The Concentration Channel of the Minimum Wage

Salvatore Lo Bello
Bank of Italy

Lorenzo Pesaresi
Bank of Italy

September 25, 2026

Abstract

We study the equilibrium and welfare 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 the *concentration channel* of the minimum wage. We contribute an equilibrium model with frictional labor markets and oligopolistic product markets. We estimate the model on Italian administrative data, replicating the structure of detailed labor and product markets. We find that both the aggregate labor share and consumption are hump-shaped in the minimum wage, due to the opposing responses of price markups and wage markdowns. The optimal minimum wage directly affects 10% of workers. Endogenous product market power lowers the optimal minimum wage. We provide empirical evidence 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.

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 curbing firms' market power. This shift in focus is reflected in policy proposals advocating for raising the MW to unprecedented levels in many countries. Several studies suggest that raising the MW is an effective policy for countering firms' labor market power. When firms have high labor market power, a higher MW tends to

*This paper was previously circulated with the title "Equilibrium Effects of the Minimum Wage: The Role of Product Market Power". We thank Nir Jaimovich, Alessandro Ferrari, and Gianluca Violante for their guidance. We also thank Corina Boar, Moritz Drechsel-Grau, Jan Eeckhout, Niklas Engbom, David Hémous, Erik Hurst, Kyle Herkenhoff, Moritz Kuhn, Andrea Linarello, Guido Menzio, Claudio Michelacci, Virgiliu Midrigan, Simon Mongey, Chris Moser, Lukas Nord, Thomas Sargent, Florian Scheuer, and Josef Zweimüller for their feedback, and our discussants Manolis Galenianos, Gabriele Macci, and Aseem Patel. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Bank of Italy. The authors have no relevant financial relationships or other potential conflicts of interest to disclose.

reallocate workers to more productive firms, with small effects on unemployment (Dustmann et al., 2022; Azar et al., 2024). However, little is known about how the effects of the MW interact with product market power – that is, the ability of dominant firms to charge prices above marginal costs.[1] This gap is surprising, as strong product market power may undermine the MW's beneficial effects in terms of both efficiency—by limiting reallocation—and redistribution—by disproportionately benefiting low-labor-share firms.

In this paper we study the equilibrium and welfare 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. We depart from the existing MW literature (Engbom and Moser, 2022; Berger et al., 2025) 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 reduces the 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 the *concentration channel* of the MW.[2] We show that the concentration channel affects the qualitative response of the aggregate labor share and output to the MW.

To quantify the effects of the concentration channel, we build a novel quantitative model where heterogeneous workers and firms interact in frictional labor markets (Engbom and Moser, 2022) and oligopolistic product markets (Atkeson and Burstein, 2008). 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., sock-manufacturing 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. 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.

[1] Two exceptions are Harasztosi and Lindner (2019) and Link (2024), which document that firms 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 (e.g., see https://www.aboutamazon.com/news/policy-news-views/its-time-to-raise-the-federal-minimum-wage for Amazon's official position).

We estimate the model using administrative data from the Italian National Social Security Institute and the Italian Institute of Statistics. We estimate a two-way fixed-effects model à la Abowd et al. (1999) and classify workers into two skill groups, high- and low-skilled, based on whether their worker fixed effect is above or below the median. Labor markets are defined by worker skill group and the 1-digit industry code of the employer firm. Product markets are defined by 4-digit ATECO industry codes, which are similar in granularity to 5-digit NAICS industries in the US. For locally traded goods, product markets are further segmented by region. We discipline the vast cross-sectional heterogeneity in market structures by replicating the observed wage distributions and worker transitions in each labor market, as well as the detailed distribution of the number of competing firms across product markets within each 1-digit industry. We estimate the key parameters governing the elasticities of demand and labor supply via the simulated method of moments (SMM), targeting average product-market concentration and the average gradient of firms' labor shares with respect to their market shares. 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 assessing the actual impact of MW reforms and the scope for worker reallocation.

We use the estimated model to simulate the effects of MW reforms on the Italian economy. According to our results, aggregate labor market power decreases with the MW (up to 6.7 percentage points), driven by low-productivity firms compressing their wage markdowns. On the contrary, aggregate product market power increases with the MW (up to 2.2 percentage points), driven by the concentration channel. The rise in aggregate product market power offset, on average, one-third of the reduction in labor market power, and is driven by industries with high baseline concentration.[3] For sufficiently high MWs, the product market power response leads to an increase in *aggregate* market power with the MW. Mirroring the aggregate market power response, both the aggregate labor share and consumption are hump-shaped in the level of the MW, with a peak at 35% of affected workers. The increase in aggregate consumption is driven by productivity gains from worker reallocation (due to lower misallocation). However, the reallocation process produces some excess unemployment. It follows that the MW has the potential to raise aggregate consumption, 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. The optimal MW directly affects 10% of workers, with associated welfare gains of 0.8% in consumption equivalent units. The welfare effects of the MW are heterogeneous across worker skill types, which generates winners and losers. The welfare of low-skilled workers increases with the MW up to 7% of affected workers. As the MW increases further, the high-skilled workers are the main beneficiaries. Hence, our results offer a word of caution about the idea of raising the MW to unprecedented levels to reduce

[3]Indeed, model counterfactuals varying baseline product market concentration show that the concentration channel operates more strongly in more concentrated economies, such as the US.

(welfare) inequality.

We validate the quantitative performance of our model in two ways. First, we show that the model-implied effects of the MW on employment, prices, and reallocation are consistent with external empirical evidence from different countries. Second, we simulate industry-specific MWs in our estimated model and compare the model predictions to new empirical estimates based on fifteen years of changes in industry-specific contractual wage floors in Italy. Using firms' balance sheet data, we document a close fit between the model and the empirical responses of employment, labor shares, and profits across the firm productivity distribution.

Finally, we examine the role of endogenous markups in driving our results through two complementary approaches, one structural and one reduced-form. In the structural approach, we repeat the same counterfactual experiments in an observationally equivalent economy that retains the same demand structure but features exogenous markups. This exercise shows that ignoring endogenous markups would overstate the benefits of the MW: aggregate labor share and consumption would increase monotonically with the MW until half of the workforce is affected, and the optimal MW would increase from affecting 10% to 15% of workers, yielding 40% larger welfare gains. We therefore conclude that endogenous product market power is an important determinant of the equilibrium and welfare effects of the MW.

In the reduced-form approach, we provide empirical validation of the concentration channel using the same Italian industry-specific contractual wage floors, together with variation in concentration across product markets. We show that the firm-level labor share response decreases with concentration in the firm's product market, turning negative for high-productivity firms. As predicted by the concentration channel, this pattern is driven by the positive profit response of high-productivity firms operating in highly concentrated product markets.

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. Drechsel-Grau (2023) and Hurst et al. (2025) 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. (2025). Engbom and Moser (2022) use a quantitative wage-posting model to study the effect of the MW on earnings inequality. Our model extends the framework of Engbom and Moser (2022) by considering a generalized hiring cost function and adding product market power, as well as characterizing the optimal MW. Berger et al. (2025) study 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 the framework of Berger et al. (2025), ours 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; Edmond et al., 2023; Burstein et al., 2025). The closest papers in this literature to ours are MacKenzie (2021), Deb et al. (2022, 2023), Trottner (2023), Firooz (2025) and Wong (2026), which propose models of oligopolistic competition in sectoral product markets and oligopsonistic competition (or wage bargaining) in local labor markets. 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 were laid by Burdett and Mortensen (1998), van den Berg and Ridder (1998), Bontemps et al. (1999) and Bontemps et al. (2000). Recent advances related to our framework include Bilal and Lhuillier (2021), Gouin-Bonenfant (2022), Engbom and Moser (2022), and Gottfries and Jarosch (2023). We contribute to this literature by characterizing the wage-posting equilibrium with decreasing returns to scale in revenues, size-dependent hiring costs, and endogenous job creation.

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 direction of the 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 with discrete support $\mathcal{Z}$. Both the labor and the product market are imperfectly competitive. Firms choose their labor demand by solving the following profit maximization problem:

$$\max_{\ell} \; p(y)y - w(\ell)\ell - \kappa \;\; \text{s.t.} \;\; y = z\ell, \quad w(\ell) = \ell^{\frac{1}{\eta}}, \quad p(y) = y^{-\frac{1}{\epsilon(N)}}.$$

Firms seek to maximize current profits, which equal the difference between revenues, py, and costs. In turn, costs consist of the wage bill, $w\ell$, and some overhead costs, κ.[4] Firms operate a linear technology in labor. Labor market power stems from an upward-sloping labor supply curve with elasticity of labor supply $\eta \in (0, \infty)$. Since wages are increasing in employment, the marginal cost exceeds the wage. Formally, $MCL(\ell) = w(\ell) + w'(\ell)\ell$. Similarly, product

[4] Firm-level output, wages, and prices are expressed relative to fixed reference levels normalized to 1 and held constant across equilibria. 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).

market power stems from a downward-sloping demand curve 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., $N' > N \implies \epsilon(N') > \epsilon(N)$ (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. 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 consists of a double wedge between the optimal price and (productivity-adjusted) wage:

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

Equation (1) 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., $\mu(N) \equiv \frac{p(z)z}{MCL(z)} = \frac{\epsilon(N)}{\epsilon(N)-1} \geq 1$. 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.[5] Henceforth, we will refer to the price markup and wage markdown simply as markup and markdown, respectively. The optimality condition 1, along with the labor supply and product demand constraints, also characterizes the firm-specific labor share in operating revenues, $LS(z)$, and output, $y(z)$:

$$LS(z) \equiv \frac{w(z)\ell(z)}{p(z)z\ell(z)} = \frac{1}{\mu(N)\psi}, \quad y(z) = z^{\epsilon\frac{1+\eta}{\epsilon+\eta}} \left(\frac{1}{\mu(N)\psi}\right)^{\frac{\eta\epsilon}{\epsilon+\eta}}.$$

Notice that both the labor share and output are decreasing in the firm's market power, that is, the product between markup and markdown. The labor share is constant across the productivity distribution, while output increases with productivity. We restrict attention to parameter vectors such that firm-level employment is increasing in the elasticity of demand.[6]

We conclude the characterization of the baseline equilibrium by defining two key aggregate variables, that is, the aggregate labor share in the operating revenues, LS, and the aggregate output, Y:

$$LS \equiv \frac{\sum_{z\in\mathcal{Z}} w(z)\ell(z)\hat{\gamma}(z)}{\sum_{z\in\mathcal{Z}} p(z)y(z)\hat{\gamma}(z)} = \frac{1}{\mu(N)\psi}, \quad Y \equiv N\sum_{z\in\mathcal{Z}} y(z)\hat{\gamma}(z) = NZ^{\epsilon\frac{1+\eta}{\epsilon+\eta}} \left(\frac{1}{\mu(N)\psi}\right)^{\frac{\eta\epsilon}{\epsilon+\eta}},$$

[5] The existing literature uses the term *markdown* either to denote the ratio between MRPL and wage – e.g., Gottfries and Jarosch (2023) – or the ratio between wage and MRPL – e.g., Berger et al. (2025). For ease of interpretation, we adopt the former formulation that grants a direct link between the notions of markdown and labor market power.

[6] We report the formal parametric restriction in Lemma A.1 in Appendix A.

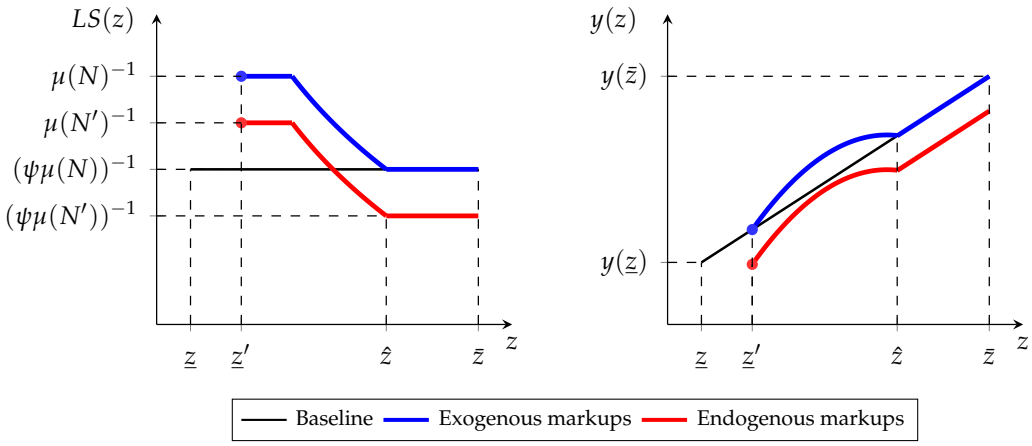
where $Z \equiv \left(\sum_{z\in\mathcal{Z}} z^{\epsilon\frac{1+\eta}{\epsilon+\eta}}\hat{\gamma}(z)\right)^{\frac{\epsilon+\eta}{\epsilon(1+\eta)}}$ is an aggregate productivity index. Hence, aggregate market power acts as a uniform tax on aggregate output.

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 $w(\ell) = \max\left\{\underline{w}, \ell^{\frac{1}{\eta}}\right\}$. Thus, the labor supply curve is piecewise: it is perfectly flat for low enough employment and increasing thereafter. This means that firms with low employment are effectively wage takers, as if they operated in perfect competition in the labor market.

In response to the MW, some low-productivity firms exit the market, as they can no longer break even. 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., $\mu(N') = \mu(N)$, as in the existing literature (Clemens, 2021). Figure 1 displays in blue the adjustment of the firm-level labor share and output to the MW across the productivity distribution when markups are exogenous.[7] In this case, the effects of the MW are shaped by the labor market power response only. A binding MW pushes constrained firms to pay higher wages by compressing

Figure 1: Cross-sectional impact of the MW



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). 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.

[7] For simplicity, in Figure 1 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.

Second, we factor in the endogenous markup response – our original mechanism. Figure 1 displays in red the adjustment to the MW when markups are endogenous. In this case, the effects of the MW are shaped by the direction of the 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 more rigid demand curve, i.e., $\epsilon(N') < \epsilon(N)$. Hence, they find it optimal to exercise their higher product market power by raising their 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 the *concentration channel* of the MW.[8] 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 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, the responses of both aggregate labor share and output are qualitatively ambiguous and crucially depend on the direction of the market power response – that is, whether the reduction in the aggregate markdown outweighs the increase in the aggregate markup, or vice versa. The output response additionally reflects firm exit and the direct effect of the change in demand elasticity on surviving firms' output.

3 Quantitative Model

In this section, we develop a novel quantitative model to assess the equilibrium and welfare effects of the MW when product market power is endogenous and varies with market competition. The model combines a wage-posting framework that generalizes Engbom and Moser (2022) with a Bayes-Cournot pricing framework that generalizes Atkeson and Burstein (2008). We consider segmented labor markets by worker skill and industry, characterized by search-and-matching frictions and on-the-job search. The source of labor market power is matching frictions, which capture geographical distance and incomplete information.[9] We consider a continuum of product markets, in which a finite number of firms engage in

[8] 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 magnify the concentration channel. However, Section 8 provides evidence that the concentration channel occurs primarily through the product market. Our concentration channel is related to the insight of Haucap et al. (2001) that collective wage setting may reduce product market competition if large firms are over-represented in employers' associations.

[9] Endogenous markdowns can stem from firms' granularity in their labor markets, search frictions, or both. According to Berger et al. (2023), search frictions are the quantitatively most relevant determinant of markdowns. Unlike in frictionless models such as Berger et al. (2025), 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.

oligopolistic competition. The source of product market power is the imperfect substitutability across firms' varieties and the granularity of firms in their product market.

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 into $I \in \mathbb{N}$ industries. For each industry i, we consider a continuum of labor markets of unit measure, indexed by a. Within each labor market, we consider a continuum of product markets, or *sectors*, indexed by $k(a_i)$. Let the final (consumption) good be the numeraire.

Workers. Workers are hand-to-mouth agents with preferences exhibiting constant relative risk aversion (CRRA) over a consumption good. Workers differ in their permanent skill type a and industry i.[10] Let L_i denote the share of workers in industry i and $\Xi_i(a)$ denote the industry-specific skill distribution. Employed workers own the firms and receive rebates of aggregate profits in proportion to their relative wage. In addition, they provide labor services for firms' overhead costs, for which they are remunerated at the economy-wide average wage $\bar{W}$.

Final good producers. Final good producers operate a Cobb-Douglas (CD) production function that aggregates up industry goods into a homogeneous consumption good. In turn, each industry good is a double-nested aggregator with constant elasticity of substitution (CES) over 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.[11] All firms operate a linear technology in labor. Following Engbom and Moser (2022), labor productivity is a composite of worker skill and firm physical productivity, combining multiplicatively. 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, each offering some

[10] 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.

[11] Formally, the effective productivity of a firm with physical productivity z and skill requirement $\hat{a}$ hiring workers of skill a is given by $z_{\hat{a}} = z\ \mathbb{1}\{a = \hat{a}\}$.

wage piece rate w per skill unit.[12] As in Burdett and Mortensen (1998), offered wage piece rates follow 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 cumulative distribution function of employed workers who earn piece rates less than or equal to 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 (i.e., 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 labor-market-specific unemployment rate. The total number of meetings occurring at any point in time is governed by a constant-returns-to-scale (CRS) meeting 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 $\lambda_{a_i} = \lambda(\theta_{a_i}) = \mathcal{M}(1, \theta_{a_i})$. To ensure finite meeting rates, we specialize to a CES meeting function $\mathcal{M}(S_{a_i}, V_{a_i}) = \chi \left[\beta V_{a_i}^{-\iota} + (1-\beta) S_{a_i}^{-\iota}\right]^{-\frac{1}{\iota}}$ with substitution parameter ι, vacancy share β, and meeting efficiency χ.[13]

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 $\hat{a}$. Firms operating in the same product market source workers from one labor market but sell their good to all consumers (via final good producers). In equilibrium, a number $N_{k(a_i)} \leq \bar{N}_{k(a_i)}$ of potential firms operate. Let $\Gamma_{k(a_i)}(z)$ denote the productivity distribution of potential firms. To ensure equilibrium uniqueness, we assume that all the product markets sourcing from the same labor market are ex ante identical. Formally, they share the same probability mass function for the number of potential firms, $\bar{N}_{k(a_i)} \sim \mathcal{N}_{a_i}$, and the same productivity distribution of potential firms, $\Gamma_{k(a_i)}(z) = \Gamma_{a_i}(z)$. Given that $\bar{N}_{k(a_i)} < \infty$, firms are granular in their product market, i.e., they account for a positive share of sectoral production, and engage in oligopolistic competition à la Bayes-Cournot.

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

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

[13] Finite meeting rates are a necessary condition for our equilibrium uniqueness proof.

Hamilton-Jacobi-Bellman (HJB) equations:

$$rU_{a_i} = \mathscr{U}(ab_{a_i}) + \lambda(\theta_{a_i}) \int \{W_{a_i}(w) - U_{a_i}\}^+ dF_{a_i}(w), \tag{2}$$

$$rW_{a_i}(w) = \mathscr{U}(e_a(w)) + s_{a_i}\lambda(\theta_{a_i}) \int \{W_{a_i}(w') - W_{a_i}(w)\}^+ dF_{a_i}(w') + \delta_{a_i}(U_{a_i} - W_{a_i}(w)), \tag{3}$$

where the flow utility is $\mathscr{U}(x) = x^{1-\vartheta}/(1-\vartheta)$ for $\vartheta \neq 1$ and $\mathscr{U}(x) = \log x$ for $\vartheta = 1$, $\vartheta > 0$ is the CRRA coefficient, r is the instantaneous discount rate, $e_a(w) = aw + \Pi(aw) + \mathcal{O} + T$ 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 her share of aggregate profits $\Pi(aw)$, payments for overhead services $\mathcal{O}$, and transfers T. 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 $\tau \in [0,1]$. As in Kaplan et al. (2018), employed workers are further rebated a share $\mathcal{S}(G(aw))$ of after-tax aggregate profits proportional to their relative wage, where $G(aw)$ is the economy-wide wage distribution. Formally, $\Pi(aw) = \mathcal{S}(G(aw))(1-\tau)\bar{\Pi}$, where the employment-weighted integral of $\mathcal{S}(G(aw))$ is normalized to 1 and $\bar{\Pi}$ denotes aggregate profits. 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:

$$\max_{\{y_{jk(a_i)}\}} PY - \sum_{i=1}^{I} \int_0^1 \int_0^{K_{a_i}} \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)} \, dk(a_i) \, da_i, \tag{4}$$

$$\text{s.t.} \quad Y = \prod_{i=1}^{I} Y_i^{\alpha_i}, \tag{5}$$

$$Y_i = K_i^{-\frac{1}{\rho-1}} \left(\int_0^1 \int_0^{K_{a_i}} Y_{k(a_i)}^{\frac{\rho-1}{\rho}} \, dk(a_i) \, da_i \right)^{\frac{\rho}{\rho-1}}, \tag{6}$$

$$Y_{k(a_i)} = \left(\sum_{j=1}^{N_{k(a_i)}} y_{jk(a_i)}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}. \tag{7}$$

The final good Y is a CD aggregator of industry goods Y_i, with weights $\alpha_i \in (0,1)$, $\sum_{i=1}^{I} \alpha_i = 1$. Each industry good Y_i is a CES aggregator of sector-specific goods $Y_{k(a_i)}$, with elasticity of substitution $\rho > 1$. For each industry i and skill requirement a, there is a measure K_{a_i} of sectors

(product markets) indexed by $k(a_i)$. Let $K_i \equiv \int K_{a_i}\, da_i$ be the total measure of sectors within each industry i. The normalization by K_i in front of the aggregator (6) removes love-of-variety effects associated with differences in the measure of sectors across industries. Finally, each sector-specific good $Y_{k(a_i)}$ is a CES aggregator of firm-level varieties $y_{jk(a_i)}$, produced by a finite number $N_{k(a_i)}$ of competing firms, with elasticity of substitution $\sigma > \rho$.

3.4 Firm's problem

Firms choose their wage and size policies under uncertainty about the number and productivity type of competitors within their product market. This information structure ensures that, within a given labor market, the wage and size policies are *functions* – rather than correspondences – of the firm's productivity.

For a given distribution of the number of potential firms across product markets $\mathcal{N}_{a_i}$, a randomly selected potential firm forms beliefs about the total number of potential firms in its product market according to the size-biased distribution $\mathcal{N}^F_{a_i}(n) = \frac{n\mathcal{N}_{a_i}(n)}{\sum_m m\mathcal{N}_{a_i}(m)}$. Intuitively, a market with n potential firms is n times as likely to contain the randomly selected firm as an otherwise identical market with a single potential firm. Larger markets therefore receive proportionally more weight from the firm's perspective than when product markets are sampled equally. Conditional on a realized number of potential firms $\bar{N}_{a_i} = n$, the firm faces $n-1$ potential rivals whose productivities are independently drawn from the productivity distribution Γ_{a_i}. Given an output policy $y_{a_i}(z)$ and activity cutoff $\underline{z}_{a_i}$, define the realized competitors' CES-output contribution as $B_{a_i} = \sum_{m=1}^{\bar{N}_{a_i}-1} \mathbb{1}\{z_m \geq \underline{z}_{a_i}\} y_{a_i}(z_m)^{\frac{\sigma-1}{\sigma}} \sim \mathcal{B}_{a_i}$.

Following the equilibrium search literature, we reduce the firm's dynamic problem to a static profit maximization problem by assuming steady-state behavior (Burdett and Mortensen, 1998).[14] A firm with productivity z in industry i with skill requirement a solves the following profit maximization problem:

$$\pi_{a_i}(z) = \max_{\substack{w \geq \underline{w}_{a_i}, \\ v \geq 0}} \mathbb{E}_{B_{a_i}}\left[p_{k(a_i)}(y; B_{a_i}) y_{a_i}(w, v; z)\right] - aw\ell_{a_i}(w, v) - ac_{a_i}(w, v) - a\kappa_{a_i}\bar{W}\mathbb{1}_{\{v>0\}} \quad (8)$$

$$\text{s.t. } y_{a_i}(w, v; z) = az\ell_{a_i}(w, v), \quad (9)$$

$$\ell_{a_i}(w, v) = \frac{v}{V_{a_i}} \frac{\lambda(\theta_{a_i})[u_{a_i} + s_{a_i}(1-u_{a_i})G_{a_i}(w)]}{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})[1 - F_{a_i}(w)]}, \quad (10)$$

$$p_{k(a_i)}(y; B_{a_i}) = \left(\frac{y}{Y_{k(a_i)}(y; B_{a_i})}\right)^{-\frac{1}{\sigma}} \left(\frac{Y_{k(a_i)}(y; B_{a_i})}{Y_i / K_i}\right)^{-\frac{1}{\rho}} \left(\frac{Y_i}{\alpha_i Y}\right)^{-1} P, \quad (11)$$

$$c_{a_i}(w, v) = \bar{c}_{a_i} h_{a_i}(w)^{\phi} \frac{v^{1+\zeta}}{1+\zeta}, \quad (12)$$

[14] In Appendix B.9 we develop the dynamically consistent version of the firm's problem and show that the two formulations yield the same results in the *timeless* limit, i.e., when $r \to 0$.

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 and $B_{a_i} \sim \mathcal{B}_{a_i}$ is the CES-output contribution of the firm's competitors within its product market. Conditional on the firm producing y and on a B_{a_i} realization, sectoral output equals $Y_{k(a_i)}(y; B_{a_i}) = (y^{(\sigma-1)/\sigma} + B_{a_i})^{\sigma/(\sigma-1)}$. Hence, firms choose their size by solving a Bayes–Cournot game in quantities before observing the realized number and productivity of their competitors, anticipating that their price will adjust to clear the product market contingent on the realized set of competitors.

Firms maximize expected stationary profits by leveraging two control variables, namely a wage piece rate w – subject to the MW constraint – and a mass of vacancies v. Expected profits equal expected revenues net of labor costs, $aw\ell_{a_i}(w, v)$, hiring costs, $ac_{a_i}(w, v)$, and overhead costs, $a\kappa_{a_i}\bar{W}$. The latter are paid only if the firm is active, i.e., it posts a positive mass of vacancies.[15] 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, but differs from it along three key dimensions. First, following Engbom and Moser (2022), all the labor-market variables scale with worker skill a. This is the simplest way to generate wage dispersion due to worker and firm heterogeneity separately. 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 (12). Third, and most importantly, the elasticities of labor supply and product demand are no longer constant parameters.

The labor supply curve (10) 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 the 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 to unemployment or poaching from higher-paying firms), times the stationary employment, $\ell_{a_i}(w, v)$. Crucially, the firm can control its stationary employment not only by posting vacancies, but also through the wage policy: the higher the posted wage, the higher the vacancy-filling rate and the lower the quit rate, that is, the higher the stationary employment.

The product demand curve (11) follows from 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 $\frac{1}{\sigma} - \frac{1}{\rho}$. See Appendix B.2 for details on the derivation of the labor supply and product demand curves.

The hiring cost (piece rate) function (12) 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

[15] Since activity decisions are made under uncertainty, some firms may end up making negative *realized* profits.

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). Vacancy posting costs are convex for a wage posting equilibrium with endogenous job creation to be well-defined.

Limit cases. The firm's problem (8) 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.[16] On the other hand, the model nests the price-setting model of Atkeson and Burstein (2008) when the labor market is frictionless, overhead costs are absent, and the distribution of competitors' CES-output contribution, $\mathcal{B}_{a_i}$, is degenerate. Formally, if $\bar{c}_{a_i} = 0$ (or, equivalently in this limit, if the meeting function is infinitely efficient, i.e., $\chi \to \infty$), $\kappa_{a_i} = 0$, and $B_{a_i} = \bar{B}_{a_i}$ almost surely, firms face no search frictions and know their competitors' aggregate contribution when choosing output. Their problem therefore reduces to the standard Cournot game with nested CES demand.

3.5 Equilibrium characterization

We now define and characterize the rank-preserving stationary equilibrium of our economy. An equilibrium is rank preserving if, within each labor market, wages are monotonically increasing in productivity.

Each labor market a_i supplies a continuum of ex ante symmetric product markets, while each product market contains a finite number of competitors. The key tractability assumption is that each firm's wage and size policies are chosen under uncertainty about the number and productivity types of its product-market competitors. Product markets sourcing from the same labor market are therefore identical in primitives *and* in firms' wage and size policy functions, irrespective of their realized product-market structure. [17]

Definition (Stationary equilibrium)
For given primitives and probability mass function $\mathcal{N}_{a_i}$ for the number of potential firms $\bar{N}_{a_i}$ 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 (2) and employment (3);*
- *Output demand functions $y^d_{jk(a_i)}(p_{jk(a_i)}), \forall j, \forall a_i$, that solve the final good producers' problem (4), with the final good used as numeraire ($P = 1$);*

[16] 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).

[17] If firms chose wages as a function of the realized product-market structure, the firm-specific rank in the equilibrium wage distribution would depend on both the exogenous productivity rank *and* the endogenous competitive environment. As a result, no rank-preserving equilibrium would generally exist.

