The Concentration Channel of the Minimum Wage*

Salvatore Lo Bello — Bank of Italy

Lorenzo Pesaresi — University of Zurich

November 4, 2024

Latest version here

Abstract

We study the equilibrium effects of the minimum wage when product market power is endogenous and varies with market competition. A higher minimum wage reallocates workers from small to large firms. Large firms gain market share and increase their price markups. We call this mechanism *concentration channel* of the minimum wage. We contribute a model with frictional labor markets, to allow for worker reallocation, and oligopolistic product markets, to account for the response of price markups. We estimate the model on Italian social security data, replicating the structure of detailed labor and product markets. Our counterfactuals suggest that both the aggregate labor share and value added are hump-shaped in the minimum wage, due to the opposing responses of price markups and wage markdowns. The optimal minimum wage, which trades off reallocation gains against employment losses, equals 70% of the current median wage. If product market power were exogenous, both the optimal minimum wage and the associated welfare gains would be higher. We provide novel empirical evidence from Italian firms supporting the concentration channel.

JEL Codes: D43, E24, E25, E64, J31, J38, J42, J64.

Keywords: Minimum Wage, Wage Distribution, Worker and Firm Heterogeneity, Equilibrium Search Model, Reallocation, Oligopolistic Competition, Endogenous Markups.

*This paper was previously circulated with the title "Equilibrium Effects of the Minimum Wage: The Role of Product Market Power". The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Bank of Italy. Pesaresi is grateful to Alessandro Ferrari, Nir Jaimovich and Gianluca Violante for their guidance and support. We thank Mark Aguiar, Manuel Amador, Pol Antrás, Gaetano Basso, Ariel Burstein, Sydnee Caldwell, Carlos Carrillo-Tudela, Lorenzo Casaburi, Emanuela Ciapanna, Diego Comin, Paola Conconi, Giancarlo Corsetti, Rafael Dix-Carneiro, Moritz Drechsel-Grau, Jan Eeckhout, Nicola Fuchs-Schunden, Xavier Gabaix, Axel Gottfries, Basile Grassi, David Hémous, Kyle Herkenhoff, Leo Kaas, Moritz Kuhn, Ivan Lagrosa, Philip Lane, Andrea Linarello, Attila Lindner, Alexander Ludwig, Rocco Macchiavello, Ioana Marinescu, Guido Menzio, Pascal Michaillat, Claudio Michelacci, Simon Mongey, Lukas Nord, Luigi Paciello, Alessandra Peter, Diego Puga, Roland Rathelot, Alfonso Rosolia, Aysegul Sahin, Thomas Sargent, Andreas Schaab, Florian Scheuer, Jan Stuhler, Roberto Torrini, Liangjie Wu and Josef Zweimüller, our discussants Manolis Galenianos, Roberto Leombruni, Gabriele Macci and Aseem Patel, as well as participants in conferences and seminars.

1 Introduction

In recent years, there has been a renewed interest in the minimum wage (MW). Its traditional goal of reducing in-work poverty has given way to more ambitious macroeconomic objectives, such as reducing income inequality and boosting the aggregate labor share (Stansbury and Summers, 2020). This shift in focus reflects in the current discussion of numerous policy proposals advocating for raising the MW to unprecedented levels. Several empirical and theoretical studies suggest that the MW is more effective when competition in the labor market is low. In such environments, displaced workers from less productive firms – those most affected by the MW – tend to reallocate to more productive firms. Hence, the employment response to the MW is muted, while the reallocation of workers enhances labor productivity (Dustmann et al., 2022). However, little is known about the role of product market power – that is, the ability of dominant firms to charge prices significantly above marginal costs – in shaping the effects of the MW.[1] This gap is surprising, as the reallocation of workers hinges precisely upon whether firms mostly adjust prices or employment in response to the MW.

In this paper we study the equilibrium effects of the MW when product market power is endogenous and varies with market competition. First, we build a stylized model with imperfect competition both on the labor and the product market. This model advances the existing MW literature by allowing firms to increase their product market power as the number of competitors decreases – echoing the very premise of antitrust laws. In this context, we show that raising the MW generally reduces labor market power of low-productivity firms, but increases product market power of high-productivity firms. For low-productivity firms, raising the MW makes labor supply more elastic, which reduces their wage markdown – an index of labor market power. On the other hand, raising the MW also induces selection on the firm side, forcing low-productivity firms to either downsize or exit the market altogether. As a result, workers reallocate towards high-productivity firms, which increase their market share. A higher market share allows high-productivity firms to raise their price markup – an index of product market power. We call this novel mechanism *concentration channel* of the MW.[2] Allowing for the concentration channel, we show that the effects of the MW on both aggregate labor share and output are qualitatively ambiguous and mirror the *net* market power response, i.e., the net effect of reduced wage markdowns and increased price markups on aggregate market power. This means that, if the concentration channel is strong enough, the MW could even *reduce* – rather than increase – the aggregate labor share. We provide a simple condition on the elasticity of firm-level price markups to the number of active firms such that both the aggregate labor share and output are hump-shaped in the MW.

To illustrate and quantify the concentration channel, we build a novel quantitative model

[1] Harasztosi and Lindner (2019) and Link (2024) document that firms directly affected by the MW are more likely to pass the cost increase onto consumer prices when product market competition is low.

[2] We see the concentration channel as a plausible explanation for the vocal support (and active lobbying) of big companies in the US for a drastic increase in the federal minimum wage. See https://www.aboutamazon.com/news/policy-news-views/its-time-to-raise-the-federal-minimum-wage for Amazon's official position.

where heterogeneous workers and firms interact in frictional labor markets and oligopolistic product markets (Engbom and Moser, 2022; Atkeson and Burstein, 2008). The economy is populated by workers, who differ in their skill and industry, and by firms, which differ in their productivity, skill requirement, and industry. Labor markets are segmented by worker skills and industries (e.g., low-skilled in manufacturing). Product markets are populated by a finite number of firms with the same skill requirements and industry (e.g., metalworking sector). Since labor markets are frictional, workers are partially locked in at their current workplace because they receive job offers only from time to time. Firms profit from this lack of outside options by marking down the wage of their workers relative to the marginal revenue product. Moreover, the allocation of workers across firms is generally inefficient since firms with different productivities compete for the same workers, thus opening the scope for an efficient reallocation. Since product markets are oligopolistic, large firms leverage their dominant position to mark up their price relative to the marginal cost. Hence, both price markups and wage markdowns are endogenous and react to changes in the competitive environment induced by policy reforms, such as the MW.

We estimate the model using administrative data from the Italian Social Security and Institute of Statistics. Upon running a two-way fixed effect regression á la Abowd et al. (1999) on matched employer-employee data, we distinguish two worker skill types (high- and low-skilled) based on the median worker fixed effect. Labor markets are defined by the skills of workers and the 1-digit industry codes of the firms they work for. Product markets are defined by 4-digit industry codes. We address the vast cross-sectional heterogeneity in observed market structures by replicating the observed wage distribution in each labor market and the number of competing firms in each product market. We estimate the key parameters governing the elasticities of demand and labor supply via the Simulated Method of Moments (SMM), by targeting the degree of concentration of 4-digit product markets and the average correlation between labor share and market share in each such markets. Within the SMM routine, we use the structure of our model to estimate the firms' productivity distribution that rationalizes the observed wage distributions. As a result, the estimated model replicates the wage distributions for each industry and worker type, which is key to replicate the actual impact of MW reforms and the scope for worker reallocation. Since the Italian economy features a few large firms and one of the lowest levels of product market concentration in the Western world – the average number of competing firms in a 4-digit product market is around $3,000$ –, we argue that our setting represents a lower bound for the effects of endogenous markups in international comparisons.

We make use of the estimated model to simulate the effects of MW reforms in the Italian economy. According to our results, the aggregate markdown decreases with the MW, driven by low-productivity firms compressing profit margins. On the contrary, the aggregate markup increases with the MW, driven by the concentration channel. The rise in aggregate markup is relatively modest (up to 1pp), occurs primarily for high MWs, and is driven by the manufacturing industry. Still, for MWs higher than 90% of the current median wage, the aggregate markup response leads to an increase in net market power with the MW. Mirroring the net market power response, both the aggregate labor share and value added are hump-shaped in the MW with a peak at 92% of the current median wage (40% of workers directly affected). Such a consumption-maximizing MW reform raises aggregate value added by 3% and the aggregate labor share by 2pp. The increase in aggregate value added is driven by productivity gains from worker reallocation (due to lower misallocation). However, the reallocation process produces some excess unemployment. Specifically, an average of one in five jobs paying less than the MW gets destroyed by the policy. It follows that the MW has the potential to raise aggregate value added, but at the cost of reducing employment.

To strike a balance between these opposing forces, we resort to a utilitarian social welfare function as a normative criterion. Absent asset markets, the optimal MW equals 70% of the current median wage (10% of workers directly affected), with associated welfare gains of 0.5% in consumption equivalent units. The welfare effects of the MW are heterogeneous across worker skill types, which generates winners and losers. In particular, welfare of low-skilled workers increases with the MW up to 65% of the current median wage. As the MW increases further, low-skilled workers bear most of the increase in aggregate unemployment, whereas the high-skilled enjoy wage gains with milder employment effects. Consequently, welfare inequality decreases with the MW up to 65% of the current median wage and increases thereafter. Hence, our results express a word of caution about the idea of raising the MW to unprecedented levels to reduce (welfare) inequality.

Then, we zoom into the role of endogenous markups in shaping the equilibrium and welfare effects of the MW. To do so, we replicate the same counterfactual experiments in an observationally equivalent economy with monopolistic competition in product markets, which features exogenous markups. This counterfactual exercise suggests that ignoring endogenous markups would result in a monotonically increasing labor share, as well as to higher consumption-maximizing and optimal MWs (from 92 to 100%, and from 70 to 75% of the current median wage, respectively). Welfare gains from implementing the optimal MW would be nearly 50% larger (0.8% rather than 0.5% in consumption equivalent units). Therefore, even in the context of the low concentrated Italian economy, endogenous product market power is an important determinant of the equilibrium and welfare effects of the MW.

Finally, we provide empirical validation of the concentration channel on Italian firms' balance sheet data, leveraging the variation induced by industry-specific contractual wage floors. We document two sets of reduced-form evidence. First, higher wage floors induce reallocation of workers towards larger and more productive firms, which increases product market concentration and average labor productivity. Second, the firm-level labor share response decreases with the concentration of the (4-digit) product market in which the firm operates. As predicted by the concentration channel, this effect is driven by the positive response of profits for high-productivity firms in highly concentrated product markets.[3]

[3]In line with the existing literature, the concentration of local labor markets affects the labor share response *positively*. This reinforces our hypothesis that the negative effect of product market concentration on the firm-level labor share reflects an increase in price markups – as opposed to wage markdowns.

Related Literature. This paper is related to three main strands of literature. First, it contributes to the literature on structural modelling of the equilibrium effects of the MW. Hurst et al. (2023) and Drechsel-Grau (2023) propose search models to study the distributional impact of the MW across heterogeneous households, while Ahlfeldt et al. (2022) and Karabarbounis et al. (2022) focus on the spatial dimension and entry decisions, respectively. The closest papers to ours in this literature are Engbom and Moser (2022) and Berger et al. (2024). Engbom and Moser (2022) uses a quantitative wage-posting model to study the effect of the MW on earnings inequality. Our model extends Engbom and Moser (2022)'s by considering a generalized hiring cost function and adding product market power, as well as characterizing the optimal MW. Berger et al. (2024) studies the welfare effects of the MW in the oligopsonistic-competition model of Berger et al. (2022), enriched by worker heterogeneity in wealth and productivity. Unlike Berger et al. (2024)'s, our framework features inefficient (frictional) unemployment, endogenous hiring costs and endogenous product market power, which react to the introduction of the MW.

Second, this paper speaks to the macroeconomic literature on oligopolistic competition in product markets (Atkeson and Burstein, 2008; Edmond et al., 2015; De Loecker et al., 2021, 2020; Burstein et al., 2021; Edmond et al., 2023). The closest papers in this literature to ours are MacKenzie (2020), Deb et al. (2022, 2023) and Firooz (2023), which propose models of oligopolistic competition in sectoral product markets and oligopsonistic competition in local labor markets or frictional labor markets with wage bargaining.[4] Our model is the first to combine oligopolistic competition in product markets with wage posting in labor markets, and can be used to study spillover effects between product and labor markets induced by a variety of policies – not limited to the MW.

Finally, this paper contributes to the literature on wage-posting models, whose foundations are laid down by Burdett and Mortensen (1998), van den Berg and Ridder (1998), Bontemps et al. (1999) and Bontemps et al. (2000). Recent advances include Moscarini and Postel-Vinay (2013), Engbom and Moser (2022), Heise and Porzio (2022), Gouin-Bonenfant (2020), Bilal and Lhuillier (2021) and Gottfries and Jarosch (2023). Since Manning (2003), this class of models has been widely used as theoretical underpinning for estimating wage markdowns. We contribute to this literature by characterizing the wage-posting equilibrium with decreasing returns to scale in the revenues, endogenous job creation and a finite number of productivity types. Moreover, we show that, whenever hiring costs are size-dependent, the equilibrium markdown depends both on the elasticity of labor supply and a hiring cost correction term. By ignoring the latter, standard estimates of markdowns are generally upward-biased.

[4] Kroft et al. (2020), Tortarolo and Zarate (2020) and Yeh et al. (2022) propose theory-grounded empirical strategies to estimate markups and markdowns separately.

2 Stylized Model

In this section we develop a stylized model to (i) formalize the necessary conditions behind the concentration channel of the MW and (ii) highlight how the equilibrium effects of the MW depend on the net market power response, i.e., the net effect of the response of labor and product market power. The model is intentionally parsimonious and reduced-form. In the next section, we will incorporate these qualitative insights into a fully-fledged, microfounded quantitative model.

Baseline Equilibrium. We consider an economy populated by a finite number N of firms with heterogeneous productivity z. Firms' productivity follows some distribution $\Gamma(z)$ with support $[\underline{z}, \bar{z}]$. Let $\hat{\gamma}(z)$ be the probability mass function of realized productivity. Both the labor and the product market are imperfectly competitive. Firms choose their labor demand by solving the following profit maximization problem:

$$\max_{\ell} \quad p(y)y - w(\ell)\ell - \kappa \tag{1}$$

$$\text{s.t.} \quad y = z\ell, \tag{2}$$

$$w(\ell) = \ell^{\frac{1}{\eta}}, \tag{3}$$

$$p(y) = y^{-\frac{1}{\epsilon(N)}}. \tag{4}$$

Firms seek to maximize current profits (1), which equal the difference between revenues, py, and costs. In turn, costs consist of the wage bill, $w\ell$, and some overhead costs, κ.[5] Throughout we assume that the overhead costs are low enough that the lowest-productivity firms make positive profits in the baseline equilibrium. Firms operate the linear technology (2), with labor being its only input. Labor market power stems from the upward-sloping labor supply curve (3), with elasticity of labor supply $\eta \in (0, \infty)$. Since wages are increasing in employment, the marginal cost exceeds the wage. Formally, $MC(\ell) = w(\ell) + w'(\ell)\ell$. Similarly, product market power stems from the downward-sloping demand curve (4), with elasticity of demand $\epsilon \in (1, \infty)$. We let the elasticity of demand be an increasing function of the number of active firms, i.e., $d\epsilon/dN > 0$ (Jaimovich, 2007). It follows that firms have more product market power the fewer the number of competitors they face – echoing the very premise of antitrust laws.[6] Hence, we interpret N as an inverse index of market *concentration*. Since prices are decreasing in output, the marginal revenue product of labor falls short of its marginal product, $p(y)z$. Formally, $MRPL(y) = p(y)z + p'(y)zy$.

The solution to the profit maximization (1) consists of a double wedge between the optimal

[5] Overhead costs are needed to induce some exit of low-productivity firms upon introducing the MW, in line with the empirical evidence (Draca et al., 2011; Dustmann et al., 2022).

[6] Suppose that consumers place high value a certain good but are indifferent among its firm-specific varieties. If there were a large number of firms, the elasticity of demand faced by each of them would approach infinity. If there were just one firm producing the good, the elasticity of demand faced by the monopolist would be rather low. As the number of firms varies, the firm-specific elasticity of demand moves between these two extremes.

price and (productivity-adjusted) wage:

$$p(z) = \underbrace{\frac{\epsilon(N)}{\epsilon(N)-1}}_{\mu(N)} \underbrace{\frac{1+\eta}{\eta}}_{\psi} \frac{w(z)}{z}. \tag{5}$$

Equation (5) allows us to single out two different sources of market power. Let μ denote the price markup over marginal costs that firms optimally charge, i.e., $\mu(N) \equiv \frac{p(z)z}{MC(z)} = \frac{\epsilon(N)}{\epsilon(N)-1} \geq 1$. The price markup is an inverse function of the elasticity of demand $\epsilon(N)$, thus representing an index of product market power. Similarly, let ψ denote the wage markdown that firms optimally charge, i.e., $\psi \equiv \frac{MRPL(z)}{w(z)} = \frac{1+\eta}{\eta} \geq 1$. The wage markdown is an inverse function of the elasticity of labor supply η, thus representing an index of labor market power.[7] Henceforth, we will refer to the price markup and wage markdown simply as markup and markdown, respectively. The optimality condition (5) also implicitly defines the firm-specific labor share in the operative revenues:

$$LS(z) \equiv \frac{w(z)\ell(z)}{p(z)z\ell(z)} = \frac{1}{\mu(N)\psi}. \tag{6}$$

Notice that the labor share is constant across the productivity distribution, as well as inversely related to the firm's net market power, that is, the product between its markup and markdown. The intuition is simple: the higher the firm's market power on any market, the lower its labor share. Exactly as the labor share, the firm's output is decreasing in its net market power. Indeed, substituting for the labor supply and product demand constraints (3)-(4) into the optimality condition (5) yields:

$$y(z) = z^{\epsilon \frac{1+\eta}{\epsilon+\eta}} \left(\frac{1}{\mu(N)\psi} \right)^{\frac{\eta\epsilon}{\epsilon+\eta}}. \tag{7}$$

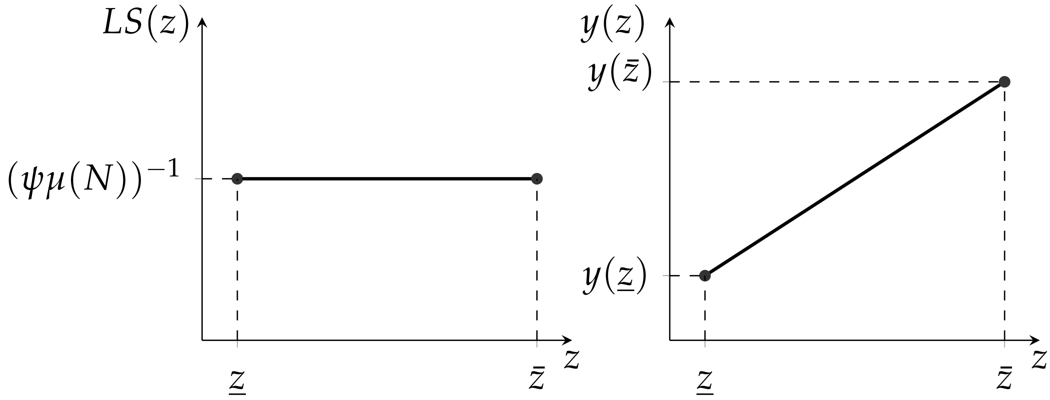
The cross-sectional distribution of output tracks productivity differences, with higher-productivity firms producing higher output, as well. Figure (1) reports the firm-level labor share (left panel) and output (right panel) plotted against the support of the productivity distribution.

We conclude the characterization of the baseline equilibrium by defining two key aggregate variables. The aggregate labor share in the operative revenues, LS, equals the average firm-level labor share:

$$LS \equiv \sum_{\underline{z}}^{\bar{z}} LS(z)\hat{\gamma}(z) = \frac{1}{\mu(N)\psi}. \tag{8}$$

[7] The existing literature uses the term *markdown* either to denote the ratio between MC and wage – e.g., Gottfries and Jarosch (2023) – or the ratio between wage and MC – e.g., Berger et al. (2024). For ease of interpretation, we adopt the former formulation that grants a direct link between the notions of markdown and labor market power.

Figure 1: Baseline labor share and output distribution



Aggregate output Y equals the sum of firm-level output:

$$Y \equiv N \sum_{\underline{z}}^{\bar{z}} y(z)\hat{\gamma}(z) = NZ^{\epsilon\frac{1+\eta}{\epsilon+\eta}} \left(\frac{1}{\mu(N)\psi}\right)^{\frac{\eta\epsilon}{\epsilon+\eta}}, \tag{9}$$

where $Z \equiv \left(\sum_{\underline{z}}^{\bar{z}} z^{\epsilon\frac{1+\eta}{\epsilon+\eta}}\hat{\gamma}(z)\right)^{\frac{\epsilon+\eta}{\epsilon(1+\eta)}}$ is an aggregate productivity index. Hence, aggregate net market power acts as a uniform tax on aggregate output, as the aggregate markup does in Edmond et al. (2023).

MW Equilibrium. We now study the response of this economy to the introduction of a binding minimum wage $\underline{w}$. Formally, we assume that the labor supply curve faced by firms is now given by:

$$w(\ell) = \max\left\{\underline{w}, \ell^{\frac{1}{\eta}}\right\}. \tag{10}$$

The labor supply curve (10) is piecewise: it is perfectly flat for low enough employment and increasing thereafter. This means that firms with low employment are effectively wage-taker, as if they operated in perfect competition.

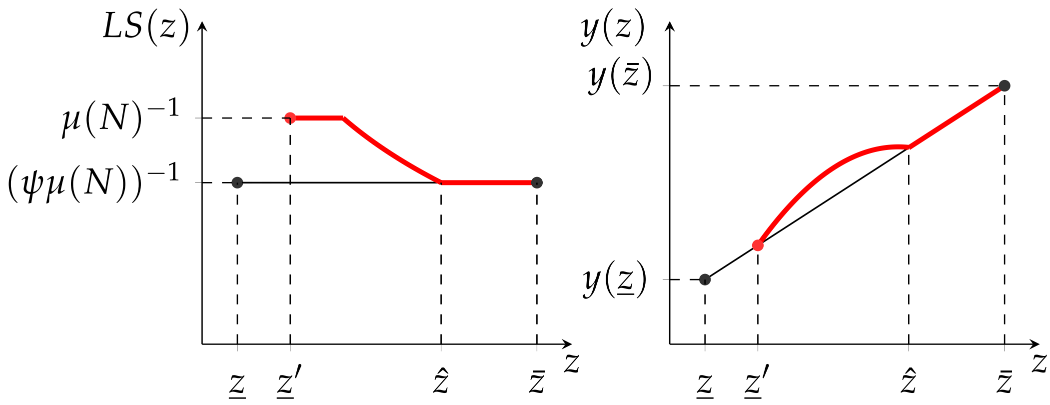
If the MW reform is large enough, some low-productivity firms exit the market, as they can no longer break even overhead costs with operating profits. Let $\underline{z}'(\underline{w})$ denote the lowest productivity level such that a firm survives. Surviving firms can be ranked in two groups based on their productivity. Relatively low-productivity firms – which end up paying the MW after the reform – are *constrained* by the MW in their wage setting. Relatively high-productivity firms – which pay higher wages than the MW – are *unconstrained*. Let $\hat{z}(\underline{w})$ denote the highest-productivity firm that is constrained by the MW.

To single out the role of the endogenous markup response, we carry out our analysis in two steps. First, we study the effects of the MW with exogenous markups, i.e., $d\mu/dN = 0$. Second, we factor in the endogenous markup response.

Figure (2) displays in red the adjustment of the firm-level labor share and output to the

MW across the productivity distribution when markups are exogenous.[8] In this case, the effects of the MW are shaped by the labor market power response only.

Figure 2: Cross-sectional impact of the MW with exogenous markups



Introducing a binding MW pushes constrained firms to pay higher wages by compressing their markdowns (as their labor supply is more elastic). In turn, the markdown reduction increases both their labor share (left panel) and their output (right panel). See Figure A.1 in Appendix A for further details on the response of constrained firms. Unconstrained firms are unaffected by the MW reform. Hence, if markups are exogenous, a higher MW raises the aggregate labor share. Moreover, if output gains from surviving firms exceed the output losses from exiting firms, the aggregate output increases with the MW, as well.

We now proceed by allowing for the endogenous markup response. Henceforth, we assume that that the support of the productivity distribution is such that firm-level employment is always increasing in the elasticity of demand (see Lemma A.3 in Appendix A for the formal condition). Figure (3) displays in red the adjustment of the firm-level labor share and output to the MW across the productivity distribution when markups are endogenous. In this case, the effects of the MW are shaped by the net market power response, i.e., the net effect of the response of labor and product market power.

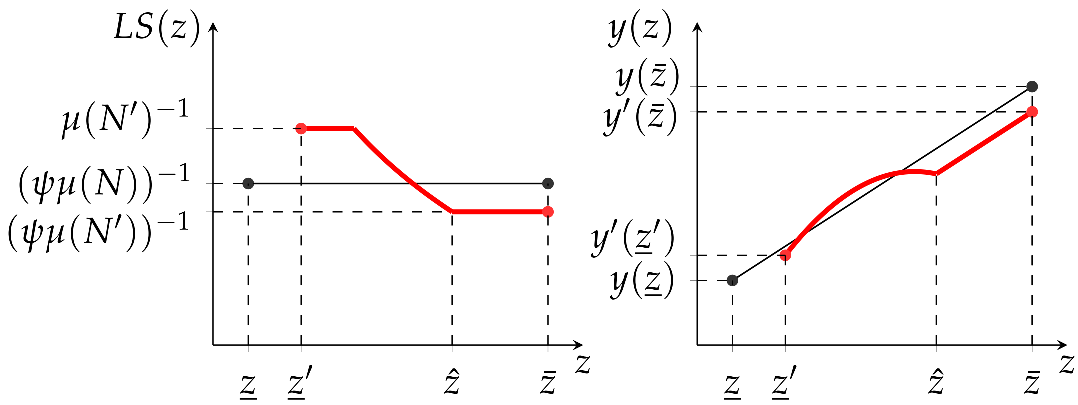
Because of the exit of some low-productivity firms, all the surviving firms face a flatter demand curve, i.e., $\epsilon(N') < \epsilon(N)$. Hence, they find it optimal to exercise their higher product market power by raising their price markups, i.e., $\mu(N') > \mu(N)$, and reducing their output. Since the positive markup response is mediated by heightened concentration (lower N) in the product market, we call this mechanism *concentration channel* of the MW.[9]

Depending on whether firms are constrained by the MW or not, the labor share and output responses are different. In particular, the responses of constrained firms are qualitatively

[8]For simplicity, in Figure (2) we plot a MW reform such that all the constrained firms that would scale down their employment exit the market and all firms that would (weakly) scale up their employment survive. See Appendix A for further details.

[9]Models of oligopsonistic competition, such as Berger et al. (2022), would predict a further "concentration channel" operating through a positive markdown response of unconstrained firms. This intuition may be formalized in reduced form by making the elasticity of labor supply, η, depend on N positively, as well. This addition would increase the magnitude of the concentration channel. However, Section 8 provides evidence that the concentration channel occurs primarily through the product market.

Figure 3: Cross-sectional impact of the MW with endogenous markups



ambiguous. On the one hand, they lose (some or all) their labor market power, showing up as a markdown reduction. On the other hand, they gain product market power, showing up as a markup increase. Instead, unconstrained firms end up having *more* market power than before, due to the increase in product market power. Therefore, both their labor share and output fall with respect to the baseline equilibrium. Overall, both the aggregate labor share and output response are generally ambiguous. The next proposition provides a simple condition on the markup elasticity to the number of firms such that both the aggregate labor share and output are hump-shaped in the MW.[10]

Proposition 1 (Aggregate effects of the MW with endogenous markups)
Introducing a binding MW in an economy with endogenous markups has the following aggregate effects:

1. *Aggregate labor share and output may increase or decrease;*
2. *If $\mu(N-1) > \mu(N)\psi \ \forall N$, then the aggregate labor share and output are hump-shaped in the MW.*

Taking stock, in this section we have studied the aggregate effects of the MW in a simple model with imperfect competition in both the labor and the product market, where product market power increases with market concentration. We have shown that the MW induces both a (well-known) reduction in labor market power of relatively low-productivity firms and a (novel) increase in product market power, due to the concentration channel, i.e., heightened product market concentration. In this environment, both the aggregate labor share and output response to the MW are qualitatively ambiguous and crucially depend on the sign of the net market power response – that is, whether the reduction in the aggregate markdown outweighs the increase in the aggregate markup, or vice versa. In Section (8) we carry out an empirical validation of the firm-level response to the MW predicted by the stylized

[10] The intuition is as follows. If the MW is small enough, no firm exits the market. Both the aggregate labor share and output increase. Once the MW is high enough to induce some exit, the aggregate labor share and output responses depend on the relative magnitude of the reduction in the aggregate markdown vis-á-vis the increase in the aggregate markup. If the latter dominates, any further increase in the MW reduces both the aggregate labor share and output.

model. Figure (14a) shows that the more concentrated product markets are in the data, the lower the firm-level labor share response to the MW. Moreover, high-productivity firms in highly concentrated markets reduce their labor share and increase profits, as predicted by the concentration channel.

Discussion. In this section we have developed the simplest theoretical framework to illustrate how the concentration channel shapes the aggregate impact of the MW by affecting the net market power response. Although useful to single out the key behavioral effects at play, the stylized model relies on strong assumptions – e.g., common baseline market power across firms, frictionless hiring – that make it unfit for quantifying the equilibrium effects of the MW. Indeed, the equilibrium effects of the MW depends not only on the net market response but also on the degree of worker reallocation, which varies by baseline market power and labor market frictions. For instance, if high-productivity firms hold high market power, reallocation is likely to be constrained and the unemployment response more pronounced. On the contrary, if labor market frictions are severe, reducing the number of active firms could alleviate congestion effects in hiring, allowing surviving firms to expand their employment and output due to lower marginal costs.

3 Quantitative Model

In this section, we develop a fully-fledged quantitative model to assess the equilibrium effects of the MW when product market power is endogenous and varies with market competition. To do so, we develop a novel quantitative model putting together a wage-posting model, that generalizes Engbom and Moser (2022), with the workhorse price-setting model of Atkeson and Burstein (2008). We consider segmented labor markets by worker skills and industry, characterized by search-and-matching frictions and on-the-job search. The source of labor market power is matching frictions, which we think as any element – mainly geographical distance and incomplete information – that prevents workers from observing all the wage offers that are posted by the firms.[11] We consider a continuum of product markets, in which a finite number of firms engage in oligopolistic competition. The source of product market power is the imperfect substitutability across firms' varieties and the granularity of firms in their product market.

Overall, the quantitative model allows us to endogenize the elasticity of labor supply, η, of the stylized model – by making it potentially heterogeneous across the productivity distribution –, and to micro-found the dependence of the aggregate markup on product market concentration. Crucially, the quantitative model delivers both endogenous markups

[11] Endogenous markdowns can stem either from idiosyncratic preferences for workplaces, when firms are granular in their labor market, or search-and-matching frictions (or, most likely, from both). According to Berger et al. (2023), search-and-matching frictions are the quantitatively most relevant determinant of markdowns. Unlike in frictionless models such as Berger et al. (2024), the MW induces reallocation of workers to more productive firms not only because of their markdown response, but also by alleviating congestion effects in hiring.

and markdowns, which react to changes in the competitive environment induced by policy reforms, such as the MW.

3.1 Environment

We study a continuous-time, stationary economy populated by a measure 1 of workers, a measure M of potential firms producing intermediate goods, and a continuum of final good producers. The economy is organized in $I \in \mathbb{N}$ industries. For each industry i, we consider a continuum of labor markets indexed by a_i. For each labor market, we consider a continuum of identical product markets, or *sectors*, indexed by $k(a_i)$.

Workers. Workers are hand-to-mouth agents with CRRA preferences over a consumption good.[12] Let $\vartheta > 0$ denote the coefficient of relative risk aversion. Workers differ in their permanent skill type a and industry i. Let $\Xi(i)$ denote the share of workers operating in industry i and $\Omega_i(a)$ denote the industry-specific skill distribution. Hence, labor markets are characterized by the skill and industry of the workers populating them. We think of skill types as reflecting both observable and unobservable invariant heterogeneity across workers. We model skills as industry-specific to reflect educational choices prior to labor market entry. Workers own the firms and get rebated a share of aggregate profits proportional to their relative wage.

Final good producers. Final good producers operate a Cobb-Douglas production function that aggregates up industry goods into a homogeneous consumption good. In turn, each industry good is a double-nested CES aggregator of firm-level varieties. Let ρ denote the elasticity of substitution across sectoral goods, and σ denote the elasticity of substitution across firm-level varieties within sectors. We assume that $\rho > 1$ and $\sigma > \rho$, meaning that goods are more substitutable within than across sectors. The final good market is perfectly competitive.

Intermediate firms. Intermediate firms differ in their physical productivity z, skill requirement $\hat{a}$ and industry i. We model a firm's skill requirement, $\hat{a}$, as the (only) worker skill type that fits its production process.[13] All firms operate a linear technology with labor being its only input. Following Engbom and Moser (2022), labor productivity is a composite of worker skill and firm physical productivity, combining in a multiplicative fashion.[14] This implies

[12] Our model abstracts from savings. As explained in Section (6), this choice allows us to compute an upper bound on the welfare loss from the increase in the unemployment rate induced by the MW.

[13] Formally, the effective productivity of a firm with physical productivity z and skill requirement $\hat{a}$ hiring workers of skill a is given by $z_{\hat{a}} = z\,\mathbb{1}\{a = \hat{a}\}$. Since firms hire only workers of their skill requirement, henceforth we omit the $\hat{a}$ subscript when denoting productivity. This assumption makes labor demand for each worker type independent of that of others, such as CRS technology (Engbom and Moser, 2022) or skill-specific capital (Berger et al., 2024).

[14] The log-linear specification of labor productivity is pervasive in the literature. None of the mechanisms outlined in the paper hinges upon this particular functional form.

that firm's output is given by:

$$y = az\ell. \tag{11}$$

Firms maximize profits by posting vacancies v and wage piece rates w per skill unit in the labor market corresponding to their skill requirement. Henceforth, we will refer to intermediate firms simply as *firms*.

Labor market structure. Labor markets are segmented by worker skill a and industry i. Both unemployed and employed workers randomly search for vacancies, which are attached an offered wage piece rate w per skill unit.[15] As in Burdett and Mortensen (1998), offered wage piece rates follows a continuous wage *offer* distribution $F_{a_i}(w)$ – an equilibrium object that is determined by firms' optimization. Let $G_{a_i}(w)$ be the resulting employment wage distribution, i.e., the density of employed workers who earn piece rates equal or lower than w. On-the-job search restrains firms' labor market power: the harder employed workers search for job offers, the more compressed the wage distribution is around the competitive wage (the marginal revenue product of labor). Each labor market a_i is characterized by search-and-matching frictions, namely an exogenous separation rate δ_{a_i}, an endogenous job finding rate per unit of search effort λ_{a_i}, and an exogenous on-the-job search effort s_{a_i} – with the search effort of unemployed workers normalized to one. Aggregate search effort is equal to $S_{a_i} = u_{a_i} + s_{a_i}(1 - u_{a_i})$, where u_{a_i} denotes the mass of unemployed workers. The total number of meetings occurring at any point in time is regulated by a CRS matching function $\mathcal{M}(S_{a_i}, V_{a_i})$, where V_{a_i} is the mass of outstanding vacancies. Define labor market tightness as $\theta_{a_i} = V_{a_i}/S_{a_i}$. By virtue of CRS, the job finding rate per unit of search effort equals $\lambda_{a_i} = \lambda(\theta_{a_i}) = \mathcal{M}(1, \theta_{a_i})$.

Product market structure. Each (intermediate) product market $k(a_i)$ is populated by a finite number of potential firms $\bar{N}_{k(a_i)}$ with the same industry i and skill requirement $\hat{a}$. In other words, firms operating in the same product market employ workers of just one type, but sell their good to all consumers (via final good producers). In equilibrium, a subset $N_{k(a_i)} \subseteq \bar{N}_{k(a_i)}$ of potential firms is operating. Let $\Gamma_{k(a_i)}(z)$ denote the realized productivity distribution of potential firms. To ensure equilibrium uniqueness, we assume that all the product markets sourcing from the same labor market are identical. Formally, they share the same number of potential firms, $\bar{N}_{k(a_i)}$, and the same (realized) productivity distribution of potential firms, $\Gamma_{k(a_i)}(z)$. Given that $N_{k(a_i)} < \infty$, firms are granular in their product market, i.e., they account for a positive share of sectoral production, and engage in oligopolistic competition á la Cournot.

[15] We model the search process as random, instead of directed, for equilibrium markdowns to be inefficient (see Section (3.7)). Hence, as in our stylized model, a higher MW can increase welfare. On the contrary, markdowns are always efficient if search is directed (absent other distortions).

3.2 Worker's problem

Let U_{a_i} and $W_{a_i}(w)$ be the value of unemployment and employment at piece rate w for a worker of skill type a in industry i. These values are defined recursively by the following Hamilton-Jacobi-Bellman (HJB) equations:

$$rU_{a_i} = \frac{(ab_{a_i})^{1-\vartheta}}{1-\vartheta} + \lambda(\theta_{a_i}) \int_{\underline{w}_{a_i}}^{\bar{w}_{a_i}} \{W_{a_i}(w) - U_{a_i}\}^+ \, dF_{a_i}(w), \tag{12}$$

$$rW_{a_i}(w) = \frac{(e_a(w)+T)^{1-\vartheta}}{1-\vartheta} + s_{a_i}\lambda(\theta_{a_i}) \int_{\underline{w}_{a_i}}^{\bar{w}_{a_i}} \{W_{a_i}(w') - W_{a_i}(w)\}^+ \, dF_{a_i}(w') + \delta_{a_i}\left(U_{a_i} - W_{a_i}(w)\right), \tag{13}$$

where r is the instantaneous interest rate, $e_a(w) = aw + \Pi(aw)$ represents the employed worker's earnings, and $\{x\}^+ \equiv \max\{x, 0\}$.

An unemployed worker of skill type a in industry i receives ab_{a_i} units of consumption good as unemployment benefits and finds a job offer at rate $\lambda(\theta_{a_i})$, that she accepts if it provides a higher value than unemployment. An employed worker of skill type a earns labor earnings aw, and receives transfers T and her share of aggregate profits $\Pi(aw)$. She finds another job opportunity at rate $s_{a_i}\lambda(\theta_{a_i})$, which she accepts if it pays better than the one she currently holds. Finally, her job is destroyed at rate δ_{a_i}, in which case she transitions into unemployment. Unemployment benefits and lump-sum transfers are financed by a proportional profit tax with rate $\tau \in [0,1]$. Formally, $T(1-u) + \sum_i \int u_{a_i} ab_{a_i} d\Omega_i(a)\Xi(i) = \tau\bar{\Pi}$, where $\bar{\Pi}$ denotes aggregate profits. As in Kaplan et al. (2018), employed workers are further rebated a share $\mathcal{S}(G(aw)) \in [0,1]$ of after-tax aggregate profits proportional to their relative wage, where $G(aw)$ is the economy-wide wage distribution. Formally, $\Pi(aw) = \mathcal{S}(G(aw))(1-\tau)\bar{\Pi}$. We interpret these profits as the profit-sharing component of worker compensation.

3.3 Final good producer's problem

Final good producers solve the following profit maximization problem:

$$\max_{\{y_{jk_i}\}} PY - \sum_{i=1}^{I} \int \int_0^{K_{a_i}} \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)} \, dk(a_i) \, d\Omega_i(a), \tag{14}$$

$$\text{s.t.} \quad Y = \prod_{i=1}^{I} Y_i^{\alpha_i}, \tag{15}$$

$$Y_i = \left(\int \int_0^{K_{a_i}} Y_{k(a_i)}^{\frac{\rho-1}{\rho}} \, dk(a_i) \, d\Omega_i(a) \right)^{\frac{\rho}{\rho-1}}, \tag{16}$$

$$Y_{k(a_i)} = \left(\sum_{j=1}^{N_{k(a_i)}} y_{jk(a_i)}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \tag{17}$$

The final good is a Cobb-Douglas aggregator of industry goods Y_i, with Cobb-Douglas weights $\alpha_i \in [0,1], \sum_{i=1}^{I} \alpha_i = 1$. As apparent from the specification of the CES aggregators (16)-(17), we assume that, for each industry i and skill requirement a, there is a measure K_{a_i} of identical product markets (sectors), indexed by $k(a_i)$, within which a finite number $N_{k(a_i)}$ of firms compete.[16]

3.4 Firm's problem

Following the equilibrium search literature, we reduce firms' dynamic problem into a static profit maximization problem by assuming steady-state behavior (Burdett and Mortensen, 1998; Bontemps et al., 1999; Engbom and Moser, 2022).[17] A firm with productivity z in industry i with skill requirement a solves the following profit maximization problem:

$$\pi_{a_i}(z) = \max_{w \geq \underline{\mathbf{w}}/a, v} \; a\left[p_{k(a_i)}(\ell)z - w\right]\ell_{a_i}(w,v) - ac_{a_i}(w,v) - a\kappa_{a_i} \tag{18}$$

$$\text{s.t.} \;\; \ell_{a_i}(w,v) = \frac{v}{V_{a_i}} \frac{\lambda(\theta_{a_i})[u_{a_i} + s_{a_i}(1-u_{a_i})G_{a_i}(w)]}{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})[1-F_{a_i}(w)]}, \tag{19}$$

$$p_{k(a_i)}(\ell) = (az\ell)^{-\frac{1}{\sigma}} \, Y_{k(a_i)}(\ell)^{\frac{1}{\sigma}-\frac{1}{\rho}} \, Y_i^{\frac{1}{\rho}-1} \, \alpha_i Y, \tag{20}$$

$$c_{a_i}(w,v) = \bar{c}_{a_i} \, h_{a_i}(w)^{\phi} \, \frac{v^{1+\zeta}}{1+\zeta}, \tag{21}$$

where $h_{a_i}(w) \equiv \delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})[1-F_{a_i}(w)]$ is the hiring rate, i.e., the ratio between new hires and stationary employment.

Firms maximize stationary profits by leveraging two control variables, namely a wage piece rate w – subject to the MW constraint – and a mass of vacancies v. Profits equal revenues net of labor costs, $aw\ell_{a_i}(w,v)$, hiring costs, $ac_{a_i}(w,v)$, and overhead costs, $a\kappa_{a_i}$. Firms face both an upward-sloping labor supply curve and a downward-sloping product demand curve. The profit maximization problem shares many similarities with that analyzed in the stylized model (1). However, three key differences stand out. First, following Engbom and Moser (2022), all the relevant variables scale with worker skill a. This is the simplest way to generate wage dispersion due to worker and firm heterogeneity separately.[18] Second, firms face search frictions in hiring. Hence, to grow larger in size, they need to post costly vacancies v subject to the hiring cost function (21). Third, and most importantly, the elasticities of labor supply

[16] The measure K_{a_i} of product markets is pinned down so as to replicate a measure M_{a_i} of firms for each (industry -skill requirement) pair.

[17] The assumption of stationary employment as labor supply curve follows from the isomorphism between a dynamic model with on-the-job search and a static model with simultaneous search by workers (Burdett and Judd, 1983). Accordingly, we can interpret firms' wage posting game as a static auction awarding workers to firms. In Online Appendix G.6, we develop the dynamically consistent version of the firm's problem and discuss the differences.

[18] The ability of the model to accommodate wage dispersion both due to worker heterogeneity (across labor markets) and firm heterogeneity (within labor markets) is critical for us to infer the firms' productivity distribution from the observed wage distribution.

and product demand are no longer parametric.

The labor supply curve (19) is determined by the stationary employment of a firm posting a wage piece rate w and vacancies v. Stationary employment balances workers' inflows to and outflows from the firm. Workers' inflows equal the probability that firm's vacancies get sampled by searching workers, v/V_{a_i}, times the measure of searching workers who meet a vacancy and would accept the firm's wage offers, $\lambda(\theta_{a_i})[u_{a_i} + s_{a_i}(1 - u_{a_i})G_{a_i}(w)]$, i.e., all unemployed workers and employed workers at lower-paying firms. Workers' outflows equal the quit rate, $\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})[1 - F_{a_i}(w)]$, i.e., the probability that an employee separates from the firm (be it due to quits or poaching from higher-paying firms), times the stationary employment, $\ell_{a_i}(w, v)$. Crucially, the firm can actively influence its stationary employment through the wage (and vacancy) policy: the higher the posted wage, the higher the filling rate and the lower the quit rate, that is, the higher the stationary employment.

The product demand curve (20) is determined by the demand function of final good producers. Since firms are granular in their product market $k(a_i)$, they internalize not only the direct effect of their employment on their price, with elasticity $-\frac{1}{\sigma}$, but also the indirect effect operating through the sectoral output, with elasticity $\frac{1}{\sigma} - \frac{1}{\rho}$.

The hiring cost (piece rate) function (21) is increasing in the hiring rate $h_{a_i}(w)$, with elasticity ϕ, and increasing and convex in vacancies, with elasticity $1 + \zeta$. Intuitively, hiring costs comprise training costs and vacancy posting costs. Training costs are captured by the hiring rate component: the higher the share of new hires relative to employment, the more costly their training is – as documented in the literature (Merz and Yashiv, 2007; Mongey and Violante, 2019; Muehlemann and Strupler Leiser, 2018). Vacancy posting costs are convex for a wage posting equilibrium with endogenous job creation to be well-defined.[19]

Limit cases. The firm's problem (18) nests two popular models as limit cases. On the one hand, it nests the wage-posting model of Engbom and Moser (2022) for $p_{k(a_i)} = 1$, $\phi = 0$ and $\kappa_{a_i} = 0$, i.e., if firms have no product market power, hiring costs are independent of the hiring rate, and there are no overhead costs.[20] On the other hand, it nests the price-setting model of Atkeson and Burstein (2008) for $\bar{c}_{a_i} = 0$ (or, interchangeably, for an infinitely efficient matching function, i.e., $\mathcal{M}(S_{a_i}, V_{a_i}) \to \infty$, $\forall a_i$) and $\kappa_{a_i} = 0$, i.e., if there are no search frictions and overhead costs.

3.5 Equilibrium characterization

Since all the product markets sourcing from the same labor market are identical, in each labor market there are finite productivity types. Even if the productivity distributions are not continuous, we now define and prove existence of a stationary equilibrium with continuous wage offer distributions. Such an equilibrium is characterized by residual dispersion of wages

[19]If vacancy posting costs were linear, firms would not find it optimal to use their wage policy for hiring.

