/0014-2921(90)90066-8. (Cited on page 64).

Bertola, Giuseppe and Richard Rogerson (1997) "Institutions and Labor Reallocation," *European Economic Review*, 41 (6), 1147–1171, 10.1016/S0014-2921(96)00048-7. (Cited on page 64).

Blanchard, Olivier and Augustin Landier (2002) "The Perverse Effects of Partial Labour Market Reform: Fixed-Term Contracts in France," *The Economic Journal*, 112 (480), F214–F244, 10.1111/1468-0297.00047. (Cited on page 64).

Boeri, Tito, Pierre Cahuc, and André Zylberberg (2015) "The Costs of Flexibility-Enhancing Structural Reforms: A Literature Review," OECD Economics Department Working Papers 1264, 10.1787/5jrs558c5r5f-en. (Cited on page 64).

Bonenkamp, Jan, Lex Meijdam, Eduard Ponds, and Ed Westerhout (2017) "Ageing-driven pension reforms," *Journal of Population Economics*, 30 (3), 953–976. (Cited on page 117).

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2019) "A distributional framework for matched employer employee data," *Econometrica*, 87 (3), 699–739. (Cited on pages 10, 31, and 36).

Borgatti, Stephen P (2005) "Centrality and network flow," *Social networks*, 27 (1), 55–71. (Cited on page 28).

Bustos, Paula (2011) "Trade Liberalization, Exports, and Technology Upgrading: Evidence on the Impact of MERCOSUR on Argentinian Firms," *The American Economic Review*, 101 (1), 304–340. (Cited on page 62).

Cahuc, Pierre, Olivier Charlot, and Franck Malherbet (2016) "Explaining the Spread of Temporary Jobs and Its Impact on Labor Turnover," *International Economic Review*, 57 (2), 533–572, 10.1111/iere.12167. (Cited on page 64).

Cahuc, Pierre and Fabien Postel-Vinay (2002) "Temporary Jobs, Employment Protection and Labor Market Performance," *Labour Economics*, 9 (1), 63–91, 10.1016/S0927-5371(01)00051-3. (Cited on page 64).

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin (2006) "Wage bargaining with on-the-job search: Theory and evidence," *Econometrica*, 74 (2), 323–364. (Cited on pages 11, 13, and 21).

Cai, Xiaoming, Pieter Gautier, and Ronald P Wolthoff (2021) "Search, screening and sorting." (Cited on page 11).

Caldwell, Sydnee and Nikolaj Harmon (2019) "Outside options, bargaining, and wages: Evidence from coworker networks," *Unpublished manuscript, Univ. Copenhagen*. (Cited on page 7).

Callaway, Brantly and Tong Li (2019) "Quantile Treatment Effects in Difference in Differences Models with Panel Data," *Quantitative Economics*, 10 (4), 1579–1618, 10.3982/QE935. (Cited on page 82).

Callaway, Brantly and Pedro H. C. Sant'Anna (2021) "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225 (2), 200–230, 10.1016/j.jeconom.2020.12.001. (Cited on page 76).

Caplin, Andrew, Victoria Gregory, Soren Leth-Petersen, Johan Saeverud, Chris Tonetti, and Gianluca Violante (2022) "Accounting for Job-to-Job Moves: Wages versus Values." (Cited on pages 10 and 26).

Cappellari, Lorenzo, Carlo Dell'Aringa, and Marco Leonardi (2012) "Temporary Employment, Job Flows and Productivity: A Tale of Two Reforms*," *The Economic Journal*, 122 (562), F188–F215, 10.1111/j.1468-0297.2012.02535.x. (Cited on pages 63 and 78).

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018) "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36 (S1), S13–S70. (Cited on pages 5 and 10).

Card, David, Jörg Heining, and Patrick Kline (2013) "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly Journal of Economics*, 128 (3), 967–1015. (Cited on page 10).

Chamley, Christophe (1986) "Optimal taxation of capital income in general equilibrium with infinite lives," *Econometrica: Journal of the Econometric Society*, 607–622. (Cited on page 117).

Chancel, Lucas, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman (2021) "World inequality report 2022." (Cited on page 113).

Chaney, Thomas (2016) "Liquidity Constrained Exporters," *Journal of Economic Dynamics and Control*, 72, 141–154, 10.1016/j.jedc.2016.03.010. (Cited on page 87).