- *Wage, vacancy, employment, and output policies* $\{w_{a_i}(z), v_{a_i}(z), \ell_{a_i}(z), y_{a_i}(z)\}$, $\forall z, \forall a_i$, *that solve the firms' problem (8) for given wage offer distribution* $F_{a_i}(w)$, *wage employment distribution* $G_{a_i}(w)$, *labor market tightness* θ_{a_i}, *distribution of competitors' CES-output contribution* $\mathcal{B}_{a_i}$, *and aggregate demand shifters;*

- *Activity cutoffs* $\{\underline{z}_{a_i}\}_{a_i}$ *that determine the marginally active firm types in each labor market* a_i. *Firms with* $z < \underline{z}_{a_i}$ *post no vacancies and produce zero output, while firms with* $z \geq \underline{z}_{a_i}$ *operate according to their equilibrium policies;*

- *A set of competitors' CES-output distributions* $\{\mathcal{B}_{a_i}\}_{a_i}$, *induced by the size-biased distributions of the number of competitors* $\mathcal{N}^F_{a_i}$, *the productivity distributions* Γ_{a_i}, *the activity cutoffs, and the equilibrium output policies;*

- *Measures* $\{F_{a_i}(w), G_{a_i}(w), u_{a_i}, V_{a_i}\}$ *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 the demand for the variety supplied by each active firm equals its output:* $y^d_{k(a_i)}\left(p_{k(a_i)}(z; B_{a_i}); B_{a_i}\right) = y_{a_i}(z) = az\ell_{a_i}(z) \quad \forall z \geq \underline{z}_{a_i}, \forall B_{a_i}, \forall a_i$;

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

$$C = Y - \sum_{i=1}^{I} L_i \int M_{a_i} \int_{\underline{z}_{a_i}}^{\infty} ac_{a_i}(z)\, d\Gamma_{a_i}(z)\, d\Xi_i(a),$$

where

$$C = \sum_{i=1}^{I} L_i \int M_{a_i} \int_{\underline{z}_{a_i}}^{\infty} \left(aw_{a_i}(z)\ell_{a_i}(z) + \pi_{a_i}(z) + a\kappa_{a_i}\bar{W}\right) d\Gamma_{a_i}(z)\, d\Xi_i(a)$$

denotes aggregate consumption, Y *is the CD aggregator defined in (5),* $ac_{a_i}(z)$ *are the firm-specific hiring costs defined in (12), and* M_{a_i} *is the measure of potential firms per worker population size within a labor market.*

The set of remaining equilibrium and consistency conditions is reported in Appendix B.5. Existence and uniqueness of the stationary equilibrium obtain under regularity conditions on the parameters governing hiring costs and demand. These conditions ensure that the expected marginal revenue product of labor (MRPL) is increasing in productivity within each labor market. See Appendix B.6 for the formal proof.

Proposition 1 (Existence and uniqueness of stationary equilibrium)
Under regularity conditions, there exists a unique rank-preserving stationary equilibrium with continuous wage offer distributions and an ordered exit rule.

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, $R^n_{a_i}$, $n = \{u, e\}$, such that all offers equal or higher than that are accepted. 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.3 for the derivation.

Final good producer's policy. Solving the profit-maximization problem (4) yields the following demand function for each firm's variety $j = 1, \ldots, N_{k(a_i)}$ in product market $k(a_i)$:

$$y_{jk(a_i)} = \left(\frac{p_{jk(a_i)}}{P_{k(a_i)}}\right)^{-\sigma} \left(\frac{P_{k(a_i)}}{P_i}\right)^{-\rho} \left(\frac{P_i}{P}\right)^{-1} \frac{\alpha_i Y}{K_i}. \tag{13}$$

Reflecting the triple-nested production function (5), the demand for the good produced by firm j in sector $k(a_i)$ negatively depends on its price vis-à-vis the sectoral price index $P_{k(a_i)}$, on the sectoral price index vis-à-vis the industry price index P_i, and on the industry price index vis-à-vis the aggregate price index P. See Appendix B.2 for the derivation. Since the final good is the numeraire, its price is normalized to 1, i.e., $P = 1$. Notice that the CD aggregator of industry-specific output prevents any reallocation of expenditure shares 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 (8) consists of the following system of first-order conditions (FOCs):

$$\ell_{a_i}(w, v) = \left(\mathbb{E}_{B_{a_i}}\left[\left(1 - \epsilon^{-1}_{k(a_i)}(y; B_{a_i})\right) p_{k(a_i)}(y; B_{a_i})\right] z - w\right) \frac{\partial \ell_{a_i}(w, v)}{\partial w} - \frac{\partial c_{a_i}(w, v)}{\partial w}, \tag{14}$$

$$\frac{\partial c_{a_i}(w, v)}{\partial v} = \left(\mathbb{E}_{B_{a_i}}\left[\left(1 - \epsilon^{-1}_{k(a_i)}(y; B_{a_i})\right) p_{k(a_i)}(y; B_{a_i})\right] z - w\right) \frac{\partial \ell_{a_i}(w, v)}{\partial v}, \tag{15}$$

where $\epsilon_{k(a_i)}(y; B_{a_i}) \equiv -\frac{\partial \ln y^d_{k(a_i)}}{\partial \ln p_{k(a_i)}}$ denotes the endogenous elasticity of demand. The FOC with respect to the wage (14) equalizes the marginal increase in the wage bill (left-hand side) to the marginal increase in expected profits by raising employment through a wage rise (right-hand side). In turn, the latter is composed of the expected profit margin earned on new employees net of the induced price response ($\epsilon^{-1}_{k(a_i)}(y; B_{a_i}) > 0$), 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 (15) equalizes the marginal increase in hiring costs (left-hand side) to the marginal increase in expected 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.

In Appendix B.4 we solve the system of FOCs and show that the wage policy function equals the expected outside option of a searching worker when meeting a firm with productivity z, that is, a weighted average of her reservation wage and the expected MRPL of the

highest-paying firm she meets with productivity lower than z.

Let $\hat{p}_{k(a_i)}(z; B_{a_i}) \equiv \frac{p_{k(a_i)}(z;B_{a_i})}{\mathbb{E}[p_{k(a_i)}(z;B_{a_i})]}$ be the ratio between the realized and the expected price and $\mathbb{E}^p_{B_{a_i}}[X] = \mathbb{E}_{B_{a_i}}[\hat{p}_{k(a_i)}(z; B_{a_i})X]$ denote the price-weighted expectation with respect to the competitors' CES-output contribution. The system of FOCs can be manipulated to recover the familiar double-wedge representation of the optimality condition between the (expected) price and productivity-adjusted wage:

$$\mathbb{E}[p_{k(a_i)}(z; B_{a_i})] = \underbrace{\frac{\epsilon^{BC}_{a_i}(z)}{\epsilon^{BC}_{a_i}(z) - 1}}_{\mu_{k(a_i)}(z)} \underbrace{\frac{1 + \nu\eta_{a_i}(z)}{\nu\eta_{a_i}(z)}}_{\psi_{a_i}(z)} \frac{w_{a_i}(z)}{z}, \tag{16}$$

where $\epsilon^{BC}_{a_i}(z) \equiv \mathbb{E}^p_{B_{a_i}}\left[\epsilon^{-1}_{k(a_i)}(z; B_{a_i})\right]^{-1}$ is the endogenous Bayes-Cournot demand elasticity and η_{a_i} is the endogenous elasticity of labor supply. Equation (16) shares the same structure as its counterpart in the stylized model (1), but features two key differences. First, both the elasticity of demand, ϵ^{BC}, and the elasticity of labor supply, η, are endogenous. This means that the degree of market power varies 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 of the hiring-rate elasticity of hiring costs, ϕ, to their vacancy elasticity, $1 + \zeta$. Intuitively, this term captures the reduction in marginal hiring costs by growing larger in employment, so it is absent from any model where hiring costs are size-independent.

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.

Equilibrium Markups. The double-nested CES demand structure within each industry, as well as the granularity of firms in their own product market, entails that the elasticity of demand faced by a firm with productivity z operating in sectoral market $k(a_i)$ reads:

$$\epsilon_{k(a_i)}(z; B_{a_i}) = \left[\frac{1}{\rho}\omega_{k(a_i)}(z; B_{a_i}) + \frac{1}{\sigma}(1 - \omega_{k(a_i)}(z; B_{a_i}))\right]^{-1}, \tag{17}$$

where $\omega_{k(a_i)}(z; B_{a_i}) \equiv \left(\frac{y_{a_i}(z)}{Y_{k(a_i)}(z;B_{a_i})}\right)^{\frac{\sigma-1}{\sigma}} = p_{k(a_i)}(z; B_{a_i})y_{a_i}(z) / \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)}y_{jk(a_i)}$ is the realized 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 the weighted harmonic mean of the elasticity of substitution across sectors, ρ, and the elasticity of substitution within sectors, σ, with weights given by the firm's market share and

its complement to one, respectively (Atkeson and Burstein, 2008).

From Equation (17) and the markup definition (16), the equilibrium Bayes-Cournot markup $\mu_{k(a_i)}(z)$ of a firm with productivity z operating in product market $k(a_i)$ is given by:

$$\mu_{k(a_i)}(z) = \mathbb{E}^{p}_{B_{a_i}} \left[\left(\frac{\sigma}{(\sigma-1)\left[1 - \frac{\sigma/\rho-1}{\sigma-1}\omega_{k(a_i)}(z; B_{a_i})\right]} \right)^{-1} \right]^{-1}. \tag{18}$$

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., $\omega_{k(a_i)}(z; B_{a_i}) \to 0 \ \forall B_{a_i}$, 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., $\omega_{k(a_i)}(z; B_{a_i}) \to 1 \ \forall B_{a_i}$, then the markup approaches $\frac{\rho}{\rho-1}$, as if the firm were operating in monopolistic competition against other sectors.

A higher MW is expected to induce some exit (or layoffs) among low-productivity firms. In the presence of labor market frictions, displaced workers from low-productivity firms will (imperfectly) reallocate towards higher-productivity and larger firms, which have higher baseline markups. Moreover, the reallocation process will make large firms gain additional market share, 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.

Equilibrium Markdowns. Workers' search behavior and the labor market structure jointly entail that the labor supply elasticity faced by a firm with productivity z operating in labor market a_i reads:

$$\eta_{a_i}(z) = \frac{2 f_{a_i}(w_{a_i}(z)) w_{a_i}(z)}{\chi_{a_i}(\theta_{a_i}) + [1 - F_{a_i}(w_{a_i}(z))]}. \tag{19}$$

The elasticity of labor supply equals twice the quit rate elasticity.[18] 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. The firm-specific component makes the quit rate elasticity positively depend on the firm's wage policy. For given rank in the wage distribution and posted wage, 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.

[18] 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).

From Equation (19) and the markdown definition (16), the equilibrium markdown, $\psi_{a_i}(z)$, of a firm with productivity z operating in labor market a_i is given by:

$$\psi_{a_i}(z) = 1 + \frac{\chi_{a_i}(\theta_{a_i}) + [1 - F_{a_i}(w_{a_i}(z))]}{2\nu f_{a_i}(w_{a_i}(z))\, w_{a_i}(z)}. \tag{20}$$

Equilibrium markdown *rates*, i.e., $\psi_{a_i}(z) - 1$, equal the inverse of twice the quit rate elasticity normalized by the hiring cost correction term $\nu > 1$. 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. See Appendix B.4 for further details.

As in our stylized model, a higher MW would reduce the markdowns of constrained firms by raising the effective wage floor they face. 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.[19] Hence, these behavioral effects will push the markdown distribution to shift leftward. Depending on the convexity of the hiring cost function, excess hiring at higher-paying firms will partially offset job losses at low-paying firms, though aggregate vacancies are expected to drop and the frictional index to rise. 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. The interested reader can find all the details and formal derivations underlying our statements in Appendix C.

A constrained social planner chooses the direction and pace of worker reallocation while facing the same search and product-market information frictions as private agents. As in the baseline equilibrium, the planner reallocates workers towards higher-productivity firms whenever an opportunity arises.

Denote as *labor wedge* the ratio between the social marginal benefit and marginal cost of vacancy posting at the firm level (Hsieh and Klenow, 2009). Abstracting from endogenous activity choices, the baseline equilibrium features two sources of inefficiency. First, positive markup rates and distorted markdowns, due to congestion effects and distorted reservation wages, make the aggregate labor wedge generally differ from one. Second, heterogeneity in firm-specific markups, due to firms' granularity in their product markets, and in markdown distortions, due to the firm-specific incidence of congestion effects and reservation wages, makes the labor wedge heterogeneous across firms. The resulting allocation of labor is

[19] This spillover effect can be interpreted as a local increase in the wage offer density through the lens of Equation (20).

inefficient because it does not minimize aggregate hiring costs per unit of final good, which we adopt as our misallocation index. We further define an *aggregate* labor wedge, which aggregates up firm-level wedges. The aggregate labor wedge is above (below) one if and only if a proportional increase (decrease) in all active firms' vacancies raises aggregate consumption locally, holding activity fixed.

The MW represents an asymmetric cost-push shock, sparking reallocation from low- to high-productivity firms. If the MW pushes the aggregate labor wedge closer (farther) to one, the aggregate vacancy distortion will ameliorate (worsen).[20] If the MW decreases (increases) the productivity-adjusted dispersion of firm-level labor wedges, misallocation tends to ameliorate (worsen).

4 Structural Estimation

In this section, we outline how we estimate the parameters of our structural model using data from the Italian National Social Security Institute (INPS) and the Italian Institute of Statistics (Istat). First, we discuss how to map the model-based notions of labor and product markets to their closest real-world counterparts. Then, we present our estimation strategy and results. Further details can be found in Appendix D.

Market definition. Estimating firms' market power involves taking a stance on the definition of labor and product markets (Rossi-Hansberg et al., 2021; Eeckhout, 2020). To benchmark the market boundaries to observable characteristics in our data, we propose market definitions based on workers' heterogeneity for labor markets and goods differentiation, along with geographical scope, for product markets.

First, labor markets in the model are segmented by workers' skills and industry and populated by a measure of firms, none of which is large relative to the market. To address worker heterogeneity, we 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). Conceptually, AKM worker fixed effects capture persistent wage components that we interpret as reflecting permanent skill heterogeneity. They therefore provide an intuitive empirical counterpart to the skill type a in the model. We split workers into two skill types, high- and low-skilled, based on the median AKM worker fixed effect. We also assign each worker to an industry based on the Italian ATECO 1-digit industry code of the firm at which he or she was most frequently employed between 2016 and 2018. 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 made before labor market entry impose hard constraints on the set of jobs available to workers, while matching frictions limit the arrival of job offers from geographically distant

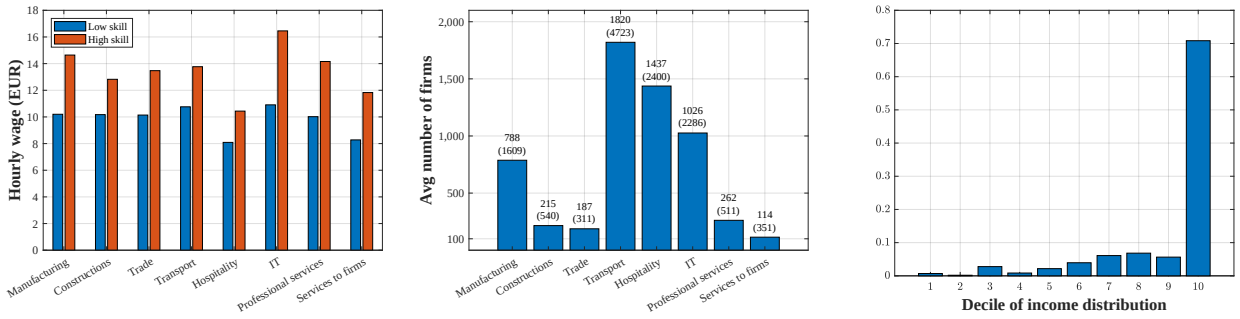
[20] Whether a reduction in the aggregate vacancy distortion is accompanied by higher or lower equilibrium vacancies and employment following a MW reform depends on the direction of the baseline inefficiency.

firms. It follows that the continuum of labor markets in the model maps into 16 segmented labor markets in the data. Average wages are increasing in our skill measure conditional on each industry, and exhibit substantial heterogeneity across industries (Figure 2a). Hence, our labor market definition seems to capture relevant structural differences.

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. We therefore think of product market boundaries in the model as comparable to ATECO 4-digit sectors – the most granular level available in our data and close to 5-digit NAICS industries in the US.[21] For local industries, we further segment product markets by region to capture the relevant pool of competitors. Figure 2b shows that the number of firms within a product market is 272 on average, with substantial variation across 1-digit industries in both mean and dispersion. Therefore, in our structural estimation we replicate not only the average number of competing firms but the entire industry-specific distribution of the number of competing firms per product market. The relatively large average number of firms per product market reflects the unusually high prevalence of small firms in the Italian economy by international standards.[22]

Figure 2: Market structure

(a) Average wage, by labor mkt (b) Firm number per prod mkt (c) Profits share, by income



Source: Authors' calculations, INPS (2016-2018), Istat (2019), and SHIW (2016). *Note*: A *product market* is defined as a 4-digit industry for national industries (manufacturing, IT, and transport) and as a 4-digit industry × region for local industries (wholesale and retail trade, constructions, hospitality, professional services, and services to firms). An *industry* is defined as a 1-digit industry (macro-sector) according to the ATECO classification. In Panel (b) numbers above bars report mean and standard deviation (in brackets) of the number of competing firms in a given product market by industry.

Discussion. By abstracting from a spatial dimension, our theoretical framework is not designed to accommodate geographical *heterogeneity* in market structure.[23] Still, we

[21] Overall, there are 615 ATECO 4-digit sectors and 698 NAICS 5-digit industries. We assume that firms operating in the same sector have common skill requirements for workers and, for multi-product firms, that the reported 4-digit industry code corresponds to the product line from which they earn the largest revenue share.

[22] Edmond et al. (2023) report an average number of 359 firms per 4-digit sector within the manufacturing industry in the US from the 2012 US Census of Manufactures (Autor et al., 2020). Although the Italian 4-digit industry classification is more granular than its US counterpart, the average number of firms per 4-digit sector within the manufacturing industry is 788.

[23] Differentiating markets based on their geographical scope would require estimating the joint spatial distribution of workers and firms across locations to pin down location-specific market sizes. We view the investigation of spatial heterogeneity as a promising avenue for future research, but beyond the scope of this paper.

account for the average effect of geographical segmentation in our estimation strategy by assuming that the empirical counterparts to our model-based markets are collections of *identical*, location-specific markets. On the one hand, this choice has no bearing on labor market equilibrium: since no firm is granular in its labor market, estimating a nationwide skill- and industry-specific labor market or a collection of identical location-specific labor markets delivers the same allocation. On the other hand, product market equilibrium is sensitive to market size. Since firms are granular in their product markets, reducing market size raises their market shares, which in turn affects their policy functions.[24]

Estimation Strategy. We estimate the model at monthly frequency by replicating a stationary equilibrium for the Italian economy, targeting empirical moments from the period 2016-2019. First, we externally set structural parameters with direct counterparts in the data. Then, we jointly estimate the remaining parameters, which are related to informative data moments by some equilibrium conditions, 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.

Externally set and directly inferred parameters. We externally set the parameters governing market structure, workers' preferences, and labor market frictions. We pin down the relative size of worker populations across labor markets by replicating the employment composition by industry and skill from INPS data (2016-2018). We account for the heterogeneous market structure across industries by replicating the empirical distribution of the number of potential (and operating) firms in a product market $k(a_i)$ for each industry i, $\mathcal{N}_{a_i} = \mathcal{N}_i$. Since a counterfactual increase in the MW can only weakly shrink the set of operating firms, assuming that all potential firms operate in the baseline equilibrium is without loss of generality for our purposes. We set the monthly discount rate r to replicate an annual interest rate of 5%. Following the macro literature, we set the coefficient of relative risk aversion, ϑ, to 1. To parametrize meeting frictions, we externally set the values of ι and β from the search literature and internally estimate χ to capture the meeting efficiency of the Italian economy.[25]

Some parameters can be directly inferred from the data. The measure of active firms in each labor market, M_{a_i}, equals the ratio between the employment share and the average firm size in each 1-digit industry i. We set the measure of product markets K_{a_i} within each labor market to ensure consistency between the measure of firms in the labor markets M_{a_i} and the

[24]The degree of international competition in product markets for *tradable* goods might also play a role for the adjustment to the MW (Harasztosi and Lindner, 2019), even if the likely direction is a priori ambiguous. On the one hand, considering larger, internationally integrated product markets is likely to widen the gap between the estimated elasticities of substitution within and across sectors needed to rationalize the observed labor share distribution across domestic firms, thus amplifying the potential markup response to the MW. On the other hand, foreign firms are unaffected by national MWs, thus limiting the extent of market share reallocation in response to the policy. Due to data limitations, we are unable to explore this trade-off quantitatively.

[25]We calibrate the elasticity of the meeting function by averaging the estimate of Blanchard and Diamond (1989) and the upper bound of the range estimated by Şahin et al. (2014). This choice produces a conservative estimate of the productivity gains from worker reallocation – *more limited* than in a model with CD meeting function.

finite number of firms populating each product market. We pick the separation rates δ_{a_i} to replicate the EN rate separately by skill type. The scalar of the hiring cost function, $\bar{c}_{a_i}$, and the on-the-job search intensity, s_{a_i}, are chosen to replicate the observed NE rate and EE rate in each labor market. Next, we proceed by discretizing the profit-rebate function $\mathcal{S}(G(aw))$ over a 10-point grid. The share of aggregate profits accruing to each decile of the wage distribution, $\mathcal{S}_d$, are chosen to replicate the empirical distribution of non-labor earnings across the income distribution reported by the Survey on Household Income and Wealth (SHIW) (Figure 2c).[26] 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 Panel A summarizes our externally set and directly inferred parameters.

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 estimate 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 estimated non-parametrically.

Let Θ be the vector of parameters still to be determined: $\Theta = \{\zeta, a_H, \chi, \phi, \sigma, \rho\}$. We choose parameter values that minimize the sum of squared percentage deviations between a set of moments estimated in actual data, m^D, and those generated by the model, $\hat{m}$. Formally, $\Theta^* = \arg\min_{\Theta \in \mathscr{P}} \frac{1}{|\mathscr{J}|} \sum_{j \in \mathscr{J}} \left(\frac{\hat{m}_j(\Theta) - m_j^D}{m_j^D} \right)^2$, where $\mathscr{P}$ denotes the parameter space and $\mathscr{J}$ is the set of targeted moments.

Although all parameters are estimated jointly, we provide heuristic arguments linking each parameter to a moment that is particularly informative for its identification. Table A.1 in Appendix reports the sensitivity matrix of parameter estimates with respect to the data moments (Andrews et al., 2017), providing formal support for these mappings. As in Engbom and Moser (2022), we estimate the convexity of the hiring cost function, ζ, to replicate the employment share of firms with more than 50 employees. We 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 normalized by the log wage variance, in the AKM regression.[27] Then, we estimate the meeting efficiency, χ, by replicating the ratio between vacant jobs and total (filled and vacant) jobs. The hiring rate elasticity of hiring costs, ϕ, is informed by the average Herfindahl-Hirschman index (HHI) across product markets, weighted by their value added. The higher ϕ, the

[26] 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 with business ownership.

[27] The AKM regression has no structural interpretation in our model due to differential labor market power across labor markets and no skill heterogeneity within each firm. Still, we proceed by simulating labor market histories generated by the stationary equilibrium of our model. Then, we cluster firms with the same expected 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.

higher the hiring cost correction term in the equilibrium markdown (20). A higher hiring cost correction term compresses the markdowns of high-productivity firms disproportionately more. Hence, for given wage distribution, the upper tail of the expected MRPL distribution is more compressed, thus lowering the weighted average HHI. Conditional on the hiring cost correction term, observed labor market moments – namely, wage distributions and worker transition rates – are sufficient statistics to identify equilibrium markdowns independently of the product market structure.

For product market parameters, we pin down the sectoral elasticity of demand, σ, by targeting an estimate of aggregate markup in Italy (Ciapanna et al., 2024). The greater 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 product market where they operate (Edmond et al., 2015). Intuitively, the lower 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. See Appendix D.3 for further details on the computation of estimation targets.

Our estimation strategy delivers a set of implied parameters. Unemployment benefits b_{a_i} are inferred from the lowest observed wages in each labor market. Since Italy has a system of contractual wage floors in place, these lowest observed wages may either reflect genuine reservation wages or binding wage floors. Because our estimation strategy interprets them as reservation wages, the resulting estimates of b_{a_i} should be viewed as upper bounds. This choice aims to limit the consumption loss upon unemployment implied by the absence of a saving technology. The parameters κ_{a_i} are set such that overhead costs equal 35% of the equilibrium operating profits of the smallest (observed) firm in each labor market. The overhead costs have to be large enough to prevent entry of lower-productivity firms in the baseline equilibrium and govern the firm exit elasticity to the MW. In Appendix E.2 we run robustness checks on the size of overhead costs and show that our choice delivers a conservative estimate of the aggregate effects of the concentration channel. In Section 7 we validate the implied firm exit response to the MW against external evidence. Finally, the CD industry weights are implied by the industry revenue shares consistent with the wage distributions. Table 1 Panel B summarizes our internally estimated parameters.

Productivity estimation. We now detail our strategy to estimate firms' productivity distributions from the observed wage distributions within the SMM routine.

Upon discretizing the wage distributions on a 125-point, equally-spaced grid, we follow a two-stage 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 (20), i.e., $\psi_{a_i}(w)$. It follows that the marginal cost piece rate of firms posting wage w in labor market a_i equals $MCL_{a_i}(w) = \psi_{a_i}(w)w$. Since, in equilibrium, firms equalize MCL to expected MRPL, this strategy allows us to characterize the model-consistent expected MRPL piece rate distribution in each labor market non-parametrically, that is, $\tilde{z} \sim \Phi_{a_i}(\psi_{a_i}(w)w)$ (Bontemps

Table 1: Parameter Estimates

Parameter	Description	Value	Target (Source)	Data	Model
		A. Externally set and directly inferred parameters			
Market structure					
I	Number of industries	8	ATECO 1-digit industries (ISTAT)		
A	Number of skill types	2	Median AKM worker FE (INPS)		
L_{a_i}	Worker population, by labor market ($= L_i \Xi_i'(a)$)	Tab. A.2	Employment composition, by industry and skill (INPS)		
M_{a_i}	Firm-to-worker population ratio, by labor market	Tab. A.2	Average firm size (ISTAT)		
$\mathcal{N}_i$	Distribution of the number of potential firms per product market, by industry	Fig. 2b	Distribution of the number of firms per product market, by 1-digit industry (ISTAT)		
K_{a_i}	Measure of product markets, by labor market	$\frac{L_i \Xi_i'(a) M_{a_i}}{N_i}$	Consistency		
Workers' preferences					
r	Discount rate	0.004	5% annual interest rate		
ϑ	Coefficient of relative risk aversion	1.000	Macro literature		
Labor market frictions					
ι	CES substitution parameter	0.289	Blanchard and Diamond (1989), Şahin et al. (2014)		
β	Vacancy share	0.5	Search literature		
δ_a	Separation rate, by skill	[0.029, 0.021]	EN rate by skill	[0.029, 0.021]	[0.029, 0.021]
$\bar{c}_{a_i}$	Scalar hiring costs	Fig. A.4a	NE rate (INPS)	0.161	0.161
s_{a_i}	On-the-job search intensity	Fig. A.4b	EE rate, by skill (INPS)	[0.013, 0.010]	[0.013, 0.010]
Profits rebate					
τ	Profits tax rate	0.662	Non-labor earnings to labor income, top income decile (SHIW)	0.708	0.708
$\mathcal{S}_{d=1}^{10}$	Profits rebate, by wage decile	Fig. 2c	Share non-labor earnings by income decile (SHIW)		
		B. Internally estimated parameters			
Labor market parameters					
ζ	Vacancy-cost curvature	1.032	Employment share 50+ firms (ISTAT)	0.427	0.403
a_H	Skill parameter, high skills	1.340	AKM sorting share of wage variance (INPS)	-0.066	-0.066
χ	Meeting efficiency	1.187	Vacancy rate (ISTAT)	0.014	0.014
ϕ	Hiring-rate elasticity of hiring costs	1.972	Weighted avg HHI, across product markets (ISTAT)	0.119	0.118
Product market parameters					
σ	Elast. of subst. within sectors	20.29	Sales-weighted harmonic avg markup rate (Ciapanna et al., 2024)	0.100	0.100
ρ	Elast. of subst. across sectors	2.500	Semi-elasticity labor share-market share (ISTAT)	-0.364	-0.365
Γ_{a_i}	Productivity distribution, by labor market	Fig. A.5	Labor-market-specific wage distributions (INPS)	Fig. A.2b	
Implied parameters					
b_{a_i}	Unemployment benefits	Fig. A.4c	Reservation wage		
κ_{a_i}	Overhead costs	Fig. A.4d	35% smallest operating profits		
α_i	CD industry weights	Tab. A.3	Industry revenue share		

Source: Authors' calculations, model, INPS, Istat, and SHIW. *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. Average firm size, the share of employment in firms with more than 50 employees, and the vacancy rate are taken from Istat data (2019). The distribution of the number of potential firms per product market, the HHI of each product market and the semi-elasticity of the labor share with respect to the product-market share are computed on the Structural Business Statistics dataset from Istat (2019). A product market corresponds to a 4-digit ATECO sector for national industries (manufacturing, IT, and transport), and to a 4-digit-ATECO-sector times region pair for local industries (trade, constructions, hospitality, professional services, and services to firms). 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. The model-based weighted average HHI and the market-share semi-elasticity of the labor share across product markets are estimated via Monte Carlo simulation. N_i is the expected number of potential firms in a product market within industry i, i.e., $N_i \equiv \mathbb{E}_{\mathcal{N}_i}[\bar{N}_i]$.

et al., 2000). Within each labor market, we consider a measure of product markets K_{a_i}, each of which is populated by $\bar{N} \sim \mathcal{N}_i$ potential (and operating) firms. The expected MRPL of each firm is an iid draw from the Φ_{a_i} distribution.

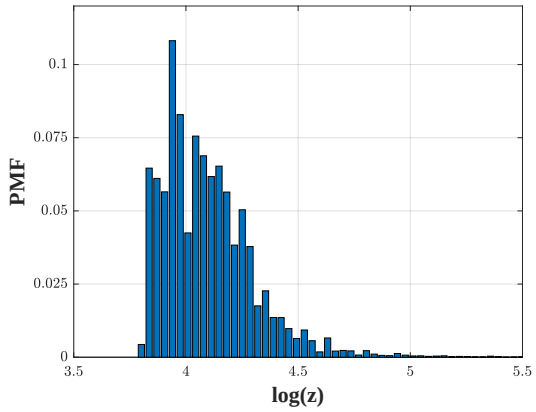
In the second stage, we notice that, for given CES elasticities of demand and expected MRPL distribution in each product market, there exists a unique combination of prices, markups, CD industry weights, and physical productivities consistent with our product market structure. In this way, we finally 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.