[20]In turn, by further letting $v = \bar{v}$, i.e., assuming exogenous job creation, the problem setup boils down to traditional wage posting-models with heterogeneous productivity (Burdett and Mortensen, 1998; Bontemps et al., 1999).

and output, i.e., firms with the same productivity pay different wages and produce different quantities.

Definition (Stationary equilibrium)
For given number of potential firms $\bar{N}_{k(a_i)}$ per product market $k(a_i)$, a stationary equilibrium of the economy consists of:

- *Reservation wages $\{R^u_{a_i}, R^e_{a_i}(w)\}, \forall a_i$, for both unemployed and employed workers, that maximize the value of unemployment (12) and employment (13);*
- *Output demand functions $y_{k(a_i)}(p_{k(a_i)}), \forall a_i$, that solve the final good producers' problem (14);*
- *Continuous residual wage offer distributions $F_{a_i}(w;z), \forall a_i$, such that firms with the same productivity z make the same profits, and associated continuous marginal revenue product distributions $\Phi_{a_i}(\tilde{z};z), \forall a_i$, such that $F_{a_i}(w;z)$ is consistent with firms' optimization;*
- *Wage and vacancy policies $w_{a_i}(\tilde{z}), v_{a_i}(\tilde{z})$, $\forall \tilde{z} \in [\underline{\tilde{z}}_{a_i}(z), \bar{\tilde{z}}_{a_i}(z)), \forall z, \forall a_i$, that solve the firms' problem (18) for given residual marginal revenue product distributions;*
- *Productivity thresholds $\left\{\underline{z}_{k(a_i)}\right\}$ that determine the marginally active firm in each product market $k(a_i)$, $\forall a_i$;*
- *Measures $\{F_{a_i}(w), G_{a_i}(w), \Phi_{a_i}(\tilde{z}), u_{a_i}, V_{a_i}\}$ and market tightness $\theta_{a_i}, \forall a_i$, that are consistent with firms' vacancy policies, balance of flows, and the meeting technology;*
- *Intermediate product market clearing conditions ensuring that supply and demand for each firm-level variety coincide:*

$$y_{k(a_i)}\left(p_{k(a_i)}(\tilde{z},z)\right) = az\ell_{a_i}(\tilde{z}) \quad \forall \tilde{z} \in [\underline{\tilde{z}}_{a_i}(z), \bar{\tilde{z}}_{a_i}(z)), \forall z, \forall a_i;$$

- *Final product market clearing condition ensuring that the aggregate resource constraint holds, that is, aggregate consumption equals aggregate value added:*[21]

$$C = Y - \sum_{i=1}^{I}\int_a \int_{\tilde{z}} M_{a_i}\, ac_{a_i}(\tilde{z})\, d\Phi_{a_i}(\tilde{z})\, d\Omega_i(a) - \sum_{i=1}^{I}\int_a M_{a_i}\, a\kappa_{a_i}\, d\Omega_i(a),$$

where $C = \sum_{i=1}^{I}\int_a \left[(1-u_{a_i})\int_{\underline{w}_{a_i}}^{\bar{w}_{a_i}} aw\, dG_{a_i}(w) + M_{a_i}\sum_{j=1}^{\mathcal{Z}_{a_i}}\pi_{k(a_i)}(z_j)\right] d\Omega_i(a)$ denotes aggregate consumption, and Y is the Cobb-Douglas aggregator defined in (15).

Existence of a stationary equilibrium requires two mild parametric assumptions. First, the elasticity of substitution across sectors (ρ) is assumed to be high enough such that the revenue function is supermodular in size and productivity. Second, the unemployment benefits (b_{a_i})

[21] We assume that all the potential firms pay the overhead cost and operate only if they can (weakly) break even, e.g., because of limited liability. This implies that the expansionary effects of the MW on aggregate consumption due to firm selection are entirely driven by worker reallocation – not by the opportunity cost of firm entry.

are assumed to be low enough such that profits of the lowest-ranked firm in each labor market – which pays the reservation wage – are positive and increasing in the rank of the wage offer distribution ($F_{a_i}(w)$).

Proposition 2 (Existence and uniqueness of stationary equilibrium with exit ordering based on losses)
There exists a symmetric stationary equilibrium with continuous wage offer distributions and exit ordering based on losses. The equilibrium is unique within the set of equilibria with nondegenerate wage distributions.

Proof. See Appendix B.8. ■

We now inspect the policy functions of the agents in our economy one at a time.

Worker's policy. Worker's problem is solved by setting a state-specific reservation wage, $R^s_{a_i}$, $s = \{u, e\}$, such that all offers equal or higher than that are accepted. Specifically, the reservation wage solves $W_{a_i}(R^u_{a_i}) = U_{a_i}$ for unemployed and is equal to the current wage for employed, i.e., $R^e_{a_i}(w) = w$. See Appendix (B.5) for the derivations.

Final good producer's policy. Solving the profits maximization problem (14) yields the following demand function for each firm's variety j in product market $k(a_i)$:

$$y_{jk(a_i)} = \left(\frac{p_{jk(a_i)}}{P_{k(a_i)}}\right)^{-\sigma} \left(\frac{P_{k(a_i)}}{P_i}\right)^{-\rho} \left(\frac{P_i}{P}\right)^{-1} \alpha_i Y \quad \forall j = 1, \ldots, N_{k(a_i)}, \forall a_i. \tag{22}$$

Reflecting the triple-nested production function (15), the demand for the good produced by firm j in sector $k(a_i)$ negatively depends on its price vis-à-vis the price index of the sector, on the price index of the sector vis-à-vis that of the industry, and on the price index of the industry vis-à-vis the aggregate price index. See Appendix (B.4) for the derivations of the price indices. Henceforth, we use the consumption good as numeraire and normalize its price to 1, i.e., $P = 1$. Notice that the Cobb-Douglas aggregator of industry-specific output prevents any reallocation of demand across industries following an asymmetric shock, such as the introduction of a MW.

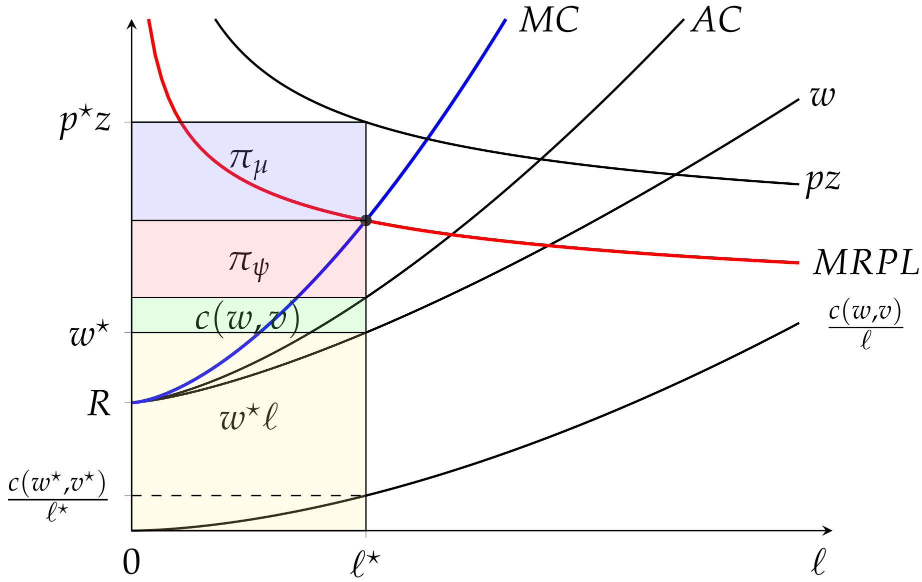
Firm's policy. The solution to the firm's profit maximization problem (18) consists of the following system of first-order conditions:

$$\ell_{a_i}(w, v) = \left[\left(1 + \epsilon^{-1}_{k(a_i)}(\ell)\right) p_{k(a_i)}(\ell)\, z - w\right] \frac{\partial \ell_{a_i}(w, v)}{\partial w} - \frac{\partial c_{a_i}(w, v)}{\partial w}, \tag{23}$$

$$\frac{\partial c_{a_i}(w, v)}{\partial v} = \left[\left(1 + \epsilon^{-1}_{k(a_i)}(\ell)\right) p_{k(a_i)}(\ell)\, z - w\right] \frac{\partial \ell_{a_i}(w, v)}{\partial v}, \tag{24}$$

where $\epsilon_{k(a_i)}(\ell)$ denotes the endogenous elasticity of demand.[22] The FOC with respect to the wage (23) equalizes the marginal increase in the wage bill (left-hand side) to the marginal increase in profits by raising employment through a wage rise (right-hand side). In turn, the latter is composed by the profit margin earned on new employees net of the induced price response ($\epsilon^{-1}_{k(a_i)}(\ell) < 0$), and by the marginal hiring cost saving from reducing worker turnover through a higher wage ($\partial c_{a_i}(w, v)/\partial w < 0$). The FOC with respect to vacancies (24) equalizes the marginal increase in the hiring cost (left-hand side) to the marginal increase in profits by raising employment through an increase in vacancies (right-hand side). Importantly, firms internalize that growing larger in size, i.e., selling larger quantities, entails cutting their price. Hence, product market power restrains optimal employment.

Figure 4: Graphical representation of firm's problem



Note: See Appendix (B.6) for the derivation of the equations behind the graph.

Figure (4) depicts the solution to the firm's problem graphically. Firms choose their employment by equating the marginal revenue product (MRPL) to the marginal cost (MC) of labor. The MRPL falls short of the marginal product by the markup factor. Hence, the light blue area in between the two curves represents profits from product market power. The MC exceeds the average cost (AC) of labor by the markdown factor. Hence, the light red area in between the two curves represents profits from labor market power. In turn, the light green area between the AC and the wage accounts for hiring costs. Finally, the light yellow area below the wage curve represents the wage bill.

After some manipulation of the system of first-order conditions, we recover the familiar double-wedge representation of the optimality condition between price and (productivity-

[22] Notice that (23) holds only where the wage offer distribution is differentiable, i.e., $supp F_{a_i}(w) \setminus \left\{ \{\underline{w}_{a_i}(z_n)\}_{n=1}^{\mathcal{Z}_{a_i}} \cup \bar{w}_{a_i}(z_{\mathcal{Z}_{a_i}}) \right\}$.

adjusted) wage:

$$p_{a_i}(z) = \underbrace{\frac{\epsilon_{a_i}(z)}{\epsilon_{a_i}(z)-1}}_{\mu_{a_i}(z)} \underbrace{\frac{1+\nu\eta_{a_i}(z)}{\nu\eta_{a_i}(z)}}_{\psi_{a_i}(z)} \frac{w_{a_i}(z)}{z}, \tag{25}$$

where η_{a_i} denotes the endogenous elasticity of labor supply. [23] Equation (25) shares the same structure as its counterpart in the stylized model (5), but features two key differences. First, both the elasticity of demand, ϵ, and the elasticity of labor supply, η, are endogenous. This means that the degree of market power is differentiated both across markets and across firms within a given market. Second, the elasticity of labor supply is multiplied by the hiring cost correction term $\nu \equiv 1 + \frac{\phi}{2(1+\zeta)}$, which depends on the ratio between the elasticity of hiring costs with respect to the hiring rate, ϕ, and with respect to vacancies, $1+\zeta$. Intuitively, this term captures the net reduction in hiring costs by growing larger in employment, so it is absent from any model where hiring costs are size-independent.

Finally, let $\tilde{z}_{a_i}(z) \equiv p_{a_i}(z)z/\mu_{a_i}(z)$ denote the equilibrium MRPL(s) of a firm with productivity z operating in labor market a_i. In Appendix (B.9), we show that the wage policy function equals the expected outside option of a searching worker:

$$w_{a_i}(\tilde{z}) = \left(1-\mathcal{P}_{a_i}(\tilde{z})\right) R^u_{a_i} + \mathcal{P}_{a_i}(\tilde{z})\mathbb{E}\left[\max\{MRPL\}|MRPL < \tilde{z}, m_{a_i}(\tilde{z}) = 1\right], \tag{26}$$

where $\mathcal{P}_{a_i}(\tilde{z}) \equiv \Pr\left(m_{a_i}(\tilde{z}) = 1\right)$ and $m_{a_i}(\tilde{z}) \equiv \mathbb{1}\{\text{meet at least one firm w/ MRPL } \leq \tilde{z}\}$ is an indicator of alternative job offers. Equation (26) is best understood in terms of static allocation of workers to firms. Accordingly, wage offers sampled by employed workers over the course of their tenure at a given firm are interpreted as the (appropriately discounted) pool of alternative job offers from which a worker can statically choose. With probability $1-\mathcal{P}_{a_i}(\tilde{z})$, the worker does not meet any firm with MRPL lower than $\tilde{z}$, in which case the worker's outside option when meeting a firm with MRPL $\tilde{z}$ boils down to her reservation wage. With complementary probability, the worker meets at least one firm with lower MRPL than $\tilde{z}$, in which case the worker's outside option when meeting a firm with MRPL $\tilde{z}$ equals the maximum wage the best of such firms can afford paying, that is, the highest expected MRPL among firms with lower MRPL than $\tilde{z}$.

3.6 Inspecting Markups and Markdowns

We now turn to our main objects of interest, that is, the equilibrium markups and markdowns. In particular, our focus will be on understanding how they would react to the introduction of a MW.

[23] Notice that all the policy functions with respect to productivity z are set-valued due to residual wage dispersion. In other words, firms with same productivity pay different wages in equilibrium, so the wage policy function is set-valued. The price, markup, and markdown policy functions are one-to-one for given productivity *and* wage.

Equilibrium Markups. Let $\mathcal{C}_{k(a_i)}(z) \equiv \psi_{a_i}(z) w_{a_i}(z)$ denotes the equilibrium MC(s) of a firm with productivity z operating in product market $k(a_i)$. The double-nested CES demand structure within each industry, as well as the granularity of firms in their own product market, entail that the elasticity of demand faced by a firm with productivity z and marginal cost $\mathcal{C}$ operating in sectoral market $k(a_i)$ reads:

$$\epsilon_{k(a_i)}(\mathcal{C}, z) = \left[-\frac{1}{\rho} sh_{k(a_i)}(\mathcal{C}, z) - \frac{1}{\sigma}(1 - sh_{k(a_i)}(\mathcal{C}, z)) \right]^{-1}, \tag{27}$$

where $sh_{k(a_i)}(\mathcal{C}, z) \equiv p_{k(a_i)}(\mathcal{C}, z) y_{k(a_i)}(\mathcal{C}, z) / \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)}$ is the sectoral market share of the firm. Hence, the firm-level elasticity of demand depends on how large the firm is relative to its product market in terms of revenues. Formally, it equals a harmonic average between the elasticity of demand across sectors, $-1/\rho$, and the elasticity of demand within sectors, $-1/\sigma$, weighted by the market share and its complement to one, respectively (Atkeson and Burstein, 2008). The fact that the firm's market share influences its elasticity of demand reflects the granular nature of product market competition in our model. Intuitively, Equation (27) provides a microfoundation to the positive dependence of the elasticity of demand on the number of active firms assumed in the stylized model in Section (2): for given productivity distribution, the larger the number of firms, the lower the firm's market share and the higher the elasticity of demand (in magnitude).

From Equation (27) and the markup definition (25), the equilibrium markup $\mu_{k(a_i)}(\mathcal{C}, z)$ of a firm with productivity z and marginal cost $\mathcal{C}$ operating in product market a_i is given by:

$$\mu_{k(a_i)}(\mathcal{C}, z) = \frac{\sigma}{(\sigma - 1)\left[1 - \frac{\sigma/\rho - 1}{\sigma - 1} sh_{k(a_i)}(\mathcal{C}, z)\right]}. \tag{28}$$

The equilibrium markup is increasing in the firm's market share, as well as moving between two intuitive bounds. When the firm is small relative to its product market, i.e., $sh_{k(a_i)}(\mathcal{C}, z) \to 0$, the markup approaches $\frac{\sigma}{\sigma-1}$, as if the firm were operating in monopolistic competition within the sector. On the contrary, when the firm is large relative to its sector, i.e., $sh_{k(a_i)}(\mathcal{C}, z) \to 1$, then the markup approaches $\frac{\rho}{\rho-1}$, as if the firm were operating in monopolistic competition against other sectors.

How would the markup distribution react to the introduction of a MW? The MW is expected to induce some exit (or layoffs) among low-productivity firms, which typically have lower baseline markups. In the presence of labor market frictions, displaced workers from low-productivity firms will (imperfectly) reallocate towards higher-productivity and larger firms, which typically have higher baseline markups. Moreover, the reallocation process will make large firms gain additional market shares, leading to higher markups through the concentration channel. Overall, the markup distribution is expected to shift rightward following the introduction of a MW, due to such reinforcing compositional and behavioral effects. Hence, in the presence of a more realistic baseline markup distribution, the concentration channel of the MW, studied in Section 2, is amplified by a compositional effect.

Equilibrium Markdowns. Let $\tilde{z}_{a_i}(z) \equiv p_{a_i}(z)z/\mu_{a_i}(z)$ denote the equilibrium MRPL(s) of a firm with productivity z operating in labor market a_i. Workers' search behavior and the labor market structure jointly entail that the labor supply elasticity faced by a firm with marginal revenue product $\tilde{z}$ operating in labor market a_i reads:

$$\eta_{a_i}(\tilde{z}) = \frac{2f_{a_i}(w_{a_i}(\tilde{z}))w_{a_i}(\tilde{z})}{\chi_{a_i}(\theta_{a_i}) + [1 - F_{a_i}(w_{a_i}(\tilde{z}))]}, \tag{29}$$

$$= \frac{w_{a_i}(\tilde{z})\ell_{a_i}(\tilde{z})}{(1+\zeta)c_{a_i}(\tilde{z})}, \tag{30}$$

where the two alternative formulations follow from Equations (23) and (24), respectively.

Despite being equivalent in equilibrium, both formulations carry instructive insights. Equation (29) focuses on the role of on-the-job search and the shape of the wage offer distribution. Specifically, the elasticity of labor supply equals twice the quit rate elasticity.[24] The quit rate elasticity depends on a common component across all firms operating in a labor market and a firm-specific component. The common component is represented by the frictional index $\chi_{a_i}(\theta_{a_i}) \equiv \frac{\delta_{a_i}}{s_{a_i}\lambda(\theta_{a_i})}$, which is an inverse measure of the speed at which employed workers climb the job ladder. The larger the frictional index, the lower the quit rate elasticity, as the competitive pressure exerted by on-the-job search is dampened.[25] The firm-specific component makes the quit rate elasticity positively depend on the firm's position in the wage offer distribution, $F_{a_i}(w_{a_i}(\tilde{z}))$, and in the wage offer density, $f_{a_i}(w_{a_i}(\tilde{z}))$, as well as on the posted wage, $w_{a_i}(\tilde{z})$. Hence, for given rank in the wage distribution and wage policy, the quit rate elasticity is increasing in the extent of *local* competition the firm is facing, as summarized by the wage offer density. Intuitively, the higher the density of the wage distribution, the larger the pool of competitors that could poach workers from the firm if it were to marginally reduce its offered wage. Equation (30) zooms into the role of hiring costs in the elasticity of labor supply. From this formulation, it is apparent that the elasticity of labor supply is decreasing in hiring costs. Moreover, if hiring costs tend to zero, then $\eta_{a_i}(\tilde{z})$ approaches infinity, i.e., the perfect competition benchmark. This is because in our model hiring costs are the source of search frictions, which in turn are the structural determinant of the upward-sloping labor supply curve faced by firms. Intuitively, Equations (29)-(30) endogenize the elasticity of labor supply taken as exogenous in the stylized model in Section (2): depending on the specific extent of hiring frictions, firms face different elasticities of labor supply.

From Equations (29)-(30) and the markdown definition (25), the equilibrium markdown,

[24] Since the quit rate elasticity equals the hiring rate elasticity in stationary equilibrium, the elasticity of labor supply amounts to twice the quit rate elasticity (Manning, 2003).

[25] As in Burdett and Mortensen (1998), on-the-job search by employed workers is the fundamental reason why firms face an upward sloping labor supply curve. Intuitively, the higher the offered wage, the higher the inflow of workers from other firms and the lower the outflow of employees towards other firms.

$\psi_{a_i}(\tilde{z})$, of a firm with marginal revenue product $\tilde{z}$ operating in labor market a_i is given by:

$$\psi_{a_i}(\tilde{z}) = 1 + \frac{\chi_{a_i}(\theta_{a_i}) + [1 - F_{a_i}(w_{a_i}(\tilde{z}))]}{2\nu f_{a_i}(w_{a_i}(\tilde{z}))\, w_{a_i}(\tilde{z})}, \tag{31}$$

$$= 1 + \frac{(1+\zeta)c_{a_i}(\tilde{z})}{w_{a_i}(\tilde{z})\ell_{a_i}(\tilde{z})}. \tag{32}$$

Owing to the two alternative formulations of the equilibrium elasticity of labor supply, the equilibrium markdown itself can be expressed in two different ways. Equation (31) shows that equilibrium *net* markdowns, i.e., $\psi_{a_i}(\tilde{z}) - 1$, equal the inverse quit rate elasticity normalized by the hiring cost correction term $\nu > 1$ (see Appendix B.10 for derivations). Specifically, the more sensitive hiring costs are to the hiring rate, the lower equilibrium markdowns are. Hence, according to our model, standard markdown estimates based on doubling the quit rate elasticity are generally upward biased. Equation (32) is the mirror image of (30) and shows that the equilibrium markdown is increasing in the ratio between hiring costs and wage bill, reflecting the key role of search frictions.

How would the markdown distribution react to the introduction of a MW? As in our stylized model, markdowns of constrained firms will reduce due to a higher (effective) reservation wage of workers. As in Burdett and Mortensen (1998), some unconstrained firms will find it optimal to compress their markdowns, as well, to preserve their rank in the equilibrium wage offer distribution.[26] Hence, these behavioral effects will push the markdown distribution to shift leftward. On the other hand, the convexity of the hiring cost function entails that excess hiring at higher-paying firms will not offset job losses at low-paying firms, which reduces aggregate vacancies and drives the frictional index up. Moreover, depending on the shape of the baseline markdown distribution, worker reallocation may bring about a further compositional effect, presumably raising the aggregate markdown. Overall, the response of the markdown distribution to the MW is a priori ambiguous.

3.7 Efficiency Properties of Baseline Equilibrium

We conclude this section by commenting on the efficiency properties of the baseline equilibrium and how they are expected to react to the introduction of a MW.

In Appendix B.13 we characterize the efficient allocation formally by solving the problem of a constrained social planner concerned with maximizing aggregate consumption. The resulting allocation would maximize a utilitarian social welfare function in a special case of our model where workers are risk neutral ($\vartheta = 0$).[27] A constrained social planner chooses the direction and pace of worker reallocation. As in the baseline equilibrium, the social planner

[26] This spillover effect can be interpreted as a local increase in the wage offer density through the lens of Equation (31).

[27] In the presence of a concave utility function, a utilitarian social planner cares about the level of aggregate consumption, the amount of uncertainty agents face, and the distribution of consumption across agents. For analytical transparency, we focus on the first of these goals only. In Section (6), we carry out a full numerical decomposition of the welfare effects induced by the MW.

reallocate workers towards higher-productivity firms whenever an opportunity arises. Unlike in the baseline equilibrium, the social planner dictates a common hiring rate across firms with the same productivity. Hence, residual dispersion in marginal products is inefficient, as in Menzio (2024).

For clarity of exposition, we now specialize on a version of our model with a large number of productivity types, hence no residual dispersion in marginal products. We proceed by comparing the optimality conditions governing efficient and equilibrium vacancy posting of a firm with productivity z operating in labor market a_i:

$$v^{\star\star}_{a_i}(z): \quad MPL_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v(z)} = \frac{\partial c_{a_i}(z)}{\partial v(z)} + E^{bs}_{a_i}(\mathbf{z} \leq z) + E^{c}_{a_i},$$
$$v^{\star}_{a_i}(z): \; MRPL_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v(z)} = \frac{\partial c_{a_i}(z)}{\partial v(z)} + w_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v(z)},$$

where $v^{\star\star}_{a_i}(z)$ denotes efficient vacancy posting and $v^{\star}_{a_i}(z)$ equilibrium vacancy posting. A constrained social planner equalizes the social marginal benefit of vacancy posting to its social marginal cost. The social marginal benefit equals the marginal product of labor ($MPL_{a_i}(z)$) resulting from vacancy posting by the firm. The social marginal cost equals the sum of the marginal increase in hiring costs, the business-stealing effects on lower-productivity firms operating in the same labor market ($E^{bs}_{a_i}(\mathbf{z} \leq z) \geq 0$), and the congestion effects on other agents (firms and workers) in the same labor market ($E^{c}_{a_i} \lesseqgtr 0$).

In equilibrium, firms equalizes the private marginal benefit of vacancy posting to its private marginal cost. The private marginal benefit equals the marginal *revenue* product of labor ($MRPL_{a_i}(z)$) resulting from higher vacancy posting by the firm. The private marginal cost equals the sum of the marginal increase in hiring costs and wages resulting from vacancy posting.

Suppose first that the reservation wage (or binding MW) is zero. Lemma (A.5) shows that the equilibrium vacancy posting is efficient if and only if firm-level varieties in a given product market are perfect substitutes ($\sigma = \infty$), so that the marginal product and marginal revenue product of labor coincide ($MPL_{a_i}(z) = MRPL_{a_i}(z)$), and congestion effects are zero, i.e., $E^{c}_{a_i} = 0$.[28] The reason is as follows: since the equilibrium wage equals the expected outside option for searching workers, the wage that firms optimally set exactly compensates for the business-stealing effects they exert on lower-paying competitors. In other words, workers are paid the opportunity cost of employment at a firm with productivity z, that is, the marginal revenue product of labor that would be generated by reallocating the worker to the second-best firm from which she receives a job offer.

To study the role of the MW, let $\Delta_{a_i}(z)$ denote the *labor wedge* at the firm level, i.e., the ratio between social marginal benefit and marginal cost of vacancy posting (Hsieh and Klenow, 2009). In equilibrium, $\Delta_{a_i}(z) = \mu_{a_i}(z) \Big/ \left(1 + \frac{D_{a_i}(z)R_{a_i} - E^{c}_{a_i}}{MC_{a_i}(z)}\right)$, where $MC_{a_i}(z) \equiv \psi_{a_i}(z) w_{a_i}(z)$

[28] Congestion effects are zero in the (knife-edge) case in which the negative marginal effect of vacancy posting on aggregate consumption due to lower meeting rate per vacancy is perfectly balanced by the positive marginal effect due to higher job finding rate.

is the equilibrium marginal cost and $D_{a_i}(z) > 0, D'_{a_i}(z) \geq 0$. The MW influences the labor wedge by increasing the reservation wage R_{a_i}. The baseline equilibrium features two sources of inefficiency. First, the existence of positive net markups (due to imperfect substitutability across firm-level varieties) and distorted markdowns (due to congestion effects and positive reservation wages) make the average labor wedge generally differ from one.[29] If the MW pushes the average labor wedge closer (farther) to one, this inefficiency will ameliorate (worsen) by showing up as an increase (decrease) in aggregate vacancies and employment.

Second, heterogeneity in markups (due to firms' granularity in their product market) and markdown distortions (due to firm-specific impact of congestion effects and reservation wages) across firms entails that the allocation of labor is inefficient, because it does not minimize aggregate hiring costs for given final output. Intuitively, misallocation arises from the fact that the labor wedge is heterogeneous across firms. If the MW decreases (increases) the productivity-adjusted dispersion of the firm-level labor wedge, this inefficiency will ameliorate (worsen) by showing up as a decrease (increase) in aggregate hiring costs per unit of final good, which we will adopt as a misallocation index.

4 Structural Estimation

In this section we outline how we estimate the parameters of our structural model using data from the Italian National Institute for Social Security (INPS) and the Italian Institute of Statistics (Istat). First, we discuss how to link the model-based notions of labor and product markets into their (closest) real-world counterparts. Then, we present our estimation strategy and results. Finally, we discuss the baseline distribution of markups and markdowns implied by the estimated model.

Market definition. Estimating firms' market power involves taking a stance on the definition of labor and product markets. This is a challenging task as industrial organization and geographical scope vary significantly across markets (Rossi-Hansberg et al., 2021; Eeckhout, 2020). Aware of the inevitable simplifications a macroeconomic perspective entails, we propose market definitions based on workers' heterogeneity for labor markets and goods differentiation for product markets.

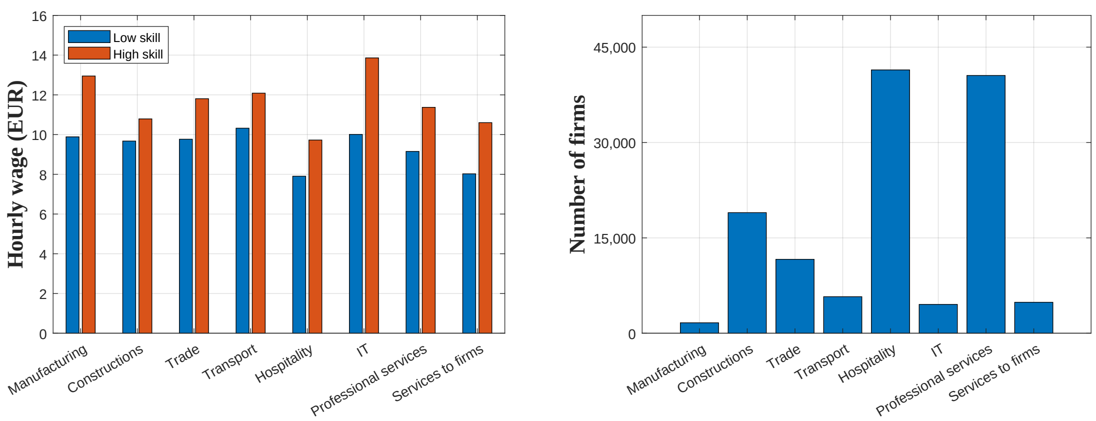
First, we notice that model-based labor markets are segmented by workers' skills and industry, as well as populated by a measure of firms – none of them being large relative to the market. Moreover, matching frictions within each labor market capture the average effect of geographical distance and incomplete information on the job finding rate.[30] Therefore, we

[29] Given that firms face hiring costs, efficient net markdowns are positive. Still, equilibrium markdowns are distorted with respect to their efficient benchmark because wage posting with endogenous job creation fails to internalize the congestion effects induced by vacancy posting

[30] If each labor market were composed by a collection of identical local labor markets with granular firms, matching frictions would capture (in reduced-form) the extent of geographical mismatch between job seekers and vacancies (Shimer, 2007). Intuitively, a common (finite) job finding rate would exactly account for the foregone job opportunities due to geographical mismatch between job seekers and vacancies as long as local

run an Abowd-Kramarz-Margolis (AKM) regression on our matched employer-employee data from INPS spanning the longest time period available to us (1990-2018) and compute worker fixed effects (Abowd et al., 1999).[31] Conceptually, AKM worker fixed effects capture permanent heterogeneity in skills that reflects in wage differences, thus representing an intuitive empirical counterpart of the skill type a in the model. We proceed by splitting workers into two skill types (high- and low-skilled) based on the median AKM worker fixed effect. Moreover, we assign each worker to an industry based on the ATECO 1-digit industry code of the firm in which he or she has been most frequently employed between 2016 and 2018.[32] Hence, we define labor markets as segmented by AKM worker fixed effects and 1-digit industry. The idea is that individual ability and educational choices prior to labor market entry constitute hard constraints to one's labor market, while matching frictions prevents workers from receiving job offers from firms located in distance places. It follows that the continuum of labor markets in the model maps into 16 segmented labor markets in the data. Figure (5a) shows that average wages are increasing in our skill measure conditional on each industry, and exhibit substantial heterogeneity across industries. Hence, our labor market definition seems to capture relevant structural differences.

(a) Average wage, by labor market (b) Avg firms in (4-digit) sectors, by industry



Source: Authors' calculations, INPS (2016-2018); Istat (2019). *Note*: Sectors and industries are defined as 4-digit and 1-digit industry according to the ATECO classification.

Second, we observe that model-based product markets are populated by a finite number of firms that compete oligopolistically. Conceptually, they should be mapped to relatively narrowly defined products. Following the literature (Grassi et al., 2017), we therefore think of product market boundaries in the model as comparable to the ATECO 4-digit industry codes – the most granular level we have access to. It follows that the average number of firms

labor markets are sufficiently homogeneous. We see the investigation of spatial heterogeneity as a promising avenue for future research, but beyond the scope of the paper.

[31] See Appendix (C.1) and (C.2) for details on the matched employer-employee data used and on the estimation of the AKM equation.

[32] ATECO industry codes are designed by Istat in compliance with the European NACE industry classification standards (the counterpart of the NAICS in the US). Some industries in which the number of observations is too small are merged with the closest industry with enough observations. Specifically, "mining" is merged with "manufacturing", "other services" with "services to firms", "real estate" with "professional services", and "energy" and "water and disposal" with "construction". Results are unchanged if we simply drop such observations.

within a 4-digit sector is $2,963$ – a large number in international comparison, explained by the disproportionate incidence of small firms in the Italian economy. Moreover, as Figure (5b) shows, the average number of firms per sector masks substantial heterogeneity across 1-digit industries – an important element we will take care of in our structural estimation.

Discussion. We adopt a labor market definition based on AKM worker fixed effects and industry codes by combining the approaches of Engbom and Moser (2022) and Bontemps et al. (2000). Although our social security data do not contain information on occupation, we are confident that our labor market definition is highly correlated with its occupational content. Alternative popular approaches to labor market definition include some geographical dimension (Berger et al., 2022) or rely on a data-driven definition based on workers' flows (Huitfeldt et al., 2023; Jarosch et al., 2024). Since our model does not feature granular firms in their labor market and the average effect of geographical distance is accounted for by matching frictions, we refrain from introducing a further geographical segmentation – in line with the equilibrium search literature. Due to data limitations and the dimensionality of the problem, we are unable to define labor markets based on workers' flows.

We adopt a product market definition based on industry codes. The reason is twofold. First, absent price data and/or exogenous variation to identify demand elasticities, the alternative approach of defining market structures based on estimated markups, as in Deb et al. (2023) and De Loecker et al. (2021), is more contentious (Bond et al., 2021; Berger et al., 2022; De Ridder et al., 2022). Second, and most importantly, using a product market definition with a direct empirical counterpart allows us to validate our theoretical predictions in the data in a model-consistent way – an exercise we carry out in Section (8). Our approach to product market definition has two main limitations. First, the geographical scope of product markets can vary significantly across industries. On the one hand, the product markets for tradable goods arguably extends beyond the national boundaries. On the other hand, the product markets for nontradable goods are likely more geographically restricted. Second, multi-product firms operate in multiple product markets, as well, so it is unclear what market share is relevant for pricing purposes. Since we do not have access to establishment-level data, we are unable to compute precise concentration measures for geographically distinct markets, but can only use proxies based on the location of firms' headquarters.[33] We deal with multi-product firms by relying on the 4-digit industry code they apply to – presumably corresponding to the product from which they raise the highest revenues.

Estimation Strategy. We estimate the model by replicating a stationary equilibrium for the Italian economy, targeting empirical moments from the period 2016-2019. Our objective is to use the estimated model to assess the equilibrium effects of introducing a nation-wide MW in the Italian economy (where none is currently in place).

[33]As a robustness check, we ran our structural estimation and counterfactuals by segmenting product markets in nontradable industries by province. It turns out that the aggregate effects of the MW are rather robust to different product market definitions.

First, we externally set structural parameters related to market stucture, workers' preferences, labor market frictions, and profits rebates. Then, the remaining parameters are estimated jointly using the Simulated Method of Moments (SMM). An original contribution of our strategy is the non-parametric estimate of firms' productivity from the wage distribution in the presence of both endogenous markups and markdowns. To do so, we follow a two-stage procedure within the SMM routine. In the first stage, we make use of the equilibrium markdown function to nonparametrically estimate the marginal revenue productivity (MRPL) distributions that rationalize the labor-market-specific wage distributions. We then use the equilibrium markup function, along with the observed number of competing firms in 4-digit product markets, to infer the underlying physical productivity distributions from the MRPL distributions. We adopt this estimation strategy to guarantee the best possible fit to the wage distributions, which is key to replicate the actual bite of MW reforms and the cross-sectional distribution of labor market power.

Externally set parameters. We externally set the parameters governing market structure, workers' preferences, labor market frictions, and profits rebates. As already discussed, we consider $I = 8$ industries and $A = 2$ skill types, which give rise to 16 labor markets. We pin down the relative size of worker population across labor markets by replicating the employment composition by industry and skill from INPS data (2016-2018). We consider $K = 567$ product markets, which equals the number of 4-digit sectors we observe in our Istat data. We account for the heterogeneous market structure across industries highlighted in Figure (5b) by assigning a number of potential (and operating) firms to a product market $k(a_i)$ equal to the average in the respective industry i, $\bar{N}_{k(a_i)} = \bar{N}_i$. The measure of active firms in each skill-specific labor market, M_{a_i}, equals the ratio between the employment rate and the average number of employees per firm in each 1-digit industry i. We read off the latter from Istat data. Our estimation strategy allows backing out the (un)employment rate externally as follows. As in Engbom and Moser (2022), we take the job finding rate $\lambda(\theta_{a_i})$ as an auxiliary parameter and set it equal to average observed NE rate in all labor markets.[34] Moreover, we pick the separation rates δ_{a_i} to replicate the EN rate separately by skill type. As a result, the skill-specific unemployment rate is pinned down by the stationarity condition $u_{a_i} = \frac{\delta_{a_i}}{\delta_{a_i}+\lambda(\theta_{a_i})}$.[35]

Turning to workers' preferences, we set the instantaneous interest rate r to 0.004 and assume a monthly frequency, implying an annual interest rate of 4%. Following the macro literature, we set the coefficient of relative risk aversion, ϑ, to 1. To model labor market frictions, we adopt a CES matching function $\mathcal{M}(S,V) = \chi\left[\beta V^{-\iota} + (1-\beta)S^{-\iota}\right]^{-\frac{1}{\iota}}$ with substitution

[34] Even if the job finding rate is an equilibrium outcome, we can treat it as an auxiliary parameter in the estimation because the scalar of the hiring cost function, $\bar{c}_{a_i}$, can rationalize any positive value of $\lambda(\theta_{a_i})$ in equilibrium.

[35] In our baseline equilibrium, differences in the unemployment rate across labor markets arise because of heterogeneous separation rates, being the job finding rate fixed. This is consistent with the prominent role of separation rates in accounting for cross-sectional heterogeneity in unemployment documented in the literature (Gervais et al., 2016).

parameter ι, vacancy share α and meeting efficiency χ.[36] We externally set the values of ι and β from Şahin et al. (2014) and internally estimate χ to capture the meeting efficiency of the Italian economy.

Next, we proceed by discretizing the profits rebate function $\mathcal{S}(G(aw))$ over a 10-point grid. It follows that workers of the same decile in the wage distribution get rebated the same amount of profits. This allows us to identify the share of aggregate profits accruing to each decile of the wage distribution, $\mathcal{S}_d$, with the empirical share of non-labor earnings across the income distribution reported by the Survey on Household Income and Wealth (SHIW).[37] The profits rebate function is heavily skewed towards workers with high wages (see Figure A.4). Specifically, workers in the top decile of the wage distribution receive more than 70% of aggregate profits. It follows that the distribution of value added between profits and labor share has profound distributional consequences, with the aggregate labor share being a synthetic (inverse) measure of income inequality. Finally, we pin down the profits tax rate, τ, to replicate the ratio of non-labor earnings to labor income in the top decile of the income distribution. Table (1) summarizes our externally set parameters.

Table 1: Externally set parameters

Parameter	Description	Value	Source/Target
Market structure			
I	Number of industries	8	ATECO 1-digit industries (ISTAT)
A	Number of skill types	2	Median AKM worker FE (INPS)
$d(i)\omega_i(a)$	Worker population, by labor market	Tab. A.2	Employment composition, by industry and skill (INPS)
K	Number of product markets	567	ATECO 4-digit sectors (ISTAT)
M_{a_i}	Firm-to-worker population ratio, by skill	Tab. A.2	Average firm size (ISTAT)
$\bar{N}_i$	Number of potential firms in sector, by industry	Fig. (5b)	Avg number of firms in 4-digit sectors, by 1-digit industry (ISTAT)
Workers' preferences			
r	Discount rate	0.004	4% annualized int. rate
ϑ	Coeff relative risk aversion	1.000	Elminejad et al. (2022)
Labor market frictions			
ι	CES subst parameter	0.152	Şahin et al. (2014)
β	Vacancy share	0.576	Şahin et al. (2014)
δ_a	Separation rate, by skill	[0.029,0.021]	EN rate by skill
Profits rebate			
τ	Profits tax rate	0.654	Non-labor earnings to labor income, top income decile (SHIW)
$\mathcal{S}_{d=1}^{10}$	Profits rebate, by wage decile	Fig. (A.4)	Share non-labor earnings by income decile (SHIW)

Source: Model, INPS, Istat, and SHIW. *Note*: Separation rates for each skill type and the employment composition across industries and skill types are computed on matched employer-employee data (2016-2018) from INPS. Average firm size and the average number of firms in 4-digit Ateco sectors are taken from Istat data (2019). The ATECO industry classification is taken from Istat. The share of non-labor earnings by income decile is taken from the 2016 wave of the Survey on Household Income and Wealth (SHIW) administered by the Bank of Italy.

[36] We choose this more general functional form rather than the standard Cobb-Douglas specification for conservativeness about potential nonlinearities in the elasticity of firms' meeting rate to labor market tightness (see Appendix (B.1) for further details). It follows that the scope for productivity gains from worker reallocation is *more limited* than in a model with Cobb-Douglas matching function.

[37] In the data, we define non-labor income as the sum of dividend payments and all other types of compensation (both fixed and variable) that are associated to business ownership.

Internally estimated parameters. The remaining parameters are estimated internally via SMM. We normalize the skill parameter of low-skilled workers, a_L, to 1 and estimate the skill parameter of high-skilled workers, a_H, internally. We leave the shape of the productivity distributions unrestricted and identify them from the labor-market-specific wage distributions. Hence, while being jointly determined with the other parameters in the SMM routine, the productivity distributions are identified non-parametrically. We clean the raw wage distributions as follows. We start by trimming the lowest 1% of each distribution to guard against outliers, as well as picking the lowest wage of the left-trimmed distribution as reservation wage – an auxiliary moment to identify the value of home production. Next, we treat the share tr of wages to right-trim as a parameter to estimate.[38] To let the data inform our choice, we internally estimate the share of right trimming by targeting the employment share of top employers, i.e., firms with more than 250 employees. Finally, we map the measure of firms in the labor markets to the finite number of firms in the product markets by normalizing to 1 the number of firms at the lowest value of the (discretized) weighted MRPL densities inferred from the observed wage distributions. See Appendix C.4 for further details.

Let Θ be the vector of parameters still to be determined: $\Theta = \{\bar{c}_{a_i}, \zeta, a_H, \chi, \phi, s_{a_i}, \sigma, \rho, tr\}$. We choose parameter values that minimize the sum of weighted squared percentage deviations between a set of moments estimated in actual data, μ_m, and those generated by the model, $\hat{\mu}_m$:

$$\Theta^* = \arg\min_{\Theta \in \overline{\Theta}} \sum_{m \in \mathcal{M}} \left(\frac{\hat{\mu}_m(\Theta) - \mu_m}{\mu_m} \right)^2,$$

where $\overline{\Theta}$ denotes the parameter space and $\mathcal{M}$ is the set of targeted moments.

As in Engbom and Moser (2022), we estimate the scalar and convexity of the hiring cost function, $\bar{c}_{a_i}$ and ζ, to replicate the average observed NE rate in all labor markets and the employment share of firms with more than 50 employees, respectively. Shifting hiring cost up reduces vacancy posting and, as a result, the job finding rate. Raising the convexity of hiring cost restrains vacancy posting by large firms, thus compressing their employment share. Similarly, we pin down the on-the-job search intensity, s_{a_i}, to replicate the EE rate for each skill type. Higher on-the-job search intensity raises the rate at which employed workers receive job offers, thus making job-to-job transitions more frequent. This is a key variable governing the extent of frictions in each labor market and, by Equation (31), of labor market power. Turning to our original targets, we first pin down the skill parameter of high-skilled workers, a_H, by replicating the share of log-wage variance accounted for by worker-firm sorting, i.e., twice the covariance between worker and firm fixed effects, in the AKM regression.[39] The higher the skill parameter of high-skilled workers (relative

[38]The reason is twofold. First, the structure of our model predicts that firms commit to a wage rate for all their employees and that larger firms pay higher wages: it follows that a small share of high observed wages (e.g., of top managers) would be associated with extremely big firms, whose size may be unrealistic. Second, the empirical literature documents that high earners are typically more likely to bargain over their wages, as opposed to the latter being posted by the firm (Hall and Krueger, 2012; Brenzel et al., 2014). These observations suggest to remove some wages at the top of the distribution.

[39]Despite wages being log-linear in worker's skill and firm's wage piece rate, the AKM regression has no

to low-skilled), the lower firm's contribution to observed wage dispersion and, hence, the lower worker-firm sorting. As we will exploit residual wage dispersion to estimate firms' productivity, it is crucial for us to purge observed wage differentials by the component due to workers' skills. Then, we identify the meeting efficiency, χ, by replicating the vacancy share, i.e., the ratio between unfilled jobs and total (filled and unfilled) jobs. The higher the meeting efficiency, the lower the labor market tightness for given job finding rate, that is, the lower the mass of unfilled jobs (vacancies). The hiring rate elasticity of hiring costs, ϕ, is informed by the average Herfindahl-Hirschman index (HHI) in 4-digit sectors, weighted by the sectoral value added. The higher ϕ, the higher the hiring cost correction term in the equilibrium markdown (31). By acting as a shifter in the net markdown function, a higher hiring cost correction term compresses the markdowns of high-productivity firms disproportionately more. Hence, for given wage distribution, the revenue productivity distribution is more skewed to the right. It follows that product market concentration rises, as summarized by the weighted average HHI.

For product market parameters, we identify the sectoral elasticity of demand, σ, by targeting an estimate of aggregate markup in Italy (Ciapanna et al., 2022). The larger the elasticity of demand, the lower the markup of all firms. Next, we estimate the elasticity of demand across sectors, ρ– or, more precisely, the difference between σ and ρ – by targeting the semi-elasticity of firms' labor share to their market share in the 4-digit sector where they operate.[40] Intuitively, the higher ρ (with respect to σ), the higher the markup of large firms and the more negative the labor share gradient with respect to the market share.