Chari, V. V., Lawrence J. Christiano, and Patrick J. Kehoe (1994) "Optimal Fiscal Policy in a Business Cycle Model," *Journal of Political Economy*, 102 (4), 617–652, 10.1086/261949. (Cited on pages 127 and 142).

Cingano, Federico, Marco Leonardi, Julián Messina, and Giovanni Pica (2016) "Employment Protection Legislation, Capital Investment and Access to Credit: Evidence from Italy," *The Economic Journal*, 126 (595), 1798–1822, 10.1111/ecoj.12212. (Cited on page 63).

Ciurila, Nicoleta (2017) "How should pension benefits be linked to pension contribution," *Mimeo*. (Cited on page 117).

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger (2009) "Taxing capital? Not a bad idea after all!," *American Economic Review*, 99 (1), 25–48. (Cited on pages 117 and 142).

Daruich, Diego, Sabrina Di Addario, and Raffaele Saggio (2023) "The Effects of Partial Employment Protection Reforms: Evidence from Italy," *The Review of Economic Studies*, rdad012, 10.1093/restud/rdad012. (Cited on pages 59, 60, 63, 66, 71, 75, 77, 78, and 85).

De La Roca, Jorge and Diego Puga (2017) "Learning by working in big cities," *The Review of Economic Studies*, 84 (1), 106–142. (Cited on page 10).

Devicienti, Francesco and Bernardo Fanfani (2021) "Firms' Margins of Adjustment to Wage Growth: The Case of Italian Collective Bargaining," *SSRN Electronic Journal*, 10.2139/ssrn.3888551. (Cited on page 79).

Dhingra, Swati and John Morrow (2019) "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," *Journal of Political Economy*, 127 (1), 196–232, 10.1086/700732. (Cited on page 91).

Di Addario, Sabrina L, Patrick M Kline, Raffaele Saggio, and Mikkel Sølvsten (2021) "It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages.,"Technical report, National Bureau of Economic Research. (Cited on page 10).

Dieppe, Alistair (2021) *Global productivity: Trends, drivers, and policies*: World Bank Publications. (Cited on page 113).

Dixit, Avinash K. and Joseph E. Stiglitz (1977) "Monopolistic Competition and Optimum Product Diversity," *The American Economic Review*, 67 (3), 297–308. (Cited on pages 62 and 86).

Dolado, Juan J., Salvador Ortigueira, and Rodolfo Stucchi (2016) "Does Dual Employment Protection Affect TFP? Evidence from Spanish Manufacturing Firms," *SERIEs*, 7 (4), 421–459, 10.1007/s13209-016-0150-9. (Cited on pages 63 and 64).

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp Vom Berge (2021) "Reallocation Effects of the Minimum Wage," *The Quarterly Journal of Economics*, 137 (1), 267–328, 10.1093/qje/qjab028. (Cited on page 59).

Engbom, Niklas (2020) “Labor market fluidity and human capital accumulation,” *New York University mimeo*. (Cited on page 38).

Engellandt, Axel and Regina T. Riphahn (2005) “Temporary Contracts and Employee Effort,” *Labour Economics*, 12 (3), 281–299, 10.1016/j.labeco.2003.11.006. (Cited on page 64).

Erosa, Andrés, Luisa Fuster, and Gueorgui Kambourov (2012) “Labor supply and government programs: A cross-country analysis,” *Journal of Monetary Economics*, 59 (1), 84–107. (Cited on page 141).

Erosa, Andrés and Martin Gervais (2002) “Optimal taxation in life-cycle economies,” *Journal of Economic Theory*, 105 (2), 338–369. (Cited on page 117).

Faccini, Renato (2014) “Reassessing Labour Market Reforms: Temporary Contracts as a Screening Device,” *The Economic Journal*, 124 (575), 167–200, 10.1111/ecoj.12072. (Cited on page 64).

Fan, Yanqin and Zhengfei Yu (2012) “Partial Identification of Distributional and Quantile Treatment Effects in Difference-in-Differences Models,” *Economics Letters*, 115 (3), 511–515, 10.1016/j.econlet.2012.01.001. (Cited on page 82).

Fanfani, Bernardo (2022) “The Employment Effects of Collective Bargaining,” *LABORatorio R. Revelli Working Paper no. 184*. (Cited on page 66).