Parameter estimates and model fit. The last two columns of Table 1 compare the estimation targets in the data with those implied by the estimated model. The model replicates all targeted moments closely. Note that the implied hiring cost correction term equals 1.49, thus representing an important determinant of equilibrium markdowns. The estimated skill parameter for high-skilled workers is 1.34, indicating that significant variation in average wages across labor markets can be attributed to worker heterogeneity, as opposed to productivity heterogeneity among firms operating in different markets. The parameters governing the elasticity of demand are estimated to be $\sigma = 20.29$ and $\rho = 2.50$. Coupled with the low level of concentration of the Italian product markets, our estimated elasticity of substitution across sectors is rather high by international standards (Edmond et al., 2015; Atkin et al., 2015; De Loecker et al., 2021; Edmond et al., 2023; Burstein et al., 2025). This suggests that markups of Italian firms are not as sensitive to product market concentration. For this reason, we show in the next section that our estimates of the concentration channel of the MW in Italy can be interpreted as a lower bound for more concentrated economies, such as the US.

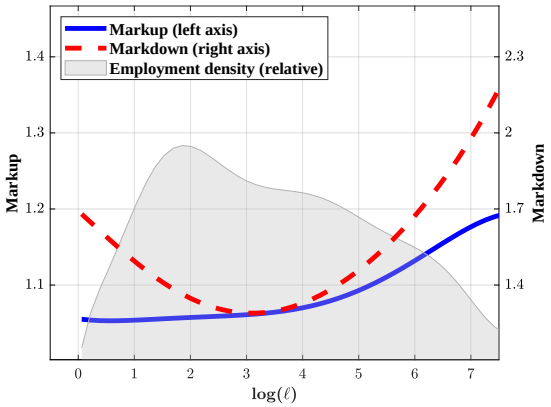
Figure 3a shows that the aggregate productivity distribution is right-skewed, with a longer upper tail than implied by a log-normal shape. This pattern masks substantial heterogeneity across industries (see Figure A.5 in the Appendix).

Figure 3: Productivity and market power

(a) Productivity distribution



(b) Market power, by firm size



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

Finally, we focus on the estimated distributions of markups and markdowns by firm size, reported in Figure 3b. The cross-sectional distribution of market power carries two main insights. First, the largest firms (more than 1,000 employees) are estimated to have both higher markups and markdowns than their competitors. 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. In terms of magnitudes, the aggregate markdown exceeds the aggregate markup (1.47 versus 1.09) – even though the former does not translate one-to-one into pure profits due to hiring costs (see Figure A.6 in Appendix for the respective profit shares).

Interestingly, our estimate of the aggregate markdown is remarkably close to those of Berger et al. (2022) and Yeh et al. (2022) for the US.

Untargeted moments. We validate our estimated model by comparing a set of relevant untargeted moments with their model counterparts, summarized in Appendix Figure A.3. First, we show that the estimated model closely matches the shape of the empirical distribution of firm-level value added. Second, we validate our productivity estimation strategy based on inverting the wage distribution by highlighting the close fit of the estimated model to the empirical wage gradient with respect to productivity and the worker-turnover gradient with respect to firm size. Finally, we emphasize that the distinctive prediction of our model that low-productivity firms enjoy greater labor market power than mid-productivity ones aligns well with the hump-shaped relationship between firm-level labor share and revenues within 1-digit industries.[28] Overall, these exercises show that the model captures key features of the Italian production structure. In Section 7 we provide further quantitative validation of the estimated model with respect to the economy's response to the MW.

5 Equilibrium Effects of the Minimum Wage

In this section we study the equilibrium effects of introducing a nationwide MW in the Italian economy, where none is currently in place. To do so, we use the estimated model as a laboratory to simulate stationary equilibria under alternative values of the MW parameter $\underline{\mathbf{w}}$.[29] Following the literature, we measure MW reforms through their *Kaitz index*, i.e., the ratio between the MW and the pre-reform median wage. Since the Kaitz index is not a sufficient statistic for the MW bite, we couple this statistic with the share of workers directly affected. As a benchmark, 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 (see Figure A.7 in Appendix).

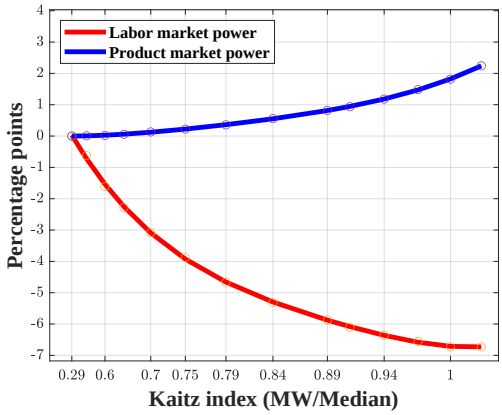
We start by assessing how market power reacts to the MW in Figure 4a. Labor market power (in red) decreases with the MW at a progressively slower rate. This reduction is driven by markdown compression among firms affected by the MW, either directly because they are constrained by the reform or indirectly through spillover effects. Since displaced workers from the most affected firms reallocate towards larger firms – which have higher baseline markdowns, on average – the reduction in labor market power progressively slows down due to a compositional effect. Overall, labor market power decreases by 6.7 percentage points from the baseline equilibrium to a Kaitz index of 100%, where it troughs. Product market power (in blue) increases with the MW at a progressively faster rate. This reflects both the concentration

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

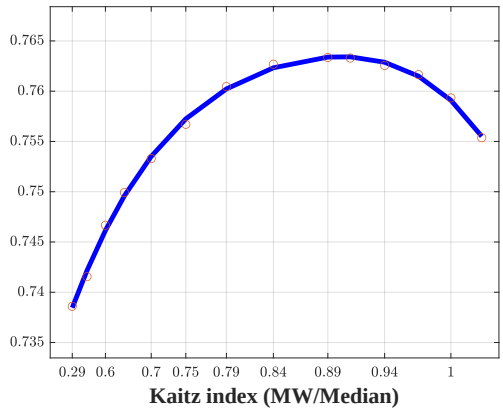
[29] This comparative statics exercise is informative about the effects of the MW in the medium term (while keeping existing contractual wage floors fixed), while being silent on the transitional dynamics of the labor market towards the new equilibrium (short-term adjustment), as well as on the potential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

Figure 4: Market power effects

(a) Market power response



(b) Aggregate labor share



Source: Model. *Note*: The labor market power response corresponds to the change in aggregate markdown weighted by the baseline markup net of a hiring cost adjustment term. The product market power response corresponds to the change in aggregate markup weighted for the baseline markdown. See Equation (46) in Appendix for the formal definitions. Aggregate labor share is defined as $LS \equiv \frac{\mathcal{W}+\mathcal{FC}}{\mathcal{C}}$, where $\mathcal{W}$ is the aggregate wage bill and $\mathcal{FC}$ are the aggregate overhead costs. See Appendix B.7 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 and red solid lines their 4^{th}-order polynomial fit.

channel of the MW (behavioral effect) and worker reallocation towards higher-markup firms (compositional effect).[30] Overall, product market power increases by 2.2 percentage points from the baseline equilibrium to the highest MW considered, where it peaks. Hence, the increase in product market power undoes, on average, one third of the reduction in labor market power. However, its importance grows as the MW increases, eventually becoming the dominant force for the largest MW reforms (see Figure A.9e in Appendix). This is reflected in the muted and hump-shaped response of the aggregate labor share in Figure 4b, peaking at a Kaitz index of 89% (35% of workers directly affected). As in the stylized model, the aggregate labor share response is driven by the net market power response. On the one hand, markdown compression is the dominant force behind the initial increase in the aggregate labor share. On the other hand, the flattening and eventual decline of the aggregate labor share are driven by the concentration channel, which raises markups. Overall, the aggregate labor share increases by 1.7 percentage points from the baseline equilibrium to the highest MW considered, with a 2.5-percentage-point increase at peak, corresponding to a Kaitz index of 89%.

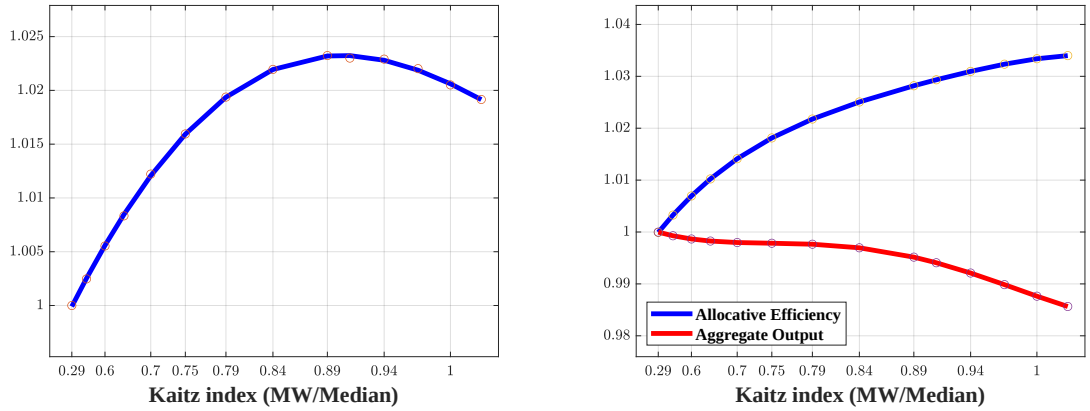
The product market power response is driven by industries with the highest baseline concentration, as measured by their average HHI across realized product markets. We estimate that an average baseline HHI one standard deviation above its mean is associated with a 34% higher elasticity of sectoral markups with respect to the share of directly affected workers (from 0.030 to 0.040). Similarly, it is associated with a lower elasticity of the sectoral labor share to the MW by 5% (from 0.056 to 0.053). This motivates our use of the baseline HHI as an ex ante indicator of the strength of the concentration channel in Section 8. Figure A.8 in

[30] Notice that the product market power response is estimated conservatively. This is because aggregate profits decline with the MW (see Figure A.9d in Appendix). Thus, a model where the mass of potential firms were pinned down by a free entry condition would entail *fewer* firms than ours.

Appendix shows that the aggregate impact of the concentration channel increases sharply with the baseline average HHI. Hence, we expect the concentration channel to operate even more strongly in economies with more concentrated product markets, such as the US.

Figure 5: Aggregate effects of the MW

(a) Aggregate consumption (b) Allocative efficiency and aggregate output



Source: Model. *Note*: Aggregate output is final good production (Y). Aggregate consumption is defined as $C = Y - \mathcal{HC} = \mathcal{E}Y$, $\mathcal{HC}$ is total hiring costs and $\mathcal{E} \equiv 1 - \frac{\mathcal{HC}}{Y}$ is an index of allocative efficiency. Panel (a) reports the response of aggregate consumption; Panel (b) reports the response of allocative efficiency (in blue) and aggregate output (in red). All variables are normalized to 1 in the baseline equilibrium without a MW. See Appendix B.7 for the model-consistent aggregation. The x-axis is scaled so as to reflect the share of directly affected workers. The y-axis reports levels relative to the baseline value. The red circular markers represent the simulated data, the blue solid lines their 4^{th}-order polynomial fit.

We now focus on the aggregate effects of the MW. Figure 5a reports the aggregate consumption response to the MW. Just like the labor share, aggregate consumption is hump-shaped in the MW, with a peak at a Kaitz index of 89%. Overall, aggregate consumption increases by 1.9% from the baseline equilibrium to the highest MW considered, with an increase by 2.3% at peak. Figure 5b decomposes the aggregate consumption response into aggregate output and allocative efficiency (one minus the misallocation index, i.e., aggregate hiring costs per unit of output). Interestingly, aggregate output (in red) *decreases* with the MW – only mildly for Kaitz indices up to 84%, though. On the other hand, allocative efficiency steadily increases with the MW. We conclude that aggregate consumption rises with the MW because high-productivity firms experience lower congestion in hiring.[31]

Finally, Figure A.9 in Appendix further decomposes the aggregate output response into labor productivity and headcount employment. It turns out that the aggregate output decline is driven by a reduction in headcount employment, only partially offset by an increase in labor productivity. According to our estimates, an average of one in six jobs paying less than the MW is destroyed by the policy.

[31] By abstracting from this channel, Berger et al. (2025) find virtually no effect of the MW on aggregate consumption. This is consistent with our finding of a muted aggregate output response.

6 Welfare Analysis

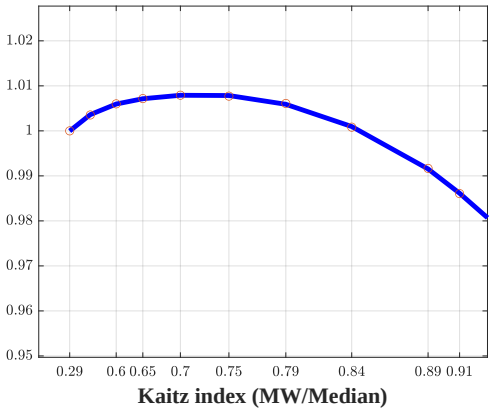
In the previous section we have shown that introducing a MW generally raises both aggregate consumption and the unemployment rate. Hence, the welfare effects of the MW are a priori ambiguous. On the one hand, lifetime consumption is higher (*level effect*). On the other hand, higher unemployment risk entails higher consumption volatility (*uncertainty effect*).[32] 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 normative 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. The welfare-maximizing MW, $\underline{\mathbf{w}}^{\star}$, is defined as:

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

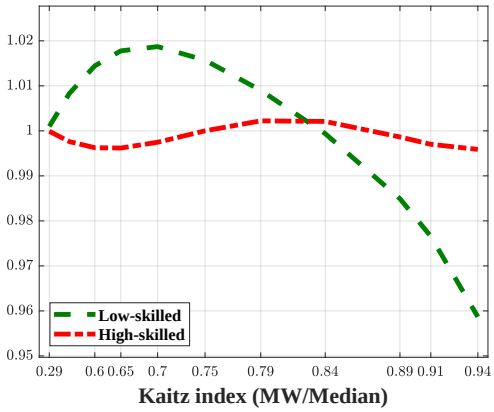
where U_{a_i} and $W_{a_i}(w)$ are the labor-market-specific asset values of unemployment and employment at wage w, respectively, defined in (2)-(3).[33] Henceforth, we refer to the welfare-maximizing MW, $\underline{\mathbf{w}}^{\star}$, as *optimal* MW.

Figure 6: Welfare effects of the MW

(a) Aggregate welfare



(b) Welfare effects, by skill



Source: Model. *Note*: Panel (a) reports the consumption-equivalent percentage variation of the aggregate utilitarian welfare across MW reforms. The red circular markers represent the simulated data, the blue solid lines their 7^{th}-order polynomial fit. Panel (b) decomposes the utilitarian welfare response into the welfare responses of low- and high-skilled workers. The displayed lines represent the 7^{th}-order polynomial fit of the simulated data. All aggregate variables are normalized to 1 in the baseline equilibrium without a MW. The x-axis is scaled so as to reflect the share of directly affected workers. The y-axis reports levels relative to the baseline value.

Figure 6a shows that the optimal MW equals a Kaitz index of 70% (10% of workers directly affected). The optimal MW brings about a welfare gain of 0.8% in consumption-equivalent units, in the face of a 1.2% increase in aggregate consumption. For Kaitz indices up to 75%, the increase in aggregate consumption (level effect) dominates the increase in consumption

[32] As our model abstracts from intertemporal consumption-savings decisions, workers have no means to mitigate individual consumption volatility.

[33] In the estimated model, workers set their reservation wage with positive time discounting, while firms adopt steady-state behavior, i.e., behave as if there is no time discounting. Quantitatively, our results are robust to letting workers adopt steady-state behavior, as well.

volatility due to 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 60%. The optimal MW balances efficiency and redistribution. See Appendix F for the formal decomposition.

To investigate the determinants of the distributional effect, Figure 6b reports the welfare effects of the MW across skill types. The 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 79% (20% of workers directly affected). Kaitz indices up to 79% redistribute consumption shares towards low-skilled workers. As a result, welfare inequality drops, which is reflected in a positive distributional effect. On the contrary, higher Kaitz indices hit significantly harder low-skilled than high-skilled workers due to larger job losses, which leads to a negative distributional effect. Therefore, our results offer a word of caution about the idea of raising the MW to unprecedented levels in order to reduce inequality, as in MaCurdy (2015). In fact, setting the MW at Kaitz indices higher than 79% would end up *increasing* welfare inequality.

7 Quantitative Validation

In this section we validate the quantitative performance of our model along two dimensions. First, we compare the model-implied effects of the MW on employment, prices, and reallocation to external empirical evidence from different countries. Second, we provide novel evidence from Italy by exploiting variation in industry-specific contractual wage floors.

External validation. We first benchmark the model-implied effects of the MW against empirical estimates from other countries. Table 2 reports the empirical moment, the corresponding model-implied moment, and the MW reform used in the model to ensure comparability. The model performs well across three key margins: employment, prices, and reallocation.

First, we benchmark the employment effects of the MW to existing evidence focusing on the model-consistent category of *minimum wage workers*, that is, low-skilled workers in hospitality. The model implies an employment elasticity in the ballpark of available estimates from other countries.

Second, we study the price effects of the MW. The model-implied price pass-through and price elasticity in hospitality align well with empirical evidence from the US and other countries (Wiltshire et al., 2026; Pérez Pérez, 2026).

Finally, the model generates reallocation effects similar to those documented by Dustmann et al. (2022) and Aaronson et al. (2018) for Germany and the US, respectively. Namely, the model replicates closely the reallocation share of wage gains at the worker level and the elasticity of firm exit to the MW. See Appendix G for additional details on the replication exercises.

Table 2: External validation of model-implied minimum wage effects

Margin	External evidence	Empirical object	Empirical estimate	Model analogue	MW reform in the model (*population at risk*)
Employment	US and international evidence (Belman and Wolfson, 2019; Dube and Lindner, 2024; Dube and Zipperer, 2024)	Employment elasticity among MW workers; own-wage employment elasticity	MW elasticity: $[-0.13, -0.07]$ Own-wage elasticity: -0.25	MW elasticity: $[-0.06, -0.04]$ Own-wage elasticity: $[-0.33, -0.28]$	52–65% Kaitz index (*low-skilled workers in hospitality*)
Prices	2016–2020 state-level MWs in California and New York (Wiltshire et al., 2026; Ashenfelter and Jurajda, 2022)	Price pass-through in US McDonald's restaurants for a reform inducing a 0.70 wage elasticity	0.50	0.50	Adjacent reform selected to match a wage elasticity of 0.70 (*firms in low-skill-hospitality sectors*)
Prices	US and international evidence (Pérez Pérez, 2026)	Price elasticity to the MW using sectoral price indices	MW elasticity: $[0.03, 0.11]$	Own-wage elasticity: $[0.04, 0.06]$	52–65% Kaitz index (*low-skilled workers in hospitality*)
Reallocation (workers)	2015 MW reform in Germany, with 15% of workers affected (Dustmann et al., 2022)	Share of wage gains due to moves to higher-paying establishments	$[14.1, 16.8]$%	17.8%	15.4% of baseline workers affected (*affected workers*)
Reallocation (firms)	2001–2006 state-level MW reforms in the US; MW increases ranged from 6.2% to 26.2% (Aaronson et al., 2018)	Excess firm exit response in limited-service restaurants and hotels/motels	LS restaurants: 0.137 (0.049) Hotels/motels: 0.456 (0.120)	$[0.133, 0.640]$	MW increases closest to 6.2–26.2% from the first MW equilibrium where over half of hospitality workers are paid within 1.5 times the MW (*firms in hospitality sectors*)

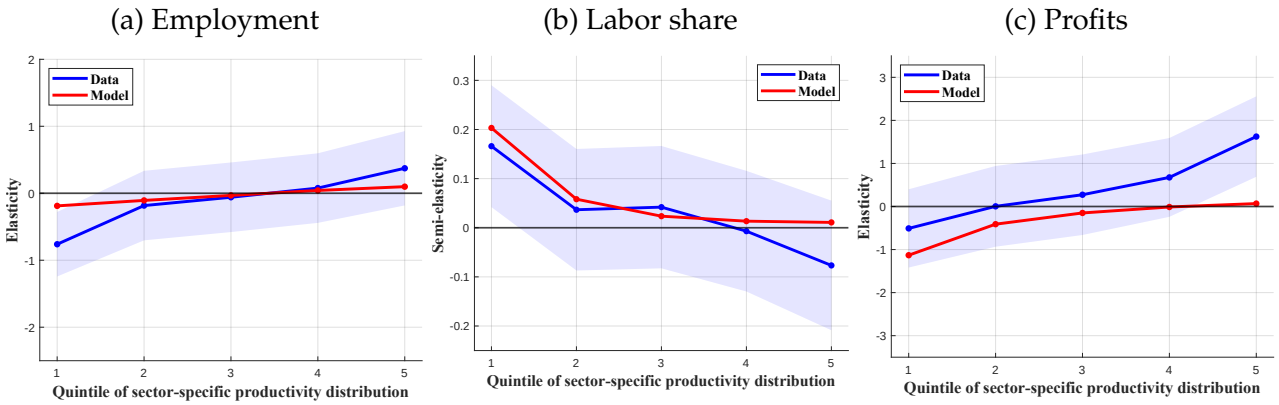
Source: External evidence and model simulations. *Notes*: The table compares empirical estimates from the literature with model-implied moments computed using comparable minimum wage reforms. Standard errors are reported in parentheses. The model ranges are the minimum and maximum elasticities across the reported reforms. The own-wage elasticity is the percent change in the outcome induced by a percent change in wages caused by the minimum wage. The population at risk is the share of affected workers, i.e., the CDF of the baseline wage distribution evaluated at the counterfactual MW. Price pass-through is the price elasticity with respect to the wage divided by the baseline value-added labor share. The selected pass-through reform moves from a Kaitz index of 94.1% to 97.1% and yields a wage elasticity of 0.706. The price-pass-through interval is selected using the wage-elasticity estimate of Ashenfelter and Jurajda (2022). For the comparison with Aaronson et al. (2018), the baseline is the first equilibrium in which more than half of hospitality employment is paid within 1.5 times the MW. The model range is computed across all policy-grid reforms between the MW-increase endpoints closest to 6.2% and 26.2% relative to that baseline.

Validation from contractual wage floors in Italy. We complement the external validation with novel evidence from Italy. Absent a nationwide MW, we exploit fifteen years of variation in industry-specific contractual wage-floor indices reflecting national collective bargaining agreements. These contractual wage floors generate staggered changes in the effective MW across industries, allowing us to compare the model-implied cross-sectional response across firms with its empirical counterpart in Italian firm-level data.[34] See Appendix H.3 for additional details on the data sources.

We proceed by simulating the response to industry-specific MWs affecting 2.44% of workers in our model—the average bite of contractual wage floors in the data. We then assign

[34] By comparing empirical and model-implied responses to industry-specific MWs, we implicitly interpret the former as causal effects. Still, we acknowledge that changes in contractual wage floors may be endogenous to industry performance. As long as potential endogeneity biases only the average industry-level response, the relative response across (invariant) firm productivity classes would raise fewer endogeneity concerns.

Figure 7: Firm-level response by productivity, industry-specific MW



Source: CERVED (2005-2020), INPS and Istat data, model. *Note*: Panels (a)–(c) compare, respectively, the empirical and model-implied average elasticity of firm-level employment, the average semi-elasticity of firm-level labor share, and the average elasticity of firm-level profits to an industry-specific MW by productivity bins, both computed from the regression model (21). Observations are weighted by firm employment. The outcome variables are measured at the end of the calendar year corresponding to the first year in which the wage-floor increase is in effect. Time effects are interacted with 2-digit industry and an indicator for product-market concentration being above the economy-wide value-added-weighted average. As in the quantitative model, a product market corresponds to a 4-digit ATECO sector for national industries (manufacturing, IT, and transport), and to a 4-digit-ATECO-sector times region pair for local industries (trade, constructions, hospitality, professional services, and services to firms). Standard errors are clustered at the firm-level. Shaded areas represent 90% confidence intervals.

firms, both in the data and in the model, to their most frequent quintile of the 4-digit-sector-year distribution of log value added per worker—a proxy of productivity.[35] We estimate the following regression model:

$$Y_{i,t} = \sum_{q=1}^{5} \beta_q \log MW_{s_3(i),t} \times \mathbb{1}_{\{Q(i)=q\}} + \phi_{t(i)} + \alpha_i + \epsilon_{i,t}, \tag{21}$$

where $Y_{i,t} = \{\log \ell_{i,t}, LS_{i,t}, \log \pi_{i,t}\}$ is the natural logarithm of employment, the labor share – defined as the ratio of total labor costs to value added –, or the natural logarithm of profits of firm i at time t, $\log MW_{s_3(i),t}$ is the natural logarithm of the average contractual wage floor of the 3-digit industry in which firm i operates, $\phi_{t(i)}$ are 2-digit-industry time effects (differentiated by high vs low product-market concentration), α_i are firm fixed effects, and $\epsilon_{i,t}$ is an idiosyncratic error term. Figure 7 compares the estimated firm-level (semi-) elasticities across the productivity distribution from regression (21) with those obtained by estimating the same fixed-effects specification on model-generated data. Panel 7a shows that the model successfully generates MW-induced worker reallocation from low- to high-productivity firms. Panel 7b shows that the model-implied labor-share response across the productivity distribution aligns well with the empirical response. Finally, Panel 7c shows that the model also delivers a plausible response of profits.

[35] Using each firm's modal productivity quintile over the sample allows us to retain firms entering the market after 2005, thus providing a more stable measure of persistent productivity.

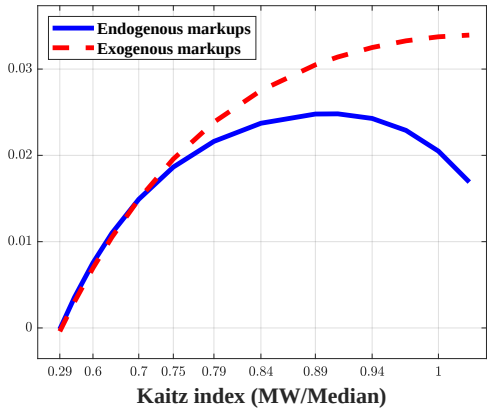
8 The Role of Endogenous Markups

In the previous two sections we have tackled positive and normative questions related to 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. To do so, we adopt two complementary approaches, one structural and one reduced-form.

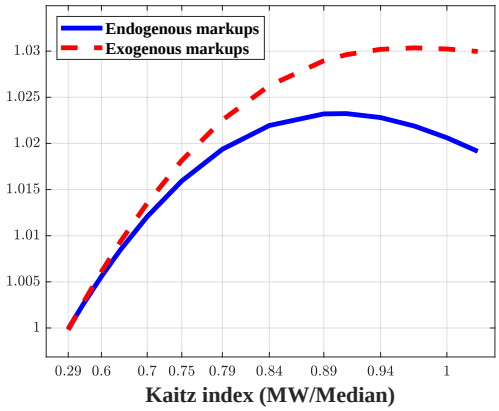
Structural approach. We repeat the same policy experiments of Section 5 in an observationally equivalent economy – that is, matching the same empirical moments as our benchmark – that retains the same demand structure but where firms charge constant, exogenous markups.[36] See Appendix H for further details on the alternative economy.

Figure 8: Aggregate response: Endogenous vs exogenous markups

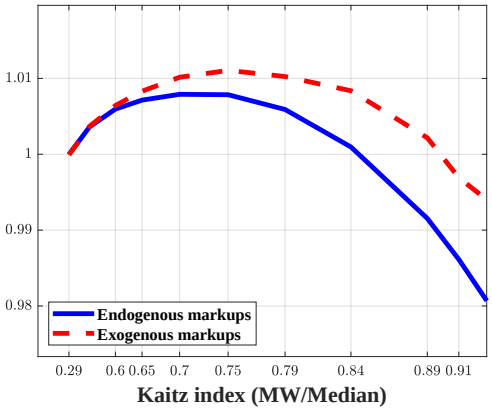
(a) Aggregate labor share response



(b) Aggregate consumption



(c) Aggregate welfare



Source: Model. *Note*: Panel (a) compares the changes in the aggregate labor shares, while Panels (b) and (c) compare the relative change of aggregate consumption and aggregate welfare, normalized to 1 in the baseline equilibrium without a MW. The *endogenous markup* economy in solid blue is our benchmark economy with oligopolistic competition in product markets, while the *exogenous markup* economy in dashed red is an observationally identical economy to our benchmark, with same demand structure but exogenous markups. The displayed lines represent the 4^{th}-order polynomial fit of the simulated data. The x-axis is scaled so as to reflect the share of directly affected workers.

[36] We interpret differential results across the benchmark and the alternative economy as *structural* counterfactuals, i.e., what would happen by ignoring endogenous markups when estimating the equilibrium effects of the MW. In Appendix H we also study an alternative economy with exogenous markups keeping the baseline productivity distribution constant and letting the wage distribution adjust. Results are consistent across the two alternative economies.

Figure 8 reports the response of the aggregate labor share, consumption, and welfare across the two economies. Our benchmark 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 increase by more and monotonically with the MW if markups were exogenous. Second, aggregate consumption would peak at a Kaitz index of 100% (instead of 89%) if markups were exogenous. Aggregate consumption is higher in the economy with exogenous markups across the entire spectrum of MW reforms. The consumption-maximizing MW brings about 40% greater consumption gains than in the benchmark economy. Finally, the economy with exogenous markups calls for a higher optimal MW. The optimal MW equals a Kaitz index of 75% (rather than 70%), affecting 15% of workers (rather than 10%), with associated welfare gains of 1.1% (rather than 0.8%) consumption-equivalent units. Hence, if markups were exogenous, the optimal MW would yield 40% greater welfare gains than in the benchmark economy. Intuitively, with endogenous product market power, firms are more reluctant to grow larger in size. This curbs the reallocation process behind the productivity gains of the MW. Overall, we conclude that endogenous markups matter for quantifying the equilibrium and welfare effects of MW reforms.