Our estimation strategy delivers some implied parameters. First, unemployment benefits b_{a_i} are identified by the labor-market-specific reservation wage. Intuitively, the higher the unemployment benefits, the higher the reservation wage equalizing the value of unemployment to the value of employment. Next, the overhead costs κ_{a_i} is set to 90% of the operating profits generated in equilibrium by the smallest (observed) firm in each labor market. Intuitively, these fixed costs have to be large enough to prevent entry of additional firms in equilibrium. Hence, we set them so that net profits are close to zero for the marginally active firm.[41] Finally, the Cobb-Douglas industry weights are implied by the industry revenue shares consistent with the wage distributions, as described in the next paragraph. Table (2) summarizes our internally estimated parameters.

structural interpretation in our model due to differential labor market power across labor markets and no skill heterogeneity within each firm. To circumvent these issues, we proceed by simulating labor market histories generated by the stationary equilibrium of our model. Then, we assign the same identifier to firms with the same MRPL – the sufficient statistic for wage posting decisions in our model. Finally, we run the same AKM regression in the data and in the model. As a result, the sorting share of log-wage variance constitutes a legitimate reduced-form target. See Appendix C.2 for additional details.

[40] Our identification of ρ follows Edmond et al. (2015) and Fig. V of Autor et al. (2020). Since we can compute market shares only in 2019, we cannot control for firm fixed effects. Therefore, our identification exploits the cross-sectional variation within 4-digit sectors. The underlying assumption is that, in such narrowly defined markets, the shape of the production function is the same across firms. Edmond et al. (2023) uses a similar strategy by targeting the regression coefficient of sector-level labor share on sector-level HHI.

[41] We choose not to set the net profits of marginally active firms exactly to zero for firms' exits to happen smoothly in our counterfactual exercises. Indeed, Appendix D.2 highlights the importance of the exit channel

Table 2: Internally estimated parameters

Parameter	Description	Value	Target (Source)	Data	Model
Labor market parameters					
$\bar{c}_{a_i}$	Scalar hiring costs	Fig. (A.3a)	NE rate (INPS)	0.161	0.161
ζ	Vacancy convexity hiring costs	0.772	Employment share 50+ firms (ISTAT)	0.427	0.441
s_{a_i}	On-the-job search intensity	Fig. (A.3b)	EE rate, by skill (INPS)	[0.013, 0.010]	[0.013, 0.010]
a_H	Skill parameter, high skills	1.320	AKM sorting share of wage variance (INPS)	-0.066	-0.066
χ	Meeting efficiency	1.465	Vacancy rate (ISTAT)	0.011	0.011
ϕ	Hiring rate elasticity hiring costs	1.715	Weighted avg HHI, 4-digit (ISTAT)	0.077	0.077
tr	Right trimming wage distributions	0.087	Employment share 250+ firms (ISTAT)	0.262	0.255
Product market parameters					
σ	Elast. of subst. within sectors	9.171	Aggregate net markup (Ciapanna et al., 2022)	0.139	0.139
ρ	Elast. of subst. across sectors	3.049	Semi-elasticity labor share-market share (ISTAT)	-0.450	-0.449
Γ_{k_i}	Productivity distribution, by product market	Fig. (A.6)	Labor-market-specific wage distributions (INPS)		
Implied parameters					
b_{a_i}	Home production	Fig. (A.3c)	Reservation wage		
κ_{a_i}	Overhead costs	Fig. (A.3d)	90% smallest operating profits		
α_i	CD industry weights	Tab. (A.3)	Industry revenue share		

Source: Authors' calculations, model, INPS and Istat. *Note*: Labor market transition rates, wage distributions, and industrial and skill composition of wage distributions are computed on matched employer-employee data (2016-2018) from INPS. The share of employment in firms with more than 50/250 employees and the vacancy rate are taken from Istat data (2019). The HHI of 4-digit Ateco sectors and the semi-elasticity of the labor share with respect to the market share at 4-digit level are computed on the Structural Business Statistics dataset from Istat (2019).

Productivity estimation. We now detail our strategy to estimate firms' productivity distributions from the (trimmed) observed wage distributions within the SMM routine. First of all, we discretize the wage distributions on a 125-point, equally-spaced grid. We then assume that residual wage dispersion is small enough relative to the distance between grid points, so that each wage grid point corresponds to a productivity type.[42]

Then, we follow a two-step estimation procedure. In the first stage, we notice that the observed wage distributions, hiring cost correction terms, and worker transition rates are sufficient statistics to identify the equilibrium markdowns per wage level in each labor market according to (31), i.e., $\psi_{a_i}(w)$. It follows that the marginal cost of firms posting wage w in labor market a_i equals $MC_{a_i}(w) = \psi_{a_i}(w)w$. Since, in equilibrium, firms equalize MC to MRPL, this strategy allows us to characterize the model-consistent MRPL distribution in each labor market non-parametrically, that is, $\tilde{z} \sim \Phi_{a_i}(\psi_{a_i}(w)w)$ – in the spirit of the wage inversion procedure of Bontemps et al. (2000). Within each labor market, we consider a number of product markets such that each of them is populated by the average number of competing firms across 4-digit sectors in each industry, as reported in Figure (5b). Next, we assign firms to product markets so that each product market sourcing from the same labor market has the same MRPL distribution.

In the second stage, we notice that, for given CES elasticities of demand and MRPL distribution in each product market, there exists a unique combination of prices, markups,

for the aggregate effects of the MW.

[42]Technically, we estimate the discretized version of the limit case of our baseline equilibrium for a large number of productivity types.

Cobb Douglas industry weights, and physical productivities consistent with our product market structure (see the system of equations (98) in Appendix). In this way, we can finally back out the productivity distributions that rationalize the observed wage distributions through the lens of our model.

This estimation strategy is specifically geared towards guaranteeing the best possible fit to the labor-market-specific wage distributions, whose shape is critical for both the bite of the MW reforms and the scope for worker reallocation. See Appendix C.4 for further details.

Parameter estimates and model fit. The last two columns of Table (2) compare the estimation targets in the data with those implied by the estimated model. Our estimated model is successful in replicating all the targeted moments accurately.

Starting from the labor market parameters (reported in Figure (A.3)), the estimated scalars of hiring costs are increasing in skill for given industry. On the other hand, the estimated on-the-job search efficiency is negatively correlated with skill. Overhead costs appear to be small in magnitude and rather uniform across labor markets. Unemployment benefits average 52% of the reservation wage across labor markets. This is below the Italian UI replacement rate (75%). This hints at a potential role for non-monetary costs of unemployment in shaping the reservation wage The hiring cost function is estimated to be less than quadratic in vacancies (elasticity 1.77), as well as strongly increasing in the hiring rate (elasticity 1.72). This implies that the hiring cost correction term equals 1.48, thus representing a quantitatively important determinant of equilibrium markdowns. The skill parameter for high-skilled workers required to replicate the sorting share of wage variance is 1.32, indicating that nearly all of the variation in average wages across labor markets can be attributed to worker heterogeneity, rather than productivity heterogeneity among firms operating in different markets. [43]

The parameters governing the elasticity of demand are estimated to be $\sigma = 9.17$ and $\rho = 3.05$. In the light of the low level of concentration of the Italian product markets, our estimated elasticity of substitution across sectors is remarkably high. This suggests that markups of Italian firms are not as sensitive to product market concentration in international comparison (Edmond et al., 2015; Atkin et al., 2015; Grassi et al., 2017; Edmond et al., 2023; Burstein et al., 2021; De Loecker et al., 2021). For this reason, we expect our estimates of the concentration channel of the MW in Italy to provide a lower bound for other countries. Figure (A.6) in Appendix reports the estimated productivity distributions by industry. In spite of substantial variation across industries, the productivity distributions generally appear more right-skewed than log-normal. The same pattern emerges when aggregating up industry-level distributions into the aggregate productivity distribution, as reported in Figure (6b).

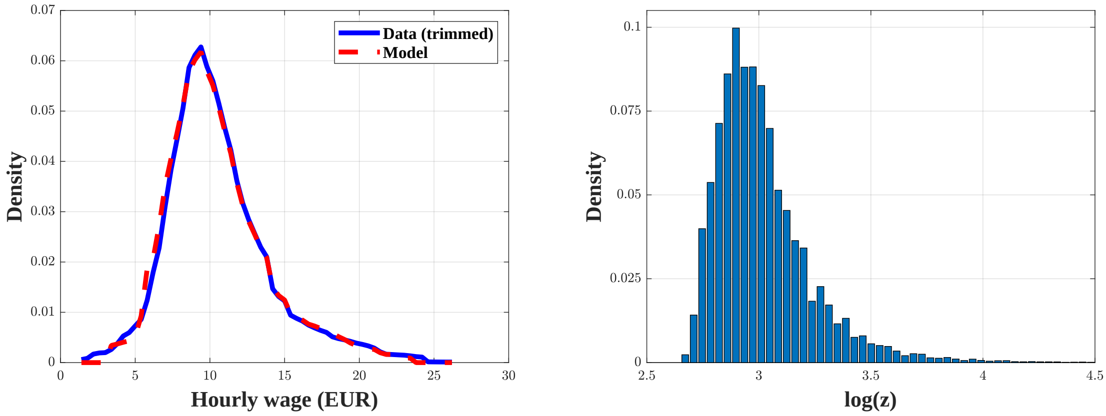
Our estimation strategy grants a virtually perfect fit to the observed wage distribution, which allows us to replicate the actual bite of MW reforms (Figure 6a). Crucially, our estimated model replicate well not only the aggregate wage distribution, but also the labor-market-

[43] Indeed, wages in high-skilled labor markets are on average exactly 32% higher than in low-skilled labor markets. Within labor markets, all wage dispersion is due to productivity differences across firms.

specific wage distributions, which are a main determinant of the equilibrium markdown function (see Figure (A.5))

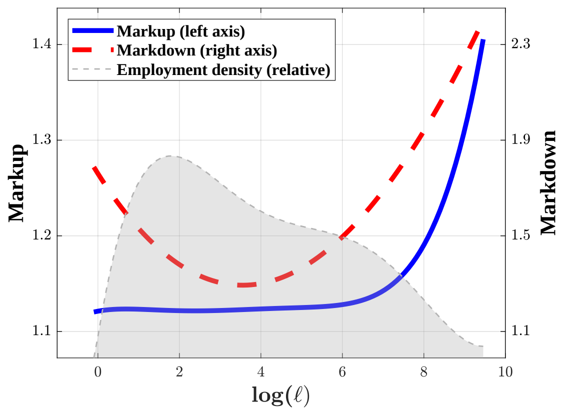
Figure 6: Aggregate wage and productivity distribution

(a) Wage distribution (b) Productivity distribution



Source: INPS data (2016-2018) and model. *Note*: Panel (a) reports the economy's employment wage distribution ($G(w)$). The wage distribution in the data (in solid blue) aggregates the 16 trimmed labor-market-specific wage distributions. See Fig. (A.2) for a comparison between the raw and trimmed employment wage distribution. Panel (b) reports the economy's distribution of log-productivity ($\hat{\Gamma}(log(z))$), which is a mixture of industry-level productivity distributions.

Figure 7: Market power distribution, by firm size



Source: Model. *Note*: The figure reports the average markup (in solid blue) and markdown (in dashed red) for given firm employment size in the baseline equilibrium of the estimated model. The area in gray represents a linear transformation of the density of employment across firm employment size such that the relative density is preserved.

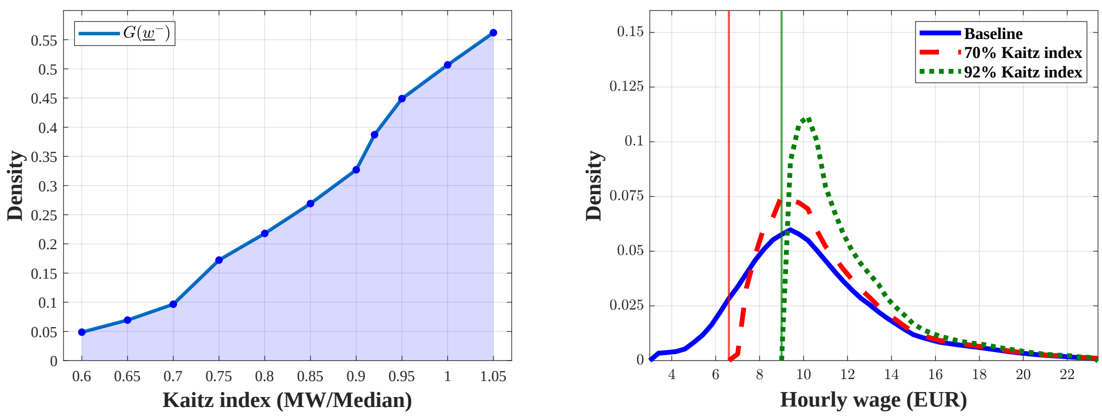
Finally, we focus on the estimated distributions of markups and markdowns by firm size, reported in Figure (7). The cross-sectional distribution of market power carries two main insights. First, the largest firms ($> 1,000$ employees) are estimated to have both higher markups and markdowns than their competitors. This means that any policy reallocating workers towards such firms is expected to reduce the aggregate labor share via a compositional effect. Second, markdowns are estimated to be U-shaped in firm size, with small firms extracting more rents than mid-sized ones in the labor market. This is a distinctive prediction of our model that aligns well with the robust (and totally untargeted) empirical pattern that, within a 1-digit industry, the average labor share in the first tercile of employment is lower than that in the second tercile (see Tab. (A.4)). [44] In terms of magnitudes, the aggregate markdown exceeds the aggregate markup (1.47 versus 1.14). Two remarks are in order. First, our estimate of the aggregate markdown is remarkably close to those of Berger et al. (2022) and Yeh et al. (2022). This entails that the extent of labor market power in Italy is comparable to the US. Second, unlike markups, markdowns in our model do *not* translate one-to-one into pure profits. The reason is that part of the profit margin ensured by markdowns covers hiring costs. Figure (A.7) provides a decomposition of firms' revenues into the different sources of profits and costs. It shows that profits from labor and product market power account for very similar shares of total revenues.

5 Equilibrium Effects of the Minimum Wage

In this section we study the equilibrium effects of the MW in our model economy. To do so, we make use of the estimated model as a laboratory to simulate the response of the Italian economy to a wide range of MW reforms. Our simulations span the entire spectrum of MW reforms from 0 to 50% of directly affected workers.[45] In what follows, we compare stationary equilibria of our model by varying the MW parameter $\underline{\mathbf{w}}$.[46]

Figure 8: MW bite and wage distribution by reform

(a) Share directly affected workers by reform (b) Wage distribution response



Source: Model. *Note*: Panel (a) plots the share of directly affected workers (i.e., workers earning strictly less than the mandated MW in the baseline equilibrium) against the Kaitz index of the MW reform. Panel (b) reports the density of the aggregate employment wage distribution $G(w)$ in the baseline equilibrium (solid blue line), after a 70%-Kaitz-index MW reform (red dashed line), and after a 92%-Kaitz-index MW reform (green dotted line).

Following the literature, we measure MW reforms through their *Kaitz index*, i.e., the

[44] Models of oligopolistic and oligopsonistic competition where product and labor market boundaries coincide, such as Deb et al. (2023), imply identical cross-sectional behavior of markups and markdowns.

[45] Due to right trimming, the estimated model underestimates the actual median wage. For this reason, we set our largest reform one grid point above the model-implied median (55% of workers directly affected).

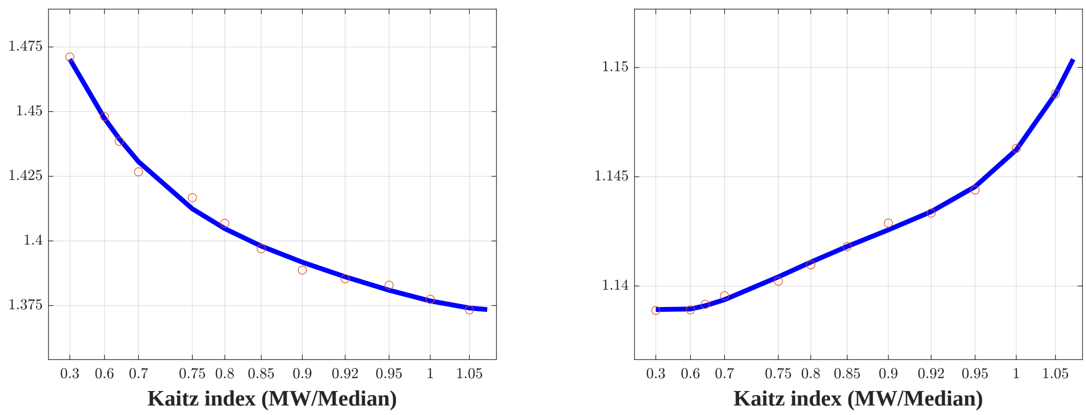
[46] This comparative statics exercise is informative about the effects of the MW in the medium term, while being silent on the transitional dynamics of the labor market towards the new equilibrium (short-term adjustment), as well as on the potential substitution away from labor, automation, and entry decisions induced by the MW (long-term adjustment).

ratio between the MW and the pre-reform median wage. However, the Kaitz index is not a sufficient statistic to assess the actual impact of the MW – specifically, the share of workers directly affected – because it ignores the shape of the baseline wage distribution. Figure (8a) provides a key to map the Kaitz index of each counterfactual reform to the share of workers directly affected, thus making international comparisons easier. For instance, the 2015 MW reform in Germany directly affected 15% of workers (Dustmann et al., 2022), which would correspond to a Kaitz index of approximately 75% in our model economy. Figure (8b) reports the response of the wage distribution for two MW reforms, namely Kaitz indices of 70% (in dashed red) and 92% (in dotted green), which will turn out to be important benchmarks. As expected, the MW induces a rightward shift of the wage distribution. This happens for two reasons. First, constrained firms raise their offered wage in compliance to the MW. In turn, this induces spillover effects higher up in the wage distribution, as some unconstrained firms also raise their offered wages to preserve their rank in the wage offer distribution. Hence, wages increase due to such behavioral effects – especially in the bottom half of the distribution. Second, most constrained firms lay off workers, who (imperfectly) reallocate towards higher-paying firms. Therefore, the wage distribution shifts rightward also because of positive selection of higher-paying firms and higher unemployment among low-earning workers, that is, a compositional effect (see Table (A.6) for a decomposition).

We now turn to studying the aggregate effects of the MW reforms on the Italian economy. We start by considering the response of the aggregate markup and markdown, that is, how market power reacts to the MW.

Figure 9: Market power effects

(a) Aggregate markdown response (b) Aggregate markup response



Source: Model. *Note*: Aggregate markdown and markup are cost-weighted averages of the respective firm-level variables, where weights are given by the wage bill. See Appendix (B.11) for the model-consistent aggregation. The x-axis is scaled so as to reflect the share of directly affected workers. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

Figure (9a) shows that the aggregate markdown is decreasing in the MW at a progressively slower rate. This size-dependent pattern results from the net effect of two opposing forces. On the one hand, firms that are affected by the MW – either directly (constrained) or indirectly (due to spillover effects) – reduces their markdowns. This behavioral effect drives down the aggregate markdown. On the other hand, displaced workers from the most affected firms reallocate towards larger firms, which have higher baseline markdown on average (see Figure (7)). This compositional effect progressively slows down the reduction of the aggregate markdown as the MW increases. Overall, the aggregate markdown decreases by 9.8pp (20.7% net markdown reduction) from the baseline equilibrium to the highest MW considered, where it troughs.

Figure (9b) shows that the aggregate markup is increasing in the MW at a progressively faster rate. This reflects both the concentration channel of the MW (behavioral effect) and worker reallocation towards higher-markup firms (compositional effect).[47] Overall, the aggregate markup increases by 1.0pp (7.1% net markup increase) from the baseline equilibrium to the highest MW considered, where it peaks. The aggregate markup response is driven by the manufacturing industry, which is characterized by the highest product market concentration (see Figure (A.8a)). Indeed, the markup response is stronger in more concentrated product markets, as measured by their HHI. We estimate that an increase in the baseline HHI of one standard deviation (starting from the mean) raises the elasticity of the sectoral net markup to the MW by 24%.

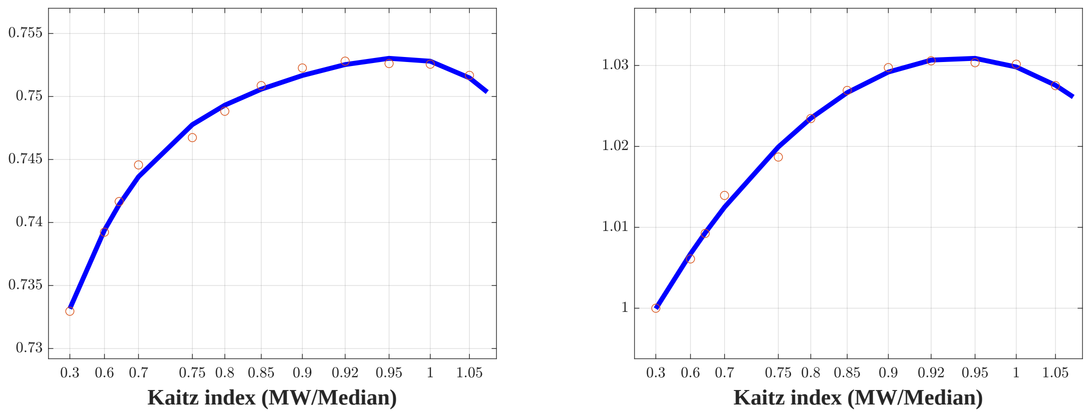
In Appendix D we show how to decompose the change in aggregate market power (adjusted for hiring costs) due to markdowns and markups. It turns out that, on average, the aggregate markup response accounts for a quarter of the (gross) aggregate market power response. However, the importance of the aggregate markup response is strongly MW-size-dependent. For Kaitz indices until 80%, the aggregate markdown response dwarfs the aggregate markup response. On the other hand, for larger MW reforms, the two responses are of the same order of magnitude (see Figure (A.14b)) – with the aggregate markup response even dominating for the highest MW reforms.

We now focus on the aggregate labor share and value added response to the MW. Figure (10a) reports the aggregate labor share response to the MW. From our counterfactual experiments, we uncover a hump-shaped response of the aggregate labor share to the MW, peaking at a Kaitz index of 92%. As in the stylized model, the aggregate labor share response is driven by the net market power response. Hence, a joint reading of Figures (9)-(10a) allows us to single out the determinants of the hump-shaped pattern of the aggregate labor share. On the one hand, the behavioral effect restraining the aggregate markdown is the dominating force behind the initial increase of the aggregate labor share. On the other hand, the flattening and eventual decline of the aggregate labor share is driven by the concentration channel raising the aggregate markup. Overall, the aggregate labor share increases by 1.9pp from the baseline equilibrium to the highest MW considered, with an increase by 2.0pp at peak in correspondence to a Kaitz index of 92%. In the manufacturing industry, where the concentration channel is quantitatively more relevant, the labor share response is more muted and the humped shape more pronounced (see Figure (A.8b)). We estimate that an increase in the

[47] Notice that the aggregate markup response is estimated conservatively. This is because aggregate profits reduce with the MW (see Fig. (A.9b)). Thus, a model where the mass of potential firms were pinned down by a free entry condition would entail more firm exit than ours.

Figure 10: Aggregate effects of the MW

(a) Aggregate labor share response (b) Aggregate value added response



Source: Model. *Note*: Aggregate labor share is defined as $LS \equiv \frac{WL}{C}$ and aggregate value added is defined as $C = Y - \mathcal{HC} - \mathcal{FC}$, where $\mathcal{HC}$ is total hiring costs and $\mathcal{FC}$ total overhead costs. See Appendix (B.11) for the model-consistent aggregation. The x-axis is scaled so as to reflect the share of directly affected workers. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

baseline HHI of one standard deviation (starting from the mean) reduces the elasticity of the sectoral labor share to the MW by 22%.

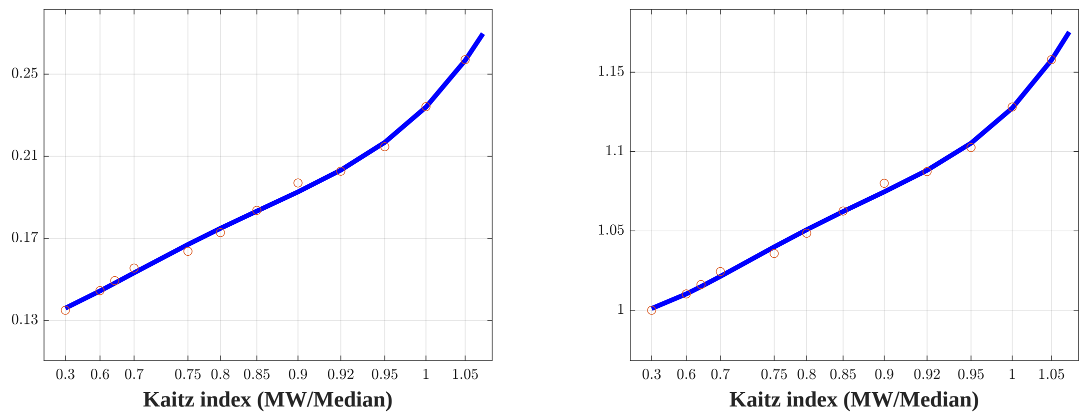
Figure (10b) reports the aggregate value added response to the MW. Exactly as for the labor share, the aggregate value added is hump-shaped in the MW, with a peak at a Kaitz index of 92%. Overall, the aggregate value added increases by 2.8% from the baseline equilibrium to the highest MW considered, with an increase by 3.1% at peak in correspondence to a Kaitz index of 92%. Notice that the hump-shaped profile of the aggregate labor share and value added response to the MW was the main result coming from our stylized model (see Proposition (1). Hence, this result remains valid even when accounting for compositional effects.

Next, we explore the determinants of the positive value added response. Figure (A.9a) decomposes the value added response into (gross) output and aggregate hiring costs. While output is virtually constant up to Kaitz indices of 100%, aggregate hiring costs decrease with the MW. This means that aggregate hiring costs per unit of final good – our index of misallocation – reduce, as well. Indeed, Figure (A.10b) shows that producing one unit of final good after implementing a Kaitz index of 92% requires 14% lower hiring costs than in the baseline equilibrium. Intuitively, the reduction in vacancy posting by low-productivity firms induced by a higher MW increases the vacancy filling rate of high-productivity firms, thus allowing them to grow larger in size due to lower marginal costs. Hence, the reduction in labor misallocation is the main driver behind the positive value added response.

We are now interested in breaking down the output response into employment and productivity effects. Figure (11a) shows that the unemployment rate increases with the MW in an almost linear fashion. This is due to product market power – which brings about decreasing returns in revenues – and the estimated convexity of the hiring cost function, which imply that worker reallocation following the MW happens imperfectly, i.e., produce some excess unemployment. The unemployment rate increases by 2.0 (6.8)pp in correspondence to a Kaitz index of 70 (92)%. To assess the magnitude of the (dis)employment effects, in Figure (A.10a) we plot the share of affected jobs, i.e., paying less than the MW before the reform, destroyed by each MW reform.[48] On the one hand, perfect competition on the labor market would imply that all affected jobs get destroyed, that is, a share of one. On the other hand, models of pure monopsony with exogenous job creation (or linear hiring costs) and CRS revenue function, such as Burdett and Mortensen (1998), would imply no disemployment effects, that is, a share of zero. Our results stand midway between these two opposite views, although closer to the monopsony benchmark. Interestingly, we uncover a rather stable share of 21% of affected jobs destroyed. This means that an average of 1 in 5 jobs paying less the MW is destroyed by the policy. The estimated magnitude of the employment effects aligns well with reduced-form estimates from other countries.[49]

Figure 11: Employment and productivity effects of the MW

(a) Unemployment rate response (b) Labor productivity response



Source: Model. *Note*: Labor productivity is defined as $\frac{Y}{L}$. The x-axis is scaled so as to reflect the share of directly affected workers. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

Hence, output remains stable *despite* a reduction in headcount employment. It follows that the reduction in employment is fully offset by productivity gains from worker reallocation. Figure (11b) documents that labor productivity strongly increases with the MW. Labor productivity rises by 2.4 (8.7)% in correspondence to a 70 (92)% Kaitz index. Given that employment losses are concentrated among low-skilled workers and low-productivity firms, the labor productivity response reflects both genuine productivity gains from worker reallocation and positive selection of worker-firm matches. However, only the former contribute to increasing output.

[48]The underlying assumption behind this statistic is that no jobs paying more than the MW before the reform get destroyed.

[49]To assess the (dis)employment effects of the MW, we compute the semi-elasticity of the unemployment rate to the MW among among two commonly used – and model-consistent – categories of *minimum wage workers*, that is, low-skilled workers and low-skilled workers in hospitality. We find that, when raising the MW from 60 to 70% of the current median wage (arguably the most comparable reform to those studied in the existing literature), the semi-elasticity equals $-0.10(-0.14)$ for low-skilled workers (low-skilled workers in hospitality). This figure is in line with existing reduced-form estimates (Belman and Wolfson, 2019; Dube and Lindner, 2024).

In sum, our results tell us that introducing a relatively large MW has the potential to raise aggregate value added (consumption) up to 3.1%, but at the cost of a significant surge in the unemployment rate. How to solve this trade-off optimally will be the object of the next section.

6 Welfare Analysis

In the previous section we have shown that introducing a MW generally raise both aggregate value added and the unemployment rate simultaneously. Hence, the welfare effects of the MW are a priori ambiguous. On the one hand, higher value added grants agents higher aggregate lifetime consumption (*level effect*). On the other hand, higher unemployment risk entails higher consumption volatility (*uncertainty effect*). Moreover, the MW is likely to induce some redistribution across worker types (*distributional effect*). To strike a balance between these forces, we resort to a utilitarian social welfare function, which provides us with a criterion to solve the trade-off optimally. Consistently with the steady-state behavior adopted by firms in our model economy, we focus on steady-state comparisons.[50]

Notice that, if agents were risk neutral or asset markets were complete, the welfare-maximizing MW would simply boil down to its consumption-maximizing level. Hence, a Kaitz index of 92% is the upper bound to the set of potentially efficient MW reforms. As our model abstracts from intertemporal consumption-savings decisions, workers have no means to mitigate consumption volatility. Hence, we interpret the welfare-maximizing MW as a lower bound to the set of potentially efficient MW reforms. The welfare-maximizing MW, $\underline{\mathbf{w}}^\star$, is defined as:

$$\underline{\mathbf{w}}^\star = \arg\max_{\underline{w}} \sum_{i=1}^{I} \int \left[u_{a_i}(\underline{w}) U_{a_i}(\underline{w}) + (1 - u_{a_i}(\underline{w})) \int W_{a_i}(aw; \underline{w})\, dG_{a_i}(w; \underline{w}) \right] d\Omega_i(a)\, \Xi(i), \quad (33)$$

where U_{a_i} and $W_{a_i}(w)$ are the labor-market-specific asset values of unemployment and employment at wage aw, respectively, defined in (12)-(13). Henceforth, we refer to the welfare-maximizing MW, $\underline{\mathbf{w}}^\star$, as *optimal* MW.

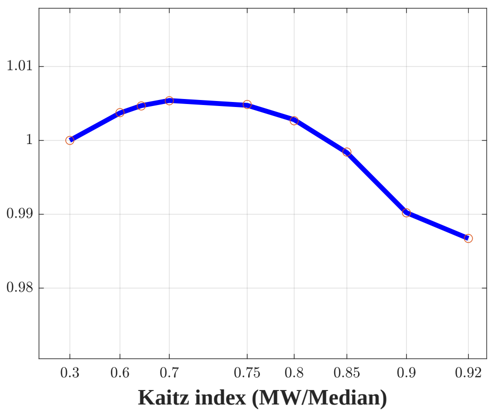
Figure (12a) shows that aggregate utilitarian welfare is maximized in correspondence to a Kaitz index of 70%. Such an optimal MW brings about a welfare gain of 0.5% in consumption equivalent units, in the face of a 1.4% increase in aggregate consumption.[51] In Appendix E we decompose the aggregate welfare response into level, uncertainty, and distributional effects (Floden, 2001). Figure (A.17) shows that, for Kaitz indices up to 75%, the increase in aggregate consumption (level effect) dominates the increase in consumption volatility due to

[50] Admittedly, our approach ignores both the short-run welfare effects induced by the transition towards the new stationary equilibrium and the long-run welfare effects induced by firm entry, capital-labor substitution, and automation incentives, which are out of the model. Since none of these effects are directly related to product market power, we consider the steady-state comparison as a satisfactory benchmark for welfare assessment.

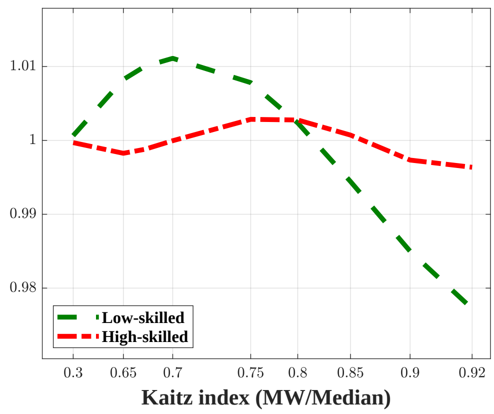
[51] Berger et al. (2024) estimates a welfare gain of 0.2% in consumption equivalent units from implementing the optimal MW in the US. The additional expansionary effect we estimate likely arises from the reduction in congestion effects induced by the MW in frictional labor markets, which Berger et al. (2024) abstracts from.

Figure 12: Welfare effects of the MW

(a) Aggregate welfare



(b) Welfare effects, by skill



Source: Model. *Note*: The The left panel a) report the consumption-equivalent percentage variation of the aggregate utilitarian welfare, i.e., the maximand of equation (33), across MW reforms in solid blue. The right panel b) decomposes the utilitarian welfare response into low-skilled and high-skilled welfare response. The x-axis is scaled so as to reflect the share of directly affected workers.

heightened unemployment risk (uncertainty effect) in welfare terms. Still, welfare gains from redistribution to poorer workers (distributional effect) are maximized at a Kaitz index of 65%. The optimal MW balances efficiency and redistribution.

To investigate the determinants of the distributional effect, Figure (12b) reports the welfare effects of the MW across skill types. Welfare of low-skilled workers is maximized at a Kaitz index of 65%. Welfare of high-skilled workers is maximized at a Kaitz index of 75%. Kaitz indices up to 65% redistribute consumption shares towards low-skilled workers. As a result, welfare inequality drops, which reflects in a positive distributional effect. On the contrary, larger Kaitz indices hit significantly harder low-skilled than high-skilled workers, which reflects in a negative distributional effect. The reason is that most of the increase in aggregate unemployment is borne by low-skilled workers themselves. Therefore, our results express a word of caution about the idea of raising the MW to unprecedented levels in order to reduce inequality. In fact, Kaitz indices higher than 65% would end up *increasing* welfare inequality.

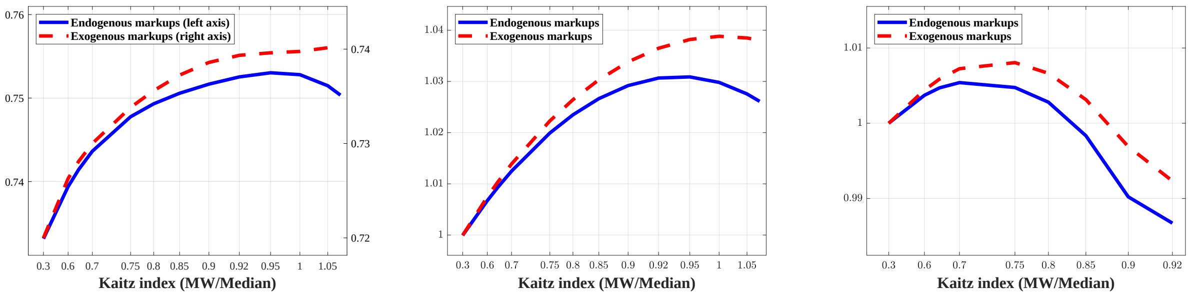
In sum, the optimal MW equals a Kaitz index of 70% and brings about aggregate welfare gains of 0.5% consumption equivalent units. The welfare effects of the MW are heterogeneous across skill types, which generates winners and losers. Welfare of low-skilled workers increases with Kaitz indices up to 65%. High-skilled workers are the relative beneficiaries of larger Kaitz indices.

7 The Role of Endogenous Markups

In the previous two sections we have tackled positive and normative questions related the introduction of a MW in an economy with endogenous markups. In this section, we aim to single out the role of endogenous markups in driving our results. With this goal in mind, we repeat the same policy experiments of Section (5) in an observationally equivalent economy – that is, matching the same empirical moments as our baseline – with monopolistic competition in sectoral markets, which implies exogenous and identical markups across firms.[52] See Appendix F for further details on the alternative economy.

Figure 13: Aggregate response: Endogenous vs exogenous markups

(a) Aggregate labor share (b) Aggregate value added (c) Aggregate welfare



Source: Model. *Note*: The *endogenous markup* economy in solid blue is our baseline economy with oligopolistic competition in sectoral markets, while the *exogenous markup* economy in dashed red is an observationally equivalent economy to our baseline with monopolistic competition in sectoral markets. The x-axis is scaled so as to reflect the share of directly affected workers. Panel (a) compares the aggregate labor share response in levels, while Panels (b) and (c) compare the percentage change of aggregate value added and aggregate welfare.

Figure (13) reports the response of the aggregate labor share, value added and welfare across the two economies. Our baseline economy with endogenous markups is depicted in solid blue, the alternative economy with exogenous markups in dashed red. From the response of aggregate variables, we draw three main takeaways. First, the aggregate labor share would be monotonically increasing in the MW if markups were exogenous. Consistently with the size-dependent response of the aggregate markup (see Figure (9b)), qualitative differences between the dynamics of the aggregate labor share across the two economies show up for Kaitz indices higher than 90%. Second, aggregate value added would peak at a Kaitz index of 100% (instead of 92%) if markups were exogenous. Aggregate value added is higher in the economy with exogenous markups across the entire spectrum of MW reforms. The consumption-maximizing MW brings about higher value added gains than in the baseline economy by 0.9pp (28%). The difference is accounted for both by a lower unemployment rate and a higher labor productivity in the economy with exogenous markups. Indeed, firms with higher markups pass through marginal cost shocks relatively more to prices than to quantities (labor demand). Since small firms have higher markups in the economy with exogenous markups, the unemployment response is lower. Since large firms have lower markups in the economy with exogenous markups, they expand more in response to a reduction in marginal hiring costs, thus contributing to a higher aggregate productivity (see Figure (A.20)). Finally, the economy with exogenous markups calls for a higher optimal MW than the baseline. Specifically, the optimal MW equals a Kaitz index of 75% (rather than

[52] The main structural difference between our baseline economy with endogenous markups and the alternative economy with exogenous markups is the map between observed wages and underlying firm's productivity. Indeed, we re-estimate the alternative model according to the same strategy outlined in Section (4) to make sure that the MW bite is exactly the same across the two economies and the simulation results are comparable. We interpret differential results across the baseline and the alternative economy as *structural* counterfactuals, i.e., what would happen by ignoring endogenous markups when estimating the equilibrium effects of the MW.

70%), with associated welfare gain of 0.8% (rather than 0.5%) consumption-equivalent units. Hence, if markups were exogenous, the optimal MW would yield nearly 50% larger welfare gains.

In Appendix F we show that the differential aggregate effects across the two economies can be further broken down into a behavioral and a compositional component. Moreover, we document that virtually all of the differential aggregate effects across the two economies stem from the manufacturing industry, whose sectoral markets are the most concentrated in the baseline estimated model. Overall, we conclude that endogenous markups matters for quantifying the equilibrium effects of MW reforms.

8 Empirical Validation

In this section we provide empirical validation to the concentration channel of the MW. The concentration channel predicts that (i) a higher MW reallocates market share to high-productivity firms, and (ii) high-productivity firms raise markups, especially in already concentrated product markets.

We test these predictions by leveraging Italian firms' balance sheet data from CERVED and exploiting the variation induced by industry-specific contractual wage floors published by Istat.[53] Absent a nationwide MW, the Italian wage setting system is based on contractual wage floors agreed through a centralized collective bargaining process, which apply to all private-sector employees. These contractual wages are set at the national level at predetermined dates. Hence, they represent an important source of variation of labor costs to firms, just like the introduction of a MW (see Appendix G.1 for further details on the institutional background). Differently from the latter, however, collective bargaining agreements set an entire array of contractual wage floors that vary by job title. As a consequence, the direct effects of an increase in contractual wages are not necessarily fully concentrated in the bottom part of the wage distribution. However, as long as the effects are heterogeneous across firms, the concentration channel is still expected to operate in the same way.[54] Potential endogeneity concerns arise because the timing and size of the adjustment may be correlated with the business performance of the respective industry. Similar concerns are often raised regarding increases in MW, as well (Neumark, 2017). On the other hand, a key advantage of using industry-specific contractual wage floors is that they allow us to control for aggregate shocks, due to the staggered nature of the collective bargaining agreements across different industries.

Overall, our estimating sample from CERVED includes about 6.2 million yearly observations covering between 326,000 and 476,000 firms each year, corresponding to more than 20% of total private-sector firms (Devicienti and Fanfani, 2021).[55] The total number of employees

[53]CERVED is an Italian business information agency, which provides data on incorporated firms.

[54]Heterogeneous effects occur e.g. if some firms have smaller profit margins or a stronger bite of contractual wage floors due to smaller *wage cushions* (Cardoso and Portugal, 2005).

[55]CERVED data cover the universe of incorporated Italian firms. Hence, large firms are over-represented with respect to the general firm population, even if the bias is less severe than in Compustat.

working at firms included in our sample ranges between 7.5 and 9.9 million, representing between 63 and 77% of total private-sector employees. Hence, our data captures the bulk of employment in the Italian private-sector economy. See Appendix G.2 for additional details on the data sources.

Reallocation effects. The first prediction of concentration channel is that a higher MW reallocates market share to high-productivity firms. Therefore, a testable implication is that the firm-level value added response is increasing in productivity. To bring this testable implication to the data, we first assign firms to their most frequent quintile of the 4-digit-sector-year distribution of log value added per worker – a proxy of productivity. We then estimate the following regression model:

$$\log VA_{i,t} = \sum_{q=1}^{5} \beta_q \log MW_{s_3(i),t} \cdot \mathbb{1}_{\{Q(i)=q\}} + \gamma_{s_2(i)} \cdot \phi_t + \alpha_i + \epsilon_{i,t}, \tag{34}$$

where $\log VA_{i,t}$ is the natural logarithm of value added of firm i at time t, $\log MW_{s_3(i),t}$ is the natural logarithm of the contractual wage floor of the 3-digit industry in which firm i operates, $\gamma_{s_2(i)} \cdot \phi_t$ are 2-digit-industry-specific time trends, α_i are firm fixed effects and $\epsilon_{i,t}$ is an idiosyncratic error term. The interaction term between the contractual wage floor and the quintile indicators allows for heterogeneous value added response by productivity bin. We find that the firm-level elasticity of value added to the wage floor varies strongly across productivity, ranging from -1.5 for the first bin to above 1 for last bin (Figure A.21), implying reallocation of market share towards high-productivity firms. These findings are consistent with a large number of empirical studies documenting a spike in firm exit across the least productive firms and significant worker reallocation towards larger firms upon introducing a MW (Draca et al., 2011; Aaronson et al., 2018; Mayneris et al., 2018; Chava et al., 2019; Dustmann et al., 2022; Devicienti and Fanfani, 2021).[56]

A mechanical implication of the heterogeneous value added response by productivity is that sectoral concentration and measured labor productivity both rise, driven by compositional effects. See Appendix G.3 for further details.

Concentration effects. The second prediction of the concentration channel is that high-productivity firms raise markups, especially in already concentrated product markets. Absent price data, we focus on the firm-level labor share response as a proxy for the market power response to guard against potential bias in estimated markups (Bond et al., 2021).[57] We single out two testable implications. First, the firm-level labor share response is decreasing in

[56]Note that our estimates represent a lower bound for reallocation, given that the estimating sample includes only surviving firms.

[57]Thanks to product-level custom data, Dodini et al. (2023) documents a reduction in markdowns and an increase in markups by large firms in response to an exogenous shift in union density, i.e. the firm-level share of unionized workers. Interestingly, the markup response is present only in sectors where significant market shares are reallocated from small to large firms. Insofar as unions bargain over pay floors with employers, exogenous shifts in union density are akin to setting a MW in our model.

product market concentration. Second, the labor share response of high-productivity firms in concentrated product markets is negative, driven by a positive profit response.

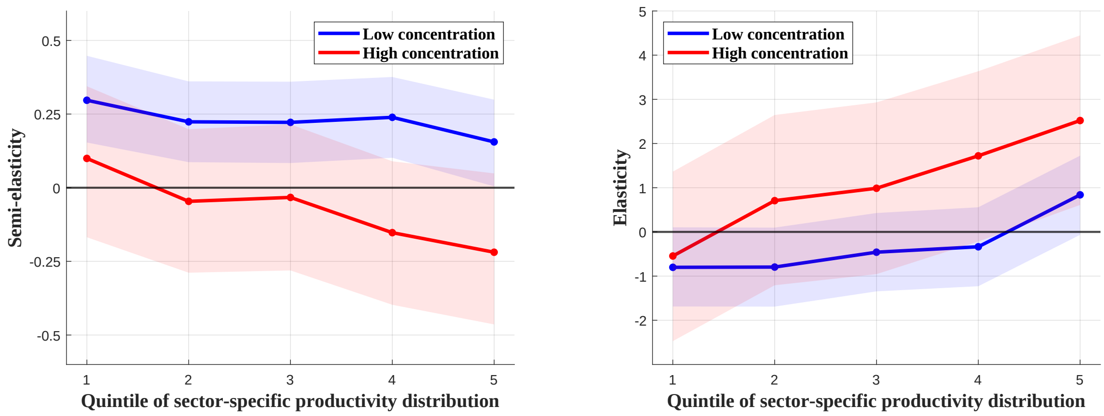
To study heterogeneous effects by productivity and product market concentration, we estimate the following model:

$$Y_{i,t} = \sum_{q=1}^{5} \beta_q \log MW_{s_3(i),t} \cdot \mathbb{1}_{\{Q(i)=q\}} + \gamma_{s_2(i)} \cdot \phi_t + \left[\sum_{q=1}^{5} \beta_q^{'} \log MW_{s_3(i),t} \cdot \mathbb{1}_{\{Q(i)=q\}} + \gamma_{s_2(i)}^{'} \cdot \phi_t \right] \cdot \mathbb{1}_{\{HHI_{s_4(i)} > \overline{HHI_{s_4}}\}} + \alpha_i + \epsilon_{i,t}, \tag{35}$$

which corresponds to the model (34) interacted with a dummy indicating whether the Herfindahl-Hirschman index (HHI) of concentration in the sector in which firm i operates is above the weighted average of the economy.[58] The outcome variables, $Y_{i,t}$, are either the labor share of firm i at time t – defined as the ratio between total labor costs and value added – or accounting profits (EBITDA). Standard errors are clustered at the firm level.