Farhi, Emmanuel (2010) “Capital taxation and ownership when markets are incomplete,” *Journal of Political Economy*, 118 (5), 908–948. (Cited on pages 114, 117, 119, 129, 132, 135, and 142).

Farhi, Emmanuel and Iván Werning (2010) “Progressive estate taxation,” *The Quarterly Journal of Economics*, 125 (2), 635–673. (Cited on page 119).

Freeman, Linton C (1978) “Centrality in social networks conceptual clarification,” *Social networks*, 1 (3), 215–239. (Cited on page 27).

Gandhi, Amit, Salvador Navarro, and David A. Rivers (2020) “On the Identification of Gross Output Production Functions,” *Journal of Political Economy*, 128 (8), 2973–3016, 10.1086/707736. (Cited on pages 60, 72, and 80).

Garibaldi, Pietro and Giovanni L. Violante (2005) “The Employment Effects of Severance Payments with Wage Rigidities*,” *The Economic Journal*, 115 (506), 799–832, 10.1111/j.1468-0297.2005.01020.x. (Cited on page 64).

Garnero, Andrea (2018) "The Dog That Barks Doesn't Bite: Coverage and Compliance of Sectoral Minimum Wages in Italy," *IZA Journal of Labor Policy*, 7 (1), 3, 10.1186/s40173-018-0096-6. (Cited on page 68).

Garriga, Carlos (2019) "Optimal fiscal policy in overlapping generations models," *Public Finance Review*, 47 (1), 3–31. (Cited on page 117).

Gnocato, Nicolò, Francesca Modena, and Chiara Tomasi (2020) "Labor Market Reforms and Allocative Efficiency in Italy," *Labour Economics*, 67, 101938, 10.1016/j.labeco.2020.101938. (Cited on page 64).

Gourieroux, Christian, Alain Monfort, and Eric Renault (1993) "Indirect inference," *Journal of applied econometrics*, 8 (S1), S85–S118. (Cited on pages 8 and 35).

Gregory, Victoria (2020) "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search,"Technical report, Federal Reserve Bank of St. Louis. (Cited on pages 5, 10, 11, 20, 36, 37, and 38).

Hall, Robert E and Andreas I Mueller (2018) "Wage dispersion and search behavior: The importance of nonwage job values," *Journal of Political Economy*, 126 (4), 1594–1637. (Cited on page 10).

Harenberg, Daniel and Alexander Ludwig (2019) "Idiosyncratic risk, aggregate risk, and the welfare effects of social security," *International Economic Review*, 60 (2), 661–692. (Cited on page 117).

Haucap, Justus and Christian Wey (2004) "Unionisation Structures and Innovation Incentives," *The Economic Journal*, 114 (494), C149–C165. (Cited on page 59).

Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M Phillips (2018) "Production and learning in teams,"Technical report, National Bureau of Economic Research. (Cited on page 38).

Hipsman, Nathaniel E (2018) "Optimal Taxation in Overlapping Generations Economies with Aggregate Risk," *Mimeo*. (Cited on pages 114, 117, 130, and 142).

Hopenhayn, Hugo and Richard Rogerson (1993) "Job Turnover and Policy Evaluation: A General Equilibrium Analysis," *Journal of Political Economy*, 101 (5), 915–938. (Cited on page 64).

Hornstein, Andreas, Per Krusell, and Giovanni L Violante (2011) "Frictional wage dispersion in search models: A quantitative assessment," *American Economic Review*, 101 (7), 2873–98. (Cited on page 5).

Hosseini, Roozbeh and Ali Shourideh (2019) "Retirement financing: An optimal reform approach," *Econometrica*, 87 (4), 1205–1265. (Cited on pages 117, 140, and 142).

Hsieh, Chang-Tai and Peter J. Klenow (2009) "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124 (4), 1403–1448. (Cited on page 64).

Huitfeldt, Ingrid, Andreas Kostol, Jan Sebastian Nimczik, and Andrea Weber (2021) "Internal labor markets: A worker flow approach." (Cited on page 10).

Ichino, Andrea, Fabrizia Mealli, and Tommaso Nannicini (2008) "From Temporary Help Jobs to Permanent Employment: What Can We Learn from Matching Estimators and Their Sensitivity?" *Journal of Applied Econometrics*, 23 (3), 305–327, 10.1002/jae.998. (Cited on page 64).