Reduced-form approach. In the previous section, we have shown that Italian contractual wage floors induce a reallocation of productive resources from low- to high-productivity firms. We now aim to provide reduced-form evidence for the concentration channel: high-productivity firms react to higher market share by raising their 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).[37]

To study heterogeneous effects by productivity and product market concentration, we estimate the model (21) interacted with a dummy indicating whether the concentration of the product market k in which firm i operates is above the economy-wide weighted average, $D_k(i) \equiv \mathbb{1}\{HHI_{k(i)} > \overline{HHI}\}$:

$$Y_{i,t} = \sum_{q=1}^{5} \left(\beta_q + \beta_q^h D_k(i)\right) \log MW_{s_3(i),t} \times \mathbb{1}_{\{Q(i)=q\}} + \phi_{t(i)} + \alpha_i + \epsilon_{i,t}. \tag{22}$$

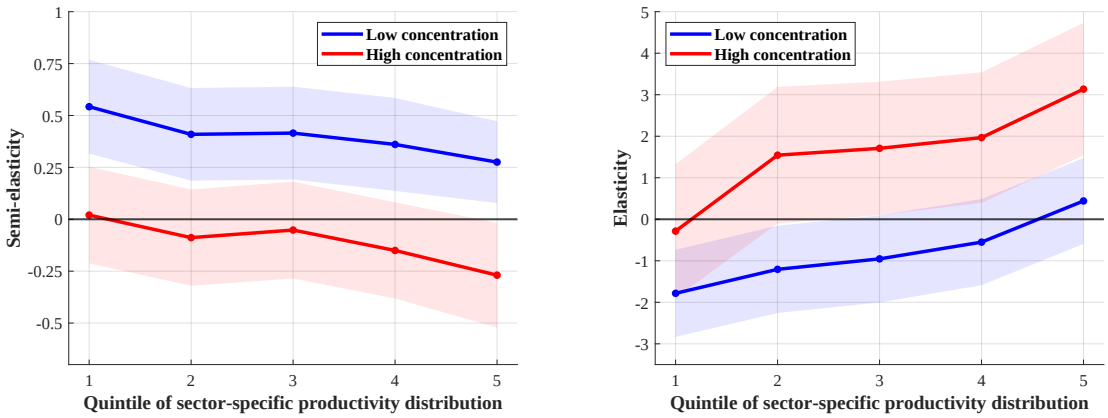
The outcome variable, $Y_{i,t}$, is either the labor share of firm i at time t or the natural logarithm of accounting profits (EBITDA). We use the same definition of product markets as in the quantitative model. Standard errors are clustered at the firm level.

The regression results are plotted in Figure 9, which uncovers two important results. First,

[37] Thanks to product-level custom data, Dodini et al. (2023) document 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. Consistently with the concentration channel, the markup response is present only in sectors where significant market share is reallocated from small to large firms. Du and Wang (2020) document an increase in manufacturing markups induced by the MW in China.

Figure 9: Firm-level response by concentration and productivity

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



Source: CERVED (2005-2020), INPS and Istat data. *Note*: The panels plot the estimated coefficients of the interaction terms between the quintile of the sector-specific productivity distributions by product market concentration and the natural logarithm of wage floor, computed from regression (22). As in the quantitative model, a product market corresponds to a 4-digit ATECO sector for national industries (manufacturing, IT, and transport), and to a 4-digit-ATECO-sector times region pair for local industries (trade, constructions, hospitality, professional services, and services to firms). The outcome variables are the firm-level labor share in Panel (a) and the natural logarithm of firm-level profits in Panel (b). Firm level profits are measured by the *Earnings Before Interest, Taxes, Depreciation, and Amortization* (EBITDA), as reported in firms' balance sheet. Observations are weighted by firm employment. The outcome variables are measured at the end of the calendar year corresponding to the first year in which the wage-floor increase is in effect. Time effects are interacted with 2-digit industry and an indicator for product-market concentration being above the economy-wide value-added-weighted average. Standard errors are clustered at the firm-level. Shaded areas represent 90% confidence intervals.

the labor share response is higher in less-concentrated product markets than in highly concentrated product markets across the entire productivity distribution. Hence, the labor share response is decreasing in product market concentration. Second, the labor share response of high-productivity firms in highly concentrated product markets is negative, driven by a positive profits response. The increase 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 H.3 we show that the response of markups (estimated from revenue data) is consistent with this interpretation, and that labor market power is unlikely to be a confounder of our results. In fact, we find that the labor share response in more concentrated labor markets is actually *higher* across the entire productivity distribution – in line with existing studies (Azkarate-Askasua and Zerecero, 2023; Popp, 2024; Azar et al., 2024). 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 and welfare effects of the minimum wage when product market power is endogenous and varies with market competition. We make two main original contributions to the literature. First, we are the first to formalize and empirically validate the concentration channel of the minimum wage. A higher minimum wage reallocates workers from small to large firms. If product market power is endogenous, large firms gain market share and increase their price markups. Second, we develop and estimate a new

structural model combining oligopolistic competition in product markets with wage posting in frictional labor markets, which can be used to study spillover effects between product and labor markets induced by a variety of policies, such as the minimum wage.

The main takeaway from this paper is that the optimal minimum wage depends on the structure of product markets. It follows that setting a high minimum wage may not be advisable if product markets feature a few dominant firms. 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 include the addition of capital and an input-output structure, which may play a role in the transmission of the minimum wage to the economy.

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. CEP Discussion Papers dp1823, Centre for Economic Performance, LSE.

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592.

Ashenfelter, O. and Jurajda, S. (2022). Minimum wages, wages, and price pass-through: The case of mcdonald's restaurants. *Journal of Labor Economics*, 40(S1):S179–S201.

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. (2024). Minimum wage employment effects and labour market concentration. *Review of Economic Studies*, 91(4):1843–1883.

Azkarate-Askasua, M. and Zerecero, M. (2023). Union and firm labor market power. Crc tr 224 discussion paper series, University of Bonn and University of Mannheim, Germany.

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. (2025). Minimum wages, efficiency, and welfare. *Econometrica*, 93(1):265–301.

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

Bilal, A. and Lhuillier, H. (2021). Outsourcing, inequality and aggregate output. Working Paper 29348, National Bureau of Economic Research.

Blanchard, O. and Diamond, P. (1989). The beveridge curve. *Brookings Papers on Economic Activity*, 20(1):1–76.

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.

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

Burstein, A. T., Carvalho, V. M., and Grassi, B. (2025). Bottom-up markup fluctuations. *The Quarterly Journal of Economics*, 140(4):2619–2684.

Ciapanna, E., Formai, S., Linarello, A., and Rovigatti, G. (2024). Measuring market power: macro- and micro-evidence from italy. *Empirical Economics*, 67(6):2677–2717.

Clemens, J. (2021). How do firms respond to minimum wage increases? understanding the relevance of non-employment margins. *Journal of Economic Perspectives*, 35(1):51–72.

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.

Deb, S., Eeckhout, J., Patel, A., and Warren, L. (2022). Fbbva lecture 2022: What drives wage stagnation: Monopsony or monopoly? *Journal of the European Economic Association*, 20(6):2181–2225.

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

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. CESifo Working Paper Series 10412, CESifo.

Du, P. and Wang, S. (2020). The effect of minimum wage on firm markup: Evidence from china. *Economic Modelling*, 86:241–250.

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

Dube, A. and Zipperer, B. (2024). Own-wage elasticity: Quantifying the impact of minimum wages on employment. NBER Working Papers 32925, National Bureau of Economic Research, Inc.

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.

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

Firooz, H. (2025). The pro-competitive consequences of trade in frictional labor markets. *Journal of International Economics*, 153(C).

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

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

Gouin-Bonenfant, E. (2022). Productivity dispersion, between-firm competition, and the labor share. *Econometrica*, 90(6):pp. 2755–2793.

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

Haucap, J., Pauly, U., and Wey, C. (2001). Collective wage setting when wages are generally binding an antitrust perspective. *International Review of Law and Economics*, 21(3):287–307.

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.

Hurst, E., Kehoe, P. J., Pastorino, E., and Winberry, T. (2025). The macroeconomic dynamics of labor market policies. Working Paper 33614, National Bureau of Economic Research.

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

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.

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. (2021). Trade and market power in product and labor markets. Staff Working Papers 21-17, Bank of Canada.

MaCurdy, T. (2015). How effective is the minimum wage at supporting the poor? *Journal of Political Economy*, 123(2):497–545.

Manning, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

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.

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

Pérez Pérez, J. (2026). The effect of the minimum wage on prices: A new survey and a meta-analysis. Technical report.

Popp, M. (2024). Minimum wages in concentrated labor markets. IZA Discussion Papers 17357, Institute of Labor Economics (IZA).

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

Trottner, F. (2023). Unbundling market power. Technical report, UC, San Diego.

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

Wiltshire, J., McPherson, C., Reich, M., and Sosinskiy, D. (2026). Minimum wage effects and monopsony explanations. *Journal of Labor Economics*, forthcoming.

Wong, H. C. (2026). Understanding high-wage firms: Monopoly, monopsony, and bargaining power. *American Economic Review*, forthcoming.

Yeh, C., Macaluso, C., and Hershbein, B. (2022). Monopsony in the us labor market. *American Economic Review*, 112(7):2099–2138.

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

Appendix

A Stylized Model

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

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

Firm-level output, wages, and prices are expressed relative to fixed reference levels normalized to 1 and held constant across equilibria. For any MW, the equilibrium number of active firms is given by $N = \sum_j \mathbb{1}\left\{\pi_j^\star(z_j; N, \underline{w}) \geq 0\right\}$, where $\pi_j^\star(z_j; N, \underline{w})$ represents maximized profits by a firm with productivity z, when the number of competing firms is N and the MW is $\underline{w}$. Since maximized profits are strictly increasing in productivity, the set of active firms is ordered by productivity. We condition on an active-set equilibrium satisfying this consistency requirement.

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:

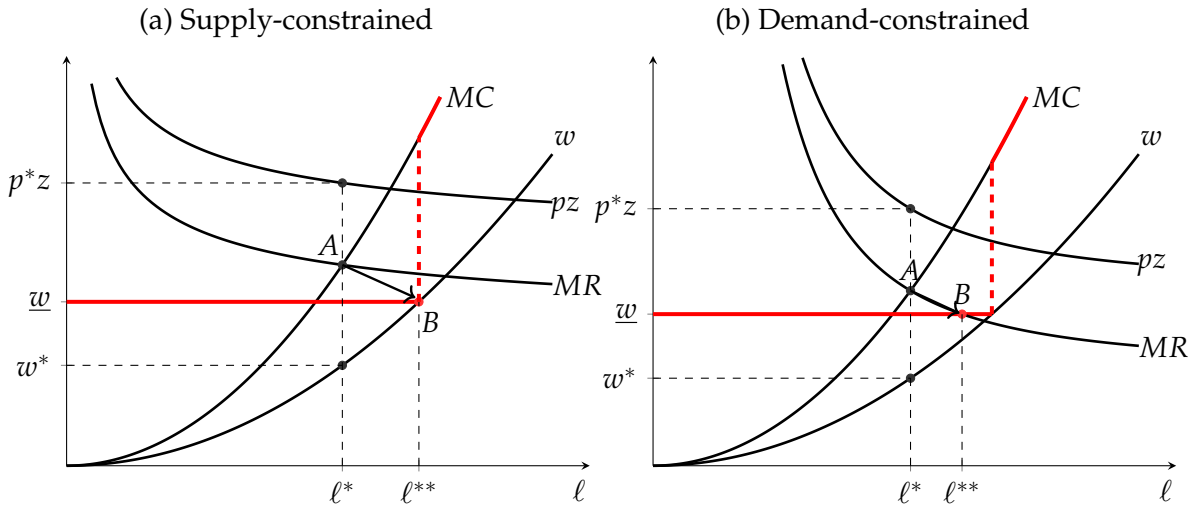
$$\begin{aligned} \ell(z) &= z^{\eta\frac{\epsilon-1}{\epsilon+\eta}}(\mu\psi)^{-\frac{\epsilon\eta}{\epsilon+\eta}}, \\ p(z) &= (z\ell(z))^{-\frac{1}{\epsilon}}, \quad w(z) = \ell(z)^{\frac{1}{\eta}}, \end{aligned} \tag{23}$$

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.1 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. The MW eliminates their labor-market power, compressing their markdown 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 that employment lies on the labor supply curve (left panel of Figure A.1). Hence, the policy

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



functions of supply-constrained firms are given by:

$$\ell(z) = z^{\eta\frac{\epsilon-1}{\epsilon+\eta}}(\mu\psi(z))^{-\frac{\epsilon\eta}{\epsilon+\eta}}, \tag{24}$$

$$p(z) = (z\ell(z))^{-\frac{1}{\epsilon}}, \quad \underline{w} = \ell(z)^{\frac{1}{\eta}}. \tag{25}$$

From (24) and (25), 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:

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

The most productive constrained firm is instead defined by the productivity level at which the constrained markdown reaches its unconstrained value:

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

Hence, firms with $z < z^{ds}(\underline{w})$ are demand-constrained, firms with $z \in [z^{ds}(\underline{w}), z^{su}(\underline{w})]$ are supply-constrained, and firms with $z > z^{su}(\underline{w})$ are unconstrained. In the notation of the main text, $\hat{z}(\underline{w}) = z^{su}(\underline{w})$.

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:

$$\ell(z) = z^{\epsilon-1}(\mu\underline{w})^{-\epsilon}, \qquad (26)$$
$$p(z) = (z\ell(z))^{-\frac{1}{\epsilon}}.$$

Demand-constrained firms can either increase or decrease their employment relative 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:

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

The cutoff satisfies $z^c(\underline{w}) = z^{ds}(\underline{w})\psi^{-\frac{\eta}{\epsilon-1}} < z^{ds}(\underline{w})$, and therefore lies in the demand-constrained region. The maximized profit of a demand-constrained firm is $\pi_D^\star(z;\underline{w},\epsilon,\kappa) = (\mu-1)\underline{w}\,\ell_D(z) - \kappa$, where $\ell_D(z) = z^{\epsilon-1}(\mu\underline{w})^{-\epsilon}$, which is strictly increasing in productivity. Figure 1 in the main text considers a reform such that $\pi_D^\star(z^c(\underline{w});\underline{w},\epsilon,\kappa) = 0$. It follows that all demand-constrained firms that would have reduced their employment relative to the baseline equilibrium exit, whereas those that weakly expand survive.

Lemma A.1 (Firm-level output response to the number of active firms)
Let $\mathscr{N} \subset \mathbb{N}$ and $\mathscr{W} \subset \mathbb{R}_+$ denote the admissible numbers of active firms and MW levels, and let $\underline{z}(\underline{w})$ be the least productive active firm. Suppose that $\epsilon(N)$ is strictly increasing in N. If

$$\underline{z}(\underline{w}) > \max\left\{\frac{\epsilon(N)}{\epsilon(N)-1}\exp\left\{-\frac{1}{\epsilon(N)-1}\right\}\underline{w},\ \left(\frac{\epsilon(N)}{\epsilon(N)-1}\frac{1+\eta}{\eta}\right)^{\frac{\eta}{1+\eta}}\exp\left\{-\frac{\epsilon(N)+\eta}{(1+\eta)(\epsilon(N)-1)}\right\}\right\}$$

then every surviving firm weakly reduces its output following a reduction in the number of active firms. Formally, for every $N', N \in \mathscr{N}$ with $N' < N$ and every fixed $\underline{w} \in \mathscr{W}$, each firm that is active under both allocations satisfies

$$y(z;N',\underline{w}) \le y(z;N,\underline{w}).$$

Proof. Let $\mu(\epsilon) \equiv \frac{\epsilon}{\epsilon-1}$, and $\psi \equiv \frac{1+\eta}{\eta}$. Unconstrained employment is $\ell_U(z;\epsilon) = z^{\eta\frac{\epsilon-1}{\epsilon+\eta}}\left[\mu(\epsilon)\psi\right]^{-\frac{\epsilon\eta}{\epsilon+\eta}}$. Differentiating with respect to ϵ gives

$$\frac{\partial \log \ell_U(z;\epsilon)}{\partial\epsilon} = \frac{\eta(1+\eta)}{(\epsilon+\eta)^2}\log z - \frac{\eta^2}{(\epsilon+\eta)^2}\log\left[\mu(\epsilon)\psi\right] + \frac{\eta}{(\epsilon+\eta)(\epsilon-1)}.$$

The first two terms capture the change in the curvature of demand holding the markup fixed (*elasticity effect*), while the last term captures the decline in the markup as demand becomes more elastic (*markup effect*). Therefore,

$$\frac{\partial \log \ell_U(z;\epsilon)}{\partial\epsilon} > 0 \quad\Longleftrightarrow\quad z > \left[\mu(\epsilon)\psi\right]^{\frac{\eta}{1+\eta}}\exp\left\{-\frac{\epsilon+\eta}{(1+\eta)(\epsilon-1)}\right\}.$$

Demand-constrained employment is $\ell_D(z;\epsilon) = z^{\epsilon-1}\left[\mu(\epsilon)\underline{w}\right]^{-\epsilon}$. Differentiating with

respect to ϵ gives $\frac{\partial \log \ell_D(z;\epsilon)}{\partial \epsilon} = \log\left(\frac{z}{\mu(\epsilon)\underline{w}}\right) + \frac{1}{\epsilon-1}$. Hence,

$$\frac{\partial \log \ell_D(z;\epsilon)}{\partial \epsilon} > 0 \quad \Longleftrightarrow \quad z > \mu(\epsilon) \exp\left\{-\frac{1}{\epsilon-1}\right\} \underline{w}.$$

The two productivity thresholds on the right-hand sides are increasing in ϵ. In fact, their log derivatives are $\frac{1}{\epsilon(\epsilon-1)^2} > 0$ and $\frac{\epsilon+\eta}{(1+\eta)\epsilon(\epsilon-1)^2} > 0$, respectively. The condition in the lemma therefore implies that both $\ell_D(z;\epsilon)$ and $\ell_U(z;\epsilon)$ are strictly increasing in ϵ for every active productivity and throughout the interval between any two admissible demand elasticities.

It remains to account for supply-constrained firms and changes across regimes. Let $\ell_F \equiv \underline{w}^\eta$. Revenue is strictly concave in employment, while the wage bill, $\max\left\{\underline{w}\ell,\ \ell^{1+\frac{1}{\eta}}\right\}$, is convex. The employment policy is therefore unique and satisfies

$$\ell^\star(z;\epsilon) = \begin{cases} \ell_D(z;\epsilon) & \text{if } \ell_D(z;\epsilon) \leq \ell_F, \\ \ell_F & \text{if } \ell_D(z;\epsilon) > \ell_F \text{ and } \ell_U(z;\epsilon) < \ell_F, \\ \ell_U(z;\epsilon) & \text{if } \ell_U(z;\epsilon) \geq \ell_F. \end{cases}$$

These cases are exhaustive. Indeed, at ℓ_F, the marginal cost of labor jumps from $\underline{w}$ to $\psi\underline{w}$. Consequently, $\ell_D \leq \ell_F$ and $\ell_U \geq \ell_F$ cannot hold simultaneously.

Both interior policies are strictly increasing in ϵ, whereas the kink policy ℓ_F is independent of ϵ. The complete policy is continuous when either interior policy reaches ℓ_F. It follows that $\ell^\star(z;\epsilon)$ is weakly increasing in ϵ, including across regime changes.

Finally, $N' < N$ implies $\epsilon(N') < \epsilon(N)$. Holding $\underline{w}$ fixed, every firm active under both allocations therefore satisfies

$$\ell^\star(z;\epsilon(N')) \leq \ell^\star(z;\epsilon(N)).$$

Since $y = z\ell$ and productivity is fixed, the same inequality holds for firm-level output. ■

B Quantitative Model

B.1 Derivation of Stationary Distributions

Let $G_{a_i}(w)$ be the cumulative distribution function of employed workers of skill a who earn wage rate less than or equal to 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:

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

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:

$$\dot{u}_{a_i} = \delta_{a_i}e_{a_i} - \lambda(\theta_{a_i})u_{a_i}, \quad \dot{e}_{a_i} = \lambda(\theta_{a_i})u_{a_i} - \delta_{a_i}e_{a_i}.$$

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 $\dot{G}_{a_i}(w) = \dot{u}_{a_i} = \dot{e}_{a_i} = 0$):

$$\begin{aligned} G_{a_i}(w) &= \frac{\lambda(\theta_{a_i})F_{a_i}(w)}{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})(1 - F_{a_i}(w))}\frac{u_{a_i}}{e_{a_i}}, \\ u_{a_i} &= \frac{\delta_{a_i}}{\delta_{a_i} + \lambda(\theta_{a_i})}. \end{aligned} \tag{27}$$

Equation (27) 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 (27), we get

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

From the previous expression, it is clear that the G_{a_i} first-order stochastically dominates 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:

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

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 at 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 endpoints of the wage offer distribution.

B.2 Derivation of Labor Supply and Product Demand Curves

Labor supply curve. The evolution of employment $\ell_{a_i}(w, v)$ of a firm hiring from market a, posting a wage piece rate w and vacancies v 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 inflow and the second term represents the outflow of workers. By evaluating the equation in stationary equilibrium, i.e., setting $\dot{\ell} = 0$, and using the fact that $q(\theta_{a_i}) = \frac{\lambda(\theta_{a_i})}{\theta_{a_i}} = \lambda(\theta_{a_i}) \frac{S_{a_i}}{V_{a_i}}$, stationary employment reads:

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

that is the labor supply curve shown in the firms' problem in the main text. Substituting $G_{a_i}(w)$ away from (27) yields:

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

where the second equality uses the stationarity condition for u_{a_i} and the definition of V_{a_i}.

Product demand curve. The Lagrangian associated to the profit maximization problem of a final good producer (4) is the following:

$$\mathcal{L} = P \prod_{i=1}^{I} \left(K_i^{-\frac{1}{\rho}} \int_0^1 \int_0^{K_{a_i}} Y_{k(a_i)}^{\frac{\rho-1}{\rho}} \, dk(a_i) \, da_i \right)^{\alpha_i \frac{\rho}{\rho-1}} - \sum_{i=1}^{I} \int_0^1 \int_0^{K_{a_i}} \sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)} y_{jk(a_i)} \, dk(a_i) \, da_i.$$

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 (11):

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

Similarly, the FOCs with respect to (fictitious) sectoral and industry good aggregators read: $P_{k(a_i)} = P\left(\frac{Y_{k(a_i)}}{Y_i/K_i}\right)^{-\frac{1}{\rho}}\left(\frac{Y_i}{\alpha_i Y}\right)^{-1}$, $P_i = P\left(\frac{Y_i}{\alpha_i Y}\right)^{-1}$.

The price indices at different levels of aggregation are pinned down by forward and backward adding-up constraints, requiring sectoral revenues to equal the expected sum of firm-level revenues within that sector, and sectoral revenues to be consistent with the final good producer's FOC for the sectoral output aggregator:

$$P_{k(a_i)} = \left(\sum_{j=1}^{N_{k(a_i)}} p_{jk(a_i)}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}, \; P_i = K_i^{-\frac{1}{1-\rho}}\left(\int_0^1\int_0^{K_{a_i}} P_{k(a_i)}^{1-\rho}dk(a_i)da_i\right)^{\frac{1}{1-\rho}}, \; P = \left(\prod_{i=1}^{I}\left(\frac{\alpha_i}{P_i}\right)^{\alpha_i}\right)^{-1}.$$

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

Substituting $\frac{p_{jk(a_i)}}{P_{k(a_i)}} = \left(\frac{y_{jk(a_i)}}{Y_{k(a_i)}}\right)^{-\frac{1}{\sigma}}$, $\frac{P_{k(a_i)}}{P_i} = \left(\frac{Y_{k(a_i)}}{Y_i/K_i}\right)^{-\frac{1}{\rho}}$, $\frac{P_i}{P} = \left(\frac{Y_i}{\alpha_i Y}\right)^{-1}$ into (11) yields the (direct) product demand curve reported in the main text (13).

B.3 Derivation of Reservation Wage

Employed workers find it optimal to accept any job offer with a piece rate exceeding that of their current job, i.e., $R^e_{a_i}(w) = w$. By setting $W_{a_i}(R^u_{a_i}) = U_{a_i}$ and solving for $R^u_{a_i}$, we can characterize the reservation wage for the unemployed workers as follows:

$$\begin{aligned} aR^u_{a_i} + \Pi(aR^u_{a_i}) + \mathcal{O} + T &= \left[(ab_{a_i})^{1-\vartheta} + (1-\vartheta)(1-s_{a_i})\lambda(\theta_{a_i})\int_{R^u_{a_i}}^{\bar{w}_{a_i}}\left(W_{a_i}(w') - U_{a_i}\right)dF_{a_i}(w')\right]^{\frac{1}{1-\vartheta}} \\ &= \left[(ab_{a_i})^{1-\vartheta} + (1-\vartheta)(1-s_{a_i})\lambda(\theta_{a_i})\int_{R^u_{a_i}}^{\bar{w}_{a_i}}\frac{1-F_{a_i}(w)}{r+\delta_{a_i}+s_{a_i}\lambda(\theta_{a_i})(1-F_{a_i}(w))}\right. \\ &\quad \left.\frac{a\left(1+\Pi'(aw)\right)}{(aw+\Pi(aw)+\mathcal{O}+T)^{\vartheta}}dw\right]^{\frac{1}{1-\vartheta}}, \quad \vartheta \neq 1, \end{aligned}$$

$$\begin{aligned} aR^u_{a_i} + \Pi(aR^u_{a_i}) + \mathcal{O} + T = \exp\Bigg\{ &\log(ab_{a_i}) + (1-s_{a_i})\lambda(\theta_{a_i})\int_{R^u_{a_i}}^{\bar{w}_{a_i}}\frac{1-F_{a_i}(w)}{r+\delta_{a_i}+s_{a_i}\lambda(\theta_{a_i})(1-F_{a_i}(w))} \\ &\frac{a\left(1+\Pi'(aw)\right)}{aw+\Pi(aw)+\mathcal{O}+T}dw\Bigg\}, \quad \vartheta = 1. \end{aligned} \tag{28}$$

Since $\Pi' \geq 0$, the left-hand side 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.

B.4 Derivation of Firms' Policy Functions

We derive the firm's policy functions in three steps. First, we characterize the labor market policies as a function of the expected MRPL. Second, we characterize the product market policies as a function of the MCL. Third, we close the system by equating the firm's expected MRPL to its MCL. Finally, we discipline the set of active firms in equilibrium.

Labor market policies as functions of expected MRPL. Fix a labor market a_i. Let $\tilde{z}$ denote a firm's expected MRPL, expressed as piece rate. Let $\Phi_{a_i}(\tilde{z})$ denote the expected MRPL distribution among potential firms sourcing from labor market a_i, with density $\varphi_{a_i}(\tilde{z})$ (under the convention that inactive firms have zero expected MRPL). We focus on rank-preserving equilibria in which wages are increasing in expected MRPL. This allows us to express the wage offer distribution $F_{a_i}(w)$ in terms of vacancy policies as follows:

$$F_{a_i}\left(w_{a_i}(\tilde{z})\right) = \frac{M_{a_i}}{V_{a_i}} \int_{\underline{z}_{a_i}}^{\tilde{z}} v(z')\, d\Phi_{a_i}(z'), \tag{29}$$

where V_{a_i}/M_{a_i} denotes average vacancies per potential firm in labor market a_i. 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 (29) provides a key to express the first-order conditions 14-15 as a system of differential equations pinning down the wage policy function.[38]

Upon taking the derivative of Equation (29) with respect to $\tilde{z}$, we recover the mapping between the density of the wage offer distribution and that of the expected MRPL distribution at each wage level:

$$f_{a_i}(w_{a_i}(\tilde{z}))w'_{a_i}(\tilde{z}) = \frac{M_{a_i}}{V_{a_i}} v_{a_i}(\tilde{z})\varphi_{a_i}(\tilde{z}). \tag{30}$$

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 expected 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 (30) delivers the mass of vacancies per expected MRPL consistent with the equilibrium wage offer distribution:

$$\mathcal{H}'_{a_i}(\tilde{z}) = \frac{M_{a_i}}{V_{a_i}} v_{a_i}(\tilde{z})\varphi_{a_i}(\tilde{z}) \iff v_{a_i}(\tilde{z}) = \frac{V_{a_i}}{M_{a_i}} \frac{\mathcal{H}'_{a_i}(\tilde{z})}{\varphi_{a_i}(\tilde{z})}. \tag{31}$$

Replacing Equations (30) and (31) into the FOCs of the firm's problem yields the following

[38] Unlike in Engbom and Moser (2022), the expected 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 expected MRPL.

system of differential equations:

$$(v): \quad \mathcal{H}'_{a_i}(\tilde{z}) = \frac{M_{a_i}}{V_{a_i}} \varphi_{a_i}(\tilde{z}) \left(\frac{\tilde{z} - w_{a_i}(\tilde{z})}{\bar{c}_{a_i}} \frac{\lambda(\theta_{a_i}) u_{a_i}(\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i}))/V_{a_i}}{(\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})\left[1 - \mathcal{H}_{a_i}(\tilde{z})\right])^{2+\phi}} \right)^{\frac{1}{\zeta}}, \tag{32}$$

$$(w): \quad w'_{a_i}(\tilde{z}) = (\tilde{z} - w_{a_i}(\tilde{z})) \frac{2\nu s_{a_i}\lambda(\theta_{a_i})\mathcal{H}'_{a_i}(\tilde{z})}{\delta_{a_i} + s_{a_i}\lambda(\theta_{a_i})\left[1 - \mathcal{H}_{a_i}(\tilde{z})\right]}. \tag{33}$$

This system of differential equations in the two functions $w_{a_i}(\tilde{z})$ and $\mathcal{H}_{a_i}(\tilde{z})$ yields the optimal wage and vacancy policies for each expected MRPL, provided the necessary boundary conditions:

$$w_{a_i}(\underline{\tilde{z}}) = \underline{w}_{a_i}, \quad \mathcal{H}_{a_i}(\underline{\tilde{z}}) = 0,$$

where $\underline{w}_{a_i} = \max\left\{R^u_{a_i}, \frac{w}{a}\right\}$ is the effective MW – that is, either the reservation wage or the mandated MW. The first condition states that the lowest-expected-MRPL firms offer the lowest possible piece rate. The second condition states that the CDF of offered wages equals 0 at the lowest-expected-MRPL firms. Finally, 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 expected-MRPL level among active firms. Equivalently, the total mass of vacancies posted by firms needs to equal the aggregate V_{a_i}.