Figure 14: Firm-level response by concentration and productivity

(a) Labor share (b) Profits (EBITDA)



Source: CERVED (2005-2020), INPS and Istat data. *Note*: The graph reports the average semi-elasticity of firm-level labor share and the elasticity of firm-level profits (EBITDA) to the MW by productivity bins, computed from the regression model (35). Standard errors are clustered at the firm-level. Shaded areas represent 90% confidence intervals.

The regression results are plotted in Figure (14), which uncovers three important results. First, the labor share response is decreasing in productivity. Second, the labor share response is higher in low concentrated markets than in highly concentrated markets across the entire productivity distribution. This confirms our first testable implication that the labor share response is decreasing in product market concentration. Third, the labor share response of high-productivity firms in highly concentrated product markets is negative, driven by a positive profits response.[59] This validates our second testable implication. The increase

[58] We measure the HHI of value added in 2019 in our Istat data, which are representative of the universe of Italian firms with employees. This is preferable to using the HHI computed within CERVED, which is biased towards relatively large firms.

[59] Only the profits response is statistically different from zero with clustered standard errors. Results are

in profits hints at a positive markup response – rather than capital-labor substitution – as the main driver of the negative labor share response. In Appendix G.5 we repeat the same analysis using common proxies of labor market concentration and show that labor market power is not a likely confounder of our results. In fact, we find that the labor share response is higher in *highly* concentrated labor markets than in low concentrated labor markets across the entire productivity distribution (see Figure (A.24)) – in line with the empirical literature (Azar et al., 2019a; Popp, 2023). We conclude that our reduced-form estimates are supportive of a significant role of the concentration channel of the MW.

9 Conclusions

This paper studies the equilibrium effects of the minimum wage when product market power is endogenous and varies with market competition. First, we developed a stylized model to formalize a novel concentration channel of the minimum wage. A higher minimum wage reallocates workers from small to large firms. Large firms gain market share and increase their price markups. Second, we contributed a structural model where heterogeneous workers and firms interact in frictional labor markets and oligopolistic product markets. The structural model features both endogenous markups and markdowns, which react to changes in the competitive environment induced by policy reforms, such as the minimum wage. Third, we made use of the estimated model on the Italian economy to simulate the effects of a wide range of minimum wage reforms. Our results suggest that both the aggregate labor share and value added are hump-shaped in the minimum wage, due to the opposing responses of markups and markdowns. The optimal minimum wage, which trades off reallocation gains against employment losses, equals 70% of the current median wage. Ignoring endogenous markups would lead to a monotonically increasing labor share, as well as to higher consumption-maximizing and optimal minimum wage. Finally, we provided empirical validation to the concentration channel on Italian firms' balance sheet data. We documented that the firm-level labor share response to the minimum wage is decreasing in product market concentration, and that this effect is driven by high-productivity firms in highly concentrated product markets raising profits.

The main takeaway from this paper is that product market structure is important for the equilibrium effects of the minimum wage. When product markets feature a few dominant firms, raising the minimum wage may not be advisable.

An interesting avenue for future research is assessing the welfare effects of the transitional dynamics to the new stationary equilibrium following the introduction of a minimum wage. Possible extensions to our framework are the addition of capital and an input-output structure, which may play a role in the transmission of the minimum wage to the economy.

unchanged if we look at the elasticity of the labor share to the MW, i.e., if we use log labor share as outcome variable. See Appendix G.4 for the other margins of adjustment.

References

Aaronson, D., French, E., Sorkin, I., and Toi, T. (2018). Industry dynamics and the minimum wage: A putty-clay approach. *International Economic Review*, 59(1).

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. (2001). Good Jobs versus Bad Jobs. *Journal of Labor Economics*, 19(1):1–21.

Adamopoulou, E. and Villanueva, E. (2022). Wage determination and the bite of collective contracts in italy and spain. *Labour Economics*, 76:102147.

Ahlfeldt, G. M., Roth, D., and Seidel, T. (2022). Optimal minimum wages.

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031.

Atkin, D., Chaudhry, A., Chaudhry, S., Khandelwal, A. K., and Verhoogen, E. (2015). Markup and cost dispersion across firms: Direct evidence from producer surveys in pakistan. *American Economic Review*, 105(5):537–44.

Autor, D., Dorn, D., Katz, L. F., and Van Reenen, J. (2020). The fall of the labor share and the rise of superstar firms. *Quarterly Journal of Economics*, pages 645–709.

Azar, J., Huet-Vaughn, E., Marinescu, I., Taska, B., and von Wachter, T. (2019a). Minimum wage employment effects and labor market concentration. Technical report, National Bureau of Economic Research.

Azar, J., Marinescu, I., and Steinbaum, M. (2019b). Measuring labor market power two ways. In *AEA Papers and Proceedings*, volume 109, pages 317–321. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Azar, J. A., Berry, S. T., and Marinescu, I. (2022). Estimating labor market power. Technical report, National Bureau of Economic Research.

Belman, D. and Wolfson, P. (2019). 15 years of research on us employment and the minimum wage. *Labour*, 33(4):488–506.

Berger, D., Herkenhoff, K., Kostøl, A. R., and Mongey, S. (2023). An Anatomy of Monopsony: Search Frictions, Amenities, and Bargaining in Concentrated Markets. In *NBER Macroeconomics Annual 2023, volume 38*, NBER Chapters. National Bureau of Economic Research, Inc.

Berger, D., Herkenhoff, K., and Mongey, S. (2022). Labor market power. *American Economic Review*, 112(4):1147–93.

Berger, D., Herkenhoff, K., and Mongey, S. (2024). Minimum wages, efficiency and welfare. Technical report.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, 60(4):889–917.

Bilal, A. G. and Lhuillier, H. (2021). Outsourcing, inequality and aggregate output. Technical report.

Biondi, F. (2022). Firm productivity and derived factor demand under variable markups. Technical report, Working paper, Katholieke Universiteit Leuven, Leuven.

Boeri, T. (2012). Setting the minimum wage. *Labour Economics*, 19(3):281–290.

Boeri, T., Ichino, A., Moretti, E., and Posch, J. (2021). Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces. *Journal of the European Economic Association*, 19(6):3249–3292.

Bond, S., Hashemi, A., Kaplan, G., and Zoch, P. (2021). Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data. *Journal of Monetary Economics*, (121):1–14.

Bontemps, C., Robin, J.-M., and Van den Berg, G. J. (1999). An empirical equilibrium job search model with search on the job and heterogeneous workers and firms. *International Economic Review*, 40(4):1039–1074.

Bontemps, C., Robin, J.-M., and Van den Berg, G. J. (2000). Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review*, 41(2):305–358.

Brenzel, H., Gartner, H., and Schnabel, C. (2014). Wage bargaining or wage posting? evidence from the employers' side. *Labour Economics*, 29:41–48.

Burdett, K. and Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, 51(4):955–969.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273.

Burstein, A., Carvalho, V. M., and Grassi, B. (2021). Bottom-up markup fluctuations. Technical report, National Bureau of Economic Research.

Cardoso, A. R. and Portugal, P. (2005). Contractual wages and the wage cushion under different bargaining settings. *Journal of Labor economics*, 23(4):875–902.

Chava, S., Oettl, A., and Singh, M. (2019). Does a one-size-fits-all minimum wage cause financial stress for small businesses. Technical report.

Ciapanna, E., Formai, S., Linarello, A., and Rovigatti, G. (2022). Measuring market power: macro and micro evidence from italy. Technical report.

Dávila, E. and Schaab, A. (2022). Welfare assessments with heterogeneous individuals. Technical report, National Bureau of Economic Research.

De Loecker, J., Eeckhout, J., and Mongey, S. (2021). Quantifying market power and business dynamism in the macroeconomy. Technical report, National Bureau of Economic Research.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics*, 135(2):561–644.

De Ridder, M., Grassi, B., and Morzenti, G. (2022). The hitchhiker's guide to markup estimation. Technical report.

Deb, S., Eeckhout, J., Patel, A., and Warren, L. (2022). What drives wage stagnation: Monopoly or monopsony? Technical report, UPF mimeo.

Deb, S., Eeckhout, J., Patel, A., and Warren, L. (2023). Market power and wage inequality. *Econometrica*, forthcoming.

Devicienti, F. and Fanfani, B. (2021). Margins of adjustment to wage growth: The case of italian collective bargaining.

Devicienti, F., Fanfani, B., and Maida, A. (2019). Collective bargaining and the evolution of wage inequality in italy. *British Journal of Industrial Relations*, 57(2):377–407.

Devicienti, F., Maida, A., and Sestito, P. (2007). Downward wage rigidity in italy: micro-based measures and implications. *The Economic Journal*, 117(524):F530–F552.

Dodini, S., Stansbury, A., and Willén, A. (2023). How Do Firms Respond to Unions? *IZA Discussion Paper No. 16697*.

Draca, M., Machin, S., and Van Reenen, J. (2011). Minimum wages and firm profitability. *American Economic Journal: Applied Economics*, 3(3):129–151.

Drechsel-Grau, M. (2023). Employment and reallocation effects of higher minimum wages. Technical report.

Dube, A. and Lindner, A. S. (2024). Minimum wages in the 21st century. Working Paper 32878, National Bureau of Economic Research.

Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., and Vom Berge, P. (2022). Reallocation effects of the minimum wage. *Quarterly Journal of Economics*, pages 267–328.

D'Amuri, F. and Nizzi, R. (2018). Recent developments of italy's industrial relations system. *E-Journal of International and Comparative Labour Studies*.

Edmond, C., Midrigan, V., and Xu, D. Y. (2015). Competition, markups, and the gains from international trade. *American Economic Review*, 105(10):3183–3221.

Edmond, C., Midrigan, V., and Xu, D. Y. (2023). How Costly Are Markups? *Journal of Political Economy*, 131(7):1619–1675.

Eeckhout, J. (2020). Comment on: Diverging trends in national and local concentration. In *NBER Macroeconomics Annual 2020, Volume 35*. NBER.

Elminejad, A., Havranek, T., and Irsova, Z. (2022). Relative Risk Aversion: A Meta-Analysis. EconStor Preprints 260586, ZBW - Leibniz Information Centre for Economics.

Engbom, N. and Moser, C. (2022). Earnings inequality and the minimum wage: Evidence from brazil. *American Economic Review*, 112(12):3803–47.

Erikson, C. L. and Ichino, A. (1994). Wage differentials in italy: market forces, institutions, and inflation.

Ferrari, A. and Queirós, F. (2022). Firm heterogeneity, market power and macroeconomic fragility. *arXiv preprint arXiv:2205.03908*.

Firooz, H. (2023). The Pro-Competitive Consequences of Trade in Frictional Labor Markets. CESifo Working Paper Series 10649, CESifo.

Floden, M. (2001). The effectiveness of government debt and transfers as insurance. *Journal of Monetary Economics*, 48(1):81–108.

Gervais, M., Jaimovich, N., Siu, H. E., and Yedid-Levi, Y. (2016). What should i be when i grow up? occupations and unemployment over the life cycle. *Journal of Monetary Economics*, 83:54–70.

Gottfries, A. and Jarosch, G. (2023). Dynamic monopsony with large firms and noncompetes. Working Paper 31965, National Bureau of Economic Research.

Gouin-Bonenfant, E. (2020). Productivity dispersion, between-firm competition, and the labor share. Technical report.

Grassi, B. et al. (2017). Io in io: Competition and volatility in input-output networks. *Unpublished Manuscript, Bocconi University*.

Hall, R. E. and Krueger, A. B. (2012). Evidence on the incidence of wage posting, wage bargaining, and on-the-job search. *American Economic Journal: Macroeconomics*, 4(4):56–67.

Harasztosi, P. and Lindner, A. (2019). Who pays for the minimum wage? *American Economic Review*, 109(8):2693—-2727.

Heise, S. and Porzio, T. (2022). Labor misallocation across firms and regions. Working Paper 30298, National Bureau of Economic Research.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India*. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Huitfeldt, I., Kostøl, A. R., Nimczik, J., and Weber, A. (2023). Internal labor markets: A worker flow approach. *Journal of Econometrics*, 233(2):661–688.

Hurst, E., Kehoe, P., Pastorino, E., and Winberry, T. (2023). The macroeconomic dynamics of labor market policies. Technical report.

Jaimovich, N. (2007). Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations. *Journal of Economic Theory*, 137(1):300–325.

Jarosch, G., Nimczik, J. S., and Sorkin, I. (2024). Granular Search, Market Structure, and Wages. *The Review of Economic Studies*, page rdae004.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.

Karabarbounis, L., Lise, J., and Nath, A. (2022). Minimum Wages and Labor Markets in the Twin Cities. NBER Working Papers 30239, National Bureau of Economic Research, Inc.

Kroft, K., Luo, Y., Mogstad, M., and Setzler, B. (2020). Imperfect competition and rents in labor and product markets: The case of the construction industry. Working Paper 27325, National Bureau of Economic Research.

Leonardi, M., Pellizzari, M., and Tabasso, D. (2019). Wage compression within the firm: Evidence from an indexation scheme. *The Economic Journal*, 129(624):3256–3291.

Link, S. (2024). The price and employment response of firms to the introduction of minimum wages. *Journal of Public Economics*, 239:105236.

MacKenzie, G. (2020). Trade and market power in product and labor markets. Technical report.

Manacorda, M. (2004). Can the scala mobile explain the fall and rise of earnings inequality in italy? a semiparametric analysis, 1977–1993. *Journal of labor economics*, 22(3):585–613.

Manning, A. (2003). In *Monopsony in motion: Imperfect competition in labor markets*, chapter 2. Princeton University Press.

Mayneris, F., Poncet, S., and Zhang, T. (2018). Improving or disappearing: Firm-level adjustments to minimum wages in china. Technical report.

Menzio, G. (2024). Search theory of imperfect competition with decreasing returns to scale. *Journal of Economic Theory*, 218:105827.

Merz, M. and Yashiv, E. (2007). Labor and the market value of the firm. *American Economic Review*, 97(4):1419–1431.

Mongey, S. and Violante, G. L. (2019). Macro Recruiting Intensity from Micro Data. NBER Working Papers 26231, National Bureau of Economic Research, Inc.

Mortensen, D. T. (1988). Equilibrium wage distributions: A synthesis. Technical report, Discussion Paper.

Mortensen, D. T. and Vishwanath, T. (1991). Information sources and equilibrium wage outcomes. Technical report, Discussion paper.

Moscarini, G. and Postel-Vinay, F. (2013). Stochastic search equilibrium. *The Review of Economic Studies*, 80(4):1545–1581.

Muehlemann, S. and Strupler Leiser, M. (2018). Hiring costs and labor market tightness. *Labour Economics*, 52(C):122–131.

Neumark, D. (2017). The employment effects of minimum wages: Some questions we need to answer. Technical report, National Bureau of Economic Research.

Popp, M. (2023). Minimum wages in concentrated labor markets. Technical report.

Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2021). Diverging Trends in National and Local Concentration. *NBER Macroeconomics Annual*, 35(1):115–150.

Shimer, R. (2007). Mismatch. *American Economic Review*, 97(4):1074–1101.

Shimer, R. and Smith, L. (2001). Matching, Search, and Heterogeneity. *The B.E. Journal of Macroeconomics*, 1(1):1–18.

Stansbury, A. and Summers, L. H. (2020). The declining worker power hypothesis: An explanation for the recent evolution of the american economy. Technical report, National Bureau of Economic Research.

Tortarolo, D. and Zarate, R. D. (2020). Imperfect competition in product and labor markets. a quantitative analysis. Technical report.

van den Berg, G. J. and Ridder, G. (1998). An empirical equilibrium search model of the labor market. *Econometrica*, 66(5):1183–1221.

Yeh, C., Macaluso, C., and Hershbein, B. J. (2022). Monopsony in the u.s. labor market. Technical report.

Şahin, A., Song, J., Topa, G., and Violante, G. L. (2014). Mismatch unemployment. *American Economic Review*, 104(11):3529–64.

Appendix

A Appendix: Stylized Model

In the presence of a binding MW, firms choose employment by solving the following profit-maximization problem:

$$\begin{aligned} \max_{\ell} \quad & p(\ell)y - w(\ell)\ell - \kappa \\ \text{s.t.} \quad & y = z\ell, \\ & w(\ell) = \max\{\underline{w}, \ell^{\frac{1}{\eta}}\}, \\ & p(\ell) = y^{-\frac{1}{\epsilon(N)}}. \end{aligned}$$

As discussed in the main text, firms are either constrained or unconstrained in their wage setting problem.

Unconstrained firms set their optimal markup, $\mu = \frac{\epsilon}{\epsilon-1}$, and their optimal markdown, $\psi = \frac{1+\eta}{\eta}$. Hence, their employment, price and wage policies are as follows:

$$\ell(z) = z^{\eta\frac{\epsilon-1}{\epsilon+\eta}}(\mu\psi)^{-\frac{\epsilon\eta}{\epsilon+\eta}}, \tag{36}$$

$$p(z) = (z\ell(z))^{-\frac{1}{\epsilon}}, \tag{37}$$

$$w(z) = \ell(z)^{\frac{1}{\eta}}, \tag{38}$$

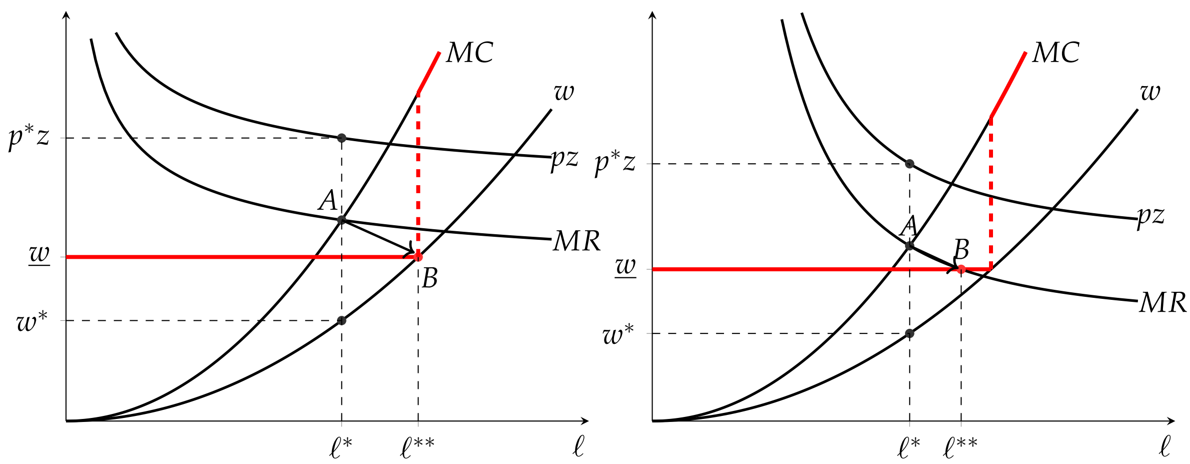
where the dependence of ϵ on N is left implicit. Unconstrained firms are affected by the MW only through the response of the elasticity of demand, mediated by the response of the number of active firms. Notice that the elasticity of demand affects output both through the shape of the demand curve and the optimal markup. Lemma A.3 derives a parametric restriction such that firm-level output is increasing in the number of active firms.

Constrained firms set their optimal markup but are not able to set their optimal markdown (because the unconstrained optimal employment would be off the labor supply curve). Figure A.1 shows that constrained firms can be further distinguished in two groups. Lower-productivity firms are demand-constrained, in that their employment is pinned down by product demand. Such firms are extracted their entire labor market power, that is, their markdown is compressed to one. Higher-productivity firms are supply-constrained, in that their employment is pinned down by labor supply. Such firms lose only part of their labor market power.

For supply-constrained firms, the constrained markdown is pinned down by the condition such that employment lies on the labor supply curve (left panel of Figure A.1). Hence, the

Figure A.1: Effects of the MW on constrained firms

(a) Supply-constrained (b) Demand-constrained



policy functions of supply-constrained firms are given by:

$$\ell(z) = z^{\eta\frac{\epsilon-1}{\epsilon+\eta}}(\mu\psi(z))^{-\frac{\epsilon\eta}{\epsilon+\eta}}, \tag{39}$$

$$p(z) = (z\ell(z))^{-\frac{1}{\epsilon}},$$

$$\underline{w} = \ell(z)^{\frac{1}{\eta}}. \tag{40}$$

From (39) and (40), it follows that the constrained markdown is given by:

$$\psi(z) = z^{\frac{\epsilon-1}{\epsilon}}\mu^{-1}\underline{w}^{-\frac{\epsilon+\eta}{\epsilon}} \in (1, \psi).$$

Since the constrained markdown is decreasing in the MW, all supply-constrained firms that would pay wages below the MW in the baseline equilibrium raise employment. Since employment is decreasing in the markup, all supply-constrained firms that would pay wages above the MW in the baseline equilibrium reduce employment. The least productive supply-constrained firm is defined by the following productivity level:

$$z^{ds}(\underline{w}) = \left(\mu\underline{w}^{\frac{\epsilon+\eta}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}}.$$

For demand-constrained firms, the constrained markdown is pinned down by the condition such that employment lies on the marginal revenue curve (right panel of Figure A.1). As a result, the markdown is always equal to one. Hence, the policy functions of demand-constrained firms are given by:

$$\ell(z) = z^{\epsilon-1}(\mu\underline{w})^{-\epsilon}, \tag{41}$$

$$p(z) = (z\ell(z))^{-\frac{1}{\epsilon}}.$$

Demand-constrained firms can either increase or decrease their employment with respect to the baseline equilibrium. For given elasticity of demand, we can define a productivity threshold z^c, such that all firms with higher (lower) productivity than z^c increase (decrease) employment, as follows:

$$z^c(\underline{w}) = (\mu^{\epsilon}\psi^{-\eta})^{\frac{1}{\epsilon-1}}\underline{w}^{\frac{\epsilon+\eta}{\epsilon-1}}.$$

Figures (2) and (3) in the main text report the cross-sectional adjustment to a MW increase such that firms at the z^c cutoff make zero profits (with the MW in place), i.e., $\pi(z^c(\underline{w});\mu,\psi) = 0$. It follows that, for given markup, all the demand-constrained firms that would have scaled down in size following the MW exit the market.

Proof of Proposition 1. The first statement follows directly from the differentiated cross-sectional responses of unconstrained, supply-constrained and demand-constrained firms (see Figure (3)): depending on the number of firms in each category, the aggregate labor share and output can either increase or decrease following a MW reform.

The hump-shaped pattern of the aggregate labor share and output (second statement) follows from the fact that the markup response is mediated by firm exit. Hence, we can prove the statement in two steps through the following lemmas.

Lemma A.1

$\exists\underline{w}^{\star}$ such that the aggregate labor share and output are increasing in the MW if MW $\leq \underline{w}^{\star}$.

Proof. Since the lowest-productivity firms make positive profits in the baseline equilibrium, a marginal increase in the MW induces a response of labor market power only – as reported in Figure (2). Since constrained firms increase their labor share and unconstrained firms keep it constant, the aggregate labor share – which is a weighted average of firm-level labor shares – increases following the MW introduction.

Aggregate output increases following the MW reform if and only if the output gains from surviving firms exceed the output losses from exiting firms. Formally,

$$\Delta Y(\underline{w}) = N\left(\sum_{\underline{z}'}^{\bar{z}}\Delta y(z)\hat{\gamma}(z) - \sum_{\underline{z}}^{\underline{z}'}y(z)\hat{\gamma}(z)\right) > 0 \iff \sum_{\underline{z}'}^{\bar{z}}\Delta y(z)\hat{\gamma}(z) > \sum_{\underline{z}}^{\underline{z}'}y(z)\hat{\gamma}(z).$$

If the MW is small enough such that no firm exits and all the constrained firms increase employment, output losses are zeroed out and output gains are positive. Formally, let $\tilde{\pi}(w(z);z)$ denote the *operating* profit function, i.e., gross of overhead costs, of a firm with productivity z that pays wage $w(z)$ (either constrained or unconstrained by a MW). If $\underline{w} \in (\underline{w}'(\underline{z}), \underline{w}^{\star}(\underline{z})] = \left(\left(\frac{\underline{z}^{\frac{\epsilon-1}{\epsilon}}}{\psi\mu}\right)^{\frac{\epsilon}{\epsilon+\eta}}, \tilde{\pi}^{-1}(\kappa;\underline{z})\right]$, then $\Delta Y(\underline{w}) > 0$. The lower bound to the set of output-enhancing MWs is given by the condition such that the MW is binding for the least productive firms. The upper bound is given by the condition that the least productive firm makes zero profits.[60] Since the least productive firms make positive profits in the baseline

[60] Recall that, throughout, we assume that demand-constrained firms that would scale down employment exit the market. Formally, $\min\left\{\left(\frac{\underline{z}^{\frac{\epsilon-1}{\epsilon}}}{\psi^{-\frac{\eta}{\epsilon}}\mu}\right)^{\frac{\epsilon}{\epsilon+\eta}}, \tilde{\pi}^{-1}(\kappa;\underline{z})\right\} = \tilde{\pi}^{-1}(\kappa;\underline{z})$.

equilibrium, i.e., $\tilde{\pi}(w(\underline{z});\underline{z}) - \kappa > 0$, it follows that the set of output-enhancing MWs is non-empty by continuity of the operating profit function with respect to the wage. ■

As the MW increases, some firms will exit the market. Specifically, we assume a finite number of potential MW values sorted on a sparse enough grid $\mathbb{W}$ such that, once the MW starts inducing some exit, the number of active firms is strictly monotonically decreasing in the MW.[61] If the markup function is such that

$$\mu(N-1) > \mu(N)\psi \ \ \forall N, \tag{42}$$

then net market power of the surviving firms increases with the MW.

Lemma A.2
Under condition (42), the aggregate labor share and output are decreasing in the MW if MW> $\underline{w}^\star$.

Proof. Let $\underline{w}^\star = \tilde{\pi}^{-1}(\kappa;\underline{z})$ as defined in the previous lemma. It follows that any MW higher than $\underline{w}^\star$ induces a reduction in the number of firms. Under condition (42), all the surviving firms increase their net market power as the number of firms decreases. Assume a sequence of increasing MW reforms. All the surviving firms respond by progressively lowering their labor share, which translates into a decreasing aggregate labor share with the MW. By Lemma A.3, unconstrained firms (in the current MW equilibrium) reduce their output. Under condition (42), firms that are supply-constrained in the current MW equilibrium become demand-constrained because of the shift in their marginal revenue curve (by Lemma A.3, Equations (39) and (40) cannot hold simultaneously under (42)). Hence, they reduce employment. To see why, notice that the marginally supply-constrained firm (from above) before the reform has productivity $\hat{z}(\underline{w}) = \left(\mu\psi\underline{w}^{\frac{\epsilon+\eta}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}}$. By Equations (40) and (41), such a firm reduces employment following the reform if and only if:

$$\begin{aligned}
\hat{z}(\underline{w})^{\epsilon'-1}(\mu'\underline{w}')^{-\epsilon'} < \underline{w}^{\eta} &\iff \left(\mu\psi\underline{w}^{\frac{\epsilon+\eta}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}(\epsilon'-1)}(\mu'\underline{w}')^{-\epsilon'} < \underline{w}^{\eta} \\
&\iff \mu' > \left(\frac{\underline{w}}{\underline{w}'}\right)(\mu\psi)^{\frac{\Delta\epsilon}{(\epsilon-1)\epsilon'}}(\underline{w})^{\frac{\Delta\epsilon}{(\epsilon-1)\epsilon'}(1+\eta)}\mu\psi.
\end{aligned}$$

Since $\underline{w} < \underline{w}'$ by construction and $\Delta\epsilon \equiv \epsilon' - \epsilon < 0$ by assumption, this condition always holds under (42). A fortiori, the same holds for the other supply-constrained firms. For instance, the marginally supply-constrained (from below) before the reform has productivity $\hat{\hat{z}}(\underline{w}) = \left(\mu\underline{w}^{\frac{\epsilon+\eta}{\epsilon}}\right)^{\frac{\epsilon}{\epsilon-1}}$. Repeating the same steps as above yields:

$$\hat{\hat{z}}(\underline{w})^{\epsilon'-1}(\mu'\underline{w}')^{-\epsilon'} < \underline{w}^{\eta} \iff \mu' > \left(\frac{\underline{w}}{\underline{w}'}\right)(\mu)^{\frac{\Delta\epsilon}{(\epsilon-1)\epsilon'}}(\underline{w})^{\frac{\Delta\epsilon}{(\epsilon-1)\epsilon'}(1+\eta)}\mu,$$

[61] The assumption of discrete number of MW values is needed to ensure strict monotonicity in the relation between MW size and number of active firms (once the first firm exits) in the presence of a discrete number of firms, whose productivity distribution is not continuous. If the distribution were continuous, we could dispense with this assumption.

which is again always true under (42). By Equation (41), employment of demand-constrained (in the current MW equilibrium *and* after the reform) firms is decreasing in the MW. Under Lemma A.3, employment of demand-constrained firms is increasing in the number of active firms. Hence, demand-constrained firms reduce their output, as well. Overall, all the surviving firms reduce employment and aggregate output needs to reduce, as well. ■

■

Lemma A.3 (Firm-level output response to the number of active firms)
If $\underline{z} > \left(\frac{\epsilon(N)}{\epsilon(N)-1}\frac{1+\eta}{\eta}\right)^{\frac{\eta}{1+\eta}} \exp\left\{-\frac{\epsilon(N)+\eta}{(1+\eta)(\epsilon(N)-1)}\right\}$ $\forall N \in \mathbb{N}$, *then unconstrained and supply-constrained firms reduce their output following a reduction in the number of active firms.*

Let $\underline{z}(\underline{w})$ *be the least productive active firm when the MW is* $\underline{w}$. *If* $\underline{z}(\underline{w}) > \frac{\epsilon(N)}{\epsilon(N)-1}\exp\{-\frac{1}{\epsilon(N)-1}\}\underline{w}\ \forall N \in \mathbb{N}$, $\forall \underline{w} \in \mathbb{W}$, *then demand-constrained firms reduce their output following a reduction in the number of active firms.*

Proof of A.3. Differentiating optimal unconstrained employment (36) with respect to ϵ yields:

$$\frac{d\ln(\ell(z))}{d\epsilon} = \frac{1+\eta}{(\epsilon+\eta)^2}\eta\ln(z) - \frac{\eta^2}{(\epsilon+\eta)^2}\ln\left(\frac{\epsilon}{\epsilon-1}\psi\right) + \frac{\eta}{\epsilon+\eta}\frac{1}{\epsilon-1},$$

where the first two terms represent the employment response to a change in the elasticity of the demand curve for given markup (*elasticity effect*) and the last term the employment response to change in markup for given elasticity of demand curve (*markup effect*). It holds that:

$$\frac{d\ln(\ell(z))}{d\epsilon} > 0 \iff z > \left(\frac{\epsilon}{\epsilon-1}\psi\right)^{\frac{\eta}{1+\eta}} \exp\left\{-\frac{\epsilon+\eta}{(1+\eta)(\epsilon-1)}\right\}.$$

Since $d\epsilon/dN > 0$ by assumption, if this condition is met by the least productive firms, then it holds for all the firms. Notice that optimal employment of supply-constrained firms (39) is the same as that of unconstrained firms up to a weakly lower markdown. Hence, the same condition guarantees that supply-constrained firms reduce their output as the number of active firms decreases.

Differentiating optimal employment of demand-constrained firms (41) with respect to ϵ yields:

$$\frac{d\ln(\ell(z))}{d\epsilon} = \ln(\frac{z}{\mu\underline{w}}) + \frac{1}{\epsilon-1},$$

where the first term represents the elasticity effect and the second term the markup effect. It holds that:

$$\frac{d\ln(\ell(z))}{d\epsilon} > 0 \iff z > \frac{\epsilon}{\epsilon-1}\exp\left\{-\frac{1}{\epsilon-1}\right\}\underline{w}.$$

Since $d\epsilon/dN > 0$ by assumption, if this condition is met by the least productive firms, then it holds for all the demand-constrained firms. ■

B Appendix: Quantitative Model

B.1 Matching Function Specification

In this Section we motivate our functional form assumption on the matching function. As explained in Section (3.7), our baseline equilibrium features congestion externalities. Moreover, our counterfactual experiments in Section (5) point to a reduction in congestion externalities as the main driver of the positive response of value added through lower misallocation. By governing the extent (as well as the scope for a reduction) of congestion externalities, the functional form of the matching function is relevant in our framework. In particular, when considering large MW reforms, it is crucial to account for potential nonlinearities in the matching function.

For this reason, we adopt a CES matching function, $\mathcal{M}(S,V) = \chi\left[\alpha V^{-\iota} + (1-\alpha)S^{-\iota}\right]^{-\frac{1}{\iota}}$, and calibrate the elasticity parameters ι and α according to the estimates of Şahin et al. (2014).[62] As pointed out by Şahin et al. (2014), the estimated parameters imply only a slight deviation from the standard Cobb-Douglas specification with vacancy elasticity of 0.5. However, we prefer the CES specification because it delivers an endogenous elasticity of firms' meeting rate with respect to labor market tightness. Formally,

$$\epsilon_{q(\theta),\theta} = -\frac{1}{1+\frac{\alpha}{1-\alpha}\theta^{-\iota}}.$$

From this formulation, it easy to see that the standard Cobb-Douglas specification – nested by setting $\iota = 0$ and $\alpha = 0.5$ – predicts a constant elasticity of 0.5. On the other hand, the CES specification implies that the elasticity is increasing in labor market tightness. As a result, the slacker the labor market is to start with, the lower the gain in terms of meeting rate of further reducing tightness. This implies that, as the MW gets larger and labor market tightness decreases, hiring cost savings from lower congestion effects progressively die out. It follows that our choice of a CES matching function is conservative in terms of positive response of value added to the MW.

B.2 Derivation of Stationary Distributions

Let $G_{a_i}(w)$ be the share of employed workers of skill a who earn wage rate equal or lower than w. This share evolves in response to inflows (unemployed workers who accept jobs with $w' \leq w$) and outflows (employed workers who accept jobs with $w' > w$ or lose their job) according to the following Kolmogorov forward equation:

$$\dot{G}_{a_i}(w) = \lambda(\theta_{a_i})u_{a_i}F_{a_i}(w) - \Big(\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1-F_{a_i}(w))\Big)e_{a_i}G_{a_i}(w). \tag{43}$$

[62] Since the estimated elasticities are based on US data, we internally estimate the efficiency of the matching function χ on our Italian data.

Similarly, the measures of employed e_{a_i} and unemployed u_{a_i} workers evolve according to labor market transitions governed by the separation and job-finding rates:

$$\dot{u}_{a_i} = \delta_{a_i} e_{a_i} - \lambda(\theta_{a_i}) u_{a_i}. \tag{44}$$

$$\dot{e}_{a_i} = \lambda(\theta_{a_i}) u_{a_i} - \delta_{a_i} e_{a_i}. \tag{45}$$

Under the assumption of constant population ($u_{a_i} + e_{a_i} = 1$), one can solve for these distributions in stationary equilibrium (i.e., setting $\dot{G}_{a_i}(w) = \dot{u}_{a_i} = \dot{e}_{a_i} = 0$):

$$G_{a_i}(w) = \frac{\lambda(\theta_{a_i}) F_{a_i}(w)}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})(1 - F_{a_i}(w))} \frac{u_{a_i}}{e_{a_i}}, \tag{46}$$

$$u_{a_i} = \frac{\delta_{a_i}}{\delta_{a_i} + \lambda(\theta_{a_i})}. \tag{47}$$

Equation (46) reveals the close connection between G_{a_i}, which corresponds to the observed employment wage distribution, and F_{a_i}, the wage offer distribution. In general, it can be shown that G_{a_i} diverges from F_{a_i}, with more and more mass being concentrated in the right part of the support of wages, when the frictional index $\chi_{a_i} \equiv \frac{\delta_{a_i}}{s_{a_i} \lambda(\theta_{a_i})}$ is low. As already noted by Burdett and Mortensen (1998), the frictional index is inversely related to the speed at which workers climb the job ladder in the model and represents a synthetic measure of the extent of labor market frictions in the economy.

We now investigate the relationship between G_{a_i} and F_{a_i}. Upon substituting the stationarity condition $\frac{u_{a_i}}{e_{a_i}} = \frac{\delta_{a_i}}{\lambda(\theta_{a_i})}$ into (46), we get

$$G_{a_i}(w) = \frac{\delta_{a_i} F_{a_i}(w)}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) \left(1 - F_{a_i}(w)\right)} \implies \frac{G_{a_i}(w)}{F_{a_i}(w)} = \frac{\delta_{a_i}}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})(1 - F_{a_i}(w))}.$$

From the previous expression, it is clear that the G_{a_i} dominates in a first-order stochastic dominance sense the F_{a_i} distribution. Formally, $G_{a_i}(w) \leq F_{a_i}(w) \;\; \forall \, w$. It can also be noted that the parameter s_{a_i} regulates the distance between the G_{a_i} and the F_{a_i} distributions. In particular, as $s_{a_i} \to 0$ for a given δ_{a_i}, we note that $G_{a_i}(w) \to F_{a_i}(w)$. Conversely, the distance between the two distributions can be made arbitrarily large by setting large values of s_{a_i}. After some manipulation, the previous expression can be rearranged as:

$$G_{a_i}(w) = \frac{F_{a_i}(w)}{1 + \chi_{a_i}(\theta_{a_i})(1 - F_{a_i}(w))}. \tag{48}$$

This formulation makes clear that the relationship between the two distributions is fully determined by $\chi_{a_i}(\theta_{a_i})$. Intuitively, if employed workers find job offers at a much faster rate than they lose their job (i.e., fall off the ladder), then more mass will be placed to higher values in the support of the wage offer distribution. Finally, one can also note that the difference between the two distributions also depends on the rank of the firm in the wage offer distribution, $F_{a_i}(w)$. Intuitively, as $F_{a_i}(w) \to 1$, then $G_{a_i}(w) \to F_{a_i}(w) = 1$, and conversely $F_{a_i}(w) \to 0$, then $G_{a_i}(w) \to F_{a_i}(w) = 0$, i.e., the two distributions need to coincide at the upper and lower point of the wage offer distribution.

B.3 Derivation of Labor Supply Curve

The evolution of employment $\ell_{a_i}(w,v)$ of a firm hiring from market a, posting a piece rate w and v vacancies is governed by the following differential equation:

$$\dot{\ell}_{a_i}(w,v) = v\, q(\theta_{a_i}) \left(\frac{u_{a_i} + s_{a_i} e_{a_i} G_{a_i}(w)}{S_{a_i}} \right) - \Big(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})(1 - F_{a_i}(w)) \Big) \ell_{a_i}(w,v),$$

where the first term represents the inflows, that is, the product between the mass of vacancies and the filling rate, while the second term represents the outflows, that is, the product between the quit rate and current employment.

By evaluating the previous equation in stationary equilibrium, i.e., setting $\dot{\ell} = 0$, and using the fact that $q_{a_i} = \frac{\lambda(\theta_{a_i})}{\theta_{a_i}} = \lambda(\theta_{a_i}) \frac{S_{a_i}}{V_{a_i}}$ (which follows from the matching function exhibiting constant returns to scale), one can derive stationay employment as follows:

$$\ell_{a_i}(w,v) = \frac{v}{V_{a_i}} \lambda(\theta_{a_i}) \frac{u_{a_i} + s_{a_i} e_{a_i} G_{a_i}(w)}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})(1 - F_{a_i}(w))},$$

that is the labor supply curve shown in the firms' problem in the main text. Using Equation (46), one can substitute G_{a_i} away and obtain:

$$\ell_{a_i}(w,v) = \frac{v}{V_{a_i}} u_{a_i} \lambda(\theta_{a_i}) \frac{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})}{\left[\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})(1 - F_{a_i}(w))\right]^2}.$$

If the equilibrium wage distributions have mass points, one needs to discipline workers' transitions across equally-paying firms. Let q_p and q_ℓ denote the probability that a firm poaches a worker from an equally paying firm and that a worker accepts a wage offer from an equally paying firm, respectively. Stationary employment reads:

$$\ell_{a_i}(w,v;q_p,q_\ell) = \frac{v}{V_{a_i}} \lambda(\theta_{a_i}) \frac{u_{a_i} + s_{a_i} e_{a_i} [G_{a_i}(w^-) + q_p \left(G_{a_i}(w) - G_{a_i}(w^-)\right)]}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) \left[1 - F_{a_i}(w) + q_\ell \left(F_{a_i}(w) - F_{a_i}(w^-)\right)\right]}.$$

Substituting for (48) yields:

$$\ell_{a_i}(w,v;q_p,q_\ell) = \frac{v}{V_{a_i}} \lambda(\theta_{a_i}) u_{a_i} \frac{h_{a_i}(w^-) h_{a_i}(w) + s_{a_i} \lambda(\theta_{a_i}) \left[F_{a_i}(w^-) h_{a_i}(w) + q_p \left(F_{a_i}(w) h_{a_i}(w^-) - F_{a_i}(w^-) h_{a_i}(w)\right)\right]}{h_{a_i}(w^-) h_{a_i}(w) [\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) \left(1 - q_\ell F_{a_i}(w^-) - (1 - q_\ell) F_{a_i}(w)\right)]}, \tag{49}$$

where $h_{a_i}(w) \equiv \delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})(1 - F_{a_i}(w))$.

Letting $q_p = q_\ell$, we recover the familiar expression for stationary employment in the equilibrium search literature:

$$\ell_{a_i}(w,v) = \frac{v}{V_{a_i}} u_{a_i} \lambda(\theta_{a_i}) \frac{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})}{\left[\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1 - F_{a_i}(w))\right]\left[\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1 - F_{a_i}(w^-))\right]}.$$

B.4 Derivation of Product Demand Curve

Consider the problem of a final good producer reported in the main text (14). The Lagrangean associated to the profit maximization problem is the following:

$$\mathcal{L} = P \prod_{i=1}^{I} \left(\int \int_0^{K_{a_i}} \left(\left(\sum_{j=1}^{N_{k(a_i)}} y_{jk(a_i)}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\rho-1}{\rho}} dk(a_i)\, d\Omega_i(a) \right)^{\alpha_i \frac{\rho}{\rho-1}} - \sum_{i=1}^{I} \int \int_0^{K_{a_i}} \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)}\, dk(a_i)\, d\Omega_i(a),$$

Taking the FOC with respect to a generic variety j in sector $k(a_i)$, we retrieve the (inverse) product demand function reported in the main text (20):

$$p_{jk(a_i)} = P \frac{\partial Y}{\partial y_{jk(a_i)}} \iff p_{jk(a_i)} = P \left(\frac{y_{jk(a_i)}}{Y_{k(a_i)}} \right)^{-\frac{1}{\sigma}} \left(\frac{Y_{k(a_i)}}{Y_i} \right)^{-\frac{1}{\rho}} \left(\frac{Y_i}{\alpha_i Y} \right)^{-1}. \tag{50}$$

Similarly, the FOCs with respect to (fictitious) sectoral and industry good aggregators read:

$$P_{k(a_i)} = P \left(\frac{Y_{k(a_i)}}{Y_i} \right)^{-\frac{1}{\rho}} \left(\frac{Y_i}{\alpha_i Y} \right)^{-1}, \quad P_i = P \left(\frac{Y_i}{\alpha_i Y} \right)^{-1}.$$

Substituting for $Y_{k(a_i)}$ and Y_i into (50) yields the (direct) product demand curve reported in the main text (22):

$$y_{jk(a_i)} = \left(\frac{p_{jk(a_i)}}{P_{k(a_i)}} \right)^{-\sigma} \left(\frac{P_{k(a_i)}}{P_i} \right)^{-\rho} \left(\frac{P_i}{P} \right)^{-1} \alpha_i Y.$$

Next, we compute the price indices at different level of aggregation by imposing the following consistency and zero-profit conditions:

$$P_{k(a_i)} Y_{k(a_i)} = \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)} \implies P_{k(a_i)} = \left(\sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)}^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \tag{51}$$

$$P_i Y_i = \int \int_0^{K_{a_i}} P_{k(a_i)} Y_{k(a_i)} dk(a_i) d\Omega_i(a) \implies P_i = \left(\int \int_0^{K_{a_i}} P_{k(a_i)}^{1-\rho} dk(a_i) d\Omega_i(a) \right)^{\frac{1}{1-\rho}}, \tag{52}$$

$$PY = \sum_{i=1}^{I} P_i Y_i \implies P = \left(\prod_{i=1}^{I} \left(\frac{\alpha_i}{P_i} \right)^{\alpha_i} \right)^{-1}. \tag{53}$$

Finally, we use the final good as numeraire and normalize the aggregate price index P to 1.

B.5 Derivation of Reservation Wage

In the labor market, the workers' problem is solved by finding the reservation wage rules. Trivially, employed workers accept all job offers with a piece rate that exceeds the one of their current job ($R^e_{a_i}(w) = w$). This reflects the absence of any switching costs and of any other characteristics of jobs (e.g., amenities). We now turn to the determination of the reservation wage for the unemployed workers. By setting $W_{a_i}(aR^u_{a_i}) = U_{a_i}$ and solving for $R^u_{a_i}$, one gets:

$$aR^u_{a_i} + \Pi(aR^u_{a_i}) + T = \left[(ab_{a_i})^{1-\vartheta} + (1-\vartheta)(1-s_{a_i})\lambda(\theta_{a_i}) \int_{aR^u_{a_i}}^{a\overline{w}_{a_i}} \left(W_{a_i}(aw') - U_{a_i}\right) \, dF_{a_i}(w') \right]^{\frac{1}{1-\vartheta}}.$$

Integrating by parts, one can simplify the integral as follows:

$$\begin{aligned}
\int_{aR^u_{a_i}}^{a\overline{w}_{a_i}} \left(W_{a_i}(aw') - U_{a_i}\right) \, dF_{a_i}(w') &= \left[W_{a_i}(aw) - U_{a_i}\right] F_{a_i}(w)\big|^{a\overline{w}}_{aR^u_{a_i}} - \int_{aR^u_{a_i}}^{a\overline{w}} W'_{a_i}(aw) \, F_{a_i}(w) \, dw \\
&= \int_{aR^u_{a_i}}^{a\overline{w}} W'_{a_i}(aw) \, dw - \int_{R^u_{a_i}}^{a\overline{w}} W'_{a_i}(aw) \, F_{a_i}(w) \, dw \\
&= \int_{aR^u_{a_i}}^{a\overline{w}} W'_{a_i}(aw) \, (1 - F_{a_i}(w)) \, dw.
\end{aligned}$$

Let us now take the derivative of $W_{a_i}(aw)$ with respect to w:

$$W'_{a_i}(aw) = \frac{1}{r + \delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1 - F_{a_i}(aw))} \frac{1 + \Pi'(aw)}{(aw + \Pi(aw) + T)^{\vartheta}}.$$

Combining this with the previous equation yields:

$$\begin{aligned}
aR^u_{a_i} + \Pi(aR^u_{a_i}) + T = \Bigg[& (ab_{a_i} + T)^{1-\vartheta} + (1-\vartheta)(1-s_{a_i})\lambda(\theta_{a_i}) \int_{R^u_{a_i}}^{\overline{w}_{a_i}} \frac{1 - F_{a_i}(w)}{r + \delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1 - F_{a_i}(w))} \\
& \frac{1 + \Pi'(aw)}{(aw + \Pi(aw) + T)^{\vartheta}} \, dw \Bigg]^{\frac{1}{1-\vartheta}}.
\end{aligned} \tag{54}$$

Since $\Pi' \geq 0$, the left-hand side of (54) is increasing in $R^u_{a_i}$ and the right-hand side is decreasing in $R^u_{a_i}$. Hence, the reservation wage for unemployed workers is unique.