Ichino, Andrea and Regina T. Riphahn (2005) "The Effect of Employment Protection on Worker Effort: Absenteeism during and after Probation," *Journal of the European Economic Association*, 3 (1), 120–143. (Cited on page 64).

İmrohoroğlu, Ayşe, Selahattin İmrohoroğlu, and Douglas H. Joines (1998) "The effect of tax-favored retirement accounts on capital accumulation," *American Economic Review*, 749–768. (Cited on page 117).

Jäger, Simon, Shakked Noy, and Benjamin Schoefer (2022) "The German Model of Industrial Relations: Balancing Flexibility and Collective Action," *The Journal of Economic Perspectives*, 36 (4), 53–80. (Cited on page 59).

Jarosch, Gregor (2021) "Searching for job security and the consequences of job loss,"Technical report, National Bureau of Economic Research. (Cited on pages 5, 10, 11, 14, 20, 21, 35, 37, and 52).

Judd, Kenneth L. (1985) "Redistributive taxation in a simple perfect foresight model," *Journal of Public Economics*, 28 (1), 59–83. (Cited on page 117).

Klemm, Alexander D, Shafik Hebous, and Christophe J Waerzeggers (2021) "Capital Income Taxation in the Netherlands," *IMF Working Papers*, 2021 (145). (Cited on pages 140 and 142).

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten (2020) "Leave-out estimation of variance components," *Econometrica*, 88 (5), 1859–1898. (Cited on page 36).

Kotlikoff, Laurence J. (2009) "The Intergenerational Conflict – Après le Deluge," in *Growth, Demography, Finance: From Economic Breakdowns to New Balances, Actes des Rencontres Economiques d'Aix-en-Provence*: Le Cercle des Economistes. (Cited on page 113).

Krueger, Dirk and Felix Kubler (2006) "Pareto-improving social security reform when financial markets are incomplete!?" *American Economic Review*, 96 (3), 737–755. (Cited on page 117).

Kugler, Adriana and Giovanni Pica (2008) "Effects of Employment Protection on Worker and Job Flows: Evidence from the 1990 Italian Reform," *Labour Economics*, 15 (1), 78–95, 10.1016/j.labeco.2006.11.002. (Cited on page 63).

Lagos, Ricardo (2006) "A Model of TFP," *The Review of Economic Studies*, 73 (4), 983–1007. (Cited on pages 59 and 64).

Lazear, Edward P. (1990) "Job Security Provisions and Employment," *The Quarterly Journal of Economics*, 105 (3), 699–726, 10.2307/2937895. (Cited on pages 59 and 64).

Levinsohn, James and Amil Petrin (2003) "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70 (2), 317–341. (Cited on pages 60, 72, and 80).

Ljungqvist, Lars (2002) "How Do Lay-off Costs Affect Employment?" *The Economic Journal*, 112 (482), 829–853. (Cited on page 64).

Lucas, Jr., Robert E. and Nancy L. Stokey (1983) "Optimal fiscal and monetary policy in an economy without capital," *Journal of Monetary Economics*, 12 (1), 55–93. (Cited on pages 117 and 125).

Makles, Anna (2012) "Stata tip 110: How to get the optimal k-means cluster solution," *The Stata Journal*, 12 (2), 347–351. (Cited on page 48).

Manova, Kalina (2013) "Credit Constraints, Heterogeneous Firms, and International Trade," *The Review of Economic Studies*, 80 (2), 711–744, 10.1093/restud/rds036. (Cited on pages 62 and 87).

McFadden, Daniel (1989) "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 995–1026. (Cited on pages 8 and 35).

Melitz, Marc J. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71 (6), 1695–1725. (Cited on pages 62 and 86).

Nimczik, Jan Sebastian (2020) "Job mobility networks and data-driven labor markets,"Technical report, Working Paper. (Cited on page 10).

Nunn, Ryan (2012) "Match quality with unpriced amenities,"Technical report, Working paper. (Cited on page 10).

OECD (2019) *Pensions at a Glance 2019*: OECD Publishing, Paris, 224, https://doi.org/https://doi.org/10.1787/b6d3dcfc-en. (Cited on page 113).