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 expected MRPL $\tilde{z}$ in the job finding process. Solving the differential Equation (33) allows us to compute the wage policy function analytically as follows:

$$w_{a_i}(\tilde{z}) = (1 - \mathcal{P}_{a_i}(\tilde{z}))\, \underline{w}_{a_i} + \mathcal{P}_{a_i}(\tilde{z}) \mathbb{E}_{\max \text{MRPL}^e | \text{MRPL}^e < \tilde{z}} \left[\text{MRPL}^e\right], \tag{34}$$

where $\mathcal{P}_{a_i}(\tilde{z}) \equiv 1 - e^{-\int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} x_{a_i}(t)dt}$ and the expectation is taken with respect to the density of the highest-expected-MRPL firm with lower expected MRPL than $\tilde{z}$, conditional on a searching worker meeting any of them, that is, $\frac{e^{-\int_s^{\tilde{z}} x_{a_i}(t)dt} x_{a_i}(s)}{1-e^{-\int_{\underline{\tilde{z}}_{a_i}}^{\tilde{z}} x_{a_i}(t)dt}}$. Equation (34) 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 expected MRPL lower than $\tilde{z}$, in which case the worker's outside option when meeting a firm with expected MRPL $\tilde{z}$ boils down to her effective MW. With complementary probability, the worker meets at least one firm with lower expected MRPL than $\tilde{z}$, in which case the worker's outside option when meeting a firm with expected MRPL $\tilde{z}$ equals the maximum wage the best of such firms can afford to pay, that is, the highest expected MRPL among firms with lower expected MRPL than $\tilde{z}$. Hence, the wage policy function equals the expected outside option of a searching worker.

Employment follows from the stationary labor supply curve (10):

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

The markdown policy is given by the ratio between expected MRPL and wage: $\psi_{a_i}(\tilde{z}) = \frac{\tilde{z}}{w_{a_i}(\tilde{z})}$. The equilibrium markdown function reported in the main text (20) can be derived from the firm's marginal cost function. Omitting the a_i subscripts, it holds that:

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

where we made use of the firm's indifference condition between vacancies and wages, the total differential of the employment constraint, and the linearity of employment in vacancies:

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

In turn, the elasticity of labor supply of a firm with expected MRPL $\tilde{z}$ in a given labor market equals:

$$\eta(\tilde{z}) \equiv \epsilon_{\ell,w}(\tilde{z}) = \frac{2f(w(\tilde{z}))w(\tilde{z})}{\chi(\theta) + [1 - F(w(\tilde{z}))]} = \frac{1}{\nu}\frac{w(\tilde{z})\ell(\tilde{z})}{(1+\zeta)c(\tilde{z})},$$

where the two alternative formulations follow from Equations (14) and (15), respectively. Owing to the two alternative formulations of the equilibrium elasticity of labor supply, the equilibrium markdown itself can be expressed in two different ways:

$$\psi(\tilde{z}) = 1 + \frac{\chi(\theta) + [1 - F\left(w(\tilde{z})\right)]}{2\nu f\left(w(\tilde{z})\right)w(\tilde{z})} = 1 + \frac{(1+\zeta)c(\tilde{z})}{w(\tilde{z})\ell(\tilde{z})}. \tag{35}$$

The second formulation shows that equilibrium markdown rates are increasing in the ratio between hiring costs and wage bill, reflecting the key role of search frictions.

Product-market policies as functions of MCL. Fix a product market $k(a_i)$. Let $\mathcal{C}$ denote a firm's MCL, expressed as piece rate, under the convention that inactive firms have zero MCL. Let $\mathcal{C} \sim \Phi^{c}_{k(a_i)}$ denote the MCL distribution among potential firms sourcing from labor market a_i. The firm's product-market policies are obtained from the Bayes-Cournot FOC.

Maximizing (8) with respect to ℓ yields:

$$\mathbb{E}_{B_{a_i}}\left[\left(1-\epsilon^{-1}_{k(a_i)}(\mathcal{C},z;B_{a_i})\right)p_{k(a_i)}(\mathcal{C},z;B_{a_i})\right]=\frac{\mathcal{C}}{z}.$$

The state-contingent price is determined by the product demand curve (11):

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

Given optimal employment, the state-contingent market share equals $\omega_{k(a_i)}(\mathcal{C},z;B_{a_i})=\left(\frac{az\ell_{a_i}(\mathcal{C},z)}{Y_{k(a_i)}\left(az\ell_{a_i}(\mathcal{C},z);B_{a_i}\right)}\right)^{\frac{\sigma-1}{\sigma}}$. In turn, the Bayes-Cournot markup reads:

$$\mu_{k(a_i)}(\mathcal{C},z)=\mathbb{E}^{p}_{B_{a_i}}\left[\left(\frac{\sigma}{(\sigma-1)\left[1-\frac{\sigma/\rho-1}{\sigma-1}\omega_{k(a_i)}(\mathcal{C},z;B_{a_i})\right]}\right)^{-1}\right]^{-1}.$$

Closing the firm policy system. For each active productivity rank z, product-market behavior yields the expected MRPL $\widetilde{z}_{a_i}(\mathcal{C},z)\equiv\frac{\mathbb{E}\left[p_{k(a_i)}\left(\mathcal{C}_{a_i}(z),z;B_{a_i}\right)\right]}{\mu_{k(a_i)}\left(\mathcal{C}_{a_i}(z),z\right)}z$ and the labor-market block yields the MCL: $\mathcal{C}_{a_i}(\widetilde{z})\equiv\psi_{a_i}(\widetilde{z})w_{a_i}(\widetilde{z})$. Equilibrium requires

$$\widetilde{z}_{a_i}(\mathcal{C}_{a_i},z)=\mathcal{C}_{a_i}\left(\widetilde{z}_{a_i}(z)\right),\qquad z\geq\underline{z}_{a_i}.$$

Equivalently, under rank-preservation, the distribution of expected MRPL induced by product-market policies must coincide with the distribution of MCL generated by the labor-market policies, under the common active productivity distribution.

Selection of active firms. For each labor market a_i, let Γ_{a_i} denote the unconditional productivity distribution of potential firms, and let $\mathcal{N}_{a_i}$ denote the distribution of the number of potential firms per product market. Selection changes neither primitive. Instead, let $\varpi_{a_i}(z)\in[0,1]$ denote the probability that a potential firm of productivity z operates.

From the perspective of a randomly selected potential firm, the number of potential firms in its product market follows the size-biased law $\mathcal{N}^F_{a_i}(n)=\frac{n\mathcal{N}_{a_i}(n)}{\mathbb{E}_{\mathcal{N}_{a_i}}[\bar{N}_{a_i}]}$. This is also the market-size law faced by a randomly selected active firm, because operation depends on productivity but not on the realized market size. Conditional on $\bar{N}_{a_i}=n$, the $n-1$ rival productivities are independent draws from Γ_{a_i}. For each rival position, define an activity dummy $D_m\mid Z_m\sim\text{Bernoulli}(\varpi_{a_i}(Z_m))$. The realized CES-output contribution of the firm's competitors is then $B_{a_i}(\varpi_{a_i})=\sum_{m=1}^{\bar{N}_{a_i}-1}D_m\,y_{a_i}(Z_m)^{\frac{\sigma-1}{\sigma}}\sim\mathcal{B}_{a_i}(\varpi_{a_i})$. An inactive rival therefore contributes zero output while retaining its position in the potential market structure. Selection changes $\mathcal{B}_{a_i}$ through the operating probabilities and output policies, but leaves Γ_{a_i} and $\mathcal{N}_{a_i}$ unchanged.

For a candidate operating rule ϖ_{a_i}, define the expected operating profits of a potential firm of productivity z by $\pi^o_{a_i}(z;\varpi_{a_i}) \equiv \sup_{\substack{w\geq \underline{w}_{a_i} \\ v>0}} \left\{ \mathbb{E}_{B_{a_i}(\varpi_{a_i})} \left[p_{k(a_i)}(y;B_{a_i})y \right] - aw\ell_{a_i}(w,v) - ac_{a_i}(w,v) - a\kappa_{a_i}\bar{W} \right\}$. Since a potential firm may choose to remain inactive, it operates whenever its optimized operating profits are positive and remains inactive whenever they are negative (it may mix when they are exactly zero).

The ordering of operating profits follows from the firm's problem. Let $m_{a_i}(y;B) \equiv \partial[p_{k(a_i)}(y;B)y]/\partial y > 0$ denote the marginal revenue product (MRP). At every interior operating choice, the envelope theorem gives

$$\frac{\partial \pi^o_{a_i}(z;\varpi_{a_i})}{\partial z} = a\ell_{a_i}(z)\, \mathbb{E}_{B_{a_i}(\varpi_{a_i})} \left[m_{a_i}(y(z);B_{a_i}) \right] > 0.$$

Thus, for any candidate rival-output law, operating profits are strictly increasing in productivity. Any equilibrium active set must consequently be an upper interval of the productivity distribution.

An equilibrium operating rule $\varpi^\star_{a_i}$ is a fixed point: it generates the rival-output law entering operating profits and, for every z, satisfies

$$\varpi^\star_{a_i}(z) \in \begin{cases} \{0\}, & \pi^o_{a_i}(z;\varpi^\star_{a_i}) < 0, \\ [0,1], & \pi^o_{a_i}(z;\varpi^\star_{a_i}) = 0, \\ \{1\}, & \pi^o_{a_i}(z;\varpi^\star_{a_i}) > 0. \end{cases} = \begin{cases} 0, & z < \underline{z}^\star_{a_i}, \\ \vartheta^\star_{a_i}, & z = \underline{z}^\star_{a_i}, \\ 1, & z > \underline{z}^\star_{a_i}, \end{cases}$$

where the equality follows from strict profit ordering, which implies that there is a cutoff $\underline{z}^\star_{a_i}$ and a boundary operating share $\vartheta^\star_{a_i} \in [0,1]$. If the cutoff is interior, continuity implies $\pi^o_{a_i}\left(\underline{z}^\star_{a_i};\varpi^\star_{a_i}\right) = 0$. [39]

B.5 Additional Equilibrium Conditions

We now complement the equilibrium definition reported in the main text with the remaining equilibrium conditions:

- The aggregate measure of potential firms M satisfies $M = \sum_{i=1}^{I} L_i \int M_{a_i} d\Xi_i(a)$;
- Consistency between the measure of firms in the labor and the product markets requires $L_i M_{a_i} d\Xi_i(a) = K_{a_i} \mathbb{E}_{\mathcal{N}_{a_i}}[\bar{N}_{a_i}] da_i$;
- The aggregate unemployment rate satisfies $u = \sum_{i=1}^{I} L_i \int u_{a_i} d\Xi_i(a)$;
- The economy-wide average worker wage satisfies $\bar{W} = \frac{\sum_{i=1}^{I} L_i \int M_{a_i} \int_{\underline{z}_{a_i}}^{\infty} a w_{a_i}(z) \ell_{a_i}(z)\, d\Gamma_{a_i}(z)\, d\Xi_i(a)}{1-u}$;

[39] When Γ_{a_i} is atomless, the cutoff type has zero mass and the value of $\vartheta^\star_{a_i}$ is immaterial. If the productivity distribution has an atom at the cutoff, $\vartheta^\star_{a_i}$ is the fraction of potential firms at that productivity that operate. It changes neither the unconditional productivity distribution nor the number of potential-firm positions.

- Payments for overhead services rebated to employed workers satisfy $(1-u)\mathcal{O} = \sum_{i=1}^{I} L_i \int M_{a_i} \int_{\underline{z}_{a_i}}^{\infty} a\kappa_{a_i}\bar{W}\, d\Gamma_{a_i}(z)\, d\Xi_i(a)$.
- Aggregate profits satisfy $\bar{\Pi} = \sum_i^I L_i \int M_{a_i} \int_{\underline{z}_{a_i}}^{\infty} \pi_{a_i}(z) d\Gamma_{a_i}(z) d\Xi_i(a)$;
- Profit rebates to employed workers satisfy $\sum_{i=1}^{I} L_i \int M_{a_i} \int_{\underline{z}_{a_i}}^{\infty} \ell_{a_i}(z)\Pi\big(aw_{a_i}(z)\big)\, d\Gamma_{a_i}(z)\, d\Xi_i(a) = (1-\tau)\bar{\Pi}$;
- The government budget constraint satisfies $T(1-u) + \sum_{i=1}^{I} L_i \int u_{a_i} a b_{a_i}\, d\Xi_i(a) = \tau\bar{\Pi}$;

B.6 Existence and Uniqueness of Stationary Equilibrium

In this section, we prove Proposition 1, claiming existence and uniqueness of our rank-preserving stationary equilibrium with continuous wage offer distributions and an ordered exit rule. Rank preservation is defined in terms of wage sorting: within each labor market, the equilibrium wage schedule is strictly increasing in firm productivity. Since optimal wages are strictly increasing in expected MRPL, it is sufficient to establish that expected MRPL is strictly increasing in productivity. We operationalize this requirement through regularity conditions on the shape of the productivity distributions, as well as parametric restrictions on the hiring-cost and demand functions. Throughout, we assume fixed effective MWs in each labor market. We interpret them as labor-market-specific MWs, consistently with the Italian contractual-wage-floor system.

The proof proceeds in two stages. First, fixing a candidate vector of industry outputs $\mathbf{Y}$, we establish existence and uniqueness of a rank-preserving stationary equilibrium in every labor market, conditional on its product-demand shifter. Second, we show that the resulting labor-market equilibria generate a unique aggregate-output fixed point. To simplify notation, we omit the a_i and $k(a_i)$ subscripts when there is no ambiguity and absorb the fixed skill shifter into productivity.

Environment. Fix $\mathbf{Y}$ in a compact positive rectangle $\mathscr{Y}$, one labor market, and a meeting rate λ in a compact interval $\Lambda \equiv [\lambda_L, \lambda_H] \subset (0, \bar{\lambda})$, where $\bar{\lambda} \equiv \chi(1-\beta)^{-1/\iota}$ is the upper bound of the meeting technology. The primitive parameters satisfy $1 < \rho < \sigma < \infty$, $\chi > 0$, $\iota > 0$, $\beta \in (0,1)$, $\delta > 0$, $s \in (0,1]$, $\bar{c} > 0$, $\phi \geq 0$, and $\zeta > 0$. The support of the market-size distribution $\mathcal{N}$ is finite.

We require the potential productivity distribution to have compact positive support with a sufficiently large lower bound for labor surplus to be positive over the relevant conditional-equilibrium domain. To isolate the role of the scale of the productivity distribution, as opposed to its shape, we express the productivity distribution in a scale representation.[40] Potential

[40] For any fixed distribution with compact positive support, the scale representation is only a reparameterization. Holding g^p fixed while varying Z_0 defines the scale family used to isolate the role of productivity scale. Compact support and the regularity imposed on g^p are substantive assumptions.

productivity has density $\gamma_{Z_0}(z) = Z_0^{-1} g^p(z/Z_0)$ on $[Z_0\underline{q}, Z_0\bar{q}]$, where $0 < \underline{q} < \bar{q} < \infty$ and g^p is strictly positive and continuously differentiable.

The ordered activity rule represents the active set by a lower productivity cutoff $Z = Z_0\mathcal{Z}$. Let $A(\mathcal{Z}) \equiv \int_{\mathcal{Z}}^{\bar{q}} g^p(q)dq$ be the active share of potential firms, $M^A \equiv MA(\mathcal{Z})$ be the measure of active firms, $t \equiv z/Z$ be the normalized productivity relative to the cutoff (*rank*), and $\eta_{\mathcal{Z}} \equiv \bar{q}/\mathcal{Z}$ be the upper bound of the rank support. Conditional on activity, normalized productivity has density $g_{\mathcal{Z}}(t) \equiv \mathcal{Z} g^p(\mathcal{Z}t)/A(\mathcal{Z})$ on $[1, \eta_{\mathcal{Z}}]$. We restrict attention to cutoffs satisfying $A(\mathcal{Z}) \geq \underline{A} > 0$. This restriction rules out sequences of candidate equilibria in which the active economy vanishes.

For a candidate vacancy policy, define the normalized vacancy-density tilt by $d(t) \equiv M^A v(Zt)/V$, so that $\int g_{\mathcal{Z}}(t)d(t)dt = 1$. Fix an envelope constant $\mathfrak{L} > 1$, and let $\mathscr{D}_{\mathcal{Z}}$ be the closed subset of continuous tilts satisfying this normalization and $\mathfrak{L}^{-1} \leq d \leq \mathfrak{L}$. Let $\xi(\lambda) \equiv s\lambda/(\delta + s\lambda)$ denote a search wedge, $\mathcal{H}_d(t) \equiv \int_1^t g_{\mathcal{Z}}(q)d(q)dq$ the equilibrium wage offer distribution, and $\mathcal{D}_d(t) \equiv 1 - \xi(\lambda)\mathcal{H}_d(t)$. Since $\mathcal{H}_d \in [0,1]$, $\mathcal{D}_d \geq \mathcal{D}_0 \equiv 1 - \bar{\xi} > 0$, where $\bar{\xi} \equiv s\lambda_H/(\delta + s\lambda_H)$. Stationary employment and output are $\ell_{\lambda,d}(t) = \Lambda_\ell(\lambda)d(t)\mathcal{D}_d(t)^{-2}$ and $y_{\lambda,d}(t) = Z\Lambda_\ell(\lambda)q_d(t)$, where $\Lambda_\ell(\lambda) \equiv \delta\lambda/[M^A(\delta + \lambda)(\delta + s\lambda)]$ and $q_d(t) \equiv td(t)\mathcal{D}_d(t)^{-2}$.

The product-market expectations are those of the Bayes–Cournot problem. A firm draws the number of potential firms from the size-biased law $\mathcal{N}^F$, while each rival's productivity is drawn independently from the unconditional distribution Γ. A draw below Z is inactive and contributes zero output. Thus, defining $q_d^A(z) = q_d(z/Z)$ for $z \geq Z$ and zero otherwise, the scale-free CES contribution of the firm's rivals is $\widehat{B}_{\mathcal{Z},d} \equiv \sum_{m=1}^{\bar{N}-1} q_d^A(Z_m)^{(\sigma-1)/\sigma}$, with $\bar{N} \sim \mathcal{N}^F$ and $Z_m \sim \Gamma$.

Let $a_\sigma \equiv (\sigma - 1)/\sigma$, $b_{\rho,\sigma} \equiv 1/\rho - 1/\sigma$, and $\tau_{\rho,\sigma} \equiv (\sigma/\rho - 1)/(\sigma - 1)$. For output y and a realization B, denote MRP by $m(y; B) \equiv \partial[p(y; B)y]/\partial y$. Expected MRPL at rank t can then be written as

$$\widetilde{z}_{\lambda,\mathcal{Z},d}(t) = a_\sigma \mathcal{P}_i(\mathbf{Y})\Lambda_\ell(\lambda)^{-1/\rho} Z^{1-1/\rho} t\, \mathbb{E}\left[\widehat{m}\left(q_d(t); \widehat{B}_{\mathcal{Z},d}\right)\right], \tag{36}$$

where $\mathcal{P}_i(\mathbf{Y}) \equiv (Y_i/K_i)^{1/\rho}(Y_i/(\alpha_i Y))^{-1}$ is the aggregate demand shifter, $\omega(q_d(t); \widehat{B}_{\mathcal{Z},d}) \equiv q^{a_\sigma}/(\widehat{B} + q^{a_\sigma})$ is the realized market share, and $\widehat{m}(q; \widehat{B}) = q^{-1/\sigma}(\widehat{B} + q^{a_\sigma})^{-b_{\rho,\sigma}/a_\sigma}[1 - \tau_{\rho,\sigma}\omega(q_d(t); \widehat{B}_{\mathcal{Z},d})]$ is the firm-specific MRP component. Because $1 < \rho < \sigma$, realized MRP is strictly positive. For a candidate cutoff, let $\pi^{\mathrm{op}}(z; \mathcal{Z}, \lambda, \mathbf{Y})$ denote the firm's maximized operating profit, net of its overhead cost, in the associated conditional environment, and let $\pi^b(\mathcal{Z}; \lambda, \mathbf{Y}) \equiv \pi^{\mathrm{op}}(Z; \mathcal{Z}, \lambda, \mathbf{Y})$ denote the profit of the boundary type.

Finally, the ordered exit rule selects full activity when the lowest potential type earns nonnegative profits. Otherwise it selects the lowest cutoff at which the boundary type breaks even. The single-crossing restriction stated below guarantees that this selection is unique.

Conditional labor-market equilibrium and rank preservation. To state a primitive sufficient condition, let $\Gamma^p \equiv \bar{q}\|g^p\|_\infty/\underline{A}$ be the upper bound on the local mass of firms at any normalized productivity rank, and define $\nu_0 \equiv 1 + \phi/2$. For a realized revenue share $\omega \in [0,1]$,

define the own-output-elasticity of the realized MRP as $\varepsilon_m(\omega) \equiv -\partial \log m(y;B)/\partial \log y = 1/\sigma + b_{\rho,\sigma}\omega + a_\sigma \tau_{\rho,\sigma}\omega(1-\omega)/(1-\tau_{\rho,\sigma}\omega)$ and $\bar{\varepsilon}_m(\rho,\sigma) \equiv \max_{\omega\in[0,1]} \varepsilon_m(\omega)$. This maximum depends only on the two demand elasticities (σ and ρ).

Let $\mathfrak{p}_{-\zeta}$ collect all primitives except vacancy-cost curvature, and let x collect the common conditional-equilibrium objects—the vacancy tilt, wage policy, and rival-output law. On the compact domains just defined, bounds on derivatives of the primitive revenue, matching, hiring-cost, and productivity-density functions generate finite constants $C_z(\mathfrak{p}_{-\zeta})$ and $C_w(\mathfrak{p}_{-\zeta})$: the former bounds the response of expected MRPL to a change in the vacancy-density tilt, and the latter is the corresponding Gronwall bound for the wage equation. The same primitive bounds and the inverse of the contraction margin generate constants $\underline{\pi}_z > 0$, $C_{\pi x} < \infty$, and $C_{x\mathcal{Z}} < \infty$. $\underline{\pi}_z$ is a lower bound on the direct productivity derivative of maximized operating profit, while $C_{\pi x}C_{x\mathcal{Z}}$ bounds the indirect effect on the boundary firm's profit of changing the activity cutoff through the conditional-equilibrium objects x. These constants are constructed in the proof and do not depend on any equilibrium variable.

Regularity condition 1 (Primitive conditional stability)
The productivity scale satisfies $Z_0 > \underline{Z}_0(\mathfrak{p}_{-Z_0})$, where $\mathfrak{p}_{-Z_0}$ denotes all other primitives and the finite threshold $\underline{Z}_0$ guarantees a uniform surplus margin $\widetilde{z} - w \geq \underline{S} > 0$ on $\mathscr{Y}$, Λ, all admissible cutoffs, and $\mathscr{D}_{\mathcal{Z}}$. The primitive derivative bounds satisfy $\underline{\pi}_z > C_{\pi x}C_{x\mathcal{Z}}$. The overhead-cost parameter satisfies $0 \leq \kappa < \overline{\kappa}(\mathfrak{p}_{-\kappa})$, where the finite primitive threshold $\overline{\kappa}$ is the lowest, across the compact conditional environments, of the gross operating surplus of the boundary type at the largest admissible cutoff divided by its overhead-cost shifter.

Moreover,

$$\bar{\varepsilon}_m(\rho,\sigma)\left(1 + \frac{2\nu_0 \mathfrak{L}\bar{\xi}\Gamma^p}{\mathcal{D}_0}\right) < 1, \qquad \zeta > \zeta_C(\mathfrak{p}_{-\zeta}), \tag{37}$$

where the finite primitive threshold ζ_C is defined in the proof of Lemma A.2.

The scale parameter Z_0 guarantees positive surplus without altering the normalized shape of productivity. $\underline{\pi}_z > C_{\pi x}C_{x\mathcal{Z}}$ is a genuine activity-selection restriction: the direct productivity advantage must dominate the general-equilibrium response induced by removing lower-productivity firms. The first inequality in (37) disciplines the price-size feedback. The maximum MRP elasticity $\bar{\varepsilon}_m$ falls when within- and across-market demand become more elastic. The remaining factor bounds how strongly firm size can vary across productivity ranks: Γ^p controls the largest local mass of productivity ranks, $\bar{\xi}/\mathcal{D}_0$ the amplification generated by worker reallocation, and ν_0 the interaction between hiring costs and the wage-offer distribution. The curvature restriction makes relative vacancy posting sufficiently insensitive to changes in surplus and worker flows.

If the MRP is constant, i.e., $\rho = \sigma = \infty$ and $\bar{\varepsilon}_m = 0$, the rank restriction is then automatic: productivity raises the MRP one for one and firm size cannot feed back into the MRP. The conditional equilibrium can also be obtained by the shooting argument at the end of the proof of Lemma A.2, which works for every $\zeta > 0$. Thus the lower bound ζ_C is not needed in this limit. Positive surplus and the primitive cutoff transversality remain necessary when activity

is endogenous.

Lemma A.2 (Conditional equilibrium and rank preservation)
Under Condition 1, every candidate $\mathbf{Y}$ *and meeting rate* $\lambda \in \Lambda$ *admit a unique conditional labor–product-market equilibrium and a unique ordered active set. Expected MRPL and the wage policy are strictly increasing in productivity.*

Proof. Fix an admissible cutoff $\mathcal{Z}$. For a candidate vacancy-density tilt d, let $w_d(t)$ be the solution of the wage differential equation (33) with the MW as its lower boundary condition. The primitive smoothness assumptions make its right-hand side continuous and locally Lipschitz in the wage. By construction of $\underline{Z}_0$, the associated solution remains at least $\underline{S}$ below expected MRPL throughout the compact active support. Standard continuation of ordinary differential equations therefore gives a unique wage solution on that support.

Define transformed surplus $J_d(t) \equiv [\widetilde{z}_d(t) - w_d(t)]\mathcal{D}_d(t)^{-2\nu}$, where $\nu \equiv 1 + \phi/[2(1+\zeta)]$. On the compact domains, bounds on derivatives of the primitive revenue, matching, hiring-cost, and productivity-density functions generate a finite bound $\Omega_J(\mathfrak{p}_{-\zeta})$ on the oscillation of $\log J$ across ranks for every admissible positive-surplus fixed-point candidate. The vacancy first-order condition (32) is equivalent to

$$(\mathcal{T}d)(t) = \frac{J_d(t)^{1/\zeta}\mathcal{D}_d(t)^{-\phi/(1+\zeta)}}{\int_1^{\eta_{\mathcal{Z}}} g_{\mathcal{Z}}(q)J_d(q)^{1/\zeta}\mathcal{D}_d(q)^{-\phi/(1+\zeta)}dq}. \tag{38}$$

We establish that this map has exactly one fixed point. First, for every perturbation $\dot{d}$, $|D_d\mathcal{H}_d[\dot{d}]|_\infty = |\int_1^t g_{\mathcal{Z}}(q)\dot{d}(q)dq| \le \|\dot{d}\|_\infty$. Hence, $\|D_d \log \mathcal{D}_d[\dot{d}]\|_\infty \le (\bar{\xi}/\mathcal{D}_0)\|\dot{d}\|_\infty$. Let $\overline{\eta} \equiv \overline{q}/\underline{q}$ be a uniform upper bound on the normalized support endpoint, so every support length is at most $\overline{\eta} - 1$. Over the compact admissible domain, define the primitive upper bound of the directional derivative of the expected MRPL function with respect to a perturbation in the vacancy-density tilt as $C_z(\mathfrak{p}_{-\zeta}) \equiv \sup_{\mathcal{Z},\, d\in\mathscr{D}_{\mathcal{Z}},\, \|\dot{d}\|_\infty\le 1} \left\|D_d\widetilde{z}_{\mathcal{Z},d}[\dot{d}]\right\|_\infty$. Finite market-size support, compact productivity support, and the strict lower bound on $1 - \tau_{\rho,\sigma}\omega$ permit differentiation under the expectation in Equation (36). The primitive derivative bounds therefore imply $C_z(\mathfrak{p}_{-\zeta}) < \infty$.

Writing the wage equation as $w' = F(t, w, d)$, define the primitive upper bound of the absolute derivative of the wage gradient with respect to the wage as $L_w \equiv \sup_{\mathcal{Z},d,t,w} |F_w(t, w, d)|$, and the primitive upper bound of the absolute directional derivative of the wage gradient with respect to a perturbation in the vacancy-density tilt as $L_d \equiv \sup_{\mathcal{Z},d,t,w,\, \|\dot{d}\|_\infty\le 1} \left|D_dF(t, w, d)[\dot{d}]\right|$. Both constants are finite by the primitive smoothness and positive-surplus bounds. Gronwall's inequality then gives $\left\|D_dw_{\mathcal{Z},d}[\dot{d}]\right\|_\infty \le C_w(\mathfrak{p}_{-\zeta})\|\dot{d}\|_\infty$, where $C_w(\mathfrak{p}_{-\zeta}) \equiv L_d\left(\exp\{L_w(\overline{\eta}-1)\} - 1\right)/L_w$, with the continuous extension $C_w = L_d(\overline{\eta} - 1)$ when $L_w = 0$.