Equation 54 determines the optimal search behavior of unemployed workers. In sum, unemployed workers accept all job offers above $R^u_{a_i}$, which generally does not coincide with the value of home production b_{a_i} due to the presence of profits rebates when employed (Π) and the option value of search (the arrival rates of employed and unemployed workers generally differ because $s_{a_i} \neq 1$).

B.6 Key Equations for Firm's Problem

In this Section, we list all the key equations that compose the firm's problem. These are the equations that are plotted in Figure 4. For simplicity we omit the a_i and $k(a_i)$ subscripts. We start with the cost structure. Let TC denote total costs, MC marginal costs and AC average costs.

$$
\begin{aligned}
TC(\ell) &= w(\ell)\,\ell + c\,(w(\ell), v(\ell))\,, && (55)\\
MC(\ell) &= w(\ell) + \frac{\partial w(\ell)}{\partial \ell}\ell + \frac{\partial c\,(w(\ell), v(\ell))}{\partial v}\frac{\partial v(\ell)}{\partial \ell} + \frac{\partial c\,(w(\ell), v(\ell))}{\partial w}\frac{\partial w(\ell)}{\partial \ell}\\
&= (1+\epsilon_{w,\ell})w(\ell) + \frac{\partial c\,(w(\ell), v(\ell))}{\partial v}\frac{v(\ell)}{\ell}\epsilon_{v,\ell} + \frac{\partial c\,(w(\ell), v(\ell))}{\partial w}\frac{w(\ell)}{\ell}\epsilon_{w,\ell}, && (56)\\
AC(\ell) &= w(\ell) + \frac{c\,(w(\ell), v(\ell))}{\ell}. && (57)
\end{aligned}
$$

Equation (55) simply states that firms face both wage and hiring costs in our model. Interestingly, the marginal cost schedule (56) reflects both the labor supply elasticity – which requires wage rises in order to expand the size – and the size-dependence of hiring costs. Finally, the difference between AC and MC, which is the source of operating profits from labor market power, is made of two different terms: the first one relates to the labor supply elasticity, whereas the second one is represented by inframarginal profits associated to hiring costs.

On the revenue side, let TR denote total revenues, MRPL the marginal revenue product of labor and AR average revenue product of labor.

$$
\begin{aligned}
TR(\ell; z) &= p(y)y, && (58)\\
MRPL(\ell; z) &= [p'(y)y + p(y)]\frac{\partial y}{\partial \ell}\\
&= (1+\epsilon_{p,y}(y))p(y)\,\frac{\partial y}{\partial \ell}. && (59)\\
AR(\ell; z) &= p(y)\,\frac{\partial y}{\partial \ell}, && (60)
\end{aligned}
$$

where $y = az\ell$. Total revenues (58) are simply given by the product between the price and the quantity produced. The MRPL schedule (59) incorporates the elasticity of inverse demand, owing to the fact that final good producers' demand implies that the price needs to drop if the producer wants to increase the quantity sold. Finally, differences between AR and MRPL, which is the source of operating profits from product market power, are uniquely driven by the (negative) price elasticity of product demand.

B.7 The Role of Vacancies

Vacancies are essential to production, even though they do not enter directly into the production function. This is because, due to the separation rate δ_{a_i}, a constant inflow of labor is needed to maintain any firm size, which in turn requires at least some vacancies being posted in the labor market. In general, it can be shown that there are important complementarities between wages and vacancies. Indeed, the elasticity of labor supply to a given of these two inputs depends on the level of the other one. For instance, suppose that a firm is considering investing more in vacancies. The total size increase that is triggered by this choice will be small (large) if the offered wage lies towards the bottom (top) of the wage distribution, reflecting the fact that these additional vacancies will be effective to a different degree at poaching workers from other firms. Conversely, an increase in the piece rate, that determines an improvement in the rank of the firm (i.e., how many other firms it can poach workers from), will deliver a higher return if the stock of vacancies posted is large. Formally, firm size is supermodular in wage and vacancies:

$$\frac{\partial \ell_{a_i}(w,v)}{\partial w\,\partial v} > 0.$$

In equilibrium, firms choose the optimal mix of wages and vacancies that guarantees them to achieve their optimal size.[63]

In equilibrium, the amount of vacancies posted shapes the relationship between firm size and ranking in the wage offer distribution $F_{a_i}(w)$. Indeed, a high level of vacancies increases the likelihood that own wage offers are sampled by workers. Such a *visibility channel* mediated by vacancies is distinct – albeit complementary – to the standard *wage channel* featuring any model with upward sloping labor supply.[64]

B.8 Proof of Proposition 2

On each labor market, firms engage in a wage-posting game with a finite number of productivity types. Existence and uniqueness of a continuous equilibrium wage offer distribution have been established for the case of CRS revenue function and exogenous job creation by Mortensen (1988) and Burdett and Mortensen (1998). We now generalize the result to the case of DRS revenue function and endogenous job creation. Assume there are $n = 1, \dots, \mathcal{Z}$ productivity types. Let γ_n denote the probability mass function of productivity. To simplify notation, we omit the a_i and k_{a_i} subscripts throughout. Consider the following problem:

$$\begin{aligned}\pi(z_n) &= \max_{w,v}\,(pz_n - w)v\tilde{\ell}(F(w)) - c(F(w),v) - \kappa \\ \text{s.t.}\quad \tilde{\ell}(F(w)) &= \frac{1}{V}\frac{\chi_0}{1+\chi_0}\frac{1+\chi_1}{[1+\chi_1(1-F(w^-))][1+\chi_1(1-F(w))]},\ p = p(v\tilde{\ell}; z_n),\end{aligned}$$

where $\tilde{\ell}$ denotes the vacancy yield, i.e., stationary employment per vacancy, and $\chi_0 \equiv \frac{\lambda}{\delta}$, $\chi_1 \equiv \frac{s\lambda}{\delta}$ are (inverse) frictional indices. Since the number of productivity types is finite, no continuous equilibrium wage offer distribution exists in pure strategies. Hence, we proceed to characterize a mixed strategy Nash equilibrium (MSNE). The MSNE features residual wage dispersion for each productivity type, represented by type-specific wage offer distributions.

[63] Indeed, the firm's problem may also be rewritten in two steps. In the first step, firms minimize costs (wage and hiring costs) subject to a size constraint. In the second step, they maximize profits by finding their optimal size.

[64] Both channels are already present in the models of Moscarini and Postel-Vinay (2013) and Engbom and Moser (2022).

Formally, the wage offer distribution of firms of type n, $F_n(w)$, is defined by:

$$
\begin{aligned}
&(p_n z_n - w_n) v_n \tilde{\ell}_n - c_n - \kappa = \pi_n \quad w_n \in \text{supp}F_n(w) = [\underline{w}_n, \bar{w}_n), \\
&(p_n z_n - w_n) v_n \tilde{\ell}_n - c_n - \kappa \leq \pi_n \quad w_n \notin \text{supp}F_n(w) = [\underline{w}_n, \bar{w}_n).
\end{aligned}
$$

For each wage level, optimal vacancy posting $v_n(w_n)$ solves:

$$
\frac{\partial c(F(w_n), v)}{\partial v} = \left(\frac{p_n(v\tilde{\ell}_n)}{\mu_n(v\tilde{\ell}_n)} z_n - w_n \right) \tilde{\ell}(F(w_n)),
$$

where $\frac{\partial c(F(w_n),v)}{\partial v \partial w_n} < 0$ and $\frac{\partial \tilde{\ell}(F(w_n))}{\partial w_n} > 0$. Notice that residual dispersion in wages is associated one-to-one with residual dispersion in marginal revenue products $\tilde{z}$ (that is, firms with the same productivity have different marginal revenue products in equilibrium). Specifically, wages need to be increasing in MRPL.[65] Hence, we can express wages and vacancies as a function of $\tilde{z}$. In other words, there exists a unique type-specific distribution of $\tilde{z}$ such that firm-level optimization is consistent with the equilibrium wage offer distribution. Monotonicity of wages in $\tilde{z}$ implies that $F_n(w_n(\tilde{z})) = \int_{\underline{\tilde{z}}}^{\tilde{z}} \frac{v_n(\tilde{z})}{\bar{v}_n} \varphi_n(\tilde{z}) d\tilde{z}$. It follows that the type-specific density of $\tilde{z}$ solves $\varphi_n(\tilde{z}) = \frac{\bar{v}_n}{v_n(w_n(\tilde{z}))} w_n'(\tilde{z}) f_n(w_n(\tilde{z}))$, where $f_n(w_n) \equiv F_n'(w_n)$ and $\bar{v}_n = \int_{\underline{\tilde{z}}_n}^{\bar{\tilde{z}}_n} v_n(\tilde{z}) \varphi_n(\tilde{z}) d\tilde{z}$.

We now guess that firms with higher productivity pay higher wages, that is, $\underline{w}_n \geq \bar{w}_{n-1}$. It follows that the vacancy yield $\tilde{\ell}$ is increasing in productivity. If the elasticity of substitution across sector-level output, ρ, is high enough, the revenue function is strictly supermodular, i.e. $\frac{\partial p(v\tilde{\ell}_n) z_n v \tilde{\ell}_n}{\partial (v\tilde{\ell}_n) \partial z_n} = \frac{\partial p(v\tilde{\ell}_n)/\mu(v\tilde{\ell}_n) z_n}{\partial z_n} > 0$.[66] Since the wage offer distribution is continuous and the productivity distribution has mass points, the marginal benefit of vacancy posting jumps upward moving to a higher productivity type, whereas the marginal cost increases continuously. Hence, more productive firms find it optimal to post more vacancies. From the MSNE definition, it follows that:

$$
\begin{aligned}
\pi_n &= (p_n z_n - w_n) v_n \tilde{\ell}_n - c_n - \kappa \\
&\geq (p_{n-1} z_n - w_{n-1}) v_{n-1} \tilde{\ell}_{n-1} - c_{n-1} - \kappa \\
&> (p_{n-1} z_{n-1} - w_{n-1}) v_{n-1} \tilde{\ell}_{n-1} - c_{n-1} - \kappa = \pi_{n-1} \\
&\geq (p_{n-1} z_{n-1} - w_n) v_n \tilde{\ell}_n - c_n - \kappa,
\end{aligned}
$$

where the first and third inequalities follow from firm's optimization. Therefore, $(p_n z_n -$

[65] If it were not the case, either wage dispersion would be inconsistent with firm's optimization ($w'(\tilde{z}) = 0$) or with equal profits on the support of the wage distribution ($w'(\tilde{z}) < 0$). Notice that monotonicity of vacancies in MRPL is not required for this argument.

[66] Since product markets sourcing from the same labor market are perfectly symmetric, for given wage, the marginal revenue product is a one-to-one correspondence with productivity. Any revenue function stemming from oligopolistic competition á la Cournot with CES demand system exhibits strict supermodularity if the elasticity of substitution across sector-level output is close enough to that across firm-level output. Formally, $\frac{\partial p(y)y}{\partial \ell \partial z} = p(y) \frac{1-\mu'(y)y}{\mu^2(y)}$. Since $\lim_{\rho \to \sigma} \mu'(y) = 0$, there always exists a threshold $\underline{\rho}$ such that $\frac{\partial p(y)y}{\partial \ell \partial z} > 0 \; \forall y$ if $\rho > \underline{\rho}$. See Biondi (2022) for further details.

$p_{n-1}z_{n-1})v_n\tilde{\ell}_n \geq (p_n z_n - p_{n-1}z_{n-1})v_{n-1}\tilde{\ell}_{n-1} \iff v_n\tilde{\ell}_n > v_{n-1}\tilde{\ell}_{n-1}$. Since $v_n > v_{n-1}$ and $\tilde{\ell}_n > \tilde{\ell}_{n-1}$, our guess is verified. Hence, if $\rho > \underline{\rho}$, the wage ranking equals the productivity ranking. Let $\pi_n(w, F(w))$ denotes profits of a firm of the n-th productivity type posting wage w and the corresponding optimal vacancies. Since firms with higher productivity pay higher wages, the equilibrium wage offer distribution is characterized by the following conditions:

$$
\begin{aligned}
&\underline{w}_1 = R, \\
&\underline{w}_n = \bar{w}_{n-1}, \quad \forall n > 1, \\
&\bar{w}_n : \pi_n\left(\bar{w}_n, F(\underline{w}_n) + \gamma_n \frac{V_n}{V}\right) = \pi_n(\underline{w}_n, F(\underline{w}_n)) \\
&F_n(w) : \pi_n\left(w, F(\underline{w}_n) + \gamma_n \frac{V_n}{V} F_n(w)\right) = \pi_n(\underline{w}_n, F(\underline{w}_n)), \quad w \in [\underline{w}_n, \bar{w}_n).
\end{aligned}
$$

In words, the infimum of the support of the type-specific wage offer distribution is the workers' reservation wage for the lowest-productivity type and the supremum of the support of the wage offer distribution of the lower-productivity type for the others. The supremum of the support of the type-specific wage offer distributions is pinned down by the indifference condition between posting the lowest wage ($F_n(\underline{w}_n) = 0$) and the highest wage ($F_n(\bar{w}_n) = 1$). Similarly, the shape of the wage offer distribution is pinned down by the indifference condition among any wage in its support. The intuition is that, in equilibrium, firms of the same type need to be indifferent between extracting higher margins from a lower mass of workers or extracting lower margins from a larger mass of workers. Hence, the necessary condition for existence of the wage-posting equilibrium is that $\left.\frac{\partial \pi(w,F(w);z_1)}{\partial F(w)}\right|_{w=R} > 0$.[67] Under this necessary condition, the wage offer distribution of the lowest productivity type is non-degenerate. Given that the revenue function is supermodular in size and productivity and the market wage offer distribution is continuous, the same condition, i.e., $\left.\frac{\partial \pi_n(w,F(w))}{\partial F(w)}\right|_{w=\underline{w}_n} > 0$, holds for every productivity type, as well. Let $\varphi(\tilde{z})$ denote the market density of MRPL, which is the mixture of the type-specific MRPL distributions, i.e., $\varphi(\tilde{z}) \equiv \sum_{n=1}^{\mathcal{Z}} \varphi_n(\tilde{z})\gamma_n$. In turn, the market wage offer distribution is the mixture:

$$
F(w) = M \sum_{n=1}^{\mathcal{Z}} \gamma_n \frac{\bar{v}_n}{V} F_n(w) = M \int_{\underline{\tilde{z}}}^{\bar{\tilde{z}}} \frac{v(\tilde{z})}{V} \varphi(\tilde{z}) d\tilde{z},
$$

where $V = M \sum_{n=1}^{\mathcal{Z}} \bar{v}_n \gamma_n = M \int_{\underline{\tilde{z}}}^{\bar{\tilde{z}}} v(\tilde{z})\varphi(\tilde{z})d\tilde{z}$ and $\bar{v}_n = \int v_n(\tilde{z})\varphi_n(\tilde{z})d\tilde{z}$. Finally, for given market wage offer distribution, the reservation wage is unique (see equation (54)). Hence, we have constructed the unique wage-posting equilibrium with a continuous wage offer distribution.[68] Finally, notice that the price function in our model is sector-specific. Even

[67] Since the reservation wage R is increasing in unemployment benefits b and $\left.\frac{\partial \pi(w,F(w);z_1)}{\partial F(w)}\right|_{w=0} > 0$ since revenues are increasing in size, there always exists a threshold $\bar{b}$ such that a wage-posting equilibrium exists if $b < \bar{b}$.

[68] As firstly pointed out by Mortensen and Vishwanath (1991), in the presence of decreasing returns in revenues, the wage-posting equilibrium described so far may coexist with a competitive equilibrium featuring a degenerate

though sectors sourcing from the same labor market are structurally identical, residual wage dispersion would make them differ in terms of wage/MRPL distribution.[69] Hence, we assume that the demand curve firms internalize depends on the expected – rather than the realized – sectoral output, which is constant. As the number of productivity types grows larger, i.e. $\mathcal{Z} \to \infty$, our wage-posting equilibrium *purifies* in the sense that the type-specific wage offer distributions $F_n(w)$ degenerate into mass points, thus removing all the residual wage dispersion.

On each product market, a finite number of firms engage in a Cournot game on quantities. Due to strategic interaction, multiple equilibria with different sets of active firms may arise. Hence, to determine which firms from the set of potential firms are active, we need to specify an equilibrium selection device. To this purpose, we consider a refinement of the Nash equilibrium of a simultaneous entry game first introduced by Berry (1992) and used by De Loecker et al. (2021). Specifically, we first assume that all the potential firms operate and compute the general equilibrium accordingly. Upon computing equilibrium profits, we proceed by removing the firm which makes the highest losses, if any.[70] Iterating this procedure until all active firms make non-negative profits, we recover the equilibrium with the largest number of active firms (Ferrari and Queirós, 2022). Hence, our equilibrium selection device constitutes the most conservative choice in terms of concentration response following shocks, such as the introduction of a MW.

B.9 Derivation of Firms' Policy Functions

We derive firm's policy functions in three steps. First, we characterize labor market policies as a function of the marginal revenue product of labor. Second, we characterize product market policies as a function of effective marginal costs. Third, we characterize firm's policy functions by setting marginal revenue product equal to marginal cost.

First, for given residual distributions of MRPL in labor market a_i for each productivity type, $\Phi_{a_i}(\tilde{z};z)$, we solve for the firm's labor market policies as a function of its MRPL. We proceed by guessing (and verifying) that optimal wages are increasing in the MRPL. This allows us to express the wage offer distribution F_{a_i} in terms of vacancy policies as follows:

$$F_{a_i}\left(w_{a_i}(\tilde{z})\right) = \frac{M_{a_i}}{V_{a_i}} \int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} v(z')\, d\Phi_{a_i}(z'), \tag{61}$$

wage offer distribution. In such an equilibrium, the unique wage clears the labor market by equalizing aggregate labor demand, stemming from the optimal vacancy condition, and aggregate (wage-insensitive) labor supply, coming from frictions. By allowing for degenerate equilibrium wage offer distributions, we conjecture that the nature of the best equilibrium (i.e., granting highest profits) depends on the value of the frictional index χ_1: if frictions are low enough, the equilibrium wage offer distribution can either have a density and a mass point or degenerate on a mass point, as described in Menzio (2024) for the case of frictional product markets.

[69] This is the case because, even if there is a continuum of sectors sourcing from the same labor market, each of them is populated by a finite number of firms.

[70] Since product markets sourcing from each labor market are perfectly symmetric, we remove the corresponding firm in all product markets at a time.

where $\Phi_{a_i}(\tilde{z})$ is the MRPL distribution of potential firms in labor market a_i, and V_{a_i}/M_{a_i} denotes average vacancies per potential firm in labor market a_i.[71] Intuitively, the equilibrium wage offer distribution weights firms by their vacancy posting, which determines their *visibility*. It follows that the likelihood of a searching worker sampling a job offer up to a given wage equals the relative mass of vacancies posted by lower-paying firms. Equation (61), provides a key to express the first-order conditions (23)-(24) as a system of differential equations pinning down the wage policy function. [72]

Upon taking the derivative of equation (61) with respect to $\tilde{z}$, we find out the mapping between the density of the wage offer distribution and that of the MRPL distribution at each wage level:

$$f_{a_i}(w_{a_i}(\tilde{z}))w'_{a_i}(\tilde{z}) = \frac{M_{a_i}}{V_{a_i}} v_{a_i}(\tilde{z})\varphi_{a_i}(\tilde{z}). \tag{62}$$

Let $\mathcal{H}_{a_i}(\tilde{z}) = F_{a_i}(w_{a_i}(\tilde{z}))$ be the wage offer distribution as a function of the MRPL. This implies that $f_{a_i}(w_{a_i}(\tilde{z})) = \frac{\mathcal{H}'_{a_i}(\tilde{z})}{w'_{a_i}(\tilde{z})}$. Combining this with equation (62) delivers the mass of vacancies per MRPL consistent with the equilibrium wage offer distribution:

$$\mathcal{H}'_{a_i}(\tilde{z}) = \frac{M_{a_i}}{V_{a_i}} v_{a_i}(\tilde{z})\varphi_{a_i}(\tilde{z}) \implies v_{a_i}(\tilde{z}) = \frac{V_{a_i}}{M_{a_i}} \frac{\mathcal{H}'_{a_i}(\tilde{z})}{\varphi_{a_i}(\tilde{z})}. \tag{63}$$

Replacing equation (62) and (63) into the FOCs of the firm's problem yields the following system of differential equations:

$$(v): \quad \mathcal{H}'_{a_i}(\tilde{z}) = \frac{M_{a_i}}{V_{a_i}} \varphi_{a_i}(\tilde{z}) \left(\frac{\tilde{z} - w_{a_i}(\tilde{z})}{\bar{c}_{a_i}} \frac{\lambda(\theta_{a_i}) u_{a_i}(\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i}))/V_{a_i}}{(\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i}) \left[1 - \mathcal{H}_{a_i}(\tilde{z})\right])^{2+\phi}} \right)^{\frac{1}{\zeta}}, \tag{64}$$

$$(w): \quad w'_{a_i}(\tilde{z}) = (\tilde{z} - w_{a_i}(\tilde{z})) \frac{2\nu s_{a_i}\lambda(\theta_{a_i})\mathcal{H}'_{a_i}(\tilde{z})}{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i}) \left[1 - \mathcal{H}_{a_i}(\tilde{z})\right]}. \tag{65}$$

This system of differential equations in the two functions $w_{a_i}(\tilde{z})$ and $\mathcal{H}_{a_i}(\tilde{z})$ yield the optimal wage and vacancy policies for each MRPL, provided the necessary boundary conditions. Since the market MRPL distribution is not uniformly continuous (it exhibits discontinuities across productivity types), the boundary conditions are type-specific:

$$\lim_{\tilde{z}_{a_i} \to \underline{\tilde{z}}_{a_i}(z_1)} w_{a_i}(\tilde{z}_{a_i}) = \max\left\{R^u_{a_i}, \frac{\mathbf{w}}{a}\right\}, \lim_{\tilde{z} \to \underline{\tilde{z}}_{a_i}(z_1)} \mathcal{H}_{a_i}(\tilde{z}_1) = 0,$$

$$\lim_{\tilde{z} \to \underline{\tilde{z}}_{a_i}(z_n)} w_{a_i}(\tilde{z}_{a_i}) = w_{a_i}(\bar{\tilde{z}}_{a_i}(z_{n-1})), \lim_{\tilde{z} \to \underline{\tilde{z}}_{a_i}(z_n)} \mathcal{H}_{a_i}(\tilde{z}) = \mathcal{H}_{a_i}(\bar{\tilde{z}}_{a_i}(z_{n-1})), \quad \forall n = 2, \ldots, \mathcal{Z}_i,$$

where $\underline{\tilde{z}}_{a_i}(z_n)$ is the infimum of the type-specific MRPL distribution for the $n-$th type. The first condition states that the lowest MRPL firm offers the lowest piece rate possible, that is

[71] See Appendix (B.8) for how to recover the unique market MRPL distribution $\Phi_{a_i}(\tilde{z})$ from the type-specific wage offer distributions.

[72] Unlike in Engbom and Moser (2022), the MRPL is an equilibrium object in our model. It follows that the labor market block of our model is independent of the product market block for a given distribution of MRPL. We exploit this feature in the numerical solution algorithm of our model described in Appendix (B.12).

either the reservation wage or the minimum wage for the lowest-productivity type and the supremum of the type-specific wage offer distribution of the lower-productivity type for the others. Instead, the second condition states that the CDF of offered wages is equal to 0 at the lowest wage for each type.[73]

Let $x_{a_i}(\tilde{z}) \equiv \frac{2\nu s_{a_i}\lambda(\theta_{a_i})\mathcal{H}'_{a_i}(\tilde{z})}{\delta_{a_i}+s_{a_i}\lambda(\theta_{a_i})[1-\mathcal{H}_{a_i}(\tilde{z})]}$ be an inverse measure of search frictions faced by workers employed at a firm with MRPL $\tilde{z}$ in the job finding process. Solving the differential equation (65) allows computing the wage policy function analytically as follows:

$$w_{a_i}(\tilde{z}) = e^{-\int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} x_{a_i}(z)dz} R^u_{a_i} + \int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} \hat{z}\, e^{-\int_{\hat{z}}^{\tilde{z}} x(z)dz} x_{a_i}(\hat{z})\, d\hat{z}. \tag{66}$$

For the sake of intuition, notice that the assumption of stationary employment as labor supply function makes the firm's sequential search problem isomorphic to a simultaneous search problem, i.e., to a static allocation problem of workers to firms. According to this interpretation, the weighting factor attached to the reservation wage in Equation (66) equals the probability that a worker does not meet any firm with MRPL weakly lower than $\tilde{z}$. On the other hand, the integral of the weighting factor $e^{-\int_{\hat{z}}^{\tilde{z}} x_{a_i}(z)dz} x(\hat{z})$ attached to each MRPL level equals the probability that a worker meets at least one firm with MRPL weakly lower than $\tilde{z}$. Formally, $\int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} e^{-\int_{\hat{z}}^{\tilde{z}} x_{a_i}(z)dz} x_{a_i}(\hat{z})\, d\hat{z} = 1 - e^{-\int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} x_{a_i}(\hat{z})\, d\hat{z}}$. Hence, we can interpret the wage posting game as a first-price sealed-bid auction on workers with an unknown number of competitors pinned down by the extent of search frictions. It follows that $\frac{e^{-\int_{\hat{z}}^{\tilde{z}} x_{a_i}(z)dz} x_{a_i}(\hat{z})}{1-e^{-\int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} x_{a_i}(\hat{z})\, d\hat{z}}}$ denotes the density of the highest-paying firm with lower MRPL than $\tilde{z}$ contacted by a worker, conditional on contacting any of them. Hence, the wage function (66) equals the expected outside option of a searching worker:

$$w_{a_i}(\tilde{z}) = \Pr\left(m_{a_i}(\tilde{z}) = 0\right) R^u_{a_i} + \Pr\left(m_{a_i}(\tilde{z}) = 1\right) \mathbb{E}_{a_i}\left[\max\{MRPL\} | MRPL < \tilde{z}, m_{a_i}(\tilde{z}) = 1\right], \tag{67}$$

where $m_{a_i}(\tilde{z}) \equiv \mathbb{1}_{\{\text{meet at least one firm w/ MRPL} \leq \tilde{z}\}}$ is an indicator of successful search. The intuition is as follows. With probability $\Pr\left(m_{a_i}(\tilde{z}) = 0\right) = e^{-\int_{\hat{z}}^{\tilde{z}} x_{a_i}(z)dz}$, the worker does not meet any firm with MRPL lower than $\tilde{z}$, in which case the worker's outside option when meeting a firm with MRPL $\tilde{z}$ boils down to her reservation wage. With complementary probability, the worker meets at least one firm with lower MRPL than $\tilde{z}$, in which case the worker's outside option when meeting a firm with MRPL $\tilde{z}$ equals the maximum wage the best of such firms can afford paying, that is, the expected highest MRPL among firms with lower MRPL than $\tilde{z}$.

Of course, equation (66) holds only within productivity types (because of kinks in the

[73] In equilibrium, the labor market tightness adjusts to guarantee that $\lim_{\tilde{z}\to\bar{\tilde{z}}_{a_i}} \mathcal{H}_{a_i}(\tilde{z}) = 1$, where $\bar{\tilde{z}}_{a_i}$ is the highest MRPL level among active firms. Equivalently, the total mass of vacancies posted by firms need to equal the aggregate V_{a_i}.

market wage distribution). The market wage function is given by:

$$w_{a_i}(\tilde{z}_{a_i}(z_n)) = \left[p_n(\tilde{z}_{a_i}(z_n)) \prod_{j=1}^{n-1} p_j(\bar{\tilde{z}}_{a_i}(z_j)) \right] R^u_{a_i} + \\ p_n(\tilde{z}_{a_i}(z_n)) \sum_{j=1}^{n-1} \prod_{v=j+1}^{n-1} p_v(\bar{\tilde{z}}_{a_i}(z_v)) \int_{\underline{\tilde{z}}_{a_i}(z_j)}^{\bar{\tilde{z}}_{a_i}(z_j)} p_j(\hat{z})\hat{z} x_{j,a_i}(\hat{z}) d\hat{z} + \int_{\underline{\tilde{z}}_{a_i}(z_n)}^{\tilde{z}_{a_i}(z_n)} p_n(\hat{z})\hat{z} x_{n,a_i}(\hat{z}) d\hat{z},$$

where $p_n(\tilde{z}_{a_i}(z_n)) \equiv e^{-\int_{\underline{\tilde{z}}_{a_i}(z_n)}^{\tilde{z}_{a_i}(z_n)} x_{n,a_i}(z)dz}$. Notice that the interpretation of the market wage function is exactly the same as that proposed in (67).

The markdown policy is given by the ratio between MRPL and wage:

$$\psi_{a_i}(\tilde{z}) = \frac{\tilde{z}}{w_{a_i}(\tilde{z})}, \tag{68}$$

The employment policy is determined by the labor supply curve (19):

$$\ell_{a_i}(\tilde{z}) = \frac{v_{a_i}(\tilde{z})}{V_{a_i}} \frac{u_{a_i}\lambda(\theta_{a_i})[\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})]}{[\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1 - \mathcal{H}_{a_i}(\tilde{z}))]^2}. \tag{69}$$

Second, for given residual distribution of marginal costs $\mathcal{C}(z) \sim \Phi^c_{k(a_i)}(z)$, we solve for the firm's product market policies. Maximizing (18) with respect to ℓ yields:

$$p_{k(a_i)}(\mathcal{C}(z), z) = \mu_{k(a_i)}(\mathcal{C}(z), z)\frac{\mathcal{C}(z)}{z}. \tag{70}$$

Optimal employment is determined by the product demand curve (20):

$$p_{k(a_i)}(\mathcal{C}(z), z) = \left(az\ell_{k(a_i)}(\mathcal{C}(z), z)\right)^{-\frac{1}{\sigma}} Y_{k(a_i)} \left(\ell_{k(a_i)}(\mathcal{C}(z), z)\right)^{\frac{1}{\sigma}-\frac{1}{\rho}} Y_i^{\frac{1}{\rho}-1} \alpha_i Y, \tag{71}$$

where $Y_{k(a_i)}(\ell_{k(a_i)}(\mathcal{C}(z), z)) = \left[\sum_{j \neq i}^{N_k} \bar{y}_{jk(a_i)}^{\frac{\sigma-1}{\sigma}} + \left(a\, z\, \ell_{k(a_i)}(\mathcal{C}(z), z)\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$.[74] Optimal employment then determines both the optimal market share and markup as follows:

$$sh_{k(a_i)}(\mathcal{C}(z), z) = \frac{(az\ell_{k(a_i)}(\mathcal{C}(z), z))^{\frac{\sigma-1}{\sigma}}}{Y_{k(a_i)}(\ell_{k(a_i)}(\mathcal{C}(z), z))^{\frac{\sigma-1}{\sigma}}}, \tag{72}$$

$$\mu_{k(a_i)}(\mathcal{C}(z), z) = \frac{\sigma}{(\sigma - 1)\left[1 - \frac{\frac{\sigma}{\rho}-1}{\sigma-1} sh_{k(a_i)}(\mathcal{C}(z), z)\right]}. \tag{73}$$

Finally, firm's policy functions obtain by setting $\Phi_{a_i}(\tilde{z}, z) = \Phi^c_{k(a_i)}(\tilde{c}, z)$ and equalizing

[74] $\bar{y}_{jk(a_i)}$ is the expected output of firm j, where the expectation is taken with respect to the residual distribution of MC.

MRPL and MC:

$$\tilde{z}_{k(a_i)}(z) \equiv \frac{p_{k(a_i)}(z)}{\mu_{k(a_i)}(z)} z = \psi_{a_i}(z) w_{a_i}(z) \equiv \tilde{c}_{a_i}(z), \tag{74}$$

where all the policy functions are set-valued.

B.10 Derivation of Equilibrium Markdown Function

In this Section we show how to derive the equilibrium markdown function reported in the main text (31) starting from the firm's marginal cost function (56).

$$\begin{aligned} MC(\ell) = w(\ell) & \left(1 + \epsilon_{w,\ell} + \frac{1}{w} \frac{\partial c\,(w(\ell), v(\ell))}{\partial v} \frac{\partial v(\ell)}{\partial \ell} + \frac{1}{w} \frac{\partial c\,(w(\ell), v(\ell))}{\partial w} \frac{\partial w(\ell)}{\partial \ell}\right) \\ w(\ell) & \left(1 + \epsilon_{w,\ell} + \left(1 + \frac{\partial c}{\partial w} \frac{1}{\ell}\right) \frac{\epsilon_{v,\ell}}{\epsilon_{\ell,w}} + \frac{\partial c}{\partial w} \frac{\epsilon_{w,\ell}}{\ell}\right) \\ w(\ell) & \left(1 + \frac{\epsilon_{\ell,w}\epsilon_{w,\ell} + \epsilon_{v,\ell}}{\epsilon_{\ell,w}} + \frac{\partial c}{\partial w} \frac{1}{\ell} \frac{\epsilon_{\ell,w}\epsilon_{w,\ell} + \epsilon_{v,\ell}}{\epsilon_{\ell,w}}\right) \\ w(\ell) & \left(1 + \frac{1 + \frac{\partial c}{\partial w}\ell}{\epsilon_{\ell,w}}\right) \\ w(\ell) & \left(1 + \frac{1}{\nu \epsilon_{\ell,w}}\right), \end{aligned}$$

where we made use of the firm's indifference condition between vacancies and wages and the size constraint:

$$\frac{\partial c/\partial v}{\partial \ell/\partial v} = \frac{\partial c}{\partial v} \frac{v}{\ell} = \left(1 + \frac{1}{\ell} \frac{\partial c}{\partial w}\right) \frac{w}{\epsilon_{\ell,w}} \iff (1+\zeta) \frac{c}{w\ell} \epsilon_{\ell,w} = 1 - \frac{\phi}{2} \frac{c}{w\ell} \epsilon_{\ell,w} \implies 1 + \frac{\partial c}{\partial w} \frac{1}{\ell} = \frac{1}{\nu},$$

$$\ell = \ell(w, v) \implies \epsilon_{\ell,w}\epsilon_{w,\ell} + \epsilon_{v,\ell} = 1.$$

B.11 Aggregation

In this section we describe how to aggregate variables at different levels, according to the structure of our model.

Sectoral level. We start from the definition of sectoral revenues and wage bill:

$$\begin{aligned} P_{k(a_i)} Y_{k(a_i)} &= \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)}, \\ w_{k(a_i)} L_{k(a_i)} &= \sum_{j=1}^{N_{k(a_i)}} w_{jk(a_i)} \ell_{jk(a_i)} \end{aligned}$$

Let $Z_k(a_i) = \frac{Y_k(a_i)}{aL_{k(a_i)}}$ denote the sectoral firm productivity. We define a sectoral net market power index, $\mathcal{M}_{k(a_i)}$, by assuming that the firm-level optimal price condition (25) holds at the sectoral level, as well:

$$\begin{aligned}
P_{k(a_i)} Y_{k(a_i)} &= \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)} \\
\mathcal{M}_{k(a_i)} \frac{w_{k(a_i)}}{Z_{k(a_i)}} a Z_{k(a_i)} L_{k(a_i)} &= \sum_{j=1}^{N_{k(a_i)}} \mathcal{M}_{jk(a_i)} \frac{w_{jk(a_i)}}{z_{jk(a_i)}} a z_{jk(a_i)} \ell_{jk(a_i)} \\
\mathcal{M}_{k(a_i)} &= \sum_{j=1}^{N_{k(a_i)}} \frac{w_{jk(a_i)} \ell_{jk(a_i)}}{w_{k(a_i)} L_{k(a_i)}} \mathcal{M}_{jk(a_i)},
\end{aligned}$$

where $\mathcal{M}_{jk(a_i)} \equiv \mu_{jk(a_i)} \psi_{jk(a_i)}$ is the firm-level net market power index. Hence, the model-consistent aggregation of the net market power index is wage-bill-weighted. Along the same lines, we proceed by defining sectoral markup and markdown as their wage-bill-weighted averages:

$$\begin{aligned}
\mu_{k(a_i)} &\equiv \sum_{j=1}^{N_{k(a_i)}} \frac{w_{jk(a_i)} \ell_{jk(a_i)}}{w_{k(a_i)} L_{k(a_i)}} \mu_{jk(a_i)}, \\
\psi_{k(a_i)} &\equiv \sum_{j=1}^{N_{k(a_i)}} \frac{w_{jk(a_i)} \ell_{jk(a_i)}}{w_{k(a_i)} L_{k(a_i)}} \psi_{jk(a_i)}.
\end{aligned}$$

We notice that the sectoral net market power index equals the product between sectoral markup and markdown plus their covariance term:

$$\mathcal{M}_{k(a_i)} = \mu_{k(a_i)} \psi_{k(a_i)} + \mathbf{Cov}[\mu_{k(a_i)}, \psi_{k(a_i)}].$$

Hence, some positive correlation between firm-level markups and markdowns makes the sectoral net market power index exceed the product between sectoral markups and markdowns.[75]

Next, let the sectoral hiring cost and overhead cost equal $\mathcal{HC}_{k(a_i)} = \sum_{j=1}^{N_{k(a_i)}} ac_{jk(a_i)}$ and $\mathcal{FC}_{k(a_i)} = \sum_{j=1}^{\bar{N}_{k(a_i)}} a\kappa_{k(a_i)}$, respectively. Then, the sectoral labor share equals:

$$LS_{k(a_i)} = \frac{aw_{k(a_i)} L_{k(a_i)}}{P_{k(a_i)} Y_{k(a_i)} - \mathcal{HC}_{k(a_i)} - \mathcal{FC}_{k(a_i)}},$$

where $P_{k(a_i)}$ and $Y_{k(a_i)}$ are defined in (51)-(17).

Industry level. Since workers' ability differ within industry, industry wage bill equals $W_i L_i = \int \int_0^{K_{a_i}} aw_{k(a_i)} L_{k(a_i)} dk(a_i) d\Omega_i(a)$. Repeating the same steps as at the sectoral level

[75] In our simulations, the discrepancy is negligible, though.

yields:

$$
\begin{aligned}
\mu_i =& \int \int_0^{K_{a_i}} \frac{a w_{k(a_i)} L_{k(a_i)}}{W_i L_i} \mu_{k(a_i)} dk(a_i) d\Omega_i(a), \\
\psi_i =& \int \int_0^{K_{a_i}} \frac{w_{k(a_i)} L_{k(a_i)}}{W_i L_i} \psi_{k(a_i)} dk(a_i) d\Omega_i(a), \\
\tilde{Z}_i =& \frac{Y_i}{L_i}, LS_i = \frac{W_i L_i}{P_i Y_i - \mathcal{C}_i^h - \mathcal{C}_i^f},
\end{aligned}
$$

where P_i and Y_i are defined in (52)-(16), and $\tilde{Z}_i$ is the industry-level (total) labor productivity.

Aggregate level. Repeating the same steps as at the industry level yields:

$$
\begin{aligned}
\mu =& \sum_{i=1}^{I} \frac{W_i L_i}{WL} \mu_i, \\
\psi =& \sum_{i=1}^{I} \frac{W_i L_i}{WL} \psi_i, \\
\tilde{Z} =& \frac{Y}{L}, \\
LS =& \frac{WL}{PY - \mathcal{HC} - \mathcal{FC}},
\end{aligned}
$$

where P and Y are defined in (53)-(15), and $\tilde{Z}$ is the aggregate (total) labor productivity. Finally, aggregate value added (consumption) is defined as:

$$
C = Y - \mathcal{HC} - \mathcal{FC}.
$$

B.12 Algorithm to solve the model

Consistently with our estimation strategy, we solve the discretized version of our model with a continuum of productivity types. Hence, we let residual wage (and MRPL) dispersion vanish and work with one-to-one firms' policy functions (rather than set-valued).

The model is solved by guess and verify for the collection of MRPL functions $\tilde{z}_{k(a_i)}(z)$. Although the same solution concept applies to any set of parameters, we will add some specific comments related to an increase in the MW parameter.

To initialize the solution routine, we make an initial guess on the collection of MRPL functions, $\tilde{z}^0_{k(a_i)}(z)$. If the MW is higher than some guessed $\tilde{z}$ values (with a positive number of firms), we adjust any such points upward until they all exceed the MW. This preliminary adjustment amounts to conjecturing that all the potential firms are active. In this respect, we leverage the theoretical insight that the presence of market power on both the labor and the product market guarantees that all firms make positive operating profits for *any* MW level. Since overhead costs do not affect firms' policy functions in any way, we can therefore find a candidate equilibrium for any set of active firms. To discipline the extensive margin of adjustment (firm exit) , we check ex-post whether the candidate equilibrium is sustainable, i.e., whether any firm is making negative profits at that equilibrium, and, if not, gradually removing firms from the market.

Upon initializing the solution routine as just described, we solve the model according to the $\tilde{z}^0(z)$ guess. This is performed in four steps:

1. Solve for the equilibrium in each labor market $a_i = 1, \dots, IA$ as follows:

   – Compute the number of firms at each $\tilde{z}^0_{a_i}(z)$-value as a share of the number of active firms (which is set equal to the number of potential firms in the estimated model) to identify the discretized counterpart of the MRPL density $\phi_{a_i}(\tilde{z})$;

   – Guess the wage and wage offer distribution as a function of the guessed $\tilde{z}^0_{a_i}$ function, $w_{a_i}(\tilde{z}^0_{a_i})$ and $\mathcal{H}_{a_i}(\tilde{z}^0_{a_i})$. Apply a first-order Taylor expansion to the nonlinear system of differential equations (64)-(65) around the guessed wage and wage offer distribution functions to transform it into a linear system of differential equations. Then, use the Euler (finite difference) method to transform the linear system of differential equations into a linear system of difference equations with boundary conditions $w_{a_i}(\underline{\tilde{z}_{a_i}}) = R^u_{a_i}$ and $\mathcal{H}_{a_i}(\underline{\tilde{z}_{a_i}}) = 0$.[76];

   – Find the labor-market-specific equilibrium job finding rate by solving the linear system of difference equations: this can be performed efficiently by first identifying a lower bound and an upper bound to the equilibrium job finding rate corresponding to too much job creation, i.e., $\lim_{\tilde{z}_{a_i} \to \bar{\tilde{z}}_{a_i}} \mathcal{H}_{a_i}(\tilde{z}_{a_i}) > 1$, and too little job creation, i.e., $\lim_{\tilde{z}_{a_i} \to \bar{\tilde{z}}_{a_i}} \mathcal{H}_{a_i}(\tilde{z}_{a_i}) < 1$, respectively, and then applying a bisection algorithm within such bounds;

   – Compute the labor-market-specific policy functions:

   $$\{w_{a_i}(\tilde{z}^0_{a_i}), \mathcal{H}_{a_i}(\tilde{z}^0_{a_i}), v_{a_i}(\tilde{z}^0_{a_i}), \ell_{a_i}(\tilde{z}^0_{a_i}), \psi_{a_i}(\tilde{z}^0_{a_i})\}$$

   by making use of (65), (64), (63), (69), and (68), respectively.

2. Solve for the equilibrium in each product market $k = 1, \dots, K$ as follows:

   – Compute sectoral, industry, and aggregate output by aggregating up firm-specific output policies $y_{k(a_i)}(\tilde{z}^0_{k(a_i)}(z)) = az\ell_{k(a_i)}(\tilde{z}^0_{k(a_i)}(z))$ according to (17), (16), and (15), respectively. To avoid that the number of sectors affects industry output, we remove love-of-variety effects from the discretized industry CES aggregator, that is, industry output reads:

   $$Y_i^{\text{discr}} = K_i^{-\frac{1}{\rho-1}} \left(\sum_{k=1}^{K_{a_L i}} Y_{k(a_L i)}^{\frac{\rho-1}{\rho}} + \sum_{k=1}^{K_{a_H i}} Y_{k(a_H i)}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}};$$

[76] We check that results are unchanged by applying the Euler method directly to the nonlinear system of differential equations.

- Compute firm-specific price policies by making use of the demand constraint (20), as well as sectoral, industry, and aggregate price indices according to (51)-(53), by aggregating up firm-level prices;

- Thanks to the oligopolistically-competitive product market structure and nested CES preferences, firm-level prices are sufficient for pinning down each firm's market share by

$$sh_{k(a_i)}(\tilde{z}^0_{k(a_i)}(z)) = \frac{p_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z)\right)^{1-\sigma}}{N_{k(a_i)} \sum_{n=1}^{Z_{k(a_i)}} p_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right)^{1-\sigma} \gamma_{k(a_i)}(z_n)},$$

  and markup policy by (28);

- Compute firm-specific profits as:

$$\pi_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right) = p_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right) y_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right) - ac_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right) - a\kappa_{k(a_i)}.$$

  Compute the implied MRPLs, $\tilde{z}^{implied}_{k(a_i)}(z)$, by making use of the MRPL definition, i.e.,

$$\tilde{z}^{implied}_{k(a_i)}(z) \equiv \frac{p_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right)}{\mu_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right)} z.$$

3. Upon solving the model conditional on the $\tilde{z}^0_{k(a_i)}(z)$ guess, we proceed by verifying and potentially updating the guess. To do so, the solution algorithm goes through the following steps:

   - Let $\epsilon^0 \equiv \mathbb{E}\left[|\tilde{z}^0_{a_i}(z) - \tilde{z}^{implied}_{a_i}(z)|\right]$ denote the average convergence error, and set a sensitivity $\bar{\epsilon}$ ($= 10^{-4}$ in our simulations);[77]

   - If $\epsilon^0 > \bar{\epsilon}$, we update the initial $\tilde{z}^0$-guess via bisection, i.e. $\tilde{z}^1_{a_i}(z) = 0.5^*\tilde{z}^0_{a_i}(z) + 0.5^*\tilde{z}^{implied}_{a_i}(\tilde{z}^0_{a_i}(z))$. If the new guess $\tilde{z}^1$ features nonmonotonicity in some labor market, we reduce the weight on the implied MRPL until such nonmonotonicity disappears.[78] Repeat the steps 1-3 iteratively until finding a guess $\tilde{z}^n_{a_i}(z)$ at the $n-$th repetition such that $\epsilon^n < \bar{\epsilon}$;

   - Store the model solution for the $\tilde{z}^n_{a_i}(z)$ guess as candidate equilibrium.