——— (2021a) *Government at a Glance 2021*: OECD Publishing, Paris, 280, https://doi.org/https://doi.org/10.1787/1c258f55-en. (Cited on page 142).

——— (2021b) *Pensions at a Glance 2021*: OECD Publishing, Paris, 224, https://doi.org/https://doi.org/10.1787/ca401ebd-en. (Cited on page 141).

——— (2021c) *Taxing Wages 2021*: OECD Publishing, Paris, 651, https://doi.org/https://doi.org/10.1787/83a87978-en. (Cited on pages 141 and 142).

Opsahl, Tore, Filip Agneessens, and John Skvoretz (2010) "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social networks*, 32 (3), 245–251. (Cited on page 48).

Piketty, Thomas and Emmanuel Saez (2013) "A theory of optimal inheritance taxation," *Econometrica*, 81 (5), 1851–1886. (Cited on page 119).

Poschke, Markus (2009) "Employment Protection, Firm Selection, and Growth," *Journal of Monetary Economics*, 56 (8), 1074–1085, 10.1016/j.jmoneco.2009.10.013. (Cited on page 65).

Postel-Vinay, Fabien and Jean-Marc Robin (2002) "Equilibrium wage dispersion with worker and employer heterogeneity," *Econometrica*, 70 (6), 2295–2350. (Cited on pages 11 and 13).

Ramsey, Frank P. (1927) "A Contribution to the Theory of Taxation," *The Economic Journal*, 37 (145), 47–61. (Cited on page 113).

RUBOLINO, ENRICO (2022) "Taxing the Gender Gap: Labor Market Effects of a Payroll Tax Cut for Women in Italy." (Cited on page 26).

SAITTO, ROBERTO (2020) "Optimal taxation in an overlapping generations economy," *Mimeo*. (Cited on pages 117 and 148).

SKLAR, M. (1959) *Fonctions de Répartition À N Dimensions Et Leurs Marges*: Université Paris 8. (Cited on page 82).

SONG, JAE, DAVID J PRICE, FATIH GUVENEN, NICHOLAS BLOOM, AND TILL VON WACHTER (2019) "Firming up inequality," *The Quarterly Journal of Economics*, 134 (1), 1–50. (Cited on page 10).

SØRENSEN, PETER BIRCH AND PETER NORMAN SØRENSEN (2004) *Measuring the tax burden on capital and labor*, 6: MIT Press. (Cited on page 140).

SORKIN, ISAAC (2018) "Ranking firms using revealed preference," *The Quarterly Journal of Economics*, 133 (3), 1331–1393. (Cited on pages 5 and 10).

STEINLEY, DOUGLAS (2006) "K-means clustering: a half-century synthesis," *British Journal of Mathematical and Statistical Psychology*, 59 (1), 1–34. (Cited on page 31).

SULLIVAN, PAUL AND TED TO (2014) "Search and nonwage job characteristics," *Journal of Human Resources*, 49 (2), 472–507. (Cited on page 10).

TABER, CHRISTOPHER AND RUNE VEJLIN (2020) "Estimation of a Roy/search/compensating differential model of the labor market," *Econometrica*, 88 (3), 1031–1069. (Cited on page 10).

UNGER, FLORIAN (2021) "Credit Frictions, Selection into External Finance and Gains from Trade," *Canadian Journal of Economics/Revue canadienne d'économique*, 54 (3), 1206–1251, 10.1111/caje.12529. (Cited on page 88).

WANG, SEAN Y. (2021) "What Is the Price for Opportunity? The Effects of Employer Learning on Worker Promotions and Turnover." (Cited on page 10).

WASMER, ETIENNE (2006) "General versus Specific Skills in Labor Markets with Search Frictions and Firing Costs," *American Economic Review*, 96 (3), 811–831, 10.1257/aer.96.3.811. (Cited on page 64).

WERNING, IVAN (2007) "Optimal fiscal policy with redistribution," *The Quarterly Journal of Economics*, 122 (3), 925–967. (Cited on page 117).

Zhelobodko, Evgeny, Sergey Kokovin, Mathieu Parenti, and Jacques-François Thisse (2012) "Monopolistic Competition: Beyond the Constant Elasticity of Substitution," *Econometrica*, 80 (6), 2765–2784, 10.3982/ECTA9986. (Cited on page 62).