Let $\underline{S}$ and $\overline{S}$ be the primitive lower and upper bounds on surplus, and let $\mathcal{D}_0 \le \mathcal{D}_d(t) \le \overline{\mathcal{D}}$. Define the constants $C_J \equiv \frac{C_z + C_w}{\underline{S}} + 2\nu_0\frac{\bar{\xi}}{\mathcal{D}_0}$, $\Omega_J \equiv \log\left(\frac{\overline{S}}{\underline{S}}\right) + 2\nu_0\log\left(\frac{\overline{\mathcal{D}}}{\mathcal{D}_0}\right)$. The constant C_J bounds from above the directional derivative of the log-transformed surplus with respect to a perturbation in the vacancy-density tilt: $\left\|D_d\log J_d[\dot{d}]\right\|_\infty \le C_J\|\dot{d}\|_\infty$. The constant Ω_J

bounds from above the oscillation of $\log J$ across ranks for every admissible positive-surplus fixed-point candidate: $\operatorname{osc}(\log J_d) \leq \Omega_J$.

For any positive fixed point, Equation (38) gives $\max d / \min d \leq \exp\{\Omega_J/\zeta + [\phi/(1+\zeta)]|\log \mathcal{D}_0|\}$. The right-hand side is smaller than $\mathfrak{L}$ for all sufficiently large ζ. Because d has $g_{\mathcal{Z}}$-weighted mean one, this implies $\mathfrak{L}^{-1} < d < \mathfrak{L}$; hence every admissible fixed point belongs to $\mathscr{D}_{\mathcal{Z}}$. Moreover, as $\zeta \to \infty$, the two exponents in Equation (38) converge uniformly to zero, so $\mathcal{T}d \to 1$ uniformly on $\mathscr{D}_{\mathcal{Z}}$. Thus $\mathscr{D}_{\mathcal{Z}}$ is mapped into itself for all sufficiently large ζ. Differentiating Equation (38), including its normalizing denominator, gives the following contraction modulus:

$$\|D_d\mathcal{T}[\dot{d}]\|_\infty \leq 2\mathfrak{L}\left[\frac{C_J}{\zeta} + \frac{\phi}{1+\zeta}\frac{\bar{\xi}}{\mathcal{D}_0}\right]\|\dot{d}\|_\infty.$$

Define $\zeta_C(\mathfrak{p}_{-\zeta})$ as any finite value above which every admissible fixed point lies in the envelope $\mathscr{D}_{\mathcal{Z}}$, the envelope is invariant, and the contraction modulus, i.e., the expression multiplying $\|\dot{d}\|_\infty$, is strictly smaller than one. Since $\mathscr{D}_{\mathcal{Z}}$ is a closed subset of the Banach space of continuous functions, Banach's theorem then gives a unique positive continuous fixed point d_λ, and the preceding argument rules out another admissible fixed point outside the envelope.

We next prove rank preservation. At the fixed point, Equation (38) implies $q_\lambda(t) = Q_\lambda^{-1} t J_\lambda(t)^{1/\zeta}\mathcal{D}_\lambda(t)^{-2\nu}$, where Q_λ is common across ranks. Define $r_\lambda(t) \equiv d\log\widetilde{z}_\lambda(t)/d\log t$, $X_\lambda(t) \equiv 2\nu\xi(\lambda)tg_{\mathcal{Z}}(t)d_\lambda(t)/\mathcal{D}_\lambda(t)$, and $\mathfrak{A}_\lambda(t) \equiv \widetilde{z}_\lambda(t)/\{\zeta[\widetilde{z}_\lambda(t) - w_\lambda(t)]\}$. Since $d_\lambda \leq \mathfrak{L}$ and $tg_{\mathcal{Z}}(t) \leq \Gamma^p$, $0 \leq X_\lambda(t) \leq 2\nu_0\mathfrak{L}\bar{\xi}\Gamma^p/\mathcal{D}_0$.

The wage equation and $\mathcal{D}'_\lambda(t) = -\xi g_{\mathcal{Z}}(t)d_\lambda(t)$ imply $dw_\lambda/d\log t = [\widetilde{z}_\lambda - w_\lambda]X_\lambda$. Consequently, $d\log J_\lambda/d\log t = \{\widetilde{z}_\lambda/[\widetilde{z}_\lambda - w_\lambda]\}r_\lambda$, and differentiation of the fixed-point expression for output gives $d\log q_\lambda/d\log t = 1 + X_\lambda + \mathfrak{A}_\lambda r_\lambda$.

For a realization of rival output, let $\omega = q^{a_\sigma}/(\widehat{B} + q^{a_\sigma})$. Direct differentiation of $\widehat{m}$ gives $-\partial\log\widehat{m}/\partial\log q = \varepsilon_m(\omega)$. Differentiating under the expectation is valid by the same compactness bound used above, and therefore the elasticity of expected MRP is the MRP-weighted average $\varepsilon_B(q) = \mathbb{E}[\widehat{m}\varepsilon_m]/\mathbb{E}[\widehat{m}] \in [0, \bar{\varepsilon}_m]$. Equation (36) then gives $r_\lambda = 1 - \varepsilon_B(q_\lambda)d\log q_\lambda/d\log t$. Combining the last two identities yields the key rank equation

$$\frac{d\log\widetilde{z}_\lambda(t)}{d\log t} = \frac{1 - \varepsilon_B(q_\lambda(t))[1 + X_\lambda(t)]}{1 + \mathfrak{A}_\lambda(t)\varepsilon_B(q_\lambda(t))}. \tag{39}$$

The first inequality in Equation (37) makes the numerator strictly positive, while the denominator is strictly positive by the positive-surplus condition . Thus, expected MRPL is strictly increasing in productivity. Since surplus and the vacancy-weighted density are strictly positive, the wage equation implies $w'_\lambda(\widetilde{z}) > 0$, and therefore $dw_\lambda(\widetilde{z}_\lambda(z))/dz > 0$.

It remains to make the cutoff-comparative-statics bounds explicit. To compare functions defined on different cutoff-dependent supports, transport each interval $[1, \eta_{\mathcal{Z}}]$ affinely to $[0,1]$ and equip the resulting product space with the supremum norm for policies and a

bounded-Lipschitz norm for rival-output laws. Retain the same notation for the transported objects. Let $x_{\mathcal{Z}}(d) \equiv (d, w_{\mathcal{Z},d}, \mathcal{B}_{\mathcal{Z},d})$ collect the common conditional-equilibrium objects generated by a candidate vacancy-density tilt d, which belong to the function space $\mathcal{X}$.

Let $\bar{\varrho}_T < 1$ be a uniform upper bound on the contraction modulus, $\bar{\varrho}_T \geq 2\mathfrak{L}\left[\frac{C_J}{\zeta} + \frac{\phi}{1+\zeta}\frac{\bar{\xi}}{\mathcal{D}_0}\right]$, for all parameter values under consideration. Define $C_{T\mathcal{Z}} \equiv \sup_{\mathcal{Z},d} \|\partial_{\mathcal{Z}} \mathcal{T}_{\mathcal{Z}}(d)\|_\infty$, $C_{xd} \equiv \sup_{\mathcal{Z},d} \|D_d x_{\mathcal{Z}}(d)\|_{\mathrm{op}}$, $C^0_{x\mathcal{Z}} \equiv \sup_{\mathcal{Z},d} \|\partial_{\mathcal{Z}} x_{\mathcal{Z}}(d)\|$. The same compactness and primitive derivative bounds used above imply that these constants are finite. Differentiating the fixed-point identity $d_{\mathcal{Z}} = \mathcal{T}_{\mathcal{Z}}(d_{\mathcal{Z}})$ gives $\left\|\frac{dd_{\mathcal{Z}}}{d\bar{\mathcal{Z}}}\right\|_\infty \leq \frac{C_{T\mathcal{Z}}}{1-\bar{\varrho}_T}$. Consequently, $\left\|\frac{dx_{\mathcal{Z}}}{d\bar{\mathcal{Z}}}\right\| \leq C_{x\mathcal{Z}} \equiv C^0_{x\mathcal{Z}} + \frac{C_{xd}C_{T\mathcal{Z}}}{1-\bar{\varrho}_T}$.

For the maximized operating-profit function, define $\underline{\pi}_z \equiv \inf_{(z,x)} \frac{\partial \pi^{\mathrm{op}}(z;x)}{\partial z}$, $C_{\pi x} \equiv \sup_{(z,x)} \sup_{\|\dot{x}\|_{\mathcal{X}} \leq 1} |D_x \pi^{\mathrm{op}}(z;x)[\dot{x}]|$, where both extrema are taken over the full compact admissible domain, not only over equilibrium values. The envelope theorem gives $\frac{\partial \pi^{\mathrm{op}}(z;x)}{\partial z} = \frac{\partial y(z)}{\partial z}\mathbb{E}_B[m(y(z);B)]$. The primitive lower bounds on $\partial y/\partial z$ and realized MRP therefore imply $\underline{\pi}_z > 0$, while compactness and smoothness imply $C_{\pi x} < \infty$.

Boundary profits are $\pi^b(\mathcal{Z}) = \pi^{\mathrm{op}}(\mathcal{Z}; x_{\mathcal{Z}})$. The chain rule and the preceding definitions yield $\frac{d\pi^b(\mathcal{Z})}{d\bar{\mathcal{Z}}} = \frac{\partial \pi^{\mathrm{op}}}{\partial z} + D_x \pi^{\mathrm{op}}\left[\frac{dx_{\mathcal{Z}}}{d\bar{\mathcal{Z}}}\right] \geq \underline{\pi}_z - C_{\pi x}C_{x\mathcal{Z}}$. Condition 1 makes the final expression strictly positive. Boundary profit is therefore strictly increasing in the cutoff. If it is nonnegative under full activity, full activity is selected. Otherwise, the restriction on κ makes it positive at the largest admissible cutoff, so continuity and strict monotonicity imply a unique zero-profit cutoff.

For completeness, consider constant MRP and fix an ordered active set. Expected MRPL is then an exogenous strictly increasing schedule $R(z)$. Let γ^A denote productivity density conditional on this active set, write $\bar{z} \equiv Z_0\bar{q}$, and let J denote transformed surplus and H the vacancy-weighted productivity distribution. For a shooting parameter $\Omega > 0$, solve

$$J'_\Omega(z) = R'(z)[1 - \xi H_\Omega(z)]^{-2\nu}, \qquad H'_\Omega(z) = \Omega\gamma^A(z)J_\Omega(z)^{1/\zeta}[1 - \xi H_\Omega(z)]^{-\phi/(1+\zeta)}, \quad (40)$$

with $J_\Omega(Z) = R(Z) - \underline{w} > 0$ and $H_\Omega(Z) = 0$. While $H_\Omega \leq 1$, the right-hand side is locally Lipschitz and cooperative. Hence $H_\Omega(\bar{z})$ is continuous and strictly increasing in Ω. It converges to zero as $\Omega \downarrow 0$; moreover $J_\Omega \geq J_\Omega(Z)$ and $1 - \xi H_\Omega \leq 1$, so integration of the second equation gives $H_\Omega(\bar{z}) \geq \Omega J_\Omega(Z)^{1/\zeta}$. It therefore exceeds one for sufficiently large Ω. There is exactly one Ω such that $H_\Omega(\bar{z}) = 1$, and then Ω^{-1} is precisely the normalizing integral in the vacancy first-order condition. This proves conditional existence and uniqueness for every $\zeta > 0$ when MRP is constant. The cutoff argument in the preceding paragraph then selects the unique ordered active set. This completes the proof. ■

Labor-market clearing. We now let λ vary over Λ. The meeting function $\lambda(\theta) = \chi[\beta\theta^{-\iota} + (1-\beta)]^{-1/\iota}$ has inverse

$$\Theta(\lambda) = \frac{\beta^{1/\iota}\lambda}{[\chi^\iota - (1-\beta)\lambda^\iota]^{1/\iota}}, \qquad \frac{d\log\Theta(\lambda)}{d\log\lambda} = \frac{\chi^\iota}{\chi^\iota - (1-\beta)\lambda^\iota} > 1. \quad (41)$$

Define an aggregate vacancy normalizing constant $Q_{\lambda,i,\mathbf{Y}} \equiv \int g_\lambda(t) J_{\lambda,i,\mathbf{Y}}(t)^{1/\zeta} \mathcal{D}_{\lambda,i,\mathbf{Y}}(t)^{-\phi/(1+\zeta)} dt$ such that the vacancy-density tilt integrates to one. Aggregating vacancies gives firms' implied tightness $\Psi_{i,\mathbf{Y}}(\lambda) = \{(MA_\lambda)^\zeta \delta Q^\zeta_{\lambda,i,\mathbf{Y}} \lambda(\delta+\lambda)^\zeta / [\bar{c}(\delta+s\lambda)^{\zeta+2+\phi}]\}^{1/(1+\zeta)}$. Labor-market clearing requires $\Psi_{i,\mathbf{Y}}(\lambda) = \Theta(\lambda)$.

The primitive derivative bounds used in the preceding proof, together with the strict cutoff margin $\underline{\pi}_z - C_{\pi x} C_{x\mathcal{Z}} > 0$, generate finite constants $C_{\mathcal{Z}\lambda}(\mathfrak{p}_{-\zeta})$ and $C_Q(\mathfrak{p}_{-\zeta})$ such that, uniformly over industries, candidate output vectors, and meeting rates, $\left|\frac{d\log \mathcal{Z}_{\lambda,i,\mathbf{Y}}}{d\log\lambda}\right| \leq C_{\mathcal{Z}\lambda}(\mathfrak{p}_{-\zeta})$, $\left|\frac{d\log Q_{\lambda,i,\mathbf{Y}}}{d\log\lambda}\right| \leq \frac{C_Q(\mathfrak{p}_{-\zeta})}{\zeta}$. These are total derivatives along the unique conditional equilibrium and therefore incorporate the induced responses of the selected cutoff, vacancy-density tilt, wage policy, and rival-output law. Let $E^p_A \equiv \sup_{\mathcal{Z}} |\partial \log A(\mathcal{Z})/\partial \log \mathcal{Z}| < \infty$ and $C_A \equiv E^p_A C_{\mathcal{Z}\lambda}$, so that $|\partial \log A_\lambda / \partial \log \lambda| \leq C_A$. All three constants are formed from suprema of primitive derivatives on the fixed compact domains and the inverse of the strict cutoff and contraction margins, as described in the proof of Lemma A.3. Define $\Delta_{\mathrm{LM}} \equiv 1 - \sup_{\lambda\in\Lambda}[\lambda/(\delta+\lambda) - s\lambda/(\delta+s\lambda)]$. Maximizing the expression in brackets over $(0,\infty)$ gives $(1-\sqrt{s})/(1+\sqrt{s})$, so $\Delta_{\mathrm{LM}} \geq 2\sqrt{s}/(1+\sqrt{s}) > 0$.

Evaluating the formula for Ψ at the extrema of the compact primitive envelopes for A_λ and $Q_{\lambda,i,\mathbf{Y}}$ gives primitive lower and upper bounds $\underline{\Psi}^p_{i,\mathbf{Y}}(\lambda)$ and $\overline{\Psi}^p_{i,\mathbf{Y}}(\lambda)$ – independent of any equilibrium condition.

Regularity condition 2 (Primitive labor-market response)
The primitive revenue, matching, hiring-cost, and density functions are twice continuously differentiable on their compact domains, the strict cutoff margin in Condition 1 is uniform, and the primitive meeting-rate bracket satisfies

$$\inf_{i,\mathbf{Y}\in\mathscr{Y}} \underline{\Psi}^p_{i,\mathbf{Y}}(\lambda_L) > \Theta(\lambda_L), \qquad \sup_{i,\mathbf{Y}\in\mathscr{Y}} \overline{\Psi}^p_{i,\mathbf{Y}}(\lambda_H) < \Theta(\lambda_H), \tag{42}$$

for some $0 < \lambda_L < \lambda_H < \bar{\lambda}$. *Moreover,* $C_A < \Delta_{\mathrm{LM}}$, *and*

$$\zeta > \max\left\{\zeta_C(\mathfrak{p}_{-\zeta}), \frac{C_Q(\mathfrak{p}_{-\zeta})}{\Delta_{\mathrm{LM}} - C_A}\right\}. \tag{43}$$

This condition is the standard parametric-contraction requirement. It does not add a new economic restriction on the price–size feedback. It ensures that the unique conditional equilibrium changes smoothly with the meeting rate. The matching parameters (χ, β, ι) determine the feasible interval for λ and the slope of required tightness. The separation rate and relative on-the-job search efficiency determine Δ_{LM}; a larger s raises its primitive lower bound. The term C_A isolates the response of the extensive margin, while C_Q collects the intensive response of relative vacancy creation and the composition effect of the cutoff. Greater vacancy-cost curvature dampens the latter response.

With constant MRP, the shooting system in Equation (40) replaces the conditional contraction and therefore removes the restriction $\zeta > \zeta_C$. It does not by itself determine λ: uniqueness of labor-market clearing still requires the firms' implied tightness to rise less rapidly than

Θ. Differentiating the shooting system and applying Gronwall's inequality gives a finite primitive bound C_Q^{lin} on $|\partial \log Q_\lambda^{\text{lin}} / \partial \log \lambda|$. In the constant-MRP limit, the sufficient slope restriction becomes $C_A^{\text{lin}} + C_Q^{\text{lin}} < \Delta_{\text{LM}}$. This restriction controls worker flows and endogenous activity, not the price–size feedback, and it does not require vacancy costs to be sufficiently convex.

Lemma A.3 (Labor-market clearing)
Under Conditions 1 and 2, every labor market has a unique market-clearing meeting rate conditional on $\mathbf{Y}$.

Proof. The parameterized contraction theorem and the uniform cutoff transversality established above imply that the conditional fixed point and selected cutoff are locally Lipschitz in $\log \lambda$. They are therefore differentiable almost everywhere and absolutely continuous on every compact subinterval of $(0, \bar{\lambda})$.

Implicit differentiation of the boundary-profit equation and the strict cutoff margin give $|\partial \log \mathcal{Z}_\lambda / \partial \log \lambda| \leq C_{\mathcal{Z}\lambda}$, and hence $|\partial \log A_\lambda / \partial \log \lambda| \leq C_A$. To obtain the response bound in Equation (43), write $J_\lambda(t) = J_\lambda(1)\bar{J}_\lambda(t)$. Primitive differentiation and Gronwall's inequality give uniform bounds on the log derivatives of $J_\lambda(1)$, $\bar{J}_\lambda(t)$, and $\mathcal{D}_\lambda(t)$. Moreover, the normalized integrand $\bar{J}_\lambda(t)^{1/\zeta}\mathcal{D}_\lambda(t)^{-\phi/(1+\zeta)}$ differs uniformly from one by order $1/\zeta$, while its log derivative is also of order $1/\zeta$. When the cutoff changes, the derivative of the conditional productivity law is a signed measure with zero total mass and therefore acts only on this order-$1/\zeta$ deviation from a constant. Combining these bounds gives a finite primitive constant C_Q such that $\zeta|\partial \log Q_{\lambda,i,\mathbf{Y}} / \partial \log \lambda| \leq C_Q$ almost everywhere, uniformly across labor markets and $\mathbf{Y} \in \mathscr{Y}$. This proves, rather than assumes, the response bound used below.

Existence follows directly from the primitive bracket. Condition (42) implies $\Psi(\lambda_L) > \Theta(\lambda_L)$ and $\Psi(\lambda_H) < \Theta(\lambda_H)$ for every labor market and $\mathbf{Y} \in \mathscr{Y}$. Continuity therefore gives at least one crossing in (λ_L, λ_H). Using an interior bracket is important in the presence of overhead costs: as $\lambda \downarrow 0$, operating revenue can vanish before overhead costs do, so a uniformly positive active mass cannot in general be inferred from endpoint asymptotics.

For uniqueness, differentiation at every regular point gives

$$\frac{\partial \log \Psi(\lambda)}{\partial \log \lambda} = \frac{1}{1+\zeta}\left[1 + \zeta\frac{\lambda}{\delta+\lambda} - (\zeta+2+\phi)\frac{s\lambda}{\delta+s\lambda} + \zeta\frac{\partial \log A_\lambda}{\partial \log \lambda} + \zeta\frac{\partial \log Q_\lambda}{\partial \log \lambda}\right].$$

Using $|\partial \log A_\lambda / \partial \log \lambda| \leq C_A$, dropping the additional negative term $-(2+\phi)s\lambda/(\delta+s\lambda)$, and using the bound defining C_Q, yields $\partial \log \Psi / \partial \log \lambda \leq [1 + C_Q + \zeta(1 - \Delta_{\text{LM}} + C_A)]/(1+\zeta) < 1$ by condition (43). Equation (41) gives $d \log \Theta / d \log \lambda > 1$. Hence $\log(\Psi/\Theta)$ has a strictly negative derivative almost everywhere. Absolute continuity makes the ratio strictly decreasing, so the crossing is unique.

When MRP is constant, Equation (40) supplies the unique conditional equilibrium for each trial λ. Differentiating that initial-value problem with respect to $\log \lambda$, using $1 - \xi H \geq \mathcal{D}_0$, and applying Gronwall's inequality bound the response of its normalizer by $|\partial \log Q_\lambda^{\text{lin}} / \partial \log \lambda| \leq$

C_Q^{lin}. Cutoff transversality similarly gives $|\partial \log A_\lambda / \partial \log \lambda| \leq C_A^{\text{lin}}$. Substitution into the displayed derivative of Ψ yields $\partial \log \Psi / \partial \log \lambda < 1$ whenever $C_A^{\text{lin}} + C_Q^{\text{lin}} < \Delta_{\text{LM}}$. Because Equation (41) gives $d \log \Theta / d \log \lambda > 1$, the clearing ratio is strictly decreasing; the primitive bracket supplies the sign change. Thus constant MRP retains the positive-surplus, endpoint, and activity-transversality conditions, but neither the rank restriction nor a lower bound on ζ. ■

Aggregate closure. For a candidate $\mathbf{Y}$, Lemmas A.2 and A.3 generate a unique equilibrium in every labor market and hence a single implied industry output. Write it as $\mathscr{Y}_i(\mathcal{P}_i(\mathbf{Y}))$, and define $\mathscr{F}_i(\mathbf{Y}) \equiv \mathscr{Y}_i(\mathcal{P}_i(\mathbf{Y}))$.

For any positive candidate rectangle, extrema of the primitive technologies on the conditional envelopes give lower and upper bounds $\underline{\mathscr{B}}_i(\mathscr{Y})$ and $\overline{\mathscr{B}}_i(\mathscr{Y})$ on implied industry output. The derivative bounds used above and the inverses of the conditional-contraction, labor-clearing, and cutoff-transversality margins also generate three finite primitive constants. Let $C_{Y\mathcal{P}}^F$ bound the log response of industry output to its demand shifter when the cutoff is held fixed, let $C_{Y\mathcal{Z}}$ bound its response to the cutoff, and let $C_{\mathcal{Z}\mathcal{P}}$ bound the response of the selected cutoff to the demand shifter. Define the primitive bound on the aggregate supply response $\bar{\varepsilon}_{\mathcal{P}}^p \equiv C_{Y\mathcal{P}}^F + C_{Y\mathcal{Z}} C_{\mathcal{Z}\mathcal{P}}$. The second term is important because vacancy-cost curvature need not eliminate the response of the extensive margin. Define the primitive bound on the demand response $R_\alpha(\rho) \equiv \max_i \{1 - \alpha_i + |\alpha_i - (\rho - 1)/\rho|\}$.

Regularity condition 3 (Primitive aggregate closure)
The Cobb–Douglas shares satisfy $\alpha_i > 0$ and $\sum_i \alpha_i = 1$. There exist primitive bounds $0 < \underline{Y}_i < \overline{Y}_i < \infty$ such that, for $\mathscr{Y} \equiv \prod_i [\underline{Y}_i, \overline{Y}_i]$, $\underline{\mathscr{B}}_i(\mathscr{Y}) \geq \underline{Y}_i$ and $\overline{\mathscr{B}}_i(\mathscr{Y}) \leq \overline{Y}_i$ for every industry. Primitive technologies satisfy the compactness and smoothness conditions above, and

$$\bar{\varepsilon}_{\mathcal{P}}^p R_\alpha(\rho) < 1. \tag{44}$$

The term $R_\alpha(\rho)$ is the maximal change in an industry's demand shifter generated by a change in the vector of industry outputs, while $\bar{\varepsilon}_{\mathcal{P}}^p$ bounds the resulting supply response, including entry and exit. Thus, condition (44) requires the aggregate demand–supply feedback to be smaller than one. Greater vacancy-cost curvature weakens the fixed-cutoff response, the cutoff transversality margin controls the extensive-margin response, and the Cobb–Douglas shares and ρ govern the propagation of output changes through demand. If MRP is constant and invariant to aggregate output, neither policies nor the cutoff depend on $\mathbf{Y}$. It follows that $\bar{\varepsilon}_{\mathcal{P}}^p = 0$ and aggregate closure is immediate.

Lemma A.4 (Aggregate-output fixed point)
Under Condition 3, the conditional labor-market equilibria generate a unique aggregate-output vector.

Proof. The parameterized contraction theorem, the unique market-clearing meeting rates, and cutoff transversality imply that $\mathscr{F}$ is continuous. The primitive bounds in Condition 3

imply that $\mathscr{F}$ maps the compact convex rectangle $\mathscr{Y}$ into itself, so Brouwer's theorem gives a fixed point.

For the response bound, first hold $\mathcal{Z}$ fixed. Differentiating the conditional fixed point and the labor-market-clearing equation with respect to $\log \mathcal{P}_i$, and applying the inverses of their strict margins, gives $|\partial \log \mathscr{Y}_i / \partial \log \mathcal{P}_i|_{\mathcal{Z}} \leq C^F_{Y\mathcal{P}}$. Let $\Pi^b_i(\mathcal{Z}, \mathcal{P}_i)$ denote boundary profit after solving those fixed-cutoff blocks. The same primitive derivative bounds give $|\partial \Pi^b_i / \partial \log \mathcal{P}_i| \leq C_{\pi\mathcal{P}}$, while conditional stability gives $\mathcal{Z}\partial \Pi^b_i / \partial \mathcal{Z} \geq \underline{\pi}_{\mathcal{Z}} > 0$. The implicit-function theorem therefore yields $|\partial \log \mathcal{Z}_i / \partial \log \mathcal{P}_i| \leq C_{\pi\mathcal{P}} / \underline{\pi}_{\mathcal{Z}} \equiv C_{\mathcal{Z}\mathcal{P}}$. Since $|\partial \log \mathscr{Y}_i / \partial \log \mathcal{Z}_i| \leq C_{Y\mathcal{Z}}$, the chain rule gives $|d \log \mathscr{Y}_i / d \log \mathcal{P}_i| \leq \bar{\varepsilon}^p_{\mathcal{P}}$.

For uniqueness, Cobb–Douglas aggregation implies $\log \mathcal{P}_i(\mathbf{Y}) = \log \alpha_i + \sum_h \alpha_h \log Y_h - (\rho - 1) \log Y_i / \rho - \log K_i / \rho$. Therefore $\partial \log \mathcal{P}_i / \partial \log Y_h = \alpha_h - (\rho - 1)\mathbf{1}\{h = i\}/\rho$. At every differentiability point, the absolute row sum of the log-Jacobian of $\mathscr{F}$ is consequently bounded by $\bar{\varepsilon}^p_{\mathcal{P}} R_\alpha(\rho) < 1$. The map is locally Lipschitz and hence absolutely continuous along line segments in $\log \mathscr{Y}$. Integrating the Jacobian bound along any such segment gives the same strict global Lipschitz bound in the sup norm. Thus $\log \mathbf{Y} \mapsto \log \mathscr{F}(\mathbf{Y})$ is a contraction, and its fixed point is unique. ■

Proof of Proposition 1. Fix a candidate aggregate-output vector. Lemma A.2 gives, for every meeting rate, a unique conditional labor–product-market equilibrium and ordered active set. Lemma A.3 selects a unique market-clearing meeting rate in every labor market. Expected MRPL and wages are strictly increasing in productivity, and the positive continuous vacancy tilt maps the continuous productivity distribution into a continuous wage-offer distribution. Lemma A.4 then selects a unique aggregate-output vector and hence unique demand shifters. The resulting aggregates reproduce the policies used to construct the Bayesian rival-output and labor-market distributions. All equilibrium conditions are therefore jointly satisfied. This establishes existence and uniqueness of the symmetric, rank-preserving stationary equilibrium within the class of continuous wage-offer distributions and ordered activity rules covered by the three primitive regularity conditions.[41] ■

In Appendix D.6 we complement this theoretical derivation with a numerical check of local uniqueness for the sequence of simulated equilibria along a MW branch underlying our quantitative results.

B.7 Aggregation

This section describes aggregation across firms, product markets, labor markets, and industries. Let $N_{a_i} \equiv \mathbb{E}_{\mathcal{N}_{a_i}}[\bar{N}_{a_i}]$ be the average number of potential firms and $\varpi_{a_i}(z)$ be the operating probability defined in Appendix B.4. The expected active-firm measure in one product market is $dN^A_{a_i} \equiv N_{a_i} \varpi_{a_i}(z) d\Gamma_{a_i}(z)$.

[41] With decreasing returns in revenues, additional stationary equilibria with degenerate wage distributions may exist outside this class (Menzio, 2024).

Representative-product-market level. Product markets within a labor market a_i are ex ante identical. Let $Y_{k(a_i)} = \left[\sum_{j=1}^{N_{a_i}} D_j y_{a_i}(Z_j)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$ denote realized product-market output, where D_j is the operating indicator of potential-firm position j. The representative output is the equal-market certainty equivalent

$$\bar{Y}_{k(a_i)} \equiv \left\{\mathbb{E}_{\mathcal{N}_{a_i},\Gamma_{a_i},\varpi_{a_i}}\left[Y_{k(a_i)}^{\frac{\rho-1}{\rho}}\right]\right\}^{\frac{\rho}{\rho-1}},$$

with associated price index $\bar{P}_{k(a_i)} = \mathcal{P}_i(\mathbf{Y})\bar{Y}_{k(a_i)}^{-1/\rho}$. Consequently, expected sales in one product market satisfy the exact adding-up identity

$$\mathcal{R}_{k(a_i)} \equiv \bar{P}_{k(a_i)}\bar{Y}_{k(a_i)} = \mathcal{P}_i(\mathbf{Y})\bar{Y}_{k(a_i)}^{\frac{\rho-1}{\rho}} = \int y_{a_i}(z)\bar{p}_{k(a_i)}(z)dN_{a_i}^A,$$

where $\bar{p}_{k(a_i)}(z) \equiv \mathbb{E}_{\mathcal{B}_{a_i}(\varpi_{a_i})}\left[p_{k(a_i)}(y_{a_i}(z);B_{a_i}) \mid z\right]$ is the firm's expected price. The first equality is the equal-market CES identity, the second equality is the population counterpart of summing expected firm revenues. The law $\mathcal{B}_{a_i}(\varpi_{a_i})$ is the rival-output law induced by the size-biased market-size distribution $\mathcal{N}_{a_i}^F$, the unconditional productivity distribution, and the operating rule.