4. Check that no firm makes negative profits in the candidate equilibrium. If it is the

[77]Our guess $\tilde{z}^0_{a_i}(z)$ is common across firms with the same productivity sourcing from the same labor market, i.e., $\tilde{z}^0_{k(a_i)}(z) = \tilde{z}^0_{a_i}(z)$. Due to integer constraints, sectors sourcing from the same labor market may not be exactly identical. In those cases, we compute the implied MRPL, $\tilde{z}^{implied}_{a_i}(z)$, as the sales-weighted average of implied MRPLs of firms with the same productivity.

[78]Because of the monotonicity restriction, this simple modified bisection algorithm is preferable to Jacobian-based methods, as well as more efficient.

case, the candidate equilibrium is sustainable and the model is solved. Otherwise, an extensive margin adjustment needs to be enacted:

– Following the equilibrium refinement device of Berry (1992), the firm making lowest (negative) profits is removed from the market. Since our model features identical sectors sourcing from the same labor market, we remove one of the worst loss-making firms in each identical sector at a time.[79] Then, the algorithm restarts from step 1.

[79]If the worst loss-making firms are more than 100 in each sector, we remove 10% of them to speed up the algorithm. We check that this shortcut has no bearing on the results.

B.13 Efficiency Properties of Baseline Equilibrium

In this section we characterize the constrained efficient allocation of our baseline model. To do so, we solve the problem of a social planner that aims to maximize aggregate consumption in steady state: [80]

$$\mathcal{W} = \max_{\substack{\vec{v}^j_{a_i}(z_n)\geq 0,\\ \vec{q}_{1,a_i}(n,n')\in[0,1],\\ \vec{q}_{2,a_i}(j,j',n)\in[0,1]}} \; Y - \sum_{i=1}^{I}\int aM_{a_i}\left(\sum_{n}^{\mathcal{Z}_i}\int_0^1 c^j_{a_i}\left(\vec{q}_{1,a_i},\vec{q}_{2,a_i},V_{a_i},v_{a_i}(z_n)\right)dj\gamma_{a_i}(z_n)\right)d\Omega_i(a)\Xi(i),$$

$$\text{s.t.}\quad Y=\prod_{i=1}^{I}Y_i^{\alpha_i},\quad Y_i=\left(\int\int_0^{K_{a_i}}Y_{k(a_i)}^{\frac{\rho-1}{\rho}}dk_{a_i}\,d\Omega_i(a)\right)^{\frac{\rho}{\rho-1}}, \tag{75}$$

$$Y_{k(a_i)}=\left(N_{k(a_i)}\sum_{n=1}^{\mathcal{Z}_i}\int_0^1 y^j_{k(a_i)}(z_n)^{\frac{\sigma-1}{\sigma}}dj\,\gamma_{k(a_i)}(z_n)\right)^{\frac{\sigma}{\sigma-1}}, \tag{76}$$

$$y^j_{k(a_i)}(z_n)=az_n\ell^j_{a_i}\left(\vec{q}_{a_i}(z_n,z_{-n}),V_{a_i},v_{a_i}(z_n)\right), \tag{77}$$

$$\ell^j_{a_i}\left(\vec{q}_{a_i},V_{a_i},\vec{v}_{a_i}(z)\right)=\frac{v^j_{a_i}(z_n)}{V_{a_i}}\lambda_{a_i}u_{a_i}\frac{h^j_{a_i}(1,n)h^j_{a_i}(0,n)+s_{a_i}\lambda_{a_i}\left[(1-\bar{\mathcal{H}}^j_{2,a_i}(n))h^j_{a_i}(1,n)+\dots\right.}{h^j_{a_i}(1,n)\,h^j_{a_i}(0,n)\,h^j_{a_i}(q^j_2(n),n)} \tag{78}$$

$$\frac{\left.\dots(1-q^j_2(n))\left((1-\bar{\mathcal{H}}^j_{1,a_i}(n))h^j_{a_i}(0,n)-(1-\bar{\mathcal{H}}^j_{2,a_i}(n))h^j_{a_i}(1,n)\right)\right]}{}, \tag{79}$$

$$u_{a_i}(\lambda_{a_i})=\frac{\delta_{a_i}}{\delta_{a_i}+\lambda_{a_i}}, \tag{80}$$

$$\lambda_{a_i}(V_{a_i})=\mathcal{M}\left(1,\frac{V_{a_i}}{S_{a_i}(\lambda_{a_i}(V_{a_i}))}\right), \tag{81}$$

$$h^j_{a_i}(p,n)=\delta_{a_i}+s_{a_i}\lambda_{a_i}\left[1-p\left(1-\bar{\mathcal{H}}^j_{1,a_i}(n)\right)+(1-p)\left(1-\bar{\mathcal{H}}^j_{2,a_i}(n)\right)\right], \tag{82}$$

$$\bar{\mathcal{H}}^j_{1,a_i}\left(\vec{q}_{1,a_i},\vec{q}_{2,a_i},V_{a_i},\vec{v}_{a_i}(z_{n'})\right)=M_{a_i}\sum_{n'\neq n}q_{1,a_i}(n,n')\frac{\bar{v}_{a_i}(z_{n'})}{V_{a_i}}\gamma_{a_i}(z_{n'})+M_{a_i}\frac{\bar{v}_{a_i}(z_n)}{V_{a_i}}\gamma_{a_i}(z_n)q^j_2(n), \tag{83}$$

$$\bar{\mathcal{H}}^j_{2,a_i}\left(\vec{q}_{1,a_i},\vec{q}_{2,a_i},V_{a_i},\vec{v}_{a_i}(z_{n'})\right)=M_{a_i}\sum_{n'\neq n}q_{1,a_i}(n,n')\frac{\bar{v}_{a_i}(z_{n'})}{V_{a_i}}\gamma_{a_i}(z_{n'}), \tag{84}$$

$$q^j_2(n)=\int_0^1 q_{2,a_i}(j,j',n)\frac{v^{j'}_{a_i}(z_n)}{V_{a_i}(z_n)}dj', \tag{85}$$

$$q_{1,a_i}(n,n')+q_{a_i}(n',n)\leq 1,\ \forall n,n',\quad q_{2,a_i}(j,j',n)+q_{2,a_i}(j',j,n)\leq 1,\ \forall j,j',n, \tag{86}$$

where $V_{a_i}=M_{a_i}\sum_n\bar{v}_{a_i}(z_n)\gamma_{a_i}(z_n)$, $S_{a_i}(\lambda_{a_i})=u_{a_i}(\lambda_{a_i})+s_{a_i}(1-u_{a_i}(\lambda_{a_i}))$, and $c^j_{a_i}(z_n)=\bar{c}_{a_i}h_{a_i}(0,z_n)^{\phi}\frac{v^j_{a_i}(z_n)^{1+\zeta}}{1+\zeta}$.

The social planner seeks to maximize aggregate consumption by leveraging two sets of control variables. First, the planner mandates each firm j with productivity z and skill

[80] Hence, we abstract both from distributional concerns and elastic labor supply.

requirement a operating in industry i to post a mass of vacancies $v^j_{a_i}(z)$. Second, the planner chooses the probability $q_{1,a_i}(n,n')$ with which a worker with skill a in industry i employed in a firm with productivity z_n transitions into a firm with productivity $z_{n'}$ if given the chance, and the probability $q_{2,a_i}(j,j',n)$ with which a worker with skill a in industry i employed in a firm with productivity z_n transitions into a firm with the same productivity if given the chance. These transition probabilities give rise to a a *shadow wage offer distribution* $\bar{\mathcal{H}}^j_{a_i}\left(\vec{q}_{1,a_i},\vec{q}_{2,a_i},V_{a_k},\vec{v}^{j'}_{a_i}(z_{n'})\right)$ as defined by (83).[81] Exactly as the equilibrium wage offer distribution, the shadow wage offer distribution plays an allocative role. Intuitively, firms' ranking in the $\mathcal{H}$ distribution determines the direction of worker reallocation. The problem setup allows for residual dispersion in policy functions if identical firms are dictated different reallocation patterns, i.e., if $q^j_2(n)\neq q^{j'}_2(n)$. The stationary employment constraint (78) equals the flow-consistent firm-level employment for a generic residual transition probability $q^j_2(n)$ among equally productive firms in the presence of finite productivity types (see (49) for its equilibrium counterpart). This expression has the property that $M_{a_i}\sum_n \bar{\ell}_{a_i}(z_n)=1-u_{a_i}(\lambda_{a_i})$, thus effectively representing a (frictional) labor resource constraint.

We now substitute for the tightness constraint (81) into the Beveridge curve (80) and plug the latter into the objective function. The Lagrangean associated to the social planner problem reads:

$$\mathcal{L}(\vec{v}^j_{a_i}(z),\vec{q}_{1,a_i},\vec{q}_{2,a_i}) = Y\left(\vec{\ell}^j_{a_i}\left(\vec{q}_{1,a_i},\vec{q}_{2,a_i},V_{a_i},\vec{v}^j_{a_i}(z)\right)\right) - \sum_{i=1}^{I}\int a\Bigg[M_{a_i}\Bigg(\sum_n^{\mathcal{Z}_i}\int_0^1 c^j_{a_i}\left(\vec{q}_{1,a_i},\vec{q}_{2,a_i},V_{a_i},\vec{v}^j_{a_i}(z_n)\right)dj$$
$$\gamma_{a_i}(z_n)\Bigg) - \sum_n\sum_{n'}\xi_{1,a_i}(n,n')[1-q_{1,a_i}(n,n')-q_{a_i}(n',n)] - \sum_n\int_0^1\int_0^1\xi_{2,a_i}(j,j',n)[1-q_{2,a_i}(j,j',n)-q_{a_i}(j',j,n)]$$
$$+\int_0^1\xi^j_{v,a_i}(z_n)v^j(z_n)dj\Bigg]d\Omega_i(a)\Xi(i),$$

where $\xi^j_{v,a_i}(z_n)\geq 0$ is the Kuhn Tucker multiplier attached to the non-negativity constraint on vacancies, while $\xi_{1,a_i}(n,n')$ and $\xi_{2,a_i}(j,j',n)$ are the Lagrange multiplier attached to the adding-up constraints on the transition probabilities.

The efficient transition probability from two firms j and j' with the same productivity z_n solves:

$$\frac{\partial\mathcal{L}}{\partial q_{2,a_i}(j,j',n)} = \left[mp^j_{a_i}(z_n)\frac{\partial\ell^j_{a_i}(z_n)}{\partial q_{2,a_i}(j,j',n)} - \frac{\partial c^j_{a_i}}{\partial q_{2,a_i}(j,j',n)}\right] - \left[mp^{j'}_{a_i}(z_n)\frac{\partial\ell^{j'}_{a_i}(z_n)}{\partial q_{2,a_i}(j',j,n)} - \frac{\partial c^{j'}_{a_i}}{\partial q_{2,a_i}(j',j,n)}\right] \geq 0, \tag{87}$$

$mp^j_{a_i}(z_n)\equiv\frac{\partial Y}{\partial\ell^j_{a_i}(z_n)}$ is the marginal product of labor of firm j with productivity z_n. Condition (87) is obviously satisfied with equality if $q^j_{2,a_i}=\frac{1}{2}\ \forall j$, so that all the firms with a certain productivity in a given labor market have the same vacancy yield. In principle, the planner would be indifferent among any potentially nondegenerate distributions of q^j_{2,a_i} consistent

[81] Notice that (84) differs from (83) in that it ignores transitions between equally productive firms.

with (87) holding as equality. Since firms are structurally identical, candidate distributions should have the property that $\frac{\partial}{\partial q^{j}_{2,a_i}}\left[mp^{j}_{a_i}(z_n)\frac{\partial \ell^{j}_{a_i}(z_n)}{\partial q_{2,a_i}(j,j',n)}-\frac{\partial c^{j}_{a_i}}{\partial q_{2,a_i}(j,j',n)}\right]=0$. However, no such distributions generally exist. [82] Hence, we can specialize on the problem setup with no residual dispersion, thus dropping the j index henceforth. As a result, the stationary employment constraint (78) simplifies to its familiar expression:

$$\ell_{a_i}\left(\vec{q}_{a_i}, V_{a_i}, \vec{v}_{a_i}(z)\right)=\frac{v_{a_i}(z_n)}{V_{a_i}}\lambda_{a_i}u_{a_i}\frac{\delta_{a_i}+s_{a_i}\lambda_{a_i}}{h_{a_i}(1,n)h_{a_i}(0,n)}.$$

The efficient transition probability from a firm with productivity z_n to a firm with productivity $z_{n'}$ in labor market a_i solves:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial q_{1,a_i}(n,n')} &= \left[mp_{a_i}(z_n)\frac{\partial \ell_{a_i}(z_n)}{\partial q_{1,a_i}(n,n')}-\frac{\partial c_{a_i}}{\partial q_{1,a_i}(n,n')}\right]\gamma_{a_i}(z_n)-\left[mp_{a_i}(z_{n'})\frac{\partial \ell_{a_i}(z_{n'})}{\partial q_{1,a_i}(n',n)}-\frac{\partial c_{a_i}(z_{n'})}{\partial q_{1,a_i}(n',n)}\right]\gamma_{a_i}(z_{n'})\\ &=\left[mp_{a_i}(z_n)\left(\frac{d\tilde{\ell}_{a_i}(z_n)}{d\bar{\mathcal{H}}_{1,a_i}(n)}+\frac{d\tilde{\ell}_{a_i}(z_n)}{d\bar{\mathcal{H}}_{2,a_i}(n)}\right)-mp_{a_i}(z_{n'})\left(\frac{d\tilde{\ell}_{a_i}(z_{n'})}{d\bar{\mathcal{H}}_{1,a_i}(n')}+\frac{d\tilde{\ell}_{a_i}(z_{n'})}{d\bar{\mathcal{H}}_{2,a_i}(n')}\right)\right]\\ &\quad+\frac{\phi}{1+\zeta}\bar{c}_{a_i}\left[\frac{v_{a_i}(z_{n'})^{\zeta}}{h_{a_i}(n')^{1-\phi}}-\frac{v_{a_i}(z_n)^{\zeta}}{h_{a_i}(n)^{1-\phi}}\right]\geq 0,\end{aligned} \tag{88}$$

where $\tilde{\ell}=\ell/v$ is the vacancy yield (independent of v) and the second line follows from simplifying for $\frac{\partial \bar{\mathcal{H}}_{s,a_i}(n)}{\partial q_{1,a_i}(n,n')}=\frac{\partial \bar{\mathcal{H}}_{s,a_i}(n)}{\partial q_{1,a_i}(n,n')}=\frac{v_{a_i}(z_{n'})}{V_{a_i}}\gamma_{a_i}(z_{n'})$, $s=1,2$, and the functional form of hiring costs. Notice that (88) describes a knife-edge condition. Specifically, it holds as equality if hiring rates, marginal products, and vacancies are all equalized across all firms. However, if hiring rates and vacancies are equalized, all firms have the same employment according to (78). If firms have the same employment, marginal products are strictly increasing in productivity. Therefore, equation (88) needs to hold as inequality, thus describing a corner solution. Formally, $q_{1,a_i}(n,n')\in\{0,1\}\ \forall n\neq n'$. We proceed by guessing and verifying that the planner chooses to reallocate workers towards higher-productivity firms whenever an opportunity arises, that is, $q_{a_i}(n,n')=1 \iff z_n<z_{n'}$. Hence, the shadow wage offer distribution equals the (complement to 1 of the) vacancy-weighted productivity distribution, i.e., $\bar{\mathcal{H}}_{a_i}\left(V_{a_i},\vec{v}_{a_i}(z_{n'})\right)=\sum_{n'=n+1}^{\mathcal{Z}_i}\frac{\bar{v}_{a_i}(z_{n'})}{V_{a_i}}\gamma_{a_i}(z_{n'})$.

Efficient vacancy posting by a firm with productivity z operating in labor market a_i meets the following conditions:

[82] Assume firms with identical productivity z_n are assigned a signal $\varrho\sim\Phi_n$. Hence, the planner can condition the choice of the residual transition probability on ϱ, e.g., $q_{2,a_i}(\varrho,\varrho',n)=\mathcal{F}(\varrho/\varrho')$. If the planner chooses to do so, then $\frac{\partial q^{\varrho}_{2,a_i}}{\partial \varrho}\neq 0$. For this strategy to be efficient, it must be that $\frac{\partial}{\partial \varrho}\left[mp^{\varrho}_{a_i}(z_n)\frac{\partial \ell^{\varrho}_{a_i}(z_n)}{\partial q_{2,a_i}(\varrho,\varrho',n)}-\frac{\partial c^{\varrho}_{a_i}}{\partial q_{2,a_i}(\varrho,\varrho',n)}\right]=0$. However, since efficient vacancies are increasing in vacancy yield and marginal product of labor, the derivative will be generally positive.

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial v_{a_i}(z_n)} = & \left[mp_{a_i}(z_n) \frac{\partial \ell_{a_i}(z_n)}{\partial v_{a_i}(z_n)} - \frac{\partial c_{a_i}(z_n)}{\partial v_{a_i}(z_n)} \right] M_{a_i} \gamma_{a_i}(z_n) \\
& + \sum_{n' \leq n}^{\mathcal{Z}_i} \left[mp_{a_i}(z_{n'}) \frac{\partial \ell_{a_i}(z_{n'})}{\partial \bar{\mathcal{H}}_{a_i}(n')} - \frac{\partial c_{a_i}(z_{n'})}{\partial \bar{\mathcal{H}}_{a_i}(n')} \right] \frac{\partial \bar{\mathcal{H}}_{a_i}(n')}{\partial v_{a_i}(z_n)} M_{a_i} \gamma_{a_i}(z_{n'}) \\
& + \sum_{n'=1}^{\mathcal{Z}_i} \left[mp_{a_i}(z_{n'}) \left(\frac{\partial \ell_{a_i}(z_{n'})}{\partial V_{a_i}} + \frac{\partial \ell_{a_i}(z_{n'})}{\partial \lambda_{a_i}} \frac{\partial \lambda_{a_i}}{\partial V_{a_i}} + \frac{\partial \ell_{a_i}(z_{n'})}{\partial \bar{\mathcal{H}}_{a_i}(n')} \frac{\partial \bar{\mathcal{H}}_{a_i}(n')}{\partial V_{a_i}} \right) \right. \\
& \left. - \frac{\partial c_{a_i}(z_{n'})}{\partial \bar{\mathcal{H}}_{a_i}(n')} \frac{\partial \bar{\mathcal{H}}_{a_i}(n')}{\partial V_{a_i}} \right] \frac{\partial V_{a_i}}{\partial v_{a_i}(z_n)} M_{a_i} \gamma_{a_i}(z_{n'}) \geq 0,
\end{aligned} \tag{89}
$$

$$
\xi_{v,a_i}(z_n) v_{a_i}(z_n) = 0, \tag{90}
$$

where (89) is the FOC with respect to vacancies, and (90) the respective complementary slackness condition.

The first line of (89) represents the *direct effect* of $v_{a_i}(z_n)$ on the firm's value added, that is, the excess marginal product of labor induced by the marginal vacancy over the marginal hiring cost.

The second line of (89) represents the *business-stealing effects* of $v_{a_i}(z)$ on the value added of lower-productivity firms operating in the same labor market. Specifically, the constrained social planner internalizes how vacancy posting by firms with higher productivity reduces the employment of firms with lower productivity, i.e., how the pace of worker reallocation affects the cross-sectional distribution of employment.[83]

The third and fourth line of (89) represent the *congestion effects* of $v_{a_i}(z_n)$ on the value added produced by workers operating in the same labor market. Specifically, the social planner internalizes that vacancy posting by some firm affects the meeting rate of all the firms via the induced change in aggregate vacancies and, in turn, labor market tightness. These negative (thin-market) externalities exerted on other firms are weighed against the positive (thick-market) externalities exerted on workers, that is, the marginal effect of $v_{a_i}(z_n)$ on the job finding rate – again mediated by the induced change in aggregate vacancies and labor market tightness.

Since both the shadow wage offer distribution and aggregate vacancies are linear in firm-level vacancies, equation (89) carries two important insights. First, conditional on a given rank in the productivity distribution, the business-stealing effects of vacancy posting by firm z_n in labor market a_i are independent of the mass of vacancies it posts. Second, the congestion effects of vacancy posting by firm z in labor market a_i are independent of the mass of vacancies it posts.

Let $E^{bs}_{a_i}(n)$ and $E^{c}_{a_i}$ denote the business-stealing effects and congestion effects of $v_{a_i}(z_n)$ on value added, respectively. It follows that vacancy posting of active firms in the efficient

[83] Intuitively, the higher the share of aggregate vacancies accounted for by firms with higher productivity, the lower the net poaching rate of firms with lower productivity.

allocation solves:

$$mp_{a_i}(z_n)\frac{\partial \ell_{a_i}(z_n)}{\partial v_{a_i}(z_n)} = \frac{\partial c_{a_i}(z_n)}{\partial v_{a_i}(z_n)} + E^{bs}_{a_i}(n) + E^{c}_{a_i}. \tag{91}$$

In words, the social planner equalizes the social marginal benefit of vacancy posting by firm z_n (left-hand side) to its social marginal cost (right-hand side). The latter is composed by the marginal vacancy posting cost and the external effects.

Notice that the social marginal benefit of vacancy posting equals the product between marginal product of labor and vacancy yield. Since the marginal product of labor is increasing in productivity, efficient vacancies are supermodular in productivity and vacancy yield. This property (coupled with the supermodularity of the output function in productivity and size) allows verifying our guess that the planner reallocates workers towards higher-productivity firms when it is given the chance, as an application of Topkis's theorem.

Relation with equilibrium allocation. In equilibrium, wages – rather than marginal products – play an allocative role via worker reallocation. Hence, for the efficient allocation to be sustainable in equilibrium, wages need to be increasing in productivity, i.e., $w'_{a_i}(z) > 0$. Notice that the social planner is indifferent between the composition of marginal costs between wages and markdowns.[84] However, comparing (91) with its equilibrium counterpart (24), it is apparent that both feature an "additive markdown", i.e., the difference between marginal (revenue) product and wage, defined as $\frac{\partial c_{a_i}(z)/\partial v_{a_i}(z)}{\partial \ell_{a_i}(z)/\partial v_{a_i}(z)}$. Hence, we proceed by defining the *shadow wage* as $w^{\star\star}_{a_i}(z) \equiv \frac{E^{bs}_{a_i}(n)+E^{c}_{a_i}}{\partial \ell_{a_i}(z)/\partial v_{a_i}(z)}$ and the *shadow markdown* as $\psi^{\star\star}_{a_i}(z) \equiv 1 + \frac{\partial c_{a_i}(z)/\partial v_{a_i}(z)}{w^{\star\star}_{a_i}(z)\partial \ell_{a_i}(z)/\partial v_{a_i}(z)}$. It follows that equilibrium markdowns are efficient if and only if wages are efficient. The efficient allocation would decentralize by dictating:

$$\mu_{a_i}(z_n) = 1, \tag{92}$$

$$\psi_{a_i}(z_n) = 1 + \frac{(1+\zeta)c_{a_i}(z_n)}{w_{a_i}(z_n)\ell_{a_i}(z_n)}, \tag{93}$$

$$w_{a_i}(z_n) = \frac{E^{bs}_{a_i}(n) + E^{c}_{a_i}}{\partial \ell_{a_i}(z_n)/\partial v_{a_i}(z_n)}. \tag{94}$$

Condition (92) makes sure that posted prices are efficient, i.e., equalizing the marginal product to the marginal *revenue* product of labor, condition (93) makes sure that markdowns are efficient for given wage policy, condition (94) makes sure that the wage policy induces efficient vacancy posting, i.e., equalizing social and private marginal cost.

How does the equilibrium wage function compare to its efficient counterpart (94)? First of all, we notice that residual wage (and marginal product) dispersion is inefficient. Yet, residual wage dispersion is needed to sustain the unique (mixed-strategy) Nash equilibrium in the wage-posting game with finite number of productivity types and continuous wage

[84] Since labor supply is inelastic, it is not influenced by the composition of income between wages and profits.

distributions. This allows establishing our first lemma:

Lemma A.4 (Efficiency requires infinite productivity types)
If the number of productivity types is finite, the wage-posting equilibrium with a continuous wage offer distribution is inefficient.

Hence, we proceed by comparing efficient and equilibrium allocation in the limit case of our economy as $\mathcal{Z}_i \to \infty$. We carry out the comparison in two steps. First, we show that the equilibrium wage function internalizes the business-stealing effects of vacancy posting if the reservation wage is zero, absent any additional source of inefficiency (i.e., provided that the marginal revenue product of labor equals its marginal product and congestion effects are zero). Second, we claim that equilibrium wages are generally inefficient for given marginal product function due to congestion effects.

Lemma A.5 (Wage posting internalizes business-stealing effects)
For given marginal product function $mp_{a_i}(z)$ and a zero reservation wage ($R_{a_i} = 0$), the equilibrium wage function internalizes the business-stealing effects of vacancy posting. Formally,

$$v_{a_i}(z) = v^\star_{a_i}(z) \implies mp_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} = \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} + E^{bs}_{a_i}(\mathcal{H}_{a_i}(z)),$$

where $v^\star_{a_i}(z)$ is the equilibrium vacancy policy function and $\mathcal{H}_{a_i}(z)$ is the equilibrium wage offer distribution.

Proof. See B.13.1. ■

In general, equilibrium vacancies and wages are inefficient for three reasons: reservation wages are positive, net markups are positive, and congestion effects are different from zero. [85] Indeed, evaluating (89) at the equilibrium solution yields:

$$\frac{\partial \mathcal{L}}{\partial v_{a_i}(z)}\Big|_{v_{a_i}(z)=v^\star_{a_i}(z)} = \frac{\mu_{a_i}(z)-1}{\mu_{a_i}(z)} mp_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} + (1-\mathcal{P}_{a_i}(z))\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} R_{a_i} - E^c_{a_i} \lesseqgtr 0,$$

where $\mathcal{P}_{a_i}(z) \equiv 1 - e^{-\int_{\underline{z}}^{z} x_{a_i}(\hat{z})d\hat{z}} \in (0,1]$ and $x_{a_i}(z) \equiv \frac{2\nu s_{a_i}\lambda(\theta_{a_i})\mathcal{H}'_{a_i}(z)}{\delta_{a_i}+s_{a_i}\lambda(\theta_{a_i})[1-\mathcal{H}_{a_i}(z)]}$. Notice that the sign of the congestion effects at the equilibrium solution is a priori ambiguous, though generally different from zero. If congestion effects are positive, i.e., $E^c_{a_i} > 0$, *ceteris paribus* equilibrium marginal costs (wages times markdowns) are lower than efficient. Since congestion effects show up as an additive wedge in the first-order condition (91) and wages are increasing in productivity, lower-productivity firms are farther away from their efficient marginal cost in relative terms. As a result, low-productivity firms typically post an inefficient mass of vacancies, leading to labor misallocation as in Shimer and Smith (2001) and Acemoglu (2001).

[85] Equilibrium wages and markdowns would be the constrained efficient outcome of the wage-posting game if job creation were exogenous and firm size were pinned down by search frictions only (Moscarini and Postel-Vinay, 2013). Intuitively, firm size would be determined by workers' efficient turnover. However, insofar as firms optimize their employment size because of decreasing returns (in production or in revenues) and/or endogenous vacancy posting, equilibrium markdowns are generally inefficient.

We conclude this section by singling out the sources of inefficiency featured by the baseline equilibrium and how the introduction of a minimum wage may affect them. To do so, let $\Delta_{a_i}(z)$ denote the firm-specific *labor wedge*, that is, the ratio between social marginal benefit (left-hand side of (91)) and social marginal cost (right-hand side of (91)) of vacancy posting by a firm with productivity z_n operating in labor market a_i. In the efficient allocation, $\Delta_{a_i}(z) = 1 \; \forall z, \; \forall a_i$. In the baseline equilibrium it equals:

$$\Delta_{a_i}(z) = \frac{\mu_{a_i}(z)}{1 + \frac{E^c_{a_i} - D_{a_i}(z) R_{a_i}}{MC_{a_i}(z)}}, \tag{95}$$

where $D_{a_i}(z) \equiv (1 - \mathcal{P}_{a_i}(z)) \partial \ell_{a_i}(z) / \partial v_{a_i}(z)$, $\partial D_{a_i}\prime(z)/\partial z \geq 0$ and $MC_{a_i}(z) \equiv \psi_{a_i}(z) w_{a_i}(z)$. The baseline equilibrium is inefficient for two reasons. First, positive net markups (due to imperfect substitutability across firm-level varieties) and distorted markdowns (due to congestion effects and positive reservation wages) make the labor wedge generally differ from one in equilibrium. Positive net markups push the labor wedge to be higher than one. Hence, if the net external effects in wage setting at the equilibrium solution are positive, equilibrium aggregate employment is inefficiently low. This means that value added would increase if all the firms posted more vacancies.

Second, heterogeneous markups (due to firms' granularity in their product market) and heterogeneous markdown distortions (due to heterogeneous impact of congestion effects and reservation wages in the cross section) make the labor wedge generally differ across firms in equilibrium. This means that the economy features *misallocation* of labor across firms. Misallocation entails that the economy could produce the same amount of final output with lower aggregate hiring costs. This means that value added would increase by reallocating the equilibrium mass of equilibrium vacancies across firms. Hence, with a large number of productivity types, decentralizing the efficient allocation requires three policy instruments: (i) a size-dependent subsidy to neutralize markup distortions (Edmond et al., 2023), (ii) a linear vacancy posting tax (or subsidy) to neutralize markdown distortions due to congestion effects, and (iii) a firm-specific wage subsidy to neutralize positive reservation wages.

However, the amount of information required to implement such first-best policies is arguably beyond the possibilities of policymakers. On the other hand, policymakers can directly control the reservation wage through MW setting. Since $\partial D_{a_i}(z)/\partial z \geq 0$, the MW is expected to affect more the labor wedge of low-productivity firms. Since, for standard parametrizations, congestion effects are positive and markups are increasing in productivity, low-productivity firms typically post more vacancies than efficient. By constraining their optimal vacancy posting, a higher minimum wage is likely to reduce labor misallocation. In this sense, a minimum wage can be thought of as a second-best policy when policymakers do not have access to more targeted policy instruments.

B.13.1 Proof Lemma A.5

To establish the claim, we define a function $\mathcal{F}_{a_i}(v_{a_i}(z)) \equiv mp_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} - \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} - E^{bs}_{a_i}\left(\mathcal{H}_{a_i}(z)\right)$, which equals zero at the efficient solution, i.e., $\mathcal{F}_{a_i}(v^{\star\star}_{a_i}(z)) = 0$. In equilibrium, optimal vacancy posting solves $mp_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} - \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} - w_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} = 0$. First of all, we notice that the business-stealing effects induced by the lowest-productivity firms are zero, i.e., $E^{bs}_{a_i}\left(\mathcal{H}_{a_i}(\underline{z}_{a_i})\right) = 0$. On the other hand, the lowest-productivity firms pay the reservation wage in equilibrium. Hence, as long as the reservation wage is positive, $\mathcal{F}_{a_i}(v^{\star}_{a_i}(\underline{z}_{a_i})) \neq 0$. Hence, equilibrium vacancy posting cannot be efficient.

We proceed by rearranging the expression for the business-stealing effects of vacancy posting as follows:

$$
\begin{aligned}
E^{bs}(\mathcal{H}_{a_i}(z)) &\equiv -\int_{\underline{z}_{a_i}}^{z}\left[mp_{a_i}(\hat{z})\frac{\partial \ell_{a_i}(\hat{z})}{\partial \bar{\mathcal{H}}_{a_i}(\hat{z})} - \frac{\partial c_{a_i}(\hat{z})}{\partial \bar{\mathcal{H}}_{a_i}(\hat{z})}\right]\frac{\partial \bar{\mathcal{H}}_{a_i}(\hat{z})}{\partial v_{a_i}(\hat{z})}M_{a_i}\gamma_{a_i}(\hat{z})d\hat{z} \\
&= \int_{\underline{z}_{a_i}}^{z}\left[mp_{a_i}(\hat{z})\frac{2s_{a_i}\lambda_{a_i}\bar{\mathcal{H}}'_{a_i}(\hat{z})}{\delta_{a_i}+s_{a_i}\lambda_{a_i}\bar{\mathcal{H}}_{a_i}(\hat{z})}\frac{\ell_{a_i}(\hat{z})}{v_{a_i}(\hat{z})} + \phi\frac{s_{a_i}\lambda_{a_i}\bar{\mathcal{H}}'_{a_i}(\hat{z})}{\delta_{a_i}+s_{a_i}\lambda_{a_i}\bar{\mathcal{H}}_{a_i}(\hat{z})}\frac{\partial c_{a_i}(\hat{z})}{v_{a_i}(\hat{z})}\right]\gamma_{a_i}(\hat{z})d\hat{z} \\
&= \int_{\underline{z}_{a_i}}^{z}\left[\frac{mp_{a_i}(\hat{z})}{\nu}\frac{\ell_{a_i}(\hat{z})}{v_{a_i}(\hat{z})} + \frac{\phi}{2\nu}\frac{\partial c_{a_i}(\hat{z})}{v_{a_i}(\hat{z})}\right]x_{a_i}(\hat{z})\gamma_{a_i}(\hat{z})d\hat{z} \\
&= \int_{\underline{z}_{a_i}}^{z}\left[\frac{mp_{a_i}(\hat{z})}{\nu} + \left(1-\frac{1}{\nu}\right)\frac{\partial c_{a_i}(\hat{z})/\partial v_{a_i}(\hat{z})}{\partial \ell_{a_i}(\hat{z})/\partial v_{a_i}(\hat{z})}\right]\frac{\partial \ell_{a_i}(\hat{z})}{\partial v_{a_i}(\hat{z})}x_{a_i}(\hat{z})\gamma_{a_i}(\hat{z})d\hat{z} \\
&= \int_{\underline{z}_{a_i}}^{z}\left[\frac{mp_{a_i}(\hat{z})}{\nu} + \left(1-\frac{1}{\nu}\right)\left(mp_{a_i}(\hat{z}) - w_{a_i}(\hat{z})\right)\right]\frac{\partial \ell_{a_i}(\hat{z})}{\partial v_{a_i}(\hat{z})}x_{a_i}(\hat{z})\gamma_{a_i}(\hat{z})d\hat{z},
\end{aligned}
$$

where we made use of the following relationships and equilibrium conditions: $\frac{\partial \ell_{a_i}(\hat{z})}{\partial \bar{\mathcal{H}}_{a_i}(\hat{z})} = -\frac{2s_{a_i}\lambda_{a_i}}{\delta_{a_i}+s_{a_i}\lambda_{a_i}\bar{\mathcal{H}}_{a_i}(\hat{z})}\ell_{a_i}(\hat{z})$, $\frac{\partial \bar{\mathcal{H}}_{a_i}(\hat{z})}{\partial v_{a_i}(\hat{z})} = \frac{M_{a_i}}{V_{a_i}}\gamma_{a_i}(\hat{z})$, $\frac{\partial c_{a_i}(\hat{z})}{\partial \bar{\mathcal{H}}_{a_i}(\hat{z})} = \phi\frac{s_{a_i}\lambda_{a_i}}{\delta_{a_i}+s_{a_i}\lambda_{a_i}\bar{\mathcal{H}}_{a_i}(\hat{z})}c_{a_i}(\hat{z})$, $\frac{\partial c_{a_i}(\hat{z})}{\partial v_{a_i}(\hat{z})} = (1+\zeta)\frac{c_{a_i}(\hat{z})}{v_{a_i}(\hat{z})}$, and $\frac{\phi}{2(1+\zeta)\nu} = 1-\frac{1}{\nu}$.

Substituting for the equilibrium condition and the expression for business-stealing effects into $\mathcal{F}_{a_i}(v_{a_i}(z))$ yields:

$$
\mathcal{F}_{a_i}(z) = w_{a_i}(mp_{a_i}(z))\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} - \int_{\underline{z}_{a_i}}^{z}\left[\frac{1}{\nu}mp_{a_i}(\hat{z}) + \left(1-\frac{1}{\nu}\right)\left(mp_{a_i}(\hat{z}) - w_{a_i}(\hat{z})\right)\right]x_{a_i}(\hat{z})\frac{\partial \ell_{a_i}(\hat{z})}{\partial v_{a_i}(\hat{z})}d\hat{z}.
$$

Substituting for $\frac{\partial \ell_{a_i}(\hat{z})}{\partial v_{a_i}(\hat{z})}$ and collecting common terms, the function depends on the following three terms:

$$
\begin{aligned}
\mathcal{F}_{a_i}(z) \propto \frac{w_{a_i}(mp_{a_i}(z))}{h_{a_i}(z)^2} &- \int_{\underline{z}_{a_i}}^{z}\frac{w_{a_i}(\hat{z})}{h_{a_i}(\hat{z})^2}\frac{x_{a_i}(\hat{z})}{\nu}d\hat{z} \\
&+ \int_{\underline{z}_{a_i}}^{z}\frac{\left(mp_{a_i}(\hat{z}) - w_{a_i}(\hat{z})\right)}{h_{a_i}(\hat{z})^2}x_{a_i}(\hat{z})d\hat{z},
\end{aligned}
$$

where $h_{a_i}(z) \equiv \delta_{a_i} + s_{a_i}\lambda_{a_i}(1-\mathcal{H}_{a_i}(z))$ is the hiring rate and $x_{a_i}(z) \equiv \frac{2\nu s_{a_i}\lambda_{a_i}\mathcal{H}'(z)}{h_{a_i}(z)}$.

We now focus on the second term and apply integration by parts:

$$
\begin{aligned}
&\int_{\underline{z}_{a_i}}^{z} \frac{w_{a_i}(\hat{z})}{h_{a_i}(\hat{z})^2} \frac{x_{a_i}(\hat{z})}{\nu} d\hat{z} \\
=&\int_{\underline{z}_{a_i}}^{z} w_{a_i}(\hat{z}) \frac{\partial h_{a_i}(\hat{z})^{-2}}{\partial \hat{z}} d\hat{z} \\
=&\frac{w_{a_i}(mp_{a_i}(z))}{h_{a_i}(z)^2} - \int_{\underline{z}_{a_i}}^{z} \frac{(mp_{a_i}(\hat{z}) - w_{a_i}(\hat{z}))}{h_{a_i}(\hat{z})^2} x_{a_i}(\hat{z}) d\hat{z},
\end{aligned}
$$

where the last equality follows from the slope of the equilibrium wage function (65). Substituting back into the previous expression allows establishing the claim:

$$
v_{a_i}(z) = v^{\star}_{a_i}(z) \implies \mathcal{F}_{a_i}(z) = 0.
$$

C Appendix: Structural Estimation

C.1 Data description and sample selection

INPS data. We leverage social security data drawn from INPS. Our dataset consists of the complete contribution histories of individuals who worked as employees in private-sector for at least one period in their life between 1990 and 2018. We use a random and representative 6.5% sample of this population. In this dataset, the unit of observation is an event that generated a contribution to the pension system. Hence, it contains all labor market events (such as contract activations and terminations) with some information on demographics (gender, age) and a large amount of information on the job, such as total yearly earnings, the number of days worked within the year, the type of contract (permanent or temporary), whether it is a full- or part-time job, and others. This allows us to derive complete job histories for the workers covered in the sample, as well as to distinguish their periods of employment and non-employment.[86] Crucially for our analysis, the dataset also contains firm identifiers, which allow us to estimate the AKM equation.

We select our sample by focusing on individuals aged 25-64. We do this to avoid capturing the latest phase of education in labor market careers. Importantly, we do *not* exclude women nor part-time workers from our sample. We claim that this is potentially important for our analysis: given these groups of workers are overall paid less than average, they therefore constitute a relevant share of the population directly affected by a minimum wage reform.

Finally, we end up with a sample of about 34 million yearly observations, covering 2,6 million distinct workers and 2,3 million distinct firms. Table A.1 reports summary statistics of the main variables that we use in our analysis.

Table A.1: Summary statistics of main variables

	Statistics			
	Mean	SD	Min	Max
Age	37.90	10.70	25	64
Female	0.37	0.48	0	1
Yearly Earnings (EUR)	13,735.99	12,024.65	65	70,075
Days worked	197.50	117.68	0	365
N	33,898,735			

Source: INPS matched employer-employee data (1990-2018).

Istat data. Our data from Istat comes from two sources. First, we make use of publicly available data on the distribution of firms across 4-digit ATECO sectors, as well as on the employment share of firms with different employment size. Second, we leverage micro-data

[86] Notice that in this sample we only observe periods of employment as employees in the private sector and we consider periods of non-employment also periods potentially spent working as self-employed or as employee in the public sector. Transitions from self employment (or public sector employment) to payroll private employment and vice versa are however very low, as estimated from the Italian Labor Force Survey.

from the 2019 wave of the Structural Business Statistics to compute the HHI of 4-digit sectors and the semi-elasticity of firm-level labor share to market share. The dataset consists of the universe of Italian firms operating in the non-agricultural sector. We select our sample to cover all the 4-digit sectors in the economy excluding the financial sector. We consider as missing negative entries for "revenues", "value added", and "labor cost". We define labor share (*LS*) as the ratio between "labor cost" and "value added" at the firm-level. To guard against outliers, we winsorize the labor share at 5%. We compute the firm-level market share (*sh*) in the 4-digit sector in which it operates as the ratio between firm-level revenues and sectoral revenues.

C.2 AKM estimation

For each worker, we first select the dominant job spell within each year. We pick this by maintaining the contract with the highest level of earnings within the year. As a second criterion, for jobs with equal earnings, we pick the one in which the worker worked more days. We then compute average hourly wages by taking the ratio of yearly earnings and the imputed number of hours worked (obtained multiplying the number of days worked by 8 or 4, depending on whether the job is full or part-time). Finally, we estimate a classical AKM equation, that is a log-linear wage regression with worker and firm fixed effects:

$$\log w_{it} = \beta X_{i,t} + \alpha_i + \phi_a + \epsilon_{i,t},$$

where $X_{i,t}$ represents a vector of time-varying controls, α_i's are the workers' fixed effects and ϕ_a's are the firms' fixed effects. We estimate the AKM equation without controls, in order to be consistent with the model, where the only source of heterogeneity is a fixed unobserved productivity component. Importantly, for the AKM estimation we use the longest possible time span in our data, which is the period 1990-2018. It is important that we can estimate the wage equation on such a long period, as this allows us to minimize the risk of capturing temporary shocks to workers' careers.

C.3 Computing estimation targets

We use our Istat micro-data to compute the HHI of each 4-digit sector k as:

$$HHI_k = \sum_{j=1}^{N_{j(k)}} sh_j^2.$$

We then aggregate sector-level HHIs by weighting them for the sectoral value added VA_k to compute our HHI target:

$$HHI = \frac{\sum_{k=1}^{K} VA_k HHI_k}{\sum_{k=1}^{K} VA_k}.$$

Next, we use the same Istat dataset to estimate the following regression model:

$$LS_{j(k)} = \beta\, sh_{j(k)} + \sum_{n=1}^{K} \delta_k \mathbb{1}_{n=k} + \epsilon_{j(k)},$$

where the δ_ks are 4-digit-sector fixed effects. Finally, we normalize the estimated coefficient $\hat{\beta} = -0.274$ by the average labor share (0.610) to compute the semi-elasticity of the firm-level labor share to the market share, which we use as target in the structural estimation.

Having estimated worker fixed effects, we partition our sample from INPS in two subsamples, each of which corresponds to half of the estimated AKM worker fixed effects distribution (low- and high-skilled workers). We then measure worker transition rates separately for each subgroup. In particular, we measure the EN rate as the Poisson rate consistent with the monthly probability of moving from employment to non-employment, the job-to-job (J2J) rate as the Poisson rate consistent with the monthly probability of switching employer without a period of unemployment, and the NE rate as the Poisson rate consistent with the monthly probability of moving from non-employment to employment. We use the estimated transition rates as targets in the structural estimation.

Finally, we focus on the time period between 2016 and 2018 and assign workers to a skill type (based on their estimated AKM worker fixed effect) and to the most frequent 1-digit industry they have been employed at. We use the resulting wage distributions at industry x skill level as targets in the structural estimation.

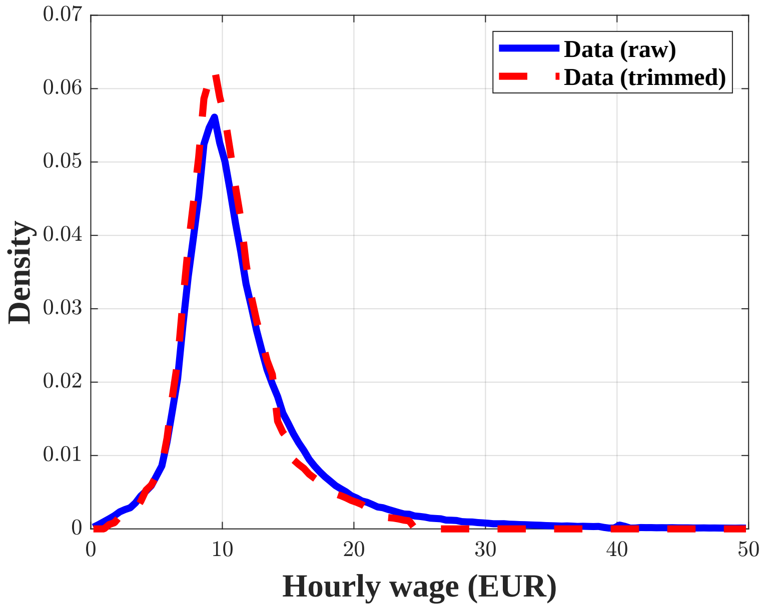
C.4 Estimating productivity from wages

In this section, we detail our strategy to estimate the productivity distribution of firms operating in a labor market from its wage distribution within the SMM routine.

Trimming wage distributions. First of all, we trim the lowest 1% and the highest 100*tr*% of each wage distribution to guard against outliers, where *tr* is a parameter to be estimated targeting the employment share of firms with more than 250 employees. Figure (A.2) reports the difference between the raw and the trimmed empirical wage distribution upon applying the optimal right trimming of $tr = 8.7\%$ delivered by our SMM estimator.

Cleaning wage distributions. In order for the trimmed empirical wage distributions to be replicated by our model, we pass them through a cleaning algorithm that achieves two goals. First, it makes sure that the wage distributions are "admissible", that is, replicable by our wage-posting model. Second, it makes sure that the number of firms in the model matches the number of firms in the data, upon normalizing the number of firms in the MRPL with lowest weighted density to 1. The second step provides us with the key to map the measure of firms in the labor markets to the finite number of firms in the product markets, while matching the actual size of the economy.

Figure A.2: Trimmed wage distribution



Source: INPS data (2016-2018) and model.

Admissibility condition. Recall the density of the MRPL distribution is structurally related to firms' wage and vacancy policies by the following relation:

$$\varphi_{a_i}(\tilde{z}) = \frac{\mathcal{H}'_{a_i}(\tilde{z})}{M_{a_i}} \frac{V_{a_i}}{v_{a_i}(\tilde{z})}. \tag{96}$$

Since $\mathcal{H}_{a_i} = F_{a_i}(w_{a_i}(\tilde{z}))$, it follows that $\mathcal{H}'_{a_i}(\tilde{z}) = w'_{a_i}(\tilde{z}) f_{a_i}(w_{a_i}(\tilde{z}))$. Let us first characterize the first term $w'_{a_i}(\tilde{z})$. Following Bontemps et al. (2000), let $K(\tilde{z})$ denote the wage policy function, i.e., $K_{a_i}(\tilde{z}) = w_{a_i}(\tilde{z})$. Using the FOC for the wage, we can write

$$K_{a_i}^{-1}(w) = w + \frac{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})[1 - F_{a_i}(w)]}{2\nu s_{a_i}\lambda(\theta_{a_i}) f_{a_i}(w)}.$$

Using the derivations in Appendix B.2, we can rewrite this as a function of the observed wage distribution G_{a_i} and its density g_{a_i}:

$$K_{a_i}^{-1}(w) = w + \frac{u_{a_i} + s_{a_i}(1 - u_{a_i})G_{a_i}(w)}{2\nu s_{a_i}(1 - u_{a_i})g_{a_i}(w)},$$

where we also have used steady-state relationships for u_{a_i}. Taking the derivative of this expression yields

$$\left(K_{a_i}^{-1}\right)'(w) = \frac{(1+2\nu)s_{a_i}(1-u_{a_i})g_{a_i}^2(w) - g'_{a_i}(w)\left[u_{a_i} + s_{a_i}(1-u_{a_i})G_{a_i}(w)\right]}{2\nu s_{a_i}(1-u_{a_i})g_{a_i}^2(w)}.$$

Let us now note that $w'_{a_i}(\tilde{z}) = K'_{a_i}\left[K_{a_i}^{-1}(w)\right] = \frac{1}{\left(K_{a_i}^{-1}\right)'(w)}$. Using this fact, along with the previously found expression for the derivative of $K_{a_i}^{-1}$ and expressing again f_{a_i} as a function of g_{a_i} and G_{a_i}, one derives the following expression for $\mathcal{H}'_{a_i}(\tilde{z})$:

$$\mathcal{H}'_{a_i}(\tilde{z}) = \frac{2\nu s_{a_i}(1-u_{a_i})g_{a_i}^2(w)}{(1+2\nu)s_{a_i}(1-u_{a_i})g_{a_i}^2(w) - g'_{a_i}(w)\left[u_{a_i} + s_{a_i}(1-u_{a_i})G_{a_i}(w)\right]} \frac{(\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i}))(1-u_{a_i})u_{a_i}g_{a_i}(w)}{\lambda(\theta_{a_i})\left[u_{a_i} + s_{a_i}(1-u_{a_i})G_{a_i}(w)\right]^2}.$$

Turning back to Equation 96, one can notice that $\varphi_{a_i}(\tilde{z}) \geq 0 \iff \mathcal{H}'_{a_i}(\tilde{z}) \geq 0$, provided that $v_{a_i}(\tilde{z}) > 0$. Inspection of the FOC for the vacancies reveals that vacancies are positive as long as the posted wage piece rate is below the firm's marginal revenue product, which must hold in equilibrium (otherwise firms are making losses):

$$v_{a_i}(\tilde{z}) = \left[\frac{\tilde{z} - w_{a_i}(\tilde{z})}{\bar{c}_{a_i}} \frac{\lambda(\theta_{a_i})u_{a_i}(\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i}))/V_{a_i}}{\left[\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1-\mathcal{H}_{a_i}(\tilde{z}))\right]^{2+\phi}}\right]^{\frac{1}{\zeta}}.$$

This implies that in order to have $\varphi_{a_i}(\tilde{z}) \geq 0$, we need $\mathcal{H}'_{a_i}(\tilde{z}) \geq 0$, which is the case if and only if:

$$(1+2\nu)s_{a_i}(1-u_{a_i})g_{a_i}^2(w) \geq g'_{a_i}(w)\left(u_{a_i} + s_{a_i}(1-u_{a_i})G_{a_i}(w)\right). \tag{97}$$

Before operating the wage inversion, that is, applying Equation 96, we need to ensure that the admissibility condition (97) is verified in our data. Hence, for each labor market, we check the admissibility condition, and find out whether there are points of the empirical wage distributions for which this condition is not met. Essentially, the condition fails when the wage density grows *too quickly*, which in some cases happens in our data in the left part of the distributions. We work around this by applying the simple following algorithm. Let $G_{a_i}^D(w)$ be the discretized version of our wage distributions, that takes values on grid points $[w_1, w_2, ...w_N]$. We proceed as follows:

1. We identify the first grid point i^Y for which the admissibility condition holds, in the left part of the distribution;

2. Starting from w_1, we move mass from the lowest grid points to the $i^Y - 2$ grid point, until the condition for $i^Y - 1$ is met. If during the process the mass of w_1 runs out, then we move to w_2;

3. When the condition for $i^Y - 1$ is met, then we turn to move mass towards $i^Y - 3$;

4. We stop when the first grid point for which the admissibility condition holds is the second one with non-zero mass;

5. At the end of this, we check again the condition over the whole distribution. For all points for which this is not verified, we progressively add mass to the previous grid points, removing it from all other points of the distribution.

In fact, the adjustment takes place exclusively in the left part of the distributions, where rapidly growing density functions cannot be generated by the model. However, as shown in Figure (A.5), the distance between the left tails of the empirical (trimmed) wage distributions and the ones generated by the model is negligible. Hence, we claim that the failure of the admissibility conditions for some data points does not represent an issue for our analysis.

Replicate actual number of firms. Due to the granular product market structure of our model, the number of firms is an important empirical moment to match. To map the measure of firms populating our labor markets to their actual number, we propose the following normalization: we let the number of firms in the MRPL with lowest weighted density be equal to 1. We find this normalization the most sensible to introduce an integer constraint in the number of firms per MRPL level. To make sure that our model replicates the actual number of firms, we implement the following algorithm:

1. Compute $M_{a_i}\varphi_{a_i}(\tilde{z})\ \forall a_i$ by inverting the (trimmed) empirical wage distributions according to the structure of the model (see next paragraph);

2. Let $\underline{\ell}$ and $\underline{\tilde{z}}$ denote the labor market and MRPL level corresponding to $\min\{M_{a_i}\varphi_{a_i}(\tilde{z})\}$. Apply the normalization: $N_{\underline{\ell}}(\underline{\tilde{z}}) = 1$.

3. Compute the number of firms in all the labor markets and MRPL values by rounding the expression $N_{a_i}(\tilde{z}) = \frac{M_{a_i}\varphi_{a_i}(\tilde{z})}{\min\{M_{a_i}\varphi_{a_i}(\tilde{z})\}} \forall a_i, \forall \tilde{z}$.

4. If the total number of firms in the model exceeds the actual number of firms with employees in Italy in 2019 ($1,555,543$), transfer the mass corresponding to the MRPL grid point with lowest number of firms to the closest, nonzero grid point to the left in the same labor market. To preserve the shape of the MRPL distribution, constrain the maximum mass to transfer to half the difference between the receiving grid point and the preceding nonzero grid point. If exceeding, split the mass over multiple grid points;

5. Iterate this procedure until the number of firms in the model is (weakly) lower than in the data. If it is lower, assess whether the error is higher in the last or second-to-last iteration step and choose the most precise one.

As shown in Figure (A.5), the distance between the right tails of the empirical (trimmed) wage distributions and the ones generated by the model is negligible. Hence, we claim that the integer constraint does not represent an issue for our analysis.

Inferring productivity from wages. Upon trimming and cleaning the wage distributions, we are able to use the structure of our model to estimate the underlying productivity distributions to the observed wage distributions. We do so in two steps.

Step 1: MRPL from wage. First, we leverage the equilibrium markdown function (31) and the mapping between wage offer distribution and wage employment distribution (48) to infer the MRPL per wage level based on the shape of the employment wage distribution. Formally,

$$\tilde{z}_{a_i}(w; G_{a_i}(w)) = \psi_{a_i}(w)w = w + \frac{u_{a_i} + s_{a_i}(1-u_{a_i})G_{a_i}(w)}{2\nu s_{a_i}(1-u_{a_i})\Delta G_{a_i}(w)}\Delta w.$$

In this way, we recover the model-consistent MRPL distributions, $\tilde{z}_{a_i} \sim \Phi_{a_i}(\psi_{a_i}(w)w)$. We then impose the integer constraint detailed in the previous paragraph to map the measure of firms into their actual number. We proceed by assigning as many firms as the average number of competing firms across 4-digit sectors in each industry to sectoral markets (or the closest number consistent with the same number of firms populating each sectoral market sourcing from the same labor market). According to the product market structure of our model, product markets sourcing from the same labor market have the same productivity distribution. Hence, we allot firms to sectoral markets in order to minimize differences across them.

Step 2: Productivity from MRPL. Second, for given CES elasticities of demand and MRPL distributions in each product market, there exists a unique combination of prices, markups, Cobb Douglas industry weights, and physical productivities consistent with our model, which solves the following system of equations:

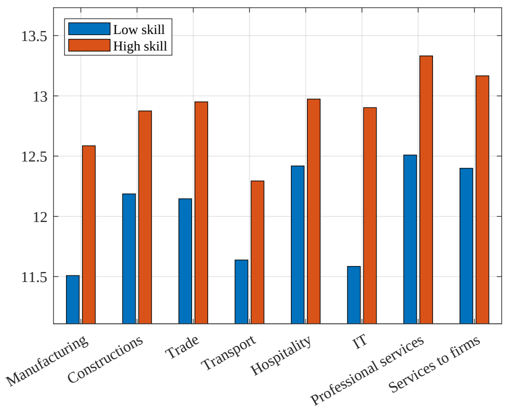
$$\begin{cases} z(\tilde{z}_{k(a_i)}) & = \frac{\mu(\tilde{z}_{k(a_i)})}{p(\tilde{z}_{k(a_i)})}\tilde{z}_{k(a_i)}, \\ p(\tilde{z}_{k(a_i)}) & = y(\tilde{z}_{k(a_i)})^{-\frac{1}{\sigma}} Y_{k(a_i)}(\tilde{\boldsymbol{z}}_{k(a_i)})^{\frac{1}{\sigma}-\frac{1}{\rho}} Y_i(\tilde{\boldsymbol{z}}_i)^{\frac{1}{\rho}-1}\alpha_i(\tilde{\boldsymbol{z}})Y(\tilde{\boldsymbol{z}}), \\ sh(\tilde{\boldsymbol{z}}_{k(a_i)}) & = \frac{p(\tilde{z}_{k(a_i)})y(\tilde{z}_{k(a_i)})}{P_{k(a_i)}(\tilde{\boldsymbol{z}}_{k(a_i)})Y_{k(a_i)}(\tilde{\boldsymbol{z}}_{k(a_i)})}, \\ \mu(\tilde{z}_{k(a_i)}) & = \frac{\sigma}{(\sigma-1)\left[1-\frac{\sigma/\rho-1}{\sigma-1}sh(\tilde{\boldsymbol{z}}_{k(a_i)})\right]}, \\ \alpha_i(\tilde{\boldsymbol{z}}) & = \frac{P_i(\tilde{\boldsymbol{z}}_i)Y_i(\tilde{\boldsymbol{z}}_i)}{Y(\tilde{\boldsymbol{z}})}, \end{cases} \tag{98}$$

where variables in bold denote vectors at the level of aggregation corresponding to their subscript, and aggregators are defined in Equations (15)-(16). Solving the system of equations (98) allows us to back out the productivity distributions that rationalize the observed wage distributions through the lens of our model, $z_{a_i} \sim \Gamma_{a_i}\left(\frac{\mu_{k(a_i)}(\tilde{z})}{p_{k(a_i)}(\tilde{z})}\tilde{z}\right)$.

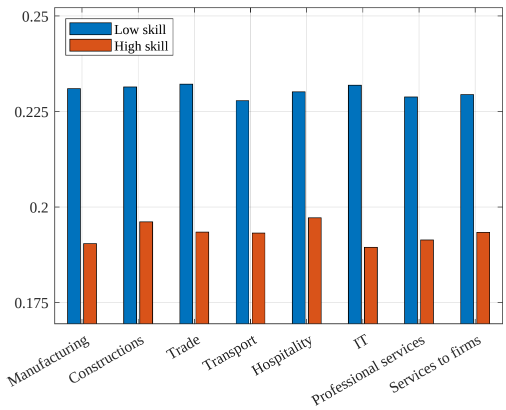
C.5 Estimation results

Figure A.3: Labor market parameters, by labor market

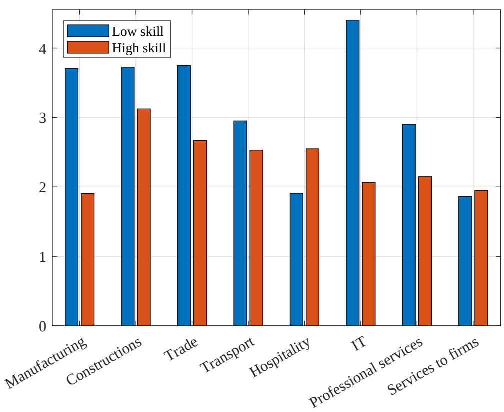
(a) Scalar hiring cost (log scale)



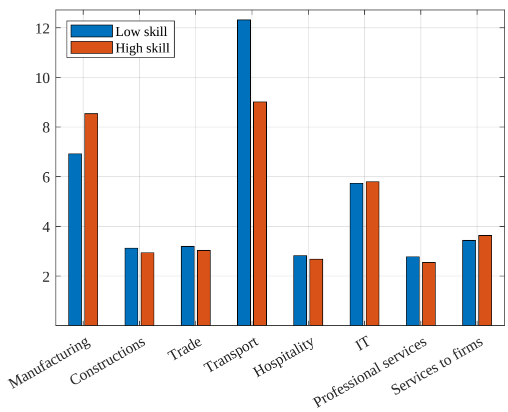
(b) OTJ search efficiency



(c) Unemployment benefits

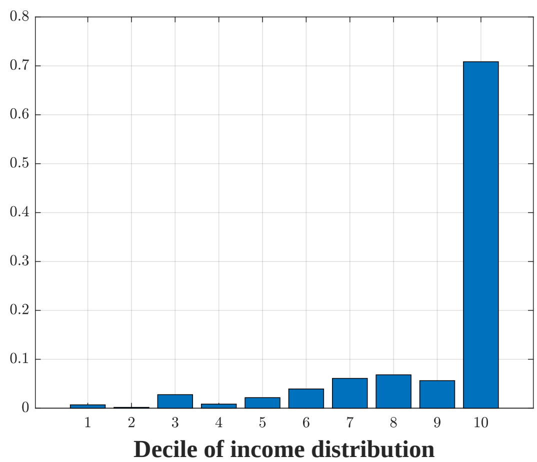


(d) Overhead cost



Source: INPS matched employer-employee data (2016-2018), model.

Figure A.4: Aggregate profits share, by income decile



Source: SHIW (2016). *Note*: The figure plots the share of aggregate profits accruing to each decile of the income distribution.

Table A.2: Workers and firms distribution across labor markets

Industry	Manufacturing		Constructions		Trade		Transport		Hospitality		IT		Professional services		Services to firms	
Skill	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
M_{a_i}	0.054	0.056	0.128	0.133	0.123	0.128	0.045	0.047	0.139	0.144	0.068	0.070	0.164	0.171	0.129	0.134
$d(i)\omega_i(a)$	0.151	0.164	0.044	0.045	0.095	0.083	0.049	0.040	0.067	0.053	0.012	0.020	0.019	0.020	0.079	0.059

Table A.3: Industry weights and empirical revenue shares

Industry	CD weight (model)	Revenue share (data)
Manufacturing	0.355	0.321
Constructions	0.085	0.133
Trade	0.179	0.319
Transport	0.090	0.052
Hospitality	0.092	0.028
IT	0.039	0.034
Professional services	0.042	0.049
Services to firms	0.119	0.065

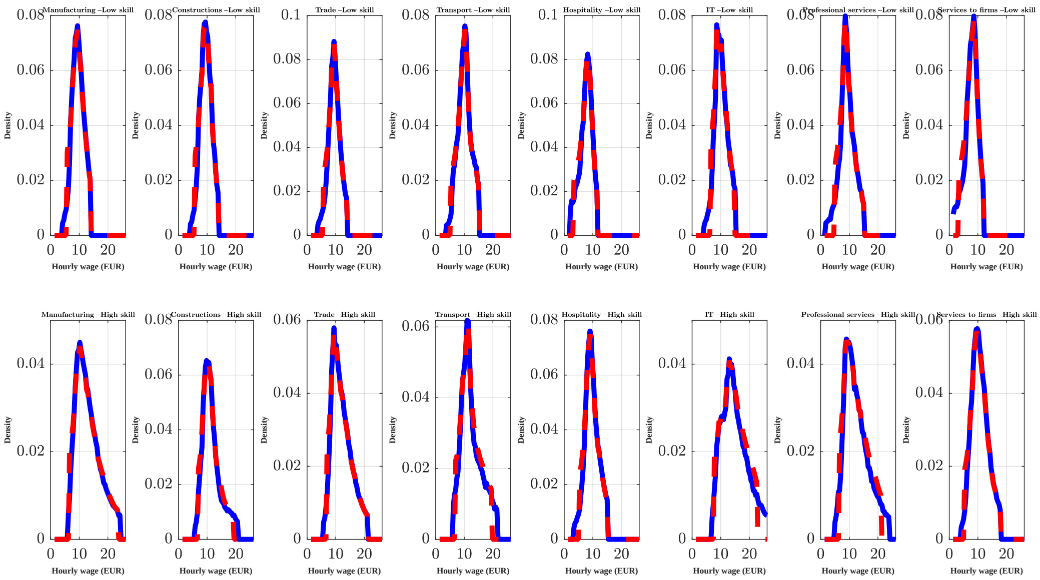
Source: Model and Istat. *Note*: The table report the model-implied Cobb-Douglas industry weights and the empirical industry revenue shares. The empirical values (second column) are computed on Istat data (2019).

Table A.4: Average labor share: first to second employment tercile

Industry	Model	Data
Manufacturing	0.928	0.838
Constructions	0.941	0.910
Trade	0.942	0.888
Transport	0.916	0.855
Hospitality	0.905	0.829
IT	0.948	0.914
Professional services	0.931	0.907
Services to firms	0.868	0.879

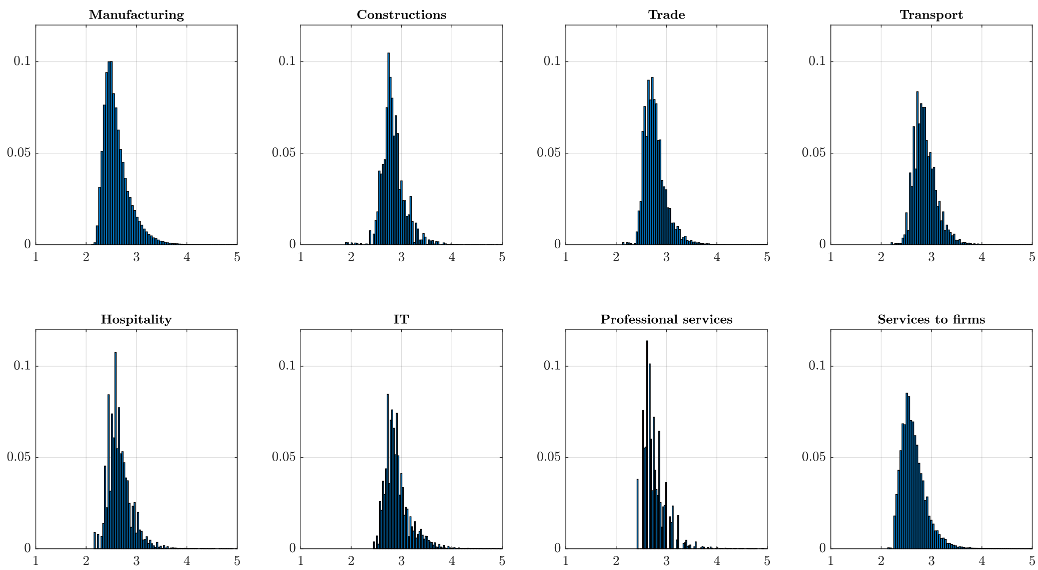
Source: Authors' calculations, model and Istat. *Note*: The table report the ratio between the average labor share in the first and the second tercile of firms ranked by employment for each 1-digit industry. The empirical values (second column) are computed on the Structural Business Statistics dataset of Istat (2019).

Figure A.5: Wage distributions, by labor market



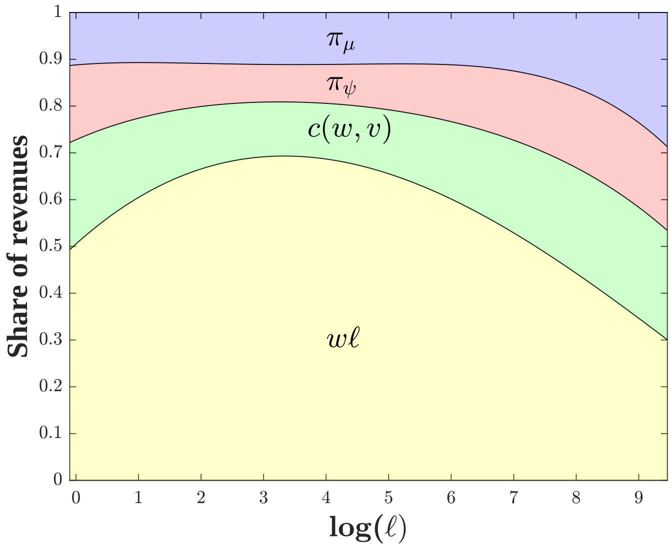
Source: INPS matched employer-employee data (2016-2018). *Note*: Distributions of hourly wage by two worker AKM fixed effects quintiles and industry. The AKM equation is estimated on the period 1990-2018 (see Appendix C.2 for details). The solid blue line represents the trimmed empirical wage distribution, the dashed red line the respective wage distribution in the model.

Figure A.6: Physical productivity distribution, by industry



Source: Model. *Note*: the charts plot the physical productivity distribution across 1-digit industries (ATECO classification), as implied by the model estimates.

Figure A.7: Revenue shares along the firm size distribution



Source: Model. *Note*: The figure reports the average composition of firm-level revenues across log employment.

D Appendix: Equilibrium Effects of the Minimum Wage

Table A.5 reports the effects on the main aggregate variables of the two policy experiments surveyed in the main text: the welfare-maximizing MW equal to the 10^{th} percentile of the original wage distribution (70% Kaitz index), and the consumption-maximizing MW equal to the 40^{th} percentile of the initial distribution (92% Kaitz index).

Table A.5: Policy experiments

Variable	Baseline	70% Kaitz index	92% Kaitz index
	Panel a. Aggregate statistics		
Value Added	1.000	1.014	1.031
Gross output	1.000	1.000	1.002
Labor share	0.733	0.745	0.753
Aggregate welfare	1.000	1.005	0.987
Unemployment rate	0.135	0.155	0.203
Labor productivity	1.000	1.024	1.087
Average wage	1.000	1.055	1.149
Average firm size	8.737	10.32	15.00
	Panel b. Market power statistics		
Aggregate markup	1.139	1.140	1.143
Aggregate markdown	1.471	1.427	1.385
	Panel c. Labor market transitions		
Job finding rate	0.161	0.136	0.098
Job separation rate	0.025	0.025	0.025

Source: Model. *Note*: The variables Value added, Gross output, Aggregate Welfare, Labor productivity, and Average wage are normalized to 1 in the baseline equilibrium.

Motivated by the labor reallocation taking place in our policy experiments, we decompose the overall change of the aggregate variables into a component that relates to the change in the firms' policy function (*behavioral effect*), and a component that reflect the change in weights of different types of firms (*compositional effect*). Formally, an aggregate variable X corresponds to a weighted average of the firm-specific variable $x(i)$, where i denotes a specific firm:

$$X = \int x(j)\,\hat{w}(j)dj,$$

where $\hat{w}(j)$ is the firm-specific weight according to the aggregation results reported in Appendix B.11. As a consequence, the change in the variable X can be decomposed as follows:

$$\Delta X = \underbrace{\int \Delta x(j)\,\hat{w}'(j)dj}_{\text{Behavioral effect}} + \underbrace{\int x(j)\,\Delta\hat{w}(j)dj}_{\text{Compositional effect}}, \tag{99}$$

where $\hat{w}'(j)$ denotes the weight after the change. Notice that the compositional effect comprises reallocation both at the intensive margin and at the extensive margin (firm exit). Table A.6 reports the decomposition results for the most important aggregate variables. We observe that the wage effects result more from workers' reallocation towards high-paying firms (56%) than from firm-level responses (44%). On the contrary the average (wage-bill-weighted) labor share is driven by and large by firm-level responses (89%). The aggregate markup response is mainly shaped by firm-level responses (56%) – that is, the behavioral concentration channel highlighted in Section (2). Finally, we detect opposing effects shaping the aggregate markdown response. On the one hand, firms react to the MW by compressing their markdowns, which reduces the aggregate markdown due to a behavioral effect. On the other hand, workers tend to reallocate towards higher-markdown firms, which increases the aggregate markdown due to a compositional effect. However, the behavioral effect largely outweighs the compositional effect.

Table A.6: Behavior vs. selection: decomposition of main aggregate effects

Variable	Percentage change (perc.)	Behavioral (perc.)	Compositional (perc.)
Baseline to consumption-maximizing MW (92% Kaitz index)			
Average wage	14.85 %	43.59 %	56.41 %
Average labor share	3.23 %	88.67 %	11.33 %
Aggregate net markup	3.21 %	55.66 %	44.34 %
Aggregate net markdown	-18.19 %	109.32 %	-9.32 %

Source: Model. *Note*: The share of change due to behavioral effects is computed by using the new policy functions but keeping the distribution constant as in the baseline; the share of change due to composition is computed by using the new distribution, but keeping the policy functions as in the baseline. See Equation (99) for the formal decomposition.

Unlike the stylized model of Section (2), the quantitative model features hiring costs. As apparent from the equilibrium markdown expression (32), labor market power affects equilibrium hiring costs. We now derive an aggregate index of net market power adjusted for hiring costs. We start by rearranging the equilibrium labor share in value added at the firm-level as follows:

$$
\begin{aligned}
LS_{a_i}(z) &\equiv \frac{aw_{a_i}(z)\ell_{a_i}(z)}{p_{k(a_i)}(z)y_{a_i}(z) - ac_{a_i}(z)} \\
&= \frac{aw_{a_i}(z)\ell_{a_i}(z)}{\mu_{k(a_i)}(z)\psi_{a_i}(z)aw_{a_i}\ell_{a_i}(z) - ac_{a_i}(z)} \\
&= \frac{aw_{a_i}(z)\ell_{a_i}(z)}{\mu_{k(a_i)}(z)\psi_{a_i}(z)aw_{a_i}\ell_{a_i}(z) - \frac{\psi_{a_i}(z)-1}{1+\zeta}aw_{a_i}\ell_{a_i}(z)} \\
&= \frac{1}{\mu_{k(a_i)}(z)\psi_{a_i}(z) - \frac{\psi_{a_i}(z)-1}{1+\zeta}},
\end{aligned}
$$

where in the first step we substitute for the optimal price (25) and in the second step for the optimal markdown in terms of hiring costs (32). From this expression, we see that the labor share depends the firm's market power in both the labor and the product market, as well as that hiring costs are a function of labor market power.

Similarly, we rearrange equilibrium value added at the firm-level as follows:

$$
\begin{aligned}
y_{a_i}(z) &\equiv p_{k(a_i)}(z) y_{a_i}(z) - a w_{a_i}(z) \ell_{a_i}(z) - a c_{a_i}(z) \\
&= \mu_{k(a_i)}(z) \psi_{a_i}(z) a w_{a_i} \ell_{a_i}(z) - a w_{a_i}(z) \ell_{a_i}(z) - a c_{a_i}(z) \\
&= \mu_{k(a_i)}(z) \psi_{a_i}(z) a w_{a_i} \ell_{a_i}(z) - a w_{a_i}(z) \ell_{a_i}(z) - \frac{\psi_{a_i}(z) - 1}{1+\zeta} a w_{a_i} \ell_{a_i}(z) \\
&= \left[\mu_{k(a_i)}(z) \psi_{a_i}(z) - \frac{\psi_{a_i}(z) - 1}{1+\zeta} - 1 \right] a w_{a_i} \ell_{a_i}(z).
\end{aligned}
$$

Hence, we define an adjusted market power index at the firm-level as:

$$
\mathcal{M}^{adj}_{a_i}(z) \equiv \mu_{k(a_i)}(z) \psi_{a_i}(z) - \frac{\psi_{a_i}(z) - 1}{1+\zeta}. \tag{100}
$$

We then apply a first-order approximation to define aggregate adjusted net market power as:

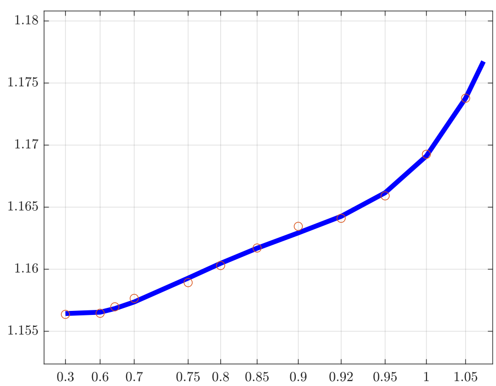
$$
\mathcal{M}^{adj} \equiv \mu \psi - \frac{\psi - 1}{1+\zeta}, \tag{101}
$$

where aggregate variables are defined in Appendix B.11. Finally, we decompose the aggregate adjusted market power response to the MW into markdown and markup component as follows:

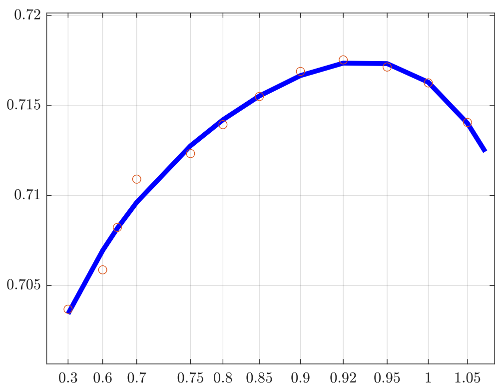
$$
\Delta \mathcal{M}^{adj}(\underline{w}') = \underbrace{\left(\mu(\underline{w}) - \frac{1}{1+\zeta} \right) \Delta \psi(\underline{w}')}_{\text{markdown response}} + \underbrace{\psi(\underline{w}) \Delta \mu(\underline{w}')}_{\text{markup response}}. \tag{102}
$$

Figure A.8: Markup and labor share in manufacturing

(a) Manufacturing markup response



(b) Manufacturing labor share response

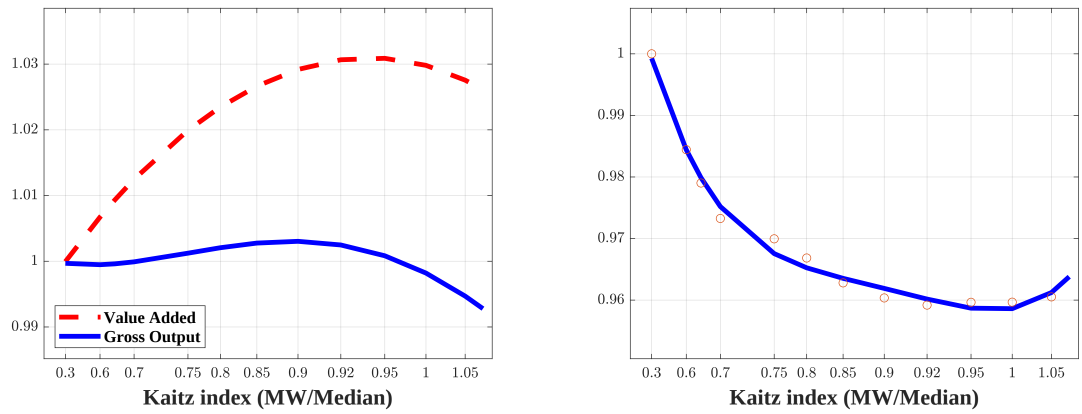


Source: Model. *Note*: Manufacturing markup is the cost-weighted average of the firm-level markup in the manufacturing industry, where weights are given by the wage bill. The x-axis is scaled so as to reflect the share of directly affected workers. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

We now explore to what extent the firms' reaction varies with their size. Indeed, as intuitive, the impact of MW reforms is typically larger for relatively small, low-paying firms, because they are directly constrained by the MW rise. Figure A.12 plots the distribution

Figure A.9: Gross output and profits

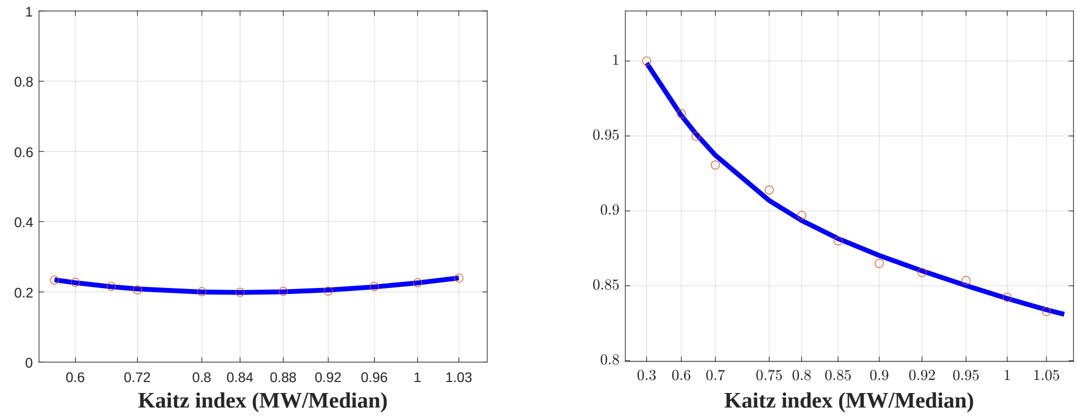
(a) Gross output vs value added response (b) Aggregate profits response



Source: Model. *Note*: Gross output is final good production (Y). Aggregate profits equal the sum of operating profits net of overhead costs. The x-axis is scaled so as to reflect the share of directly affected workers. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

Figure A.10: Employment and productivity effects of the MW – further details

(a) Share of affected jobs destroyed (b) Misallocation index response



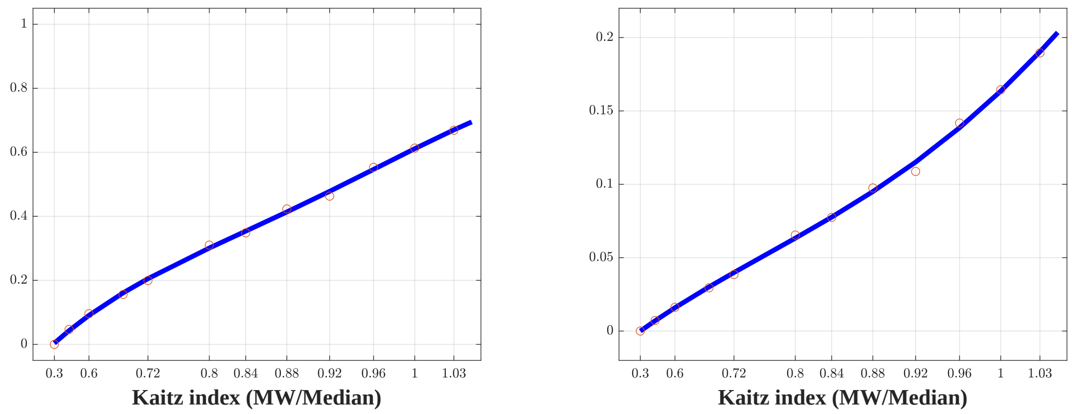
Source: Model. *Note*: The share of affected jobs destroyed is computed as the ratio of workers earning less than the MW in the baseline equilibrium and the difference in the unemployment rate between the simulated MW reform and the baseline. The misallocation index equals total hiring cost per unit of final good relative to the baseline economy. Hiring cost per unit of final good are computed as $\frac{\mathcal{C}_h}{Y}$, where $\mathcal{C}_h$ denotes total hiring costs. The x-axis is scaled so as to reflect the share of directly affected workers. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

of profits and labor costs along the firm size distribution. Increasing market power indices command a decreasing wage bill as a share of value added. In fact, we find that the average labor share is hump-shaped in firm size, reflecting the U-shaped pattern of markdowns (panel a). In response to a Kaitz index of 100%, an increase in the wage bill share becomes visible for smaller firms. This reflects both the direct effects of the reform, i.e., pay rises that are needed to meet the MW, as well as further increases triggered by increases in the *local* competition in the wage distribution. The increase in the wage bill directly erodes profits in the labor market, i.e., the red area π_{ψ}, compressing markdowns. Conversely, large firms increase profits in the product market, i.e., the blue area π_{μ}, according to the concentration channel. The response of market power indices is highlighted in Figure (A.13).

Figure (A.14) plots the average markup charged by firms after a Kaitz index of 100%

Figure A.11: Firm exit effects of the MW

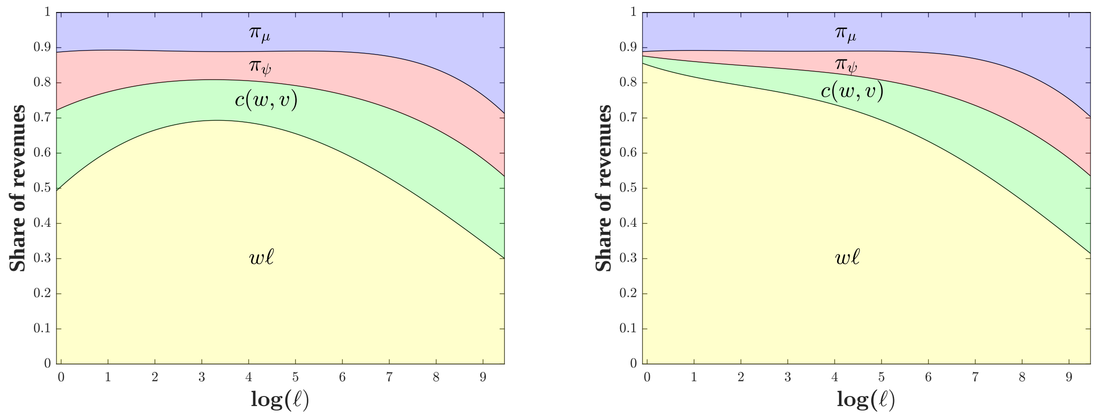
(a) Share of exiting firms, headcount (b) Share of exiting firms, employment share



Source: Model. *Note*: The *share of exiting firms firms, headcount* (Panel A) is the ratio between the number of exiting firms after a MW reform relative to the baseline economy; The *share of exiting firms firms, employment share* (Panel A) is the ratio between the baseline employment of firms exiting after a MW reform relative to total employment in the baseline economy. The x-axis is scaled so as to reflect the share of directly affected workers. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

Figure A.12: Revenue shares response by firm size

(a) Baseline (b) 100% Kaitz index

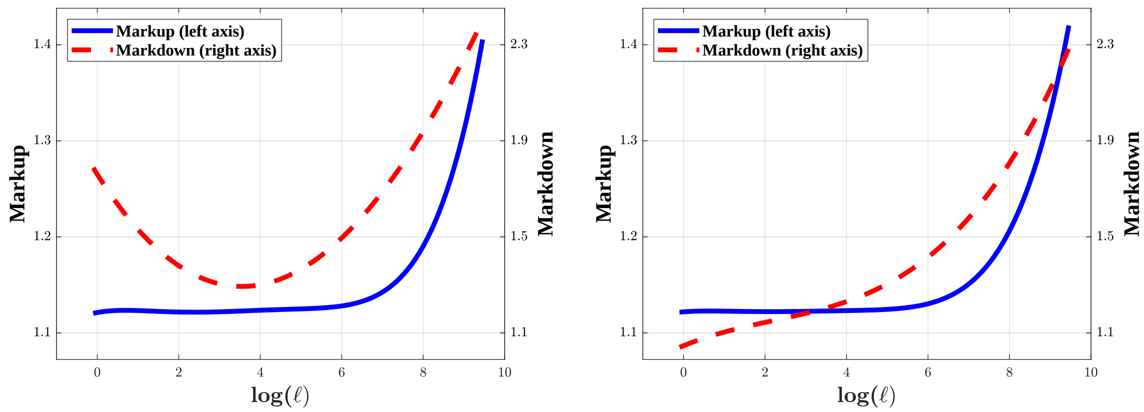


Source: Model.

against their baseline markup (panel a), and the share of the aggregate net market power index response (i.e., aggregate markup times aggregate markdown) accounted for by the aggregate markup response.

Figure A.13: Market power response by firm size

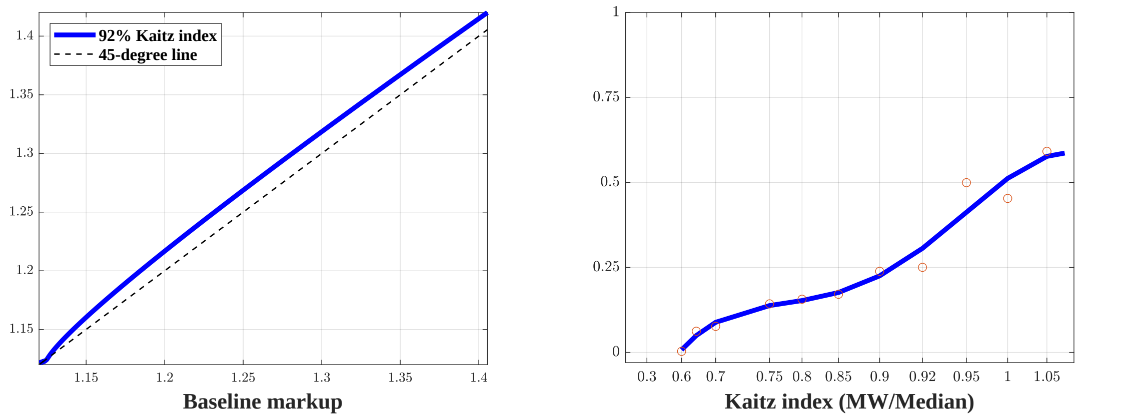
(a) Baseline (b) 100% Kaitz index



Source: Model.

Figure A.14: Markup response

(a) Markup response by firm size (b) Markup share in market power response



Source: Model. *Note*: Panel a) reports the average markup charged by firms following a 92% Kaitz index reform as a function of their average markup in the baseline equilibrium. Panel b) reports the share of the variation of the aggregate adjusted market power index (102) due to the aggregate markup response across MW reforms. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

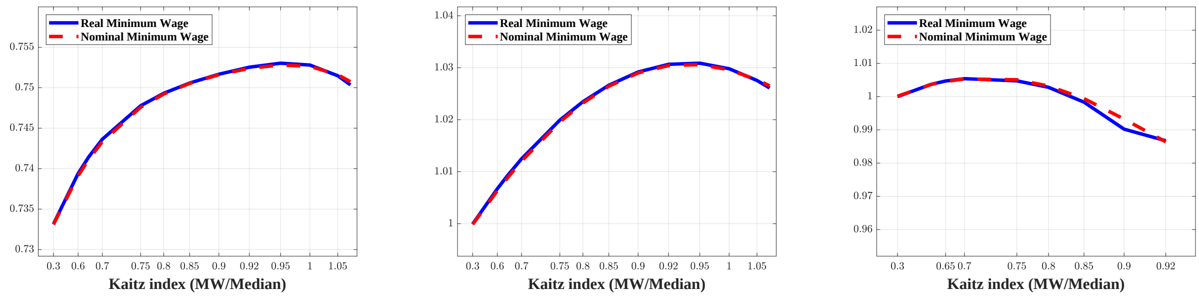
D.1 Robustness: Nominal Minimum Wage

In the main text, we analyze the response of our economy to *real* MWs, i.e. set in terms of final good. However, in the reality, mandated MWs are *nominal*, i.e. set in monetary terms. As a robustness check, we repeat our policy experiments in an economy where money, rather than the final good, is the numeraire. To do so, we keep money supply equal to the nominal output of the baseline economy, i.e., $M^s = PY$, and let the aggregate price level adjust in response to a monetary MW.

In Figure (A.15) we compare the aggregate response of the economy where the MW is set in nominal terms, money is the numeraire, and money supply is fixed, to our baseline. As apparent, setting the MW in real terms or in nominal terms does not make any significant difference.

Figure A.15: Aggregate response: Real vs nominal Minimum Wage

(a) Aggregate labor share (b) Aggregate value added (c) Aggregate welfare



Source: Model. *Note*: The solid blue line represents the response of our baseline economy to a real minimum wage, while the dashed red line represents the response of our baseline economy to a nominal minimum wage. The x-axis is scaled so as to reflect the share of directly affected workers.

D.2 The Role of the Extensive Margin Adjustment

In the main text, we consider a model where firms need to break even their overhead costs with the operating profits to be active. This assumption makes sure that some firms exit in response to MW reforms, in line with the empirical evidence. Since we model overhead costs as sunk, we appeal to some market imperfections, e.g. limited liability, to microfound the break-even condition.