Employment, payroll, and hiring costs per representative product market are

$$E_{k(a_i)} \equiv \int \ell_{a_i}(z)dN_{a_i}^A, \quad \mathcal{W}_{k(a_i)} \equiv \int aw_{a_i}(z)\ell_{a_i}(z)dN_{a_i}^A, \quad \mathcal{HC}_{k(a_i)} \equiv \int ac_{a_i}(z)dN_{a_i}^A.$$

Thus, $E_{a_i} = K_{a_i}E_{k(a_i)}$, and the average wage in the representative product market is $\bar{W}_{k(a_i)} = \mathcal{W}_{k(a_i)}/E_{k(a_i)}$. Its physical-productivity index is $Z_{k(a_i)} \equiv \frac{\bar{Y}_{k(a_i)}}{aE_{k(a_i)}}$.

To aggregate market power, let $m_{k(a_i)}(y;B) \equiv \partial[p_{k(a_i)}(y;B)y]/\partial y$ denote the realized marginal-revenue product of physical output (MRP). Accordingly, the expected MRP and the Bayes-Cournot markup are defined as $\bar{m}_{k(a_i)}(z) \equiv \mathbb{E}_{\mathcal{B}_{a_i}(\varpi_{a_i})}\left[m_{k(a_i)}(y_{a_i}(z);B_{a_i}) \mid z\right]$, $\mu_{k(a_i)}(z) \equiv \frac{\bar{p}_{k(a_i)}(z)}{\bar{m}_{k(a_i)}(z)}$. Let $\widetilde{z}_{a_i}(z) \equiv \frac{z}{a}\bar{m}_{k(a_i)}(z)$ and $\psi_{a_i}(z) \equiv \frac{\widetilde{z}_{a_i}(z)}{w_{a_i}(z)}$ denote the expected MRPL piece rate and the markdown, respectively. Since $y_{a_i}(z) = az\ell_{a_i}(z)$, the definitions above imply the Bayesian double-wedge identity

$$y_{a_i}(z)\bar{p}_{k(a_i)}(z) = \mu_{k(a_i)}(z)\psi_{a_i}(z)aw_{a_i}(z)\ell_{a_i}(z). \tag{45}$$

Equation (45) implies that the model-consistent weights for the market-power wedge are payroll weights. For any firm-level object $x(z)$, define the payroll-weighted average $\langle x\rangle_{C,k(a_i)} \equiv \frac{\int aw_{a_i}(z)\ell_{a_i}(z)x(z)dN_{a_i}^A}{\mathcal{W}_{k(a_i)}}$. The representative-product-market markup and markdown are therefore

$$\bar{\mu}_{k(a_i)} \equiv \left\langle \mu_{k(a_i)}\right\rangle_{C,k(a_i)}, \quad \bar{\psi}_{k(a_i)} \equiv \langle \psi_{a_i}\rangle_{C,k(a_i)},$$

which yield the market-power index $\bar{\mathcal{M}}_{k(a_i)} \equiv \left\langle \mu_{k(a_i)}\psi_{a_i}\right\rangle_{C,k(a_i)} = \bar{\mu}_{k(a_i)}\bar{\psi}_{k(a_i)} + \mathrm{Cov}_{C,k(a_i)}\left(\mu_{k(a_i)},\psi_{a_i}\right)$.

Integrating equation (45) therefore gives the exact identity $\mathcal{R}_{k(a_i)} = \mathcal{W}_{k(a_i)}\bar{\mathcal{M}}_{k(a_i)}$. The sales-weighted harmonic markup targeted in the structural estimation is defined as $\mu^H_{k(a_i)} \equiv \frac{\mathcal{R}_{k(a_i)}}{\int y_{a_i}(z)\bar{m}_{k(a_i)}(z)dN^A_{a_i}}$, thus differing from the model-consistent aggregation.

Expected value added, operating profits before overhead costs, and the operating labor share are

$$\mathcal{VA}_{k(a_i)} \equiv \mathcal{R}_{k(a_i)} - \mathcal{HC}_{k(a_i)}, \quad \Pi^{\text{op}}_{k(a_i)} \equiv \mathcal{R}_{k(a_i)} - \mathcal{W}_{k(a_i)} - \mathcal{HC}_{k(a_i)}, \quad LS^{\text{op}}_{k(a_i)} \equiv \frac{\mathcal{W}_{k(a_i)}}{\mathcal{VA}_{k(a_i)}}.$$

Industry level. Let $K_i \equiv \int K_{a_i}da_i$ denote the measure of product markets in industry i. The no-love-for-variety industry quantity and price indices are $Y_i = K_i^{-1/(\rho-1)}\left[\int K_{a_i}\bar{Y}^{\frac{\rho-1}{\rho}}_{k(a_i)}da_i\right]^{\frac{\rho}{\rho-1}}$ and $P_i = K_i^{-1/(1-\rho)}\left[\int K_{a_i}\bar{P}^{1-\rho}_{k(a_i)}da_i\right]^{1/(1-\rho)}$. They satisfy $P_iY_i = \int K_{a_i}\mathcal{R}_{k(a_i)}da_i$. Industry employment, payroll, and hiring costs are

$$E_i \equiv \int K_{a_i}E_{k(a_i)}da_i, \quad \mathcal{W}_i \equiv \int K_{a_i}\mathcal{W}_{k(a_i)}da_i, \quad \mathcal{HC}_i \equiv \int K_{a_i}\mathcal{HC}_{k(a_i)}da_i.$$

The corresponding industry-level market-power objects use the same payroll weights:

$$\mu_i \equiv \int K_{a_i}\frac{\mathcal{W}_{k(a_i)}}{\mathcal{W}_i}\bar{\mu}_{k(a_i)}da_i, \quad \psi_i \equiv \int K_{a_i}\frac{\mathcal{W}_{k(a_i)}}{\mathcal{W}_i}\bar{\psi}_{k(a_i)}da_i, \quad \mathcal{M}_i \equiv \int K_{a_i}\frac{\mathcal{W}_{k(a_i)}}{\mathcal{W}_i}\bar{\mathcal{M}}_{k(a_i)}da_i,$$

where $\mathcal{M}_i = \mu_i\psi_i + \text{Cov}_{C,i}(\mu,\psi)$. Industry total labor productivity and its operating labor share are

$$\widetilde{Z}_i \equiv \frac{Y_i}{E_i}, \qquad LS^{\text{op}}_i \equiv \frac{\mathcal{W}_i}{P_iY_i - \mathcal{HC}_i}.$$

Aggregate level. The final-good quantity and price indices are $Y = \prod_{i=1}^I Y_i^{\alpha_i}$, $P = \prod_{i=1}^I\left(\frac{P_i}{\alpha_i}\right)^{\alpha_i}$. Final-good cost minimization implies $PY = \sum_{i=1}^I P_iY_i$. Define $E \equiv \sum_{i=1}^I E_i$, $\mathcal{W} \equiv \sum_{i=1}^I \mathcal{W}_i$, $\mathcal{HC} \equiv \sum_{i=1}^I \mathcal{HC}_i$. Aggregate markups, markdowns, and the raw market-power wedge are

$$\mu \equiv \sum_{i=1}^I \frac{\mathcal{W}_i}{\mathcal{W}}\mu_i, \ \psi \equiv \sum_{i=1}^I \frac{\mathcal{W}_i}{\mathcal{W}}\psi_i, \ \mathcal{M} \equiv \sum_{i=1}^I \frac{\mathcal{W}_i}{\mathcal{W}}\mathcal{M}_i,$$

where $\mathcal{M} = \mu\psi + \text{Cov}_C(\mu,\psi)$. Equivalently, these objects can be obtained by applying the payroll weights directly to every active firm in every product market, i.e., recursive and direct aggregation coincide.

Let $N^A_{a_i} \equiv N_{a_i}\int \varpi_{a_i}(z)d\Gamma_{a_i}(z)$ be the expected number of active firms per product market. Since overhead labor is valued at the aggregate average wage $\bar{W} \equiv \mathcal{W}/E$, total overhead payments are $\mathcal{FC} \equiv \bar{W}\sum_{i=1}^I\int K_{a_i}N^A_{a_i}a\kappa_{a_i}da_i$. Aggregate profits, total labor productivity, and the labor share are then

$$\bar{\Pi} = PY - \mathcal{HC} - \mathcal{W} - \mathcal{FC}, \quad \widetilde{Z} \equiv \frac{Y}{E}, \quad LS \equiv \frac{\mathcal{W} + \mathcal{FC}}{PY - \mathcal{HC}}.$$

Hiring costs use final goods and are therefore subtracted from the resource constraint. Overhead payments are payments for labor services, not an additional use of final goods. Accordingly, $PC = PY - \mathcal{HC}$. Under the normalization $P = 1$, this becomes $C = Y - \mathcal{HC}$.

Hiring-cost-adjusted market power. Using equation (45) and the vacancy first-order condition, $c_{a_i}(z) = \frac{\psi_{a_i}(z)-1}{1+\zeta} w_{a_i}(z)\ell_{a_i}(z)$, the firm's operating labor share can be written as

$$LS^{\text{op}}_{a_i}(z) \equiv \frac{a w_{a_i}(z)\ell_{a_i}(z)}{y_{a_i}(z)\bar{p}_{k(a_i)}(z) - a c_{a_i}(z)} = \left[\mu_{k(a_i)}(z)\psi_{a_i}(z) - \frac{\psi_{a_i}(z)-1}{1+\zeta}\right]^{-1}.$$

This motivates the firm-level adjusted market-power index $\mathcal{M}^{\text{adj}}_{a_i}(z) \equiv \mu_{k(a_i)}(z)\psi_{a_i}(z) - \frac{\psi_{a_i}(z)-1}{1+\zeta}$. The aggregate adjusted index is its payroll-weighted average:

$$\mathcal{M}^{\text{adj}} \equiv \left\langle \mathcal{M}^{\text{adj}}_{a_i}(z) \right\rangle_C = \mathcal{M} - \frac{\psi - 1}{1+\zeta} = \mu\psi + \text{Cov}_C(\mu, \psi) - \frac{\psi - 1}{1+\zeta}.$$

This identity is exact. In particular,

$$PY - \mathcal{HC} = \mathcal{W}\mathcal{M}^{\text{adj}}, \qquad LS^{\text{op}} \equiv \frac{\mathcal{W}}{PY - \mathcal{HC}} = \frac{1}{\mathcal{M}^{\text{adj}}},$$

while the aggregate labor share also includes overhead payments in its numerator.

Finally, consider a reform from a baseline equilibrium 0 to a new equilibrium 1, and write $\Delta x \equiv x_1 - x_0$. All aggregates and covariances at each endpoint use that equilibrium's payroll weights, so that $\Delta \text{Cov}_C(\mu, \psi) \equiv \text{Cov}_{C,1}(\mu, \psi) - \text{Cov}_{C,0}(\mu, \psi)$. The exact change in adjusted market power can then be decomposed as

$$\Delta\mathcal{M}^{\text{adj}} = \underbrace{\left(\mu_0 - \frac{1}{1+\zeta}\right)\Delta\psi}_{\text{lab. mkt power response}} + \underbrace{\psi_0 \Delta\mu}_{\text{prod. mkt power response}} + \underbrace{\Delta\mu\,\Delta\psi + \Delta\,\text{Cov}_C(\mu, \psi)}_{\text{joint mkt power response}}. \quad (46)$$

The first two terms evaluate the labor- and product-market-power responses using baseline markups and markdowns. The final term collects the exact nonlinear interaction between their changes and the change in their payroll-weighted covariance.

B.8 Algorithm to solve the model

The model is solved by a guess-and-verify procedure for the collection of expected MRPL piece rate functions $\tilde{z}_{k(a_i)}(z)$. Although the same solution concept applies to any set of parameters, we 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 expected MRPL functions, $\tilde{z}^0_{k(a_i)}(z)$. If the MW exceeds some guessed $\tilde{z}$ values with positive active mass, we adjust those points upward until they satisfy the MW constraint. When solving the model from scratch, this preliminary adjustment conjectures that all potential firms are

active; along a MW path, the preceding equilibrium supplies the initial active set. In the absence of overhead costs, the interior double-wedge solution delivers nonnegative expected operating profits. Since overhead costs are additively separable and do not affect firms' policy functions, we can therefore find a candidate equilibrium conditional on any ordered active set. To discipline the extensive margin of adjustment, we then check ex post whether the candidate equilibrium is sustainable and, if not, gradually reduce activity among the lowest-productivity firms.

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, indexed by an (i, a) pair, as follows:

   - Assign to each $\tilde{z}^0_{a_i}(z)$ value its active mass under the unconditional productivity distribution, given by the potential-firm productivity probability times the corresponding activity probability. This mass is not renormalized following exit and identifies the discretized active expected MRPL piece rate measure entering the labor-market equations;

   - Guess the wage and wage offer distribution as functions 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 for $\mathcal{H}'_{a_i}(\tilde{z})$ and $w'_{a_i}(\tilde{z})$ 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 into a linear system of difference equations with boundary conditions $w_{a_i}(\underline{\tilde{z}}_{a_i}) = \underline{w}_{a_i}$ and $\mathcal{H}_{a_i}(\underline{\tilde{z}}_{a_i}) = 0$; [42]

   - 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 and an upper bound on the job finding rate corresponding to too much job creation, $\lim_{\tilde{z}\to\bar{\tilde{z}}_{a_i}} \mathcal{H}_{a_i}(\tilde{z}) > 1$, and too little job creation, $\lim_{\tilde{z}\to\bar{\tilde{z}}_{a_i}} \mathcal{H}_{a_i}(\tilde{z}) < 1$, respectively, and then applying a bisection algorithm within those bounds;

   - Compute the labor-market-specific policy functions

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

     using the differential equations for $\mathcal{H}'_{a_i}(\tilde{z})$ and $w'_{a_i}(\tilde{z})$, the functional relationship between $\mathcal{H}'_{a_i}(\tilde{z})$ and $v_{a_i}(\tilde{z})$, the markdown definition, and the labor supply curve.

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

[42] Whenever the linearized difference system is singular or ill-conditioned, produces nonfinite values, or violates the monotonicity of either the wage policy or the wage-offer distribution, the algorithm switches to solving the nonlinear difference system.

– Compute firm output policies $y_{k(a_i)}(z) = az\ell_{a_i}(\tilde{z}^0_{k(a_i)}(z))$. Compute the representative product-market output $\bar{Y}_{k(a_i)}$ under the equal-market distribution $\mathcal{N}_{a_i}$, and aggregate it into industry and aggregate output through the no-love-for-variety CES and CD aggregators. We evaluate $\bar{Y}_{k(a_i)}$ by deterministic integration, so Monte Carlo error does not enter the construction of Y_i or Y;

– Construct the exact Bayesian competitor law. For each common-random-number draw, we draw total potential market size from the size-biased law $\mathcal{N}^F_{a_i}$ and the $N-1$ rivals' productivity types from the unconditional distribution Γ_{a_i}. A rival of type z_m operates with probability $\varpi_{a_i}(z_m)$ and otherwise contributes zero output. Its realized CES-output contribution is therefore

$$B_{a_i} = \sum_{m=1}^{N-1} I_m y_{k(a_i)}(z_m)^{\frac{\sigma-1}{\sigma}}.$$

For every own type, use these draws to compute expected price $\bar{p}_{k(a_i)}(z)$, expected MRP $\bar{mrp}_{k(a_i)}(z)$, and the Bayes-Cournot markup $\mu_{k(a_i)}(z) = \bar{p}_{k(a_i)}(z)/\bar{mrp}_{k(a_i)}(z)$;

– Compute sectoral, industry, and aggregate price indices from the corresponding output objects and demand conditions. Compute firm-specific expected profits as

$$\pi_{k(a_i)}\left(\tilde{z}^0_{k(a_i)}(z_n)\right) = \bar{p}_{k(a_i)}(z_n)y_{k(a_i)}(z_n) - aw_{k(a_i)}(z_n)\ell_{a_i}(z_n) - ac_{k(a_i)}(z_n) - a\kappa_{a_i}\mathbb{E}[aw].$$

The implied expected MRPL is then obtained from its definition,

$$\tilde{z}^{\text{implied}}_{k(a_i)}(z) \equiv \frac{z}{a}\bar{mrp}_{k(a_i)}(z) = \frac{\bar{p}_{k(a_i)}(z)}{\mu_{k(a_i)}(z)}z.$$

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

– Let $\mathbf{x}^n$ stack all active values of the expected MRPL guess at iteration $n = 0, 1, 2, \ldots$, and let $T(\mathbf{x}^n)$ stack the corresponding implied values. Define the convergence error as $\varepsilon^n \equiv \|\log \mathbf{x}^n - logT(\mathbf{x}^n)\|_2$, and set a tolerance $\bar{\varepsilon}$ ($= 10^{-5}$ in our simulations);

– If $\varepsilon^n > \bar{\varepsilon}$, update the current expected MRPL guess by damped Anderson mixing. Let $\mathbf{r}^n \equiv T(\mathbf{x}^n) - \mathbf{x}^n$ and collect the most recent iterate and residual differences in ΔX^n and ΔR^n. The coefficients solve $\boldsymbol{\alpha}^n \in \arg\min_{\boldsymbol{\alpha}} \|\mathbf{r}^n - \Delta R^n \boldsymbol{\alpha}\|^2$, and the updated guess is

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \delta^n_A \mathbf{r}^n - (\Delta X^n + \delta^n_A \Delta R^n)\,\boldsymbol{\alpha}^n, \qquad \delta^n_A \in (0, 1].$$

The global damping parameter is adjusted adaptively across iterations:

$$\delta_A^{n+1} = \begin{cases} \max\{\delta_{A,\min}, d_\downarrow \delta_A^n\}, & \text{if } \varepsilon^n > \varepsilon^{n-1}, \\ \min\{\delta_{A,\max}, d_\uparrow \delta_A^n\}, & \text{if } \varepsilon^n \leq \varepsilon^{n-1}, \end{cases}$$

where $d_\downarrow \in (0,1)$, $d_\uparrow > 1$, and $\delta_{A,\min} \leq \delta_A^n \leq \delta_{A,\max}$. In our implementation, $d_\downarrow = 0.5$, $d_\uparrow = 1.05$, $\delta_{A,\min} = 10^{-6}$, and $\delta_{A,\max} = 1$. After this update, project the expected MRPL schedule back onto the feasible set, imposing monotonicity and $\tilde{z}_{a_i}(z) \geq \underline{w}_{a_i}$.[43] Repeat steps 1–3 until finding a guess such that $\varepsilon^n < \bar{\varepsilon}$;

- Store the model solution for the converged expected MRPL guess as a candidate equilibrium.

4. Check that no active firm makes negative profits in the candidate equilibrium. If this condition is satisfied, the candidate equilibrium is sustainable and the model is solved. Otherwise, implement the following extensive-margin adjustment:

   - Following the equilibrium refinement of Berry (1992), identify the labor market containing the active type with the most negative profit and reduce the mass of its lowest currently active productivity rank. Under the loss-ordering property, the triggering and adjusted ranks coincide; selecting the lowest rank also preserves the ordered active support of the discretized labor market. To speed up the algorithm, at each iteration we reduce its active mass by the larger of one firm per representative market and one percent of the rank's current active mass, without removing more than the rank's remaining mass. We check that this shortcut has no bearing on the results;

   - The adjustment changes only the boundary rank's activity probability $\varpi_{a_i}(z)$. The unconditional productivity distribution Γ_{a_i} and the distribution of potential market size $\mathcal{N}_{a_i}$ remain fixed. If the adjustment does not exhaust the rank, its activity probability lies strictly between zero and one: it represents the fraction of potential firms of that productivity that operate. Accordingly, whenever this type is drawn as a rival, it is active with probability $\varpi_{a_i}(z)$ and otherwise posts no vacancies, employs no workers, and produces no output or revenue. If the rank's remaining active mass is smaller than one firm per representative market, we set its activity probability to zero and move the lower active-support boundary to the next rank. We then reconstruct the Bayesian competitor law and restart the algorithm from step 1.

[43] Because of the monotonicity and feasibility restrictions, this damped Anderson-mixing update is preferable to Jacobian-based methods and is computationally more efficient.

B.9 Dynamic Firm's Problem

A firm with productivity z in labor market a_i chooses paths $\{w_t, v_t\}_{t\geq 0}$, taking as given the paths of aggregate labor-market objects and product-market demand shifters. Its sequential profit-maximization problem is:

$$\max_{\substack{\{w_t \geq \underline{w}_{a_i}\}_{t\geq 0}, \\ \{v_t \geq 0\}_{t\geq 0}}} \int_0^\infty e^{-rt} \left(\mathbb{E}_{B_{a_i,t}} \left[p_{k(a_i)}(y_{a_i,t}; B_{a_i,t}) y_{a_i,t}(w_t, v_t; z) \right] - a w_t \ell_{a_i,t} - a c_{a_i}(w_t, v_t) - a \kappa_{a_i} \bar{W}_t \mathbb{1}_{\{v_t > 0\}} \right) dt$$

s.t. $y_{a_i,t} = a z \ell_{a_i,t}$,

$$p_{k(a_i)}(y_t) = \left(\frac{y_t}{Y_{k(a_i)t}(y_t)} \right)^{-\frac{1}{\sigma}} \left(\frac{Y_{k(a_i)t}(y_t)}{Y_{it}/K_i} \right)^{-\frac{1}{\rho}} \left(\frac{Y_{it}}{\alpha_i Y_t} \right)^{-1},$$

$$\dot{\ell}_{a_i,t} = -\left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i,t})[1 - F_{a_i,t}(w_t)]\right) \ell_{a_i,t} + \frac{v_t}{V_{a_i,t}} \lambda(\theta_{a_i,t})[u_{a_i,t} + s_{a_i}(1 - u_{a_i,t}) G_{a_i,t}(w_t)],$$

given $\ell_0 \geq 0$ and a transversality condition.

We proceed by setting up the current-value Hamiltonian, assuming an interior solution, and dropping time indices:

$$H(w, v; \ell) = \pi^e(w, v; \ell) + \xi \left(\frac{v}{V_{a_i}} \lambda(\theta_{a_i})[u_{a_i} + s_{a_i}(1 - u_{a_i}) G_{a_i}(w)] - \left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})[1 - F_{a_i}(w)]\right) \ell \right),$$

where $\pi^e(w, v; \ell) = \mathbb{E}_{B_{a_i}} \left[p_{k(a_i)}(y_{a_i,t}; B_{a_i}) y_{a_i}(w, v; z) \right] - a w \ell_{a_i} - a c_{a_i}(w, v)$ and ξ is the co-state variable.

The first-order conditions are:

$$\frac{\partial H}{\partial w} = 0 \iff \ell + \frac{\partial c_{a_i}(w, v)}{\partial w} = \xi \left(\frac{v}{V_{a_i}} \lambda(\theta_{a_i}) s_{a_i}(1 - u_{a_i}) g_{a_i}(w) + s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w) \ell \right),$$

$$\frac{\partial H}{\partial v} = 0 \iff \frac{\partial c_{a_i}(w, v)}{\partial v} = \xi \frac{\lambda(\theta_{a_i})[u_{a_i} + s_{a_i}(1 - u_{a_i}) G_{a_i}(w)]}{V_{a_i}},$$

$$\frac{\partial H}{\partial \ell} = r\xi - \dot{\xi} \iff \tilde{z}_{k(a_i)} - w - \xi \left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})[1 - F_{a_i}(w)]\right) = r\xi - \dot{\xi},$$

$$\frac{\partial H}{\partial \xi} = \dot{\ell} \iff \dot{\ell}_t = -\left(\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})[1 - F_{a_i}(w)]\right) \ell + \frac{v}{V_{a_i}} \lambda(\theta_{a_i})[u_{a_i} + s_{a_i}(1 - u_{a_i}) G_{a_i}(w)],$$

where $\tilde{z}_{k(a_i)}$ denotes the expected MRPL piece rate. In stationary equilibrium, i.e., setting $\dot{\ell} = \dot{\xi} = 0$, the wage and vacancy policy functions read:

$$\tilde{z}_{k(a_i)} - w = \frac{r + \delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})[1 - F_{a_i}(w)]}{2 s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w)} \left(1 - \phi \frac{s_{a_i} \lambda(\theta_{a_i}) f_{a_i}(w)}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})[1 - F_{a_i}(w)]} \frac{c_{a_i}(w, v)}{\ell_{a_i}(w, v)} \right), \quad (47)$$

$$\tilde{z}_{k(a_i)} - w = \frac{r + \delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})[1 - F_{a_i}(w)]}{\delta_{a_i} + s_{a_i} \lambda(\theta_{a_i})[1 - F_{a_i}(w)]} (1 + \zeta) \frac{c_{a_i}(w, v)}{\ell_{a_i}(w, v)}. \quad (48)$$

Equations (47)-(48) boil down to their stationary counterparts in the main text in the *timeless* limit, i.e., as $r \to 0$. By putting together the two FOCs, we can compute the dynamically consistent equilibrium markdown function as follows:

$$\psi_{a_i}(\tilde{z}) = 1 + \frac{r/(s_{a_i}\lambda(\theta_{a_i})) + \chi_{a_i}(\theta_{a_i}) + [1 - F_{a_i}(w_{a_i}(\tilde{z}))]}{2\nu f_{a_i}(w_{a_i}(\tilde{z}))w_{a_i}(\tilde{z})}. \tag{49}$$

Positive time discounting raises equilibrium markdowns, without changing any qualitative results.

C Efficiency Properties of Baseline Equilibrium

In this section we derive necessary conditions for a candidate constrained-efficient allocation in a version of our quantitative model with risk-neutral consumers ($\vartheta = 0$) and no overhead costs ($\kappa_{a_i} = 0$).

The planner maximizes aggregate consumption in steady state while facing the same search and product-market information frictions as private agents. In particular, vacancy and output policies may depend on a firm's own productivity but not on the realized number or productivity of its competitors.[44]

$$\mathcal{W} = \max_{\substack{\vec{v}_{a_i}(z)\geq 0 \\ \vec{q}_{a_i}(z,z')\in[0,1]}} \left\{ Y - \sum_{i=1}^{I} L_i \int M_{a_i}\left(\int ac_{a_i}\left(v_{a_i}(z), \mathcal{H}_{a_i}(z)\right) d\Gamma_{a_i}(z) \right) d\Xi_i(a) \right\},$$

$$\text{s.t.} \quad Y = \prod_{i=1}^{I} Y_i^{\alpha_i}, \quad Y_i = K_i^{-\frac{1}{\rho-1}} \left(\int_0^1 \int_0^{K_{a_i}} \bar{Y}_{k(a_i)}^{\frac{\rho-1}{\rho}} \, dk(a_i) \, da_i \right)^{\frac{\rho}{\rho-1}},$$

$$\bar{Y}_{k(a_i)}^{\frac{\rho-1}{\rho}} = \mathbb{E}_{\substack{N\sim\mathcal{N}_{a_i} \\ Z_1,\ldots,Z_N \overset{\text{i.i.d.}}{\sim}\Gamma_{a_i}}} \left[\left(\sum_{j=1}^{N} y_{a_i}(Z_j)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\rho-1}{\rho}\frac{\sigma}{\sigma-1}} \right], \quad y_{a_i}(z) = az\ell_{a_i}\left(v_{a_i}(z), \mathcal{H}_{a_i}(z)\right),$$

$$\ell_{a_i}\left(v_{a_i}(z), \mathcal{H}_{a_i}(z)\right) = \frac{v_{a_i}(z)}{V_{a_i}} \lambda_{a_i} u_{a_i} \frac{\delta_{a_i} + s_{a_i}\lambda_{a_i}}{\left[\delta_{a_i} + s_{a_i}\lambda_{a_i}(1 - \mathcal{H}_{a_i}(z))\right]^2}, \tag{50}$$

$$u_{a_i}(\lambda_{a_i}) = \frac{\delta_{a_i}}{\delta_{a_i} + \lambda_{a_i}}, \tag{51}$$

$$\lambda_{a_i}\left(V_{a_i}\right) = \mathcal{M}_{a_i}\left(1, \frac{V_{a_i}}{S_{a_i}(\lambda_{a_i}\left(V_{a_i}\right))}\right), \tag{52}$$

$$\mathcal{H}_{a_i}(z) = 1 - M_{a_i} \int \frac{\mathfrak{q}_{a_i}(z,z')v_{a_i}(z')}{V_{a_i}} d\Gamma_{a_i}(z'), \tag{53}$$

$$\mathfrak{q}_{a_i}(z,z') + \mathfrak{q}_{a_i}(z',z) = 1 \quad \forall z \neq z', \tag{54}$$

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

where $V_{a_i} = M_{a_i} \int v_{a_i}(z) d\Gamma_{a_i}(z)$, $S_{a_i}(\lambda_{a_i}) = u_{a_i}(\lambda_{a_i}) + s_{a_i}(1 - u_{a_i}(\lambda_{a_i}))$, and $c_{a_i}(z) = \bar{c}_{a_i} [\delta_{a_i} + s_{a_i}\lambda_{a_i}(1 - \mathcal{H}_{a_i}(z))]^{\phi} \frac{v_{a_i}(z)^{1+\zeta}}{1+\zeta}$. We also let $\nu \equiv 1 + \phi/[2(1+\zeta)]$.