In Figure (A.16) we report the results of the same policy experiments in an economy without liquidity constraints, where no firms exit in response to any MW reform. As expected, firm exit is a key driver of the aggregate markup response by increasing product market concentration. Since markups react by less and exiting firms have, on average, higher labor shares than surviving firms, the aggregate labor share is monotonically increasing in the MW in the economy with no firm exit. In the same vein, unemployment rate increases by significantly less if no firms exit the market, which reflects in higher aggregate value added, as well.

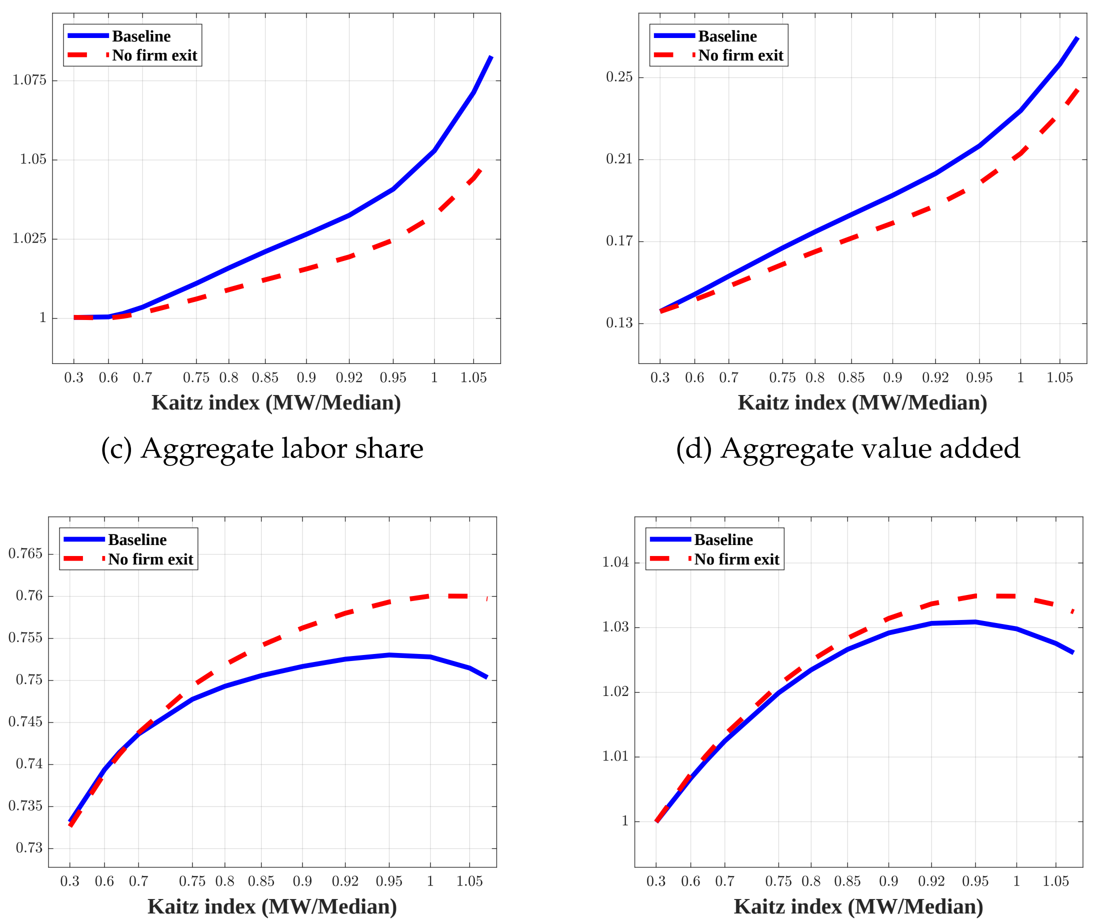
Figure A.16: Aggregate response: Baseline vs no firm exit

(a) Aggregate net markup

(b) Unemployment rate

(c) Aggregate labor share

(d) Aggregate value added



Source: Model. *Note*: The solid blue line represents the response of our baseline economy with break-even condition for exit, while the dashed red line represents the response of our baseline economy with no firm exit. The x-axis is scaled so as to reflect the share of directly affected workers.

E Appendix: Welfare Analysis

Following Floden (2001), utilitarian welfare changes from MW reforms can be decomposed into level, uncertainty, and distributional effects. Let ω_U be the consumption-equivalent welfare change from raising the MW from its baseline level ($\underline{\mathbf{w}}$) to $\underline{\mathbf{w}}'$, that is,

$$\omega_U = \left(\frac{\mathcal{W}(\underline{\mathbf{w}}')}{\mathcal{W}(\underline{\mathbf{w}})}\right)^{\frac{1}{1-\vartheta}} - 1,$$

where $\mathcal{W}$ denotes aggregate welfare, as defined in (33).

Floden (2001) shows that utilitarian welfare admits the following multiplicative decomposition:

$$\omega_U = (1 + \omega_{lev})(1 + \omega_{unc})(1 + \omega_{distr}) - 1. \tag{103}$$

We now derive the three components one at a time. In our context, the *level effect* simply boils down to the aggregate consumption response:

$$\omega_{lev} = \frac{C(\underline{\mathbf{w}}')}{C(\underline{\mathbf{w}})} - 1. \tag{104}$$

Express the value of unemployment and employment at wage w in sequence form as:

$$\begin{aligned} U_{a_i} &= \mathbb{E}_t\left[\mathcal{U}_{a_i}(b_{a_i}, \{c_s\}_{s=t+1}^{\infty})\right] \\ W_{a_i}(w) &= \mathbb{E}_t\left[\mathcal{U}_{a_i}(w, \{c_s\}_{s=t+1}^{\infty})\right], \end{aligned}$$

where $\mathcal{U}$ is the lifetime utility function associated with our CRRA instantaneous utility function. Hence, we can compute the certainty-equivalent consumption, C^e, for each expected consumption stream starting from any labor market state as:

$$\begin{aligned} C^e(U_{a_i}) &= \mathcal{U}_{a_i}^{-1}\left(\mathbb{E}_t\left[\mathcal{U}_{a_i}(b_{a_i}, \{c_s\}_{s=t+1}^{\infty})\right]\right) \\ C^e(W_{a_i}(w)) &= \mathcal{U}_{a_i}^{-1}\left(\mathbb{E}_t\left[\mathcal{U}_{a_i}(w, \{c_s\}_{s=t+1}^{\infty})\right]\right). \end{aligned}$$

Let $C^e(\underline{\mathbf{w}})$ denote the aggregate certainty-equivalent consumption, that is,

$$C^e(\underline{\mathbf{w}}) = \sum_{i=1}^{I} \int \left[u_{a_i}(\underline{w}) C^e\left(U_{a_i}(\underline{w})\right) + (1 - u_{a_i}(\underline{w})) \int C^e\left(W_{a_i}(aw; \underline{w})\right) \, dG_{a_i}(w; \underline{w})\right] d\Omega_i(a)\, d(i).$$

In turn, let C^d be the aggregate consumption-equivalent utility of certainty-equivalent consumption, that is,

$$C^d(\underline{\mathbf{w}}) = \left(\sum_{i=1}^{I} \int \left[u_{a_i}(\underline{w}) C^e\left(U_{a_i}(\underline{w})\right)^{1-\vartheta} + (1 - u_{a_i}(\underline{w})) \int C^e\left(W_{a_i}(aw; \underline{w})\right)^{1-\vartheta} \, dG_{a_i}(w; \underline{w})\right] d\Omega_i(a)\, d(i)\right)^{\frac{1}{1-\vartheta}}.$$

The *uncertainty effect* is the change in the relative aggregate certainty-equivalent consumption with respect to aggregate consumption:

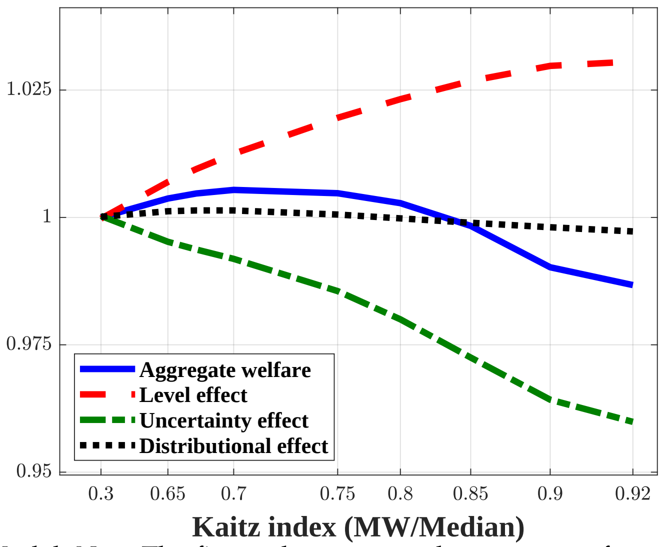
$$\omega_{unc} = \frac{C^e(\underline{\mathbf{w}}')/C(\underline{\mathbf{w}}')}{C^e(\underline{\mathbf{w}})/C(\underline{\mathbf{w}})} - 1. \tag{105}$$

Finally, the *distributional effect* is the change in the aggregate consumption-equivalent utility of certainty-equivalent consumption with respect to certainty-equivalent consumption:

$$\omega_{distr} = \frac{C^d(\underline{\mathbf{w}}')/C^e(\underline{\mathbf{w}}')}{C^d(\underline{\mathbf{w}})/C^e(\underline{\mathbf{w}})} - 1. \tag{106}$$

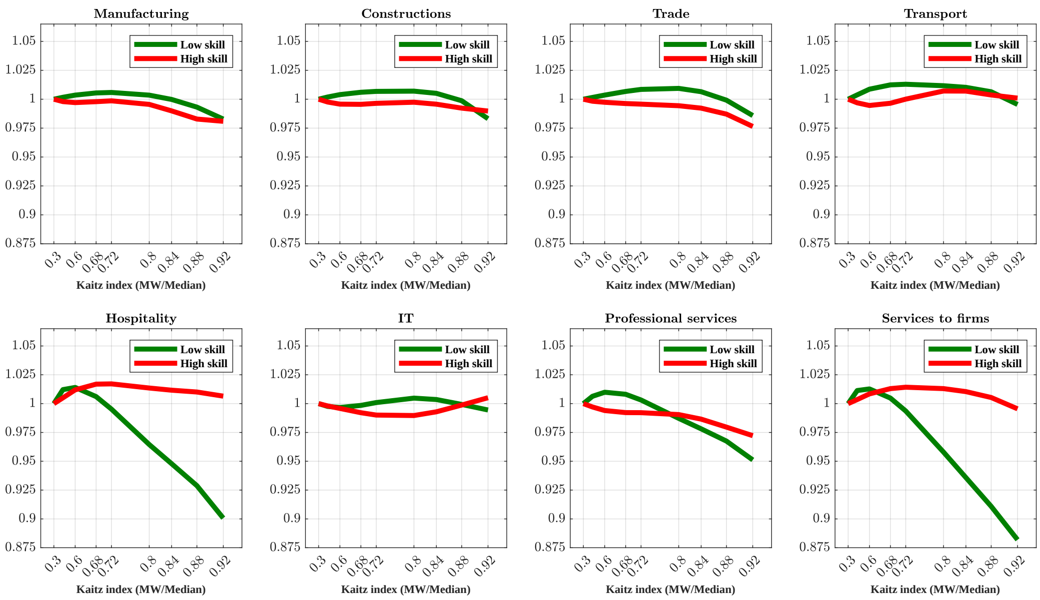
See Dávila and Schaab (2022) for further details on welfare assessments with heterogeneous agents.

Figure A.17: Welfare effects, decomposition



Source: Model. *Note*: The figure decomposes the response of aggregate utilitarian welfare into level, uncertainty, and distributional effects as defined in (104)-(106).

Figure A.18: Breaking down the distributional impact of the MW



Source: Model. *Note*: The graph reports the utilitarian welfare change for workers operating in each industry. The green lines represent low-skilled workers, the red lines high-skilled workers.

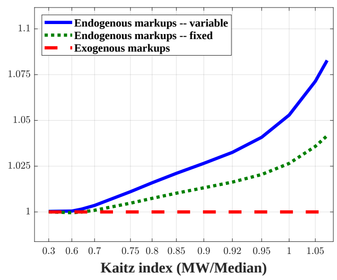
F Appendix: The Role of Endogenous Markups

We analyze the role of endogenous markups in driving the equilibrium effects of the MW in two steps. First, we single out the role of the (behavioral) concentration channel, that is, the firm-level increase in markups induced by the MW. To do so, we repeat the same policy experiments of Section (5) in our baseline economy by constraining markups to their baseline values. Hence, this alternative economy features endogenous markups (cross-sectional differences across firms) that are fixed with respect to the MW.

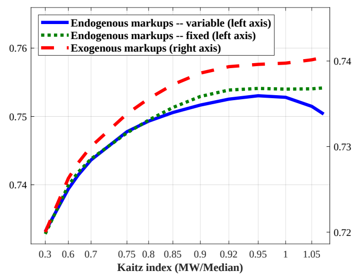
Second, we assess the full role of endogenous markups by factoring in both the firm-level markup response (concentration channel) and the differential passthrough of cost-push shocks across firms with heterogeneous markups (compositional change). With this goal in mind, we repeat the same policy experiments in an observationally equivalent economy – that is, matching the same empirical moments as our baseline – with exogenous (and identical) markups. See below for further details.

Figure A.19: Aggregate response across economies

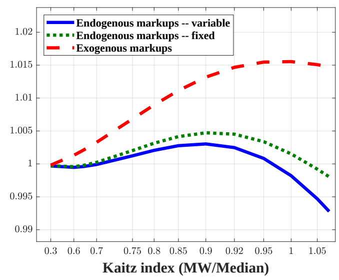
(a) Aggregate net markup



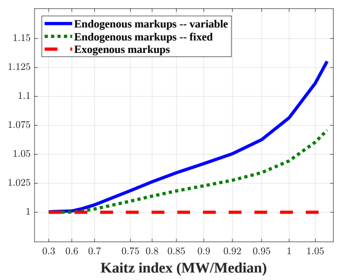
(b) Aggregate labor share



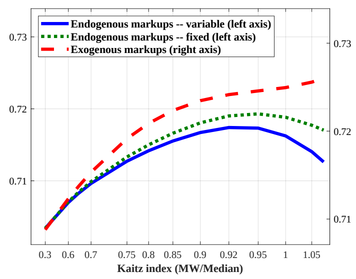
(c) Aggregate output



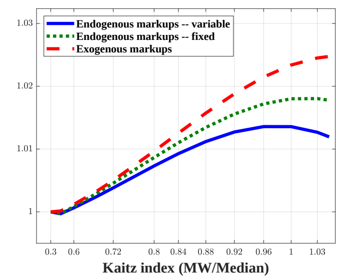
(d) Manufacturing net markup



(e) Manufacturing labor share



(f) Manufacturing output



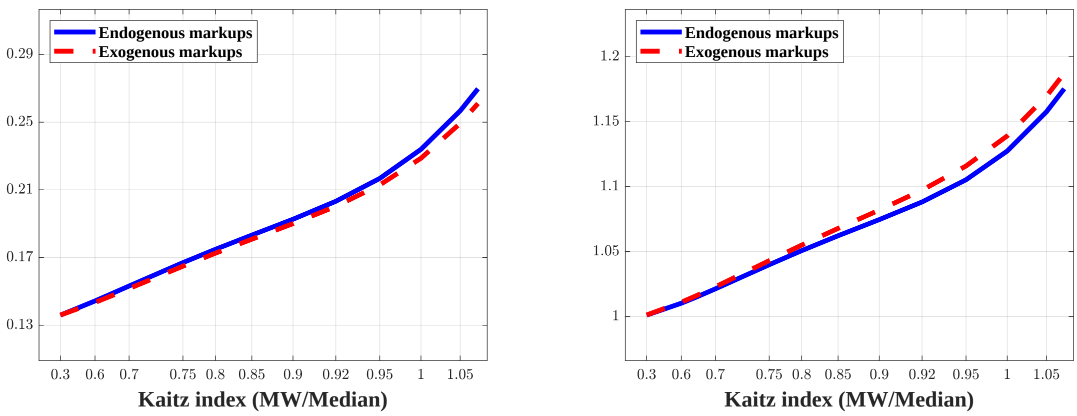
Source: Model. *Note*: The *endogenous markup – variable* economy in solid blue is our baseline economy, the *endogenous markup – fixed* economy in dotted green is our baseline economy where markups are constrained to their baseline values, the *exogenous markup* economy in dashed red features is an observationally equivalent economy to our baseline with exogenous markups. The x-axis is scaled so as to reflect the share of directly affected workers.

Figure (A.19) reports the response of net markup, labor share and output in the aggregate (first row) and in the manufacturing industry (second row) across alternative economies. Our baseline economy with endogenous and variable markups is depicted in solid blue, the alternative economy with endogenous and fixed markups in dotted green, and the alternative economy with exogenous markups in dashed red. From the response of aggregate variables, we draw three main takeaways. First, the aggregate net markup response in our baseline economy is split in almost equal shares into behavioral response (difference between the blue and the green line) and compositional change (difference between the green and the red line). We recover the same shares when decomposing the aggregate markup response in our baseline economy (see Table (A.6)). Second, the aggregate labor share would be monotonically increasing in the MW if markups did not respond. Indeed, muting the concentration channel is enough to prevent the aggregate labor share from declining for Kaitz indices higher than 92%. Third, aggregate output would increase by 1.3pp more at peak if markups were exogenous (by 0.2pp if markups were endogenous but fixed). The difference is accounted for both by a lower unemployment rate and a higher labor productivity in the economy with exogenous markups (see Figure (A.20)).

According to our simulations, a higher baseline HHI makes the elasticity of the sectoral markup (labor share) to the MW increase (decrease). Hence, the concentration channel mainly operates in concentrated sectors. Since Italian nontradable industries are remarkably low concentrated (see Figure 5b)), the aggregate effects of the concentration channel are expected to be driven by the manufacturing industry. The second row of Figure (A.19) confirms this conjecture. The net markup in manufacturing increases by as much as 11.1% in the largest MW reform, with approximately two-thirds of the increase being accounted for by the concentration channel. The labor share in manufacturing increases by less and displays a more pronounced hump shape than in the aggregate. Similarly, manufacturing output would be 40% higher at peak if markups were fixed.

Figure A.20: Employment and productivity effects: Endogenous vs exogenous markups

(a) Unemployment rate (b) Labor productivity



Source: Model. *Note*: The *endogenous markup – variable* economy in solid blue is our baseline economy, the *exogenous markup* economy in dashed red features is an observationally equivalent economy to our baseline with exogenous markups. The x-axis is scaled so as to reflect the share of directly affected workers.

F.1 Details of Alternative Economies

Endogenous Markups – Fixed. In the first alternative economy we aim to single out the effect of the behavioral concentration channel, that is, the firm-level increase in markups induced by the MW. To do so, we study the same MW reforms as in the main text in our baseline estimated model by fixing firm-level markups, i.e., not allowing firms to set their optimal price. Differences between the response of this alternative economy and the baseline can be entirely attributed to the concentration channel.

Exogenous Markups. In the second alternative economy we assume monopolistic competition on the product markets, that is, we set $\rho = \sigma$. This implies that markups are exogenous and constant across all firms. Unlike the previous case, we cannot use our baseline estimated model as benchmark. The reason is that modifying the demand structure changes the model-consistent productivity distribution. Hence, we proceed by re-estimating our baseline model according to the same strategy outline in Section (4). Since we constrain the elasticity of substitution across sectoral goods (ρ) to equal that across firm-specific varieties within sectors (σ), the model would be over-identified by using the same estimation targets. Still, for the counterfactual economy to be as close as possible to our baseline, we identify the hiring elasticity of hiring costs, ϕ, by targeting the linear (and equally weighted) combination of weighted average HHI in 4-digit sectors and the semi-elasticity of the labor share with respect to the market share.

Table (A.7) compares the differences in the estimated parameters across the two economies. As intuitive, the estimated elasticity of substitution is lower than in the baseline model to replicate the same aggregate markup target. Since we target the aggregate markup, the *average* elasticity of demand across firms remain constant, though. While the other parameters are similar across the two economies, the share of right trimming is significantly lower. This reflects the fact that firms with higher markups need to have higher productivity to rationalize the same optimal employment, which magnifies cross-sectional output differences and sectoral concentration. Hence, matching the same degree of concentration with a model with exogenous markups requires a higher share of large firms/top earners.[87]

Table A.7: Internally estimated parameters: benchmark vs monopolistic competition

Parameter	Description	Value (Benchmark)	Value (Monopolistic comp.)
a_H	Skill parameter, high skills	1.320	1.337
ζ	Vacancy convexity hiring costs	0.772	0.796
ϕ	Hiring rate elasticity hiring costs	1.715	1.745
σ	Elast. of subst. within sectors	9.171	8.193
ρ	Elast. of subst. across sectors	3.049	8.193
χ	Meeting efficiency	1.465	1.502
tr	Right trimming wage distributions	0.087	0.050
Θ	SMM loss function	0.017	0.015

[87] We check that, if we fixed the share of right trimming to equal that of the baseline model, the model with exogenous markups would be unable to replicate the estimation targets. Specifically, the SMM loss function would equal 0.140 instead of 0.026, thus making the economy no longer observationally equivalent to the baseline. We argue that the change in the right tail of the wage distribution induced by differential trimming (by less than 4%) introduces less bias than missing the other estimation targets.

Both structural models are successful in replicating the estimation targets accurately. Hence, absent a MW, the economy with exogenous markups is observationally equivalent to our baseline with endogenous markups. It follows that differential responses to the MW across the two economies can be attributed both to the endogenous markup response and the differential passthrough of the MW across the two demand structures.

G Appendix: Empirical Validation

G.1 Institutional background

The collective bargaining system in Italy consists of a large number of contracts negotiated between trade unions and employers' associations (see Boeri (2012) for the economic implications of this system). Beyond regulating other aspects of labor contracts (maximum number of hours of work, number of days off, rules for promotions and training), these agreements set wage floors that are sector and skill-specific, and typically have a duration of 3 years (2 years prior to 2009). Importantly, these contracts have a virtually universal coverage – i.e., their validity extends *erga omnes* – and are generally used by labor courts and labor inspectors as a reference for a "fair wage". As a consequence, non-compliance to contractual wages in Italy is extremely rare (Adamopoulou and Villanueva, 2022). Moreover, bargaining at the firm level is also very unusual, with the exception of a few large firms. Therefore, agreed wages represent a very important component of worker pay in the Italian economy.[88] For a more detailed description of the institutional framework see D'Amuri and Nizzi (2018).

G.2 The Data

CERVED data. We leverage administrative data on balance sheets for the universe of incorporated Italian firms covering the period 2005-2020. This dataset allows us to compute the labor share – defined as the ratio between total labor costs and value added – at the firm level, as well as to observe all typical balance sheet variables (e.g. value added, investment, profits). For each firm-year we match information on the number of employees and on the total wage bill coming from INPS (the Italian Social Security institute), so that we can derive a measure of average wage. Finally, the industry classification is used to match the relevant level of contractual wage for each year.

Contractual wages data. As explained in the previous paragraph, the Italian collective bargaining system is highly centralized. Collective agreements are signed at the nation-wide industry level, with no room for further adjustments at the local level. Contracts envisage wage floors for different worker categories. In this paper, we use data on average contractual wages at the detailed industry level (3-digit) that are made publicly available by Istat for the period 2005-2020 at the monthly frequency, computed as a weighted average of the levels set by the detailed collective contracts within each industry. Wage floors vary over time both because of renewals and the effect of pay rises agreed in the past. In principle, different firms within the same 3-digit industry may apply different labor contracts. Given that we do not

[88] Indeed, D'Amuri and Nizzi (2018) document that in the period 2005-2016 contractual wages defined at national level accounted for about 88% of overall total gross earnings. Moreover, a number of studies demonstrates the important role of collective bargaining for downward wage rigidity (Devicienti et al., 2007), wage inequality (Erikson and Ichino, 1994; Manacorda, 2004; Devicienti et al., 2019; Leonardi et al., 2019) and regional differences in employment (Boeri et al., 2021) in Italy.

have data on contractual pay at the contract level, our strategy involves using the average industry-level contractual wage.[89] In practice, this implies that shocks to labor costs at the firm level are potentially measured with error, possibly leading to a downward bias of our estimates.

G.3 Empirical Results

Reallocation effects. We use our micro-data to measure concentration by computing the revenue-based Herfindahl-Hirschman index (HHI) for each 4-digit sector, and to compute the ratio of the sectoral value added and employment as a proxy of sectoral productivity. We regress these sectoral variables against the industry-specific log wage floor:

$$Y_{s_4,t} = \beta_1 \log MW_{s_3,t} + \phi_t + \alpha_{s_4} + \epsilon_{s_4,t}, \tag{107}$$

where we control for both 4-digit sector fixed effects and time effects. As shown in Table (A.8), the estimates reveal a positive effect on both concentration and average labor productivity, consistent with the reallocation effects. On average, a 1-percent increase in the wage floor causes an increase in concentration of about 0.2pp and a rise in measured labor productivity slightly above 1 percent. We conclude that the empirical evidence supports the hypothesis that rises in wage floors imply a reallocation towards high-productivity firms and a consequent increase in the sectoral market concentration.

Table A.8: Sector-level Response to the MW

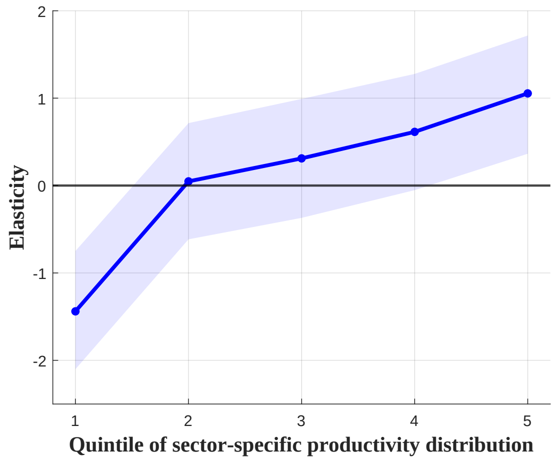
	(1) HHI (4-digit)	(2) Log labor productivity
Log wage floor	0.174	1.328**
	(0.099)	(0.485)
Time FE	Yes	Yes
Sector FE (4-digit)	Yes	Yes
N	7,381	7,381
R^2	0.871	0.893

Clustered errors at 4-digit sector-level in parentheses.
$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Source: CERVED (2005-2020) and Istat data.

[89] The 3-digit level is the most disaggregated level provided by Istat. Instead, Devicienti and Fanfani (2021) match contract-level information on wage floors to the same firm-level data. Their results are very much consistent with ours.

Figure A.21: Heterogeneous Log Value Added Response - By labor productivity



Source: CERVED (2005-2020), INPS and Istat data. *Note*: the figure plots the estimated coefficients of the interaction term between the binned level of sector-specific labor productivity and the natural logarithm of wage floor. The underlying regression is equation (34). Standard errors are clustered at the firm-level. Shaded areas represent 90% confidence intervals.

Concentration effects. We estimate the firm-level labor share response to a rise in the contractual wage floor in a given year allowing for heterogeneous effects by sectoral concentration:

$$LS_{i,t} = \beta_1 \log MW_{s_3(i),t} + \beta_2 \log MW_{s_3(i),t} \cdot HHI_{s_4(i)} + \beta_3 HHI_{s_4(i)} + \gamma_{s_2(i)} \cdot \phi_t + \alpha_i + \epsilon_{i,t}, \tag{108}$$

where $LS_{i,t}$ is the labor share of firm i at time t, defined as the ratio between total labor costs and value added. The coefficients of interest are β_1, which captures the effect of the wage floor in highly competitive markets ($HHI_{s_4(i)} \approx 0$), and β_2, which captures the gradient of the effect with respect to market concentration. Table A.9 reports the regression results.As expected, the average firm-level labor share response to the MW is positive. However, we uncover a significant and negative partial effect of baseline product market concentration on the firm-level labor share response. This implies that the higher product market concentration is, the lower the firm-level labor share response to the MW, lending support to the hypothesis that product market competition is an important determinant of the labor share response.

Table A.9: Firm-level Labor Share Response to the MW

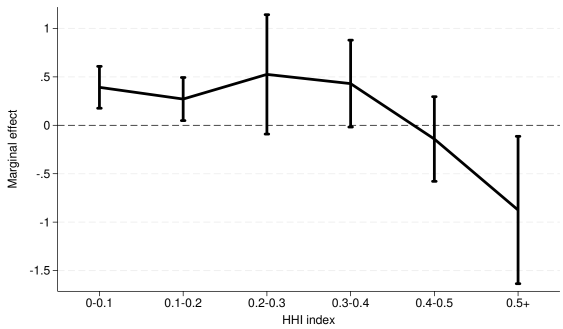
	(1) Labor share	(2) Labor share
Log wage floor	0.526**	0.522***
	(0.177)	(0.143)
Log wage floor × HHI (4-digit)		-1.086*
		(0.509)
Time × Industry (2-digit) FE	Yes	Yes
Firm FE	Yes	Yes
N	6.12e+06	6.12e+06
R^2	0.733	0.734

Standard errors clustered at the firm-level in parentheses.

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

Source: CERVED (2005-2020), INPS and Istat data. *Note*: Linear regressions of labor share, defined as the ratio between total labor costs and value added. Regression models in column (2) also controls for the 4-digit sector-specific HHI.

Figure A.22: Heterogeneous Labor Share Response - By market concentration



Source: CERVED (2005-2020), INPS and Istat data. *Note*: the figure plots the estimated coefficients of the interaction term between the binned level of HHI and the natural logarithm of wage floor. The underlying regression is the counterpart of (108) with categorical HHI bins instead of the continuous variable. Standard errors are clustered at the firm-level. Confidence intervals are at 95%.

G.4 Determinants of the labor share response

Several adjustment mechanisms at the firm level may bring about changes in the measured labor share. The data at our disposal allow us to dig deeper into the nature of these adjustments. Therefore, in Table A.10 we repeat the same regression of Column 2 of Table A.9, for different dependent variables: log average wage, log size, log value added and log profits. Our estimates show that the average effect on wages does not depend on the level of concentration. Instead, firm size and value added respond differently depending on the HHI. In particular, firms in more concentrated sectors grow by less (or even shrink) following the wage increase. The dynamics of the value added qualitatively follows the one of firm size, but it is much less pronounced. As a result, profits in high-concentration sectors *increase* with the wage floor, in stark contrast with what happens in low-concentration sectors. Taken together, these findings show that the differential reaction of the labor share is strictly associated with opposite dynamics of the profit shares at the firm level.

Table A.10: The Effect of Minimum Wages on the Determinants of the Labor Share

	(1) Log avg wage	(2) Log size	(3) Log value added	(4) Log profits
Log wage floor	0.369***	0.952*	0.293	-1.158
	(0.068)	(0.392)	(0.396)	(0.779)
Log wage floor × HHI (4-digit)	-0.030	-1.414	-0.634	1.487
	(0.192)	(1.412)	(0.532)	(1.318)
Time × Industry (2-digit) FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
N	6,155,191	6,158,901	5,774,132	4,915,300
R^2	0.932	0.988	0.978	0.952

Clustered standard errors at the firm-level in parentheses.

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

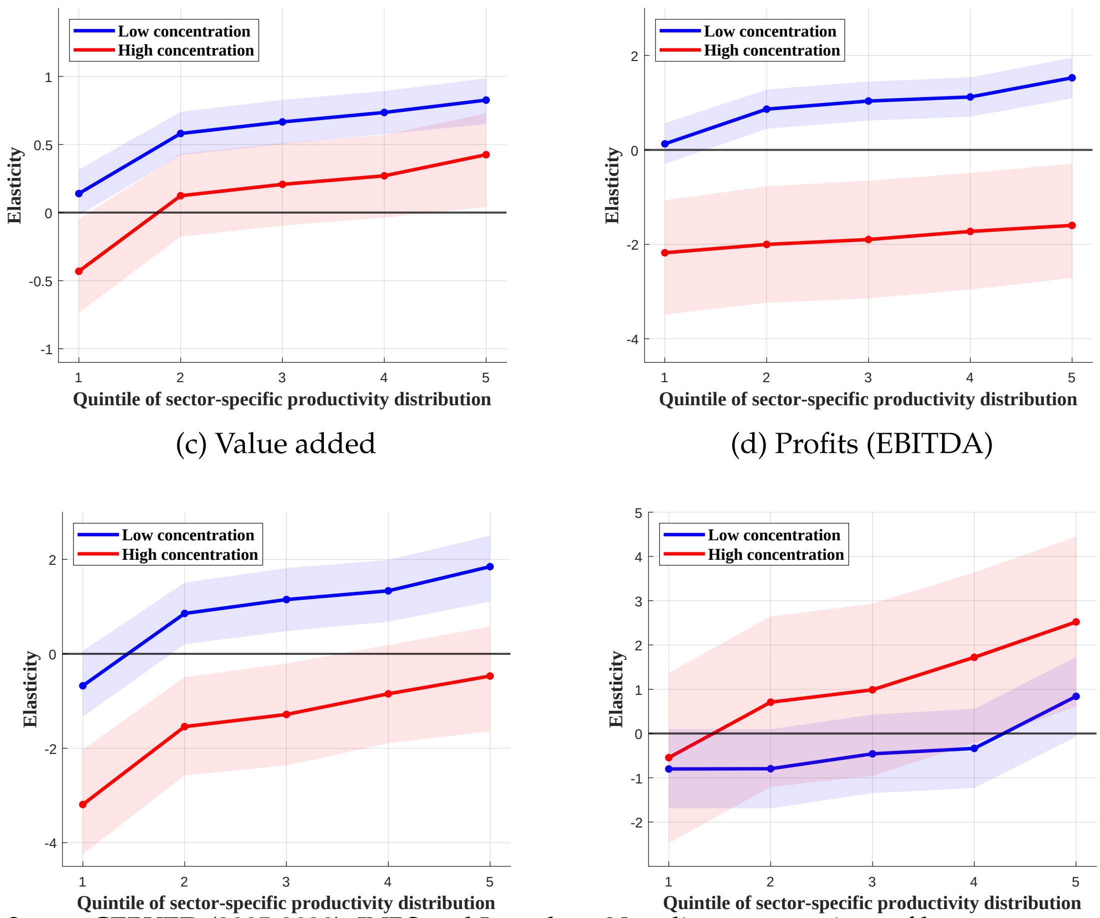
Source: CERVED (2005-2020), INPS and Istat data. *Note*: Linear regressions of log average wage, log firm size, log value added and log profits (EBITDA). Regression models in all columns also control for the industry-specific level of HHI.

We proceed by breaking down the cross-sectional labor share response into its components (Figure A.23). We uncover a positive gradient by firm productivity in the response of both firm size and value added. However, the change in value added is larger than the one of firm size for low-productivity firms, causing a drop in their profits. Instead, more productive firms experience a less than proportional variation in value added relative to the change in firm size. As a consequence, profits tend to *rise* in the upper part of the productivity distribution. If we compare these patterns by the level of concentration, we find that the wage response is relatively similar, whereas large differences arise in the reaction of firm size and value added. In particular, it is especially in high-concentration sectors that the value added response is detached from the response of firm size, determining the largest increase in profits. Note that high productivity firms in highly concentrated sectors experience a large rise in profits in the face of a reduction in value added. This implies that their profit share is unambiguously increasing.

Figure A.23: Heterogeneous Response of Labor Share Determinants

(a) Average wage (b) Employment

(c) Value added (d) Profits (EBITDA)



Source: CERVED (2005-2020), INPS and Istat data. *Note*: linear regressions of log average wage, log firm size, log value added and log profits (EBITDA). All panels plot the estimated coefficients of interaction terms between the quintile of the sector-specific productivity distributions and the natural logarithm of wage floor. Standard errors are clustered at the firm-level. Shaded areas represent 90% confidence intervals.

G.5 Product or Labor Market Concentration?

In Section (8) we have established empirically that sectoral concentration is an important predictor of the labor share response to the MW. According to our theory, sectoral concentration matters for product market competition, in that more concentration implies more market power. Mapping product markets into detailed sectors (4-digit, in our case) is a rather standard practice in the literature (Atkeson and Burstein, 2008; Grassi et al., 2017; Edmond et al., 2023; Burstein et al., 2021). However, one may still worry that the degree of sectoral concentration also affects labor market competition, especially if frictions to workers' mobility across sectors are relevant. To address this concern, we construct proxies of labor market competition and test whether and how adding them to our regression models affects the results. Following the existing literature, we work with two definitions of labor markets: province (administrative geographical units) and province-industry (2-digit).[90] As apparent, both defi-

[90]Two caveats of our definition of labor markets are in order: i) our data does not contain information on occupations, which in principle may be preferable to industries; ii) the geographical location of firms is based

nitions contain an important geographical dimension, consistent with the literature on labor market competition (e.g. Berger et al., 2022; Azar et al., 2019a,b, 2022). We start by replicating the results of Column 2 of Table A.9 adding province-time or province-industry-time effects, which control for labor market shocks. In this way, the effect of the contractual wage floor is estimated by comparing at each point in time only firms competing in the *same* labor market (i.e. using variation within the same province or province-industry cell). As shown in Table A.11, the estimates of our coefficients of interest are remarkably robust to the inclusion of these additional fixed effects (Columns 2 and 3). Next, we augment our regression model by adding an interaction term of the contractual wage floor with the wage-bill-based HHI of the labor market (province or province-industry cells) to test whether directly accounting for labor market concentration matters for our estimates. Once again, the coefficients of interest are almost unaltered by the inclusion of the additional regressors (Columns 4 and 5). Crucially, our estimates align with the existing empirical literature in suggesting that labor market concentration *increases* – rather than decreases – the firm-level labor share response (Azar et al., 2019a; Popp, 2023). Indeed, Figure A.24 shows that the labor share response is higher in *highly* concentrated labor markets than in low labor concentrated markets across the entire productivity distribution. In sum, raising the MW increases the firm-level labor share the more labor markets are concentrated to start with. The opposite holds for product market concentration. We conclude that labor market competition does not represent an important confounder of our results, confirming that product market competition importantly affects the labor share response to the minimum wage.

Table A.11: Robustness Checks: Controlling for labor market competition

	(1) Labor share	(2) Labor share	(3) Labor share	(4) Labor share	(5) Labor share
Log wage floor	0.515***	0.511***	0.406***	0.472***	0.374***
	(0.132)	(0.119)	(0.076)	(0.084)	(0.068)
Log wage floor × HHI revenues (4-digit)	-1.105*	-1.024*	-0.705**	-1.025*	-0.702**
	(0.473)	(0.408)	(0.260)	(0.405)	(0.257)
Log wage floor × HHI wage bill (prov)				3.060	
				(5.895)	
Log wage floor × HHI wage bill (prov × 1-digit)					0.336
					(0.848)
Firm FE	Yes	Yes	Yes	Yes	Yes
Time × Industry (2-digit) FE	Yes	Yes	Yes	Yes	Yes
Time × Province FE	No	Yes	No	Yes	No
Time × Industry (1-digit) × Province FE	No	No	Yes	No	Yes
N	5,613,465	5,613,462	5,612,641	5,540,053	5,539,139
R^2	0.744	0.746	0.753	0.746	0.754

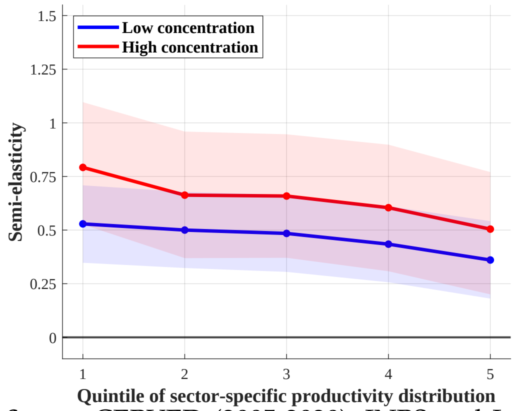
Clustered standard errors at the firm-level in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: CERVED (2005-2020), INPS and Istat data. *Note*: Linear regressions of labor share, defined as the ratio between total labor costs and value added. Regression models also control for the 4-digit sector-specific HHI.

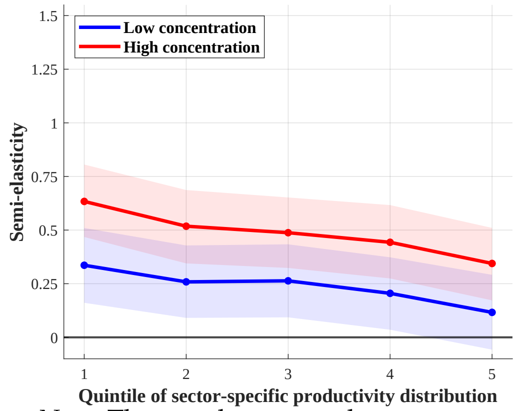
on the headquarters, which introduces some measurement error for multi-establishment firms.

Figure A.24: Firm-level Response by labor market concentration and productivity

(a) Local concentration



(b) Local concentration x 1-digit industry



Source: CERVED (2005-2020), INPS and Istat data. *Note*: The graphs report the average semi-elasticity of firm-level labor share to the MW by productivity bins, computed from the regression model (35). Panel a) uses the HHI of local labor markets (province-level) as a proxy for labor market concentration. Panel b) uses the HHI of local labor markets (province-level) per 1-digit industry as a proxy for labor market concentration. Standard errors are clustered at the firm-level. Confidence intervals are 95%.

Online Appendix

G.6 Dynamic Firm's problem

The sequential firm's profit maximization problem reads:

$$\max_{w_t \geq \underline{w}/a, v_t \geq 0} \int_0^\infty e^{-rt} \, a \left[\left(p_{k(a_i)}(y_t) z - w_t \right) \ell_t - c_{a_i}(w_t, v_t) - \kappa_{a_i} \right] dt \tag{109}$$

$$\text{s.t.} \quad \dot{\ell}_t = -\left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i t}) [1 - F_{a_i t}(w_t)] \right) \ell_t + \frac{v_t}{V_{a_i t}} \lambda(\theta_{a_i t}) [u_{a_i t} + s_{a_i}(1 - u_{a_i t}) G_{a_i t}(w_t)], \tag{110}$$

$$p_{k(a_i)}(y_t) = y_t^{-\frac{1}{\sigma}} \, Y_{k_{a_i} t}(y_t)^{\frac{1}{\sigma} - \frac{1}{\rho}} \, Y_{it}^{\frac{1}{\rho} - 1} \alpha_i Y_t, \tag{111}$$

$$y_t = az\ell_t \tag{112}$$

We proceed by setting up the current value Hamiltonian and dropping time indices:

$$H(w, v; \ell) = \pi(w, v; \ell) + \xi \left(\frac{v}{V_{a_i}} \lambda(\theta_{a_i}) [u_{a_i} + s_{a_i}(1 - u_{a_i}) G_{a_i}(w)] - \left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) [1 - F_{a_i}(w)] \right) \ell \right)$$

where $\pi(w, v; \ell) = \left(p_{k(a_i)} \left(y_{a_i}(\ell) \right) z - w \right) \ell - c_{a_i}(w, v)$ and ξ is the co-state variable.

The first-order conditions are:

$$\frac{\partial H}{\partial w} = 0 \qquad \iff \quad \ell + \frac{\partial c_{a_i}(w, v)}{\partial w} = \xi \left(\frac{v}{V_{a_i}} \lambda(\theta_{a_i}) s_{a_i} (1 - u_{a_i}) g_{a_i}(w) + s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w) \ell \right) \tag{113}$$

$$\frac{\partial H}{\partial v} = 0 \qquad \iff \quad \frac{\partial c_{a_i}(w, v)}{\partial v} = \xi \frac{\lambda(\theta_{a_i}) s_{a_i} (1 - u_{a_i}) g_{a_i}(w)}{V_{a_i}} \tag{114}$$

$$\frac{\partial H}{\partial \ell} = r\xi - \dot{\xi} \qquad \iff \quad \tilde{z}_{k(a_i)} - w - \xi \left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) [1 - F_{a_i}(w)] \right) = r\xi - \dot{\xi} \tag{115}$$

$$\frac{\partial H}{\partial \xi} = \dot{\ell} \qquad \iff \quad \dot{\ell}_t = -\left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) [1 - F_{a_i}(w)] \right) \ell + \frac{v}{V_{a_i}} \lambda(\theta_{a_i}) [u_{a_i} + s_{a_i}(1 - u_{a_i}) G_{a_i}(w)] \tag{116}$$

where $\tilde{z}_{k(a_i)} = \left(1 + \epsilon_{k(a_i)}^{-1}(\ell) \right) p_{k(a_i)} z$ denotes the MRP. In stationary equilibrium, i.e., setting $\dot{\ell} = \dot{\xi} = 0$, the wage and vacancy policy functions read:

$$\tilde{z}_{k(a_i)} - w = \frac{r + \delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) [1 - F_{a_i}(w)]}{2 s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w)} \left(1 - \phi \frac{s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w)}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) [1 - F_{a_i}(w)]} \frac{c_{a_i}(w, v)}{\ell_{a_i}(w, v)} \right), \tag{117}$$

$$\tilde{z}_{k(a_i)} - w = \frac{r + \delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) [1 - F_{a_i}(w)]}{s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w)} (1 + \zeta) \frac{c_{a_i}(w, v)}{\ell_{a_i}(w, v)} \tag{118}$$

Equation (117) boils down to (23) in the *timeless* limit, i.e., as $r \to 0$. On the other hand, equation (118) differs from (24) by the presence of the multiplier $\frac{r + \delta_{a_i} + s_{a_i} \lambda(\theta_{a_i}) [1 - F_{a_i}(w)]}{s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w)}$ to the marginal vacancy cost. Depending on whether such a multiplier is larger (lower) than one, the dynamically-consistent markdown is larger (lower) than in the stationary case reported in the main text. By putting together the two FOCs, we can compute the counterpart of the equilibrium markdown function (31) as follows:

$$\psi_{a_i}(\tilde{z}) = 1 + \frac{r/(s_{a_i}\lambda(\theta_{a_i})) + \chi_{a_i}(\theta_{a_i}) + [1 - F_{a_i}(w_{a_i}(\tilde{z}))]}{2\nu_{a_i}(w_{a_i}(\tilde{z}))f_{a_i}(w_{a_i}(\tilde{z}))w_{a_i}(\tilde{z})}. \tag{119}$$

From this expression, it is apparent that time discounting raises equilibrium markdowns and that the hiring cost correction term is now endogenous and equal to:

$$\nu_{a_i}(w_{a_i}(\tilde{z})) = 1 + \frac{\phi}{2(1+\zeta)} \frac{s_{a_i}\lambda(\theta_{a_i})f_{a_i}(w)}{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})[1 - F_{a_i}(w)]}. \tag{120}$$