To see how the information constraint enters the planner's marginal conditions, consider a potential firm drawn at random. The number of potential firms in its product market follows the size-biased law $\mathcal{N}^F_{a_i}$, and its competitors' productivity and activity draws generate the same exact distribution $\mathcal{B}_{a_i}$ faced by a private firm. Conditional on own productivity z, both the planner and the firm must therefore integrate over $B_{a_i} \sim \mathcal{B}_{a_i}$ rather than condition their vacancy or output choice on its realization.

The social planner seeks to maximize aggregate consumption by leveraging two sets of control variables. First, the planner mandates each firm with productivity z and skill requirement a operating in industry i to post a mass of vacancies $v_{a_i}(z)$. Second, the planner chooses the probability $\mathfrak{q}_{a_i}(z, z')$ with which a worker with skill a in industry i employed in a firm with productivity z transitions into a firm with productivity z' if given the chance. These transition probabilities give rise to a *shadow wage offer distribution* $\mathcal{H}_{a_i}(z)$ as defined by (53). Just like 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 stationary employment constraint (50) equals the flow-consistent firm-level employment. This expression has the property that $M_{a_i} \int \ell_{a_i}(z) d\Gamma_{a_i}(z) = 1 - u_{a_i}(\lambda_{a_i})$, thus effectively representing a (frictional) labor resource constraint.

We now substitute for the tightness constraint (52) into the Beveridge curve (51) and plug the latter into the objective function. Further substituting for the adding-up constraint (54), the Lagrangian associated to the social planner problem reads:[45]

$$\mathcal{L}(\vec{v}_{a_i}(z), \vec{q}_{a_i}) = Y\left(\overrightarrow{az\ell}_{a_i}\left(\vec{v}_{a_i}(z), \vec{\mathcal{H}}_{a_i}(z)\right)\right) - \sum_{i=1}^{I} L_i \int a \Bigg[M_{a_i} \int \left(c_{a_i}\left(\vec{v}_{a_i}(z), \vec{\mathcal{H}}_{a_i}(z)\right) - \xi^v_{a_i}(z) v_{a_i}(z) \right) d\Gamma_{a_i}(z) \Bigg] d\Xi_i(a),$$

where $\xi^v_{a_i}(z) \geq 0$ is the Kuhn–Tucker multiplier attached to the non-negativity constraint on vacancies.

At a candidate constrained optimum, a necessary condition for the transition probability from a firm with productivity z to a firm with productivity $z' > z$ in labor market a_i is the

[45] We assume that the transition probabilities respect transitivity, i.e., $\mathfrak{q}_{a_i}(z, z') = \mathfrak{q}_{a_i}(z', z'') = 1 \implies \mathfrak{q}_{a_i}(z, z'') = 1 \quad \forall$ distinct z, z', z''. Since the planner selects a complete and transitive pairwise ordering, for every pair $z < z'$ we retain $\mathfrak{q}_{a_i}(z, z')$ as the independent control and substitute $\mathfrak{q}_{a_i}(z', z) = 1 - \mathfrak{q}_{a_i}(z, z')$. For equal-productivity pairs, the orientation is immaterial; since the productivity distribution is continuous, such pairs have measure zero. The pairwise perturbations below are understood to preserve transitivity; the strict ordered corner that they identify satisfies this restriction.

following Kuhn–Tucker condition:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathfrak{q}_{a_i}(z,z')} &\propto \left[mpl_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial \mathfrak{q}_{a_i}(z,z')} - \frac{\partial c_{a_i}(z)}{\partial \mathfrak{q}_{a_i}(z,z')}\right]\gamma_{a_i}(z) - \left[mpl_{a_i}(z')\frac{\partial \ell_{a_i}(z')}{\partial \mathfrak{q}_{a_i}(z',z)} - \frac{\partial c_{a_i}(z')}{\partial \mathfrak{q}_{a_i}(z',z)}\right]\gamma_{a_i}(z') \\ &\propto \left[mpl_{a_i}(z')\frac{d\tilde{\ell}_{a_i}(z')}{d\mathcal{H}_{a_i}(z')} - mpl_{a_i}(z)\frac{d\tilde{\ell}_{a_i}(z)}{d\mathcal{H}_{a_i}(z)}\right] + \frac{\phi}{1+\zeta}s_{a_i}\lambda_{a_i}\bar{c}_{a_i}\left[\frac{v_{a_i}(z')^{\zeta}}{h_{a_i}(z')^{1-\phi}} - \frac{v_{a_i}(z)^{\zeta}}{h_{a_i}(z)^{1-\phi}}\right] \\ &\geq 0, \end{aligned} \tag{55}$$

where $mpl_{a_i}(z) \equiv \frac{1}{a}\frac{\partial Y}{\partial \ell_{a_i}(z)}$ is the firm's marginal product of labor in piece rate units and $\tilde{\ell}_{a_i} \equiv \ell_{a_i}/v_{a_i}$ is the vacancy yield (independent of v_{a_i}). The second line follows from $\partial \mathcal{H}_{a_i}(z)/\partial \mathfrak{q}_{a_i}(z,z') = -M_{a_i}v_{a_i}(z')\gamma_{a_i}(z')/V_{a_i}$, $\partial \mathcal{H}_{a_i}(z')/\partial \mathfrak{q}_{a_i}(z,z') = M_{a_i}v_{a_i}(z)\gamma_{a_i}(z)/V_{a_i}$, and the functional form of hiring costs. Paralleling the rank-preservation condition in the decentralized equilibrium, we maintain that the planner's transition index is strictly increasing in productivity. Hence, for every $z < z'$, $\frac{\partial \mathcal{L}}{\partial \mathfrak{q}_{a_i}(z,z')} > 0$, and the optimal transition rule is the ordered corner solution $\mathfrak{q}_{a_i}(z,z') = 1 \iff z < z'$. Under this rule, the shadow wage-offer distribution coincides with the vacancy-weighted productivity distribution: $\mathcal{H}_{a_i}(z) = 1 - M_{a_i}\int_z^{\bar{z}_{a_i}} \frac{v_{a_i}(z')}{V_{a_i}}\, d\Gamma_{a_i}(z')$.

At a candidate constrained optimum, the necessary Kuhn–Tucker conditions for vacancy posting by a firm with productivity z operating in labor market a_i are:

$$\begin{aligned}\mathcal{G}_{a_i}(z) \equiv\ & aM_{a_i}\left[mpl_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} - \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)}\right]\gamma_{a_i}(z) \\ &+ aM_{a_i}\int_{\underline{z}_{a_i}}^{z}\left[mpl_{a_i}(z')\frac{\partial \ell_{a_i}(z')}{\partial \mathcal{H}_{a_i}(z')} - \frac{\partial c_{a_i}(z')}{\partial \mathcal{H}_{a_i}(z')}\right]\frac{\partial \mathcal{H}_{a_i}(z')}{\partial v_{a_i}(z)}d\Gamma_{a_i}(z') \\ &+ aM_{a_i}\int_{\underline{z}_{a_i}}^{\bar{z}_{a_i}}\left[mpl_{a_i}(z')\left(\frac{\partial \ell_{a_i}(z')}{\partial V_{a_i}} + \frac{\partial \ell_{a_i}(z')}{\partial \lambda_{a_i}}\frac{\partial \lambda_{a_i}}{\partial V_{a_i}} + \frac{\partial \ell_{a_i}(z')}{\partial \mathcal{H}_{a_i}(z')}\frac{\partial \mathcal{H}_{a_i}(z')}{\partial V_{a_i}}\right)\right. \\ &\left. - \left(\frac{\partial c_{a_i}(z')}{\partial \lambda_{a_i}}\frac{\partial \lambda_{a_i}}{\partial V_{a_i}} + \frac{\partial c_{a_i}(z')}{\partial \mathcal{H}_{a_i}(z')}\frac{\partial \mathcal{H}_{a_i}(z')}{\partial V_{a_i}}\right)\right]\frac{\partial V_{a_i}}{\partial v_{a_i}(z)}d\Gamma_{a_i}(z') \leq 0, \end{aligned} \tag{56}$$

$$\mathcal{G}_{a_i}(z)v_{a_i}(z) = 0, \tag{57}$$

where $\mathcal{G}_{a_i}(z)$ is the marginal effect of vacancies on the planner's objective, net of the non-negativity multiplier. Equation (56) is the Kuhn–Tucker inequality with respect to vacancies, and (57) is the corresponding complementary slackness condition. For later use, define the marginal effect per potential firm, in a-normalized units, as $\widehat{\mathcal{G}}_{a_i}(z) \equiv \frac{\mathcal{G}_{a_i}(z)}{aM_{a_i}\gamma_{a_i}(z)}$ for $\gamma_{a_i}(z) > 0$.

The first line of (56) represents the *direct effect* of $v_{a_i}(z)$ 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 (56) 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.[46]

The third and fourth lines of (56) represent the *congestion effects* of $v_{a_i}(z)$ 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 $v_{a_i}(z)$ on the job finding rate – again mediated by the induced change in aggregate vacancies and labor market tightness.

Notice that, conditional on a candidate allocation, the direct effect of a marginal increase in $v_{a_i}(z)$ on the shadow wage offer distribution is linear in the respective vacancy perturbation, holding aggregate vacancies fixed. Likewise, its effect on aggregate vacancies is linear and does not depend on the firm's vacancy mass. This linearity property carries two important insights. First, the business-stealing effects of vacancy posting by firms in a labor market a_i vary by the firm's productivity *rank*. Second, the congestion effects are common across all firms operating in a given labor market.

Let $E^{bs}_{a_i}(z)$ and $E^{c}_{a_i}$ denote the business-stealing effects and congestion effects of $v_{a_i}(z)$ on value added, respectively. It follows that active firms at an interior candidate optimum satisfy:

$$mpl_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} = \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} + E^{bs}_{a_i}(z) + E^{c}_{a_i}. \tag{58}$$

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

Relation to the Equilibrium Allocation. In equilibrium, wages – rather than marginal products – play an allocative role via worker reallocation. Hence, for the candidate 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 decomposition of marginal costs into wages and markdowns.[47] However, comparing (58) with its equilibrium counterpart, it is apparent that both feature an "additive markdown", i.e., the difference between marginal (revenue) product and wage, defined in a-normalized units 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 $w^{\star\star}_{a_i}(z) \equiv \frac{E^{bs}_{a_i}(z)+E^{c}_{a_i}}{\partial \ell_{a_i}(z)/\partial v_{a_i}(z)}$ 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 candidate allocation can be implemented by

[46] 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.

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

dictating:[48]

$$\mu_{k(a_i)}(z) = 1, \tag{59}$$

$$\psi_{a_i}(z) = 1 + \frac{(1+\zeta)c_{a_i}(z)}{w_{a_i}(z)\ell_{a_i}(z)}, \tag{60}$$

$$w_{a_i}(z) = \frac{E^{bs}_{a_i}(z) + E^{c}_{a_i}}{\partial \ell_{a_i}(z)/\partial v_{a_i}(z)}. \tag{61}$$

Equation (59) neutralizes the Bayes-Cournot markup wedge, thereby equalizing expected MRPL and the social marginal product of labor; equation (60) makes sure that markdowns are efficient for a given wage policy; and equation (61) makes sure that the wage policy induces efficient vacancy posting by equalizing social and private marginal costs.

We conclude this section by singling out the sources of inefficiency featured by the baseline equilibrium and how the introduction of a MW may affect them. For consistency with the social planner problem setup, we further assume that workers set their reservation wage statically, so that the equilibrium reservation wage boils down to the *static* reservation wage $\hat{R}^u_{a_i} : a\hat{R}^u_{a_i} + \Pi(a\hat{R}^u_{a_i}) + T = ab_{a_i}$. The effective MW, $\underline{w}_{a_i}$, is the maximum of the static reservation wage and the mandated MW.

We carry out the comparison of the equilibrium wage function with its efficient counterpart (61) in two steps. First, we show that the equilibrium wage function internalizes the business-stealing effects of vacancy posting if the effective MW 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 due to positive markup rates and congestion effects.[49]

Lemma A.5 (Wage posting internalizes business-stealing effects)
Suppose that transitions are efficiently ordered by productivity, product-market distortions are neutralized so that $\widetilde{z}_{a_i}(z) = mpl_{a_i}(z)$, aggregate congestion effects are absent, and the effective minimum wage is zero. Then the equilibrium wage schedule internalizes the business-stealing effect of vacancy posting. Formally,

$$v_{a_i}(z) = v^{\star}_{a_i}(z) \implies mpl_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} = \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} + E^{bs}_{a_i}(z),$$

where $v^{\star}_{a_i}(z)$ is the equilibrium vacancy policy function.

Let $SMB_{a_i}(z) \equiv mpl_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)}$ and $SMC_{a_i}(z) \equiv \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} + E^{bs}_{a_i}(z) + E^{c}_{a_i}$ denote the social

[48] Here $\mu_{k(a_i)}(z) = 1$ is a pricing instruction imposed by the planner in a decentralization, not a restriction on the structural ratio of expected price to expected marginal-revenue product under the unregulated inverse-demand system.

[49] 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.

marginal benefit and social marginal cost of vacancy posting. Whenever $SMC_{a_i}(z) > 0$, define the firm-specific *labor wedge* as

$$\Delta_{a_i}(z) \equiv \frac{SMB_{a_i}(z)}{SMC_{a_i}(z)}. \tag{62}$$

At an interior candidate optimum, $\Delta_{a_i}(z) = 1$ for every active type. Inactive types satisfy the Kuhn–Tucker inequality $\Delta_{a_i}(z) \leq 1$. Define the aggregate labor wedge as

$$\overline{\Delta}^{v} \equiv \frac{\sum_{i=1}^{I} L_i \int aM_{a_i} \int v_{a_i}(z) SMB_{a_i}(z)\, d\Gamma_{a_i}(z)\, d\Xi_i(a)}{\sum_{i=1}^{I} L_i \int aM_{a_i} \int v_{a_i}(z) SMC_{a_i}(z)\, d\Gamma_{a_i}(z)\, d\Xi_i(a)}. \tag{63}$$

Formally, $\overline{\Delta}^{v}$ is the ratio between the vacancy-weighted aggregate social marginal benefit and social marginal cost of vacancy posting. Let $\mathcal{B}^v$ and $\mathcal{C}^v > 0$ denote, respectively, the numerator and denominator of (63), so that $\overline{\Delta}^{v} = \mathcal{B}^v / \mathcal{C}^v$.

To give this index a directional-welfare interpretation, consider the proportional perturbation $v^{\varepsilon}_{a_i}(z) = (1+\varepsilon) v_{a_i}(z)$ of all active vacancies, holding the activity rule and the worker transition rule fixed. Since $\left.\frac{dv^{\varepsilon}_{a_i}(z)}{d\varepsilon}\right|_{\varepsilon=0} = v_{a_i}(z)$, the directional derivative of the planner's objective along this proportional perturbation is the vacancy-weighted sum of its firm-level marginal derivatives. By construction, $SMB_{a_i}(z) - SMC_{a_i}(z)$ is the total social marginal return to an additional vacancy, including the business-stealing and congestion effects embedded in $E^{bs}_{a_i}(z)$ and $E^{c}_{a_i}$. Therefore,

$$\begin{aligned}
\left.\frac{d\mathcal{W}(\mathbf{v}^{\varepsilon})}{d\varepsilon}\right|_{\varepsilon=0} &= \sum_{i=1}^{I} L_i \int aM_{a_i} \int \left.\frac{dv^{\varepsilon}_{a_i}(z)}{d\varepsilon}\right|_{\varepsilon=0} \left[SMB_{a_i}(z) - SMC_{a_i}(z)\right] d\Gamma_{a_i}(z)\, d\Xi_i(a) \\
&= \sum_{i=1}^{I} L_i \int aM_{a_i} \int v_{a_i}(z) \left[SMB_{a_i}(z) - SMC_{a_i}(z)\right] d\Gamma_{a_i}(z)\, d\Xi_i(a) \\
&= \mathcal{B}^v - \mathcal{C}^v = \mathcal{C}^v \left(\overline{\Delta}^{v} - 1\right).
\end{aligned} \tag{64}$$

Because $\mathcal{C}^v > 0$, the aggregate wedge is above (below) one if and only if a sufficiently small proportional increase (decrease) in all active vacancies raises the planner's objective locally. This is a directional efficiency statement: $\overline{\Delta}^{v} = 1$ eliminates the welfare gain from a common proportional change in vacancies, but does not by itself rule out gains from reallocating vacancies across firms.

Lemma A.6 (Equilibrium labor wedge)
Suppose that worker transitions are efficiently ordered by productivity. Define the product-market markup wedge as $\mathfrak{m}^{\mu}_{a_i}(z) \equiv mpl_{a_i}(z) - \widetilde{z}_{a_i}(z) = \frac{\mu_{k(a_i)}(z)-1}{\mu_{k(a_i)}(z)} mpl_{a_i}(z)$, *the accumulated lower-rank markup wedge as* $\mathcal{E}^{\mu}_{a_i}(z) \equiv \frac{1}{\nu}\int_{\underline{z}_{a_i}}^{z} \mathfrak{m}^{\mu}_{a_i}(\hat{z}) \tilde{\ell}_{a_i}(\hat{z}) x_{a_i}(\hat{z})\, d\hat{z}$, *and the minimum-wage attenuation term as* $\Lambda^{u}_{a_i} \equiv \tilde{\ell}_{a_i}(\underline{z}_{a_i})$. *Finally, let* $MCL_{a_i}(z) \equiv \widetilde{z}_{a_i}(z) = mpl_{a_i}(z)/\mu_{k(a_i)}(z)$ *denote equilibrium marginal*

labor cost. Then the normalized marginal welfare effect of equilibrium vacancy posting is:

$$\widehat{\mathcal{G}}_{a_i}(z)\Big|_{v_{a_i}(z)=v^\star_{a_i}(z)} = \frac{\mu_{k(a_i)}(z)-1}{\mu_{k(a_i)}(z)} mpl_{a_i}(z)\tilde{\ell}_{a_i}(z) - \mathcal{E}^{\mu}_{a_i}(z) + \Lambda^{u}_{a_i}\underline{w}_{a_i} - E^{c}_{a_i}. \tag{65}$$

At the equilibrium allocation, the firm-specific labor wedge satisfies

$$\Delta^\star_{a_i}(z) = \frac{\mu_{k(a_i)}(z)}{1+\dfrac{\mathcal{E}^{\mu}_{a_i}(z)+E^{c}_{a_i}-\Lambda^{u}_{a_i}\underline{w}_{a_i}}{MCL_{a_i}(z)\tilde{\ell}_{a_i}(z)}}. \tag{66}$$

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 of labor (wages times markdowns) are lower than efficient. Since congestion effects show up as an additive wedge in the first-order condition 58 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).

Abstracting from endogenous activity choices, the baseline equilibrium is inefficient for two reasons. First, positive markup rates (due to imperfect substitutability across firm-level varieties) and distorted markdowns (due to congestion effects and positive static reservation wages) make the labor wedge generally differ from one in equilibrium. Positive markup rates affect the labor wedge through two opposing channels. The direct markup term raises the wedge through the numerator of (66), while the accumulated markup externality generated by lower-ranked firms raises its denominator. The net effect is generally ambiguous (except for the lowest rank). At the aggregate level, equation (64) provides the appropriate criterion: if $\overline{\Delta}^{v} > 1$, value added rises locally when all active firms proportionally increase vacancies; if $\overline{\Delta}^{v} < 1$, it rises under a proportional decrease.

Second, heterogeneous markups (due to firms' granularity in their product market) and heterogeneous markdown distortions (due to the heterogeneous proportional incidence of common congestion and static reservation-wage terms) make the labor wedge generally differ across firms in equilibrium. This means that the economy features *misallocation* of labor across firms: an infinitesimal reallocation of a fixed amount of social marginal vacancy cost from lower- to higher-wedge firms raises aggregate consumption. Quantitatively, we measure this distortion using aggregate hiring costs per unit of final output. Hence, implementing the candidate constrained-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-bill tax/subsidy to offset the reservation-wage component of the markdown wedge.

However, the information requirements to implement such first-best policies are arguably

beyond policymakers' capabilities. On the other hand, policymakers can directly control the effective minimum wage through MW setting. Since, for standard parametrizations, congestion effects are positive, a suitably chosen MW can, in principle, offset the congestion effects, so that $E^c_{a_i} - \Lambda^u_{a_i}\underline{w}_{a_i} = 0$ in the labor wedge. By constraining the optimal vacancy posting of low-productivity firms disproportionately more, a higher MW is likely to reduce labor misallocation.

Proof of Lemma A.5 To establish the claim, we define a function $\mathcal{F}_{a_i}(v_{a_i}(z)) \equiv mpl_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} - \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} - E^{bs}_{a_i}(z)$, 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 $mpl_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} - \frac{\partial c_{a_i}(z)}{\partial v_{a_i}(z)} - w_{a_i}(z)\frac{\partial \ell_{a_i}(z)}{\partial v_{a_i}(z)} = 0$. First, note that the business-stealing effects induced by the lowest-productivity firms are zero, i.e., $E^{bs}_{a_i}(\underline{z}_{a_i}) = 0$. On the other hand, the lowest-productivity firms pay the effective minimum wage in equilibrium. As long as the effective minimum 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:

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

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

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

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

ing three terms:

$$\mathcal{F}_{a_i}(z) \propto \frac{w_{a_i}(z)}{h_{a_i}(z)^2} - \int_{\underline{z}_{a_i}}^{z} \frac{w_{a_i}(\hat{z})}{h_{a_i}(\hat{z})^2} \frac{x_{a_i}(\hat{z})}{\nu} d\hat{z} - \int_{\underline{z}_{a_i}}^{z} \frac{(mpl_{a_i}(\hat{z}) - w_{a_i}(\hat{z}))}{h_{a_i}(\hat{z})^2} x_{a_i}(\hat{z}) d\hat{z},$$

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

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

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

where the last equality follows from the slope of the equilibrium wage function and by the assumption that $w_{a_i}(\underline{z}_{a_i}) = 0$. Substituting back into the previous expression allows establishing the claim:

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

■

Proof of Lemma A.6 To simplify notation, suppress the labor-market subscript a_i in the proof; in particular, write $\mu(z) \equiv \mu_{k(a_i)}(z)$. At the equilibrium allocation, $\widetilde{z}(z) = \frac{mpl(z)}{\mu(z)}$, $mpl(z) - \widetilde{z}(z) = \frac{\mu(z)-1}{\mu(z)} mpl(z) \equiv \mathfrak{m}^{\mu}(z)$. The private vacancy-posting first-order condition implies

$$\frac{\partial c(z)}{\partial v(z)} = [\widetilde{z}(z) - w(z)]\, \tilde{\ell}(z). \tag{67}$$

The planner's normalized derivative with respect to a marginal vacancy, evaluated at the equilibrium allocation, can be written as

$$\widehat{\mathcal{G}}(z)\Big|_{v(z)=v^{\star}(z)} = mpl(z)\tilde{\ell}(z) - \frac{\partial c(z)}{\partial v(z)} - E^{bs}(z) - E^{c}, \tag{68}$$

where $E^{bs}(z)$ denotes the business-stealing effect evaluated using the social marginal product $mpl(z)$, and E^{c} denotes the congestion effect.

Under ordered transitions, the business-stealing effect satisfies

$$\begin{aligned}E^{bs}(z) &= \int_{\underline{z}}^{z} \left[\frac{1}{\nu} mpl(\hat{z}) + \left(1 - \frac{1}{\nu}\right)(\widetilde{z}(\hat{z}) - w(\hat{z}))\right] \tilde{\ell}(\hat{z}) x(\hat{z})\, d\hat{z} \\ &= \int_{\underline{z}}^{z} \left[\frac{1}{\nu}\widetilde{z}(\hat{z}) + \left(1 - \frac{1}{\nu}\right)(\widetilde{z}(\hat{z}) - w(\hat{z}))\right] \tilde{\ell}(\hat{z}) x(\hat{z})\, d\hat{z} + \frac{1}{\nu}\int_{\underline{z}}^{z} \mathfrak{m}^{\mu}(\hat{z})\tilde{\ell}(\hat{z}) x(\hat{z})\, d\hat{z}.\end{aligned} \tag{69}$$

where the second equality follows from $mpl(\hat{z}) = \widetilde{z}(\hat{z}) + \mathfrak{m}^{\mu}(\hat{z})$.

Since vacancy yield is proportional to $\tilde{\ell}(z) \propto [\delta + s\lambda\,(1 - \mathcal{H}(z))]^{-2}$, and $x(z) = \frac{2\nu s\lambda \mathcal{H}'(z)}{\delta + s\lambda[1-\mathcal{H}(z)]}$, we obtain $\tilde{\ell}'(z) = \frac{x(z)}{\nu}\tilde{\ell}(z)$.

Combining this and the equilibrium wage equation gives

$$\begin{aligned}\frac{d}{dz}\left[w(z)\tilde{\ell}(z)\right] &= w'(z)\tilde{\ell}(z) + w(z)\tilde{\ell}'(z) = \left[\widetilde{z}(z) - w(z)\right]\tilde{\ell}(z)x(z) + \frac{1}{\nu}w(z)\tilde{\ell}(z)x(z) \\ &= \left[\frac{1}{\nu}\widetilde{z}(z) + \left(1 - \frac{1}{\nu}\right)\left(\widetilde{z}(z) - w(z)\right)\right]\tilde{\ell}(z)x(z). \end{aligned} \tag{70}$$

Integrating (70) from $\underline{z}$ to z, and using $w(\underline{z}) = \underline{w}$, yields

$$\int_{\underline{z}}^{z}\left[\frac{1}{\nu}\widetilde{z}(\hat{z}) + \left(1 - \frac{1}{\nu}\right)\left(\widetilde{z}(\hat{z}) - w(\hat{z})\right)\right]\tilde{\ell}(\hat{z})x(\hat{z})\,d\hat{z} = w(z)\tilde{\ell}(z) - \underline{w}\tilde{\ell}(\underline{z}). \tag{71}$$

Substituting (71) into (69) gives

$$E^{bs}(z) = w(z)\tilde{\ell}(z) - \underline{w}\tilde{\ell}(\underline{z}) + \frac{1}{\nu}\int_{\underline{z}}^{z}\mathfrak{m}^{\mu}(\hat{z})\tilde{\ell}(\hat{z})x(\hat{z})\,d\hat{z}. \tag{72}$$

Substituting equations (67) and (72) into (68) gives

$$\begin{aligned}\widehat{\mathcal{G}}(z)\Big|_{v(z)=v^{\star}(z)} &= mpl(z)\tilde{\ell}(z) - \left[\widetilde{z}(z) - w(z)\right]\tilde{\ell}(z) - w(z)\tilde{\ell}(z) + \underline{w}\tilde{\ell}(\underline{z}) - \frac{1}{\nu}\int_{\underline{z}}^{z}\mathfrak{m}^{\mu}(\hat{z})\tilde{\ell}(\hat{z})x(\hat{z})\,d\hat{z} - E^{c} \\ &= \mathfrak{m}^{\mu}(z)\tilde{\ell}(z) + \underline{w}\tilde{\ell}(\underline{z}) - \frac{1}{\nu}\int_{\underline{z}}^{z}\mathfrak{m}^{\mu}(\hat{z})\tilde{\ell}(\hat{z})x(\hat{z})\,d\hat{z} - E^{c}.\end{aligned}$$

Using the definitions of $\mathfrak{m}^{\mu}(z)$ and Λ^{u}, this proves (65).

It remains to derive the labor wedge. At the equilibrium allocation, the social marginal cost of a vacancy, is $\frac{\partial c(z)}{\partial v(z)} + E^{bs}(z) + E^{c} = \widetilde{z}(z)\tilde{\ell}(z) + \mathcal{E}^{\mu}(z) + E^{c} - \Lambda^{u}\underline{w}$. Since $\widetilde{z}(z) = MCL(z) = \frac{mpl(z)}{\mu(z)}$, we obtain

$$\Delta(z) = \frac{mpl(z)\tilde{\ell}(z)}{MCL(z)\tilde{\ell}(z) + \mathcal{E}^{\mu}(z) + E^{c} - \Lambda^{u}\underline{w}} = \frac{\mu(z)}{1 + \dfrac{\mathcal{E}^{\mu}(z) + E^{c} - \Lambda^{u}\underline{w}}{MCL(z)\tilde{\ell}(z)}}. \tag{73}$$

Restoring labor-market subscripts yields (66). ■

D Structural Estimation

D.1 Data description and sample selection

INPS data. We leverage social security data from INPS. Our dataset consists of the complete contribution histories of individuals who worked as employees in the private sector for at least one period during their lives 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. 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. 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.[50] We do so to avoid including the final stage of education in workers' labor-market histories. We end up with a sample of about 34 million yearly observations, covering 2.6 million distinct workers and 2.3 million distinct firms.

Istat data. Our data from Istat comes from two sources. First, we make use of publicly available data on the distribution of firms across ATECO industry codes and administrative regions, as well as on the employment share of firms of different sizes. Some industries in which the number of observations from the INPS dataset 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 "constructions". To obtain geographic units of more comparable size, small administrative regions are merged with nearby larger regions or with one another. Specifically, "Valle d'Aosta" is merged with "Piemonte", "Trentino - Alto Adige" with "Friuli - Venezia Giulia", "Umbria" with "Toscana", "Molise" with "Abruzzo", and "Basilicata" with "Calabria". Guided by the conventional distinction between traded and local industries based on the geographic reach of their product markets, we classify "manufacturing", "IT", and "transport" as *national* industries, and "trade", "construction", "hospitality", "professional services", and "services to firms" as *local* industries. Accordingly, we define a product market as an ATECO 4-digit sectors for national industries and as an ATECO 4-digit sector-by-region cell for local industries.

Second, we leverage micro-data from the 2019 wave of the Structural Business Statistics to compute the distribution of the number of competing firms across product markets by industry, the HHI of each product market and the semi-elasticity of firm-level labor share to product-market share. The dataset consists of the near-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 treat negative entries for "revenues", "value added", and "labor cost" as missing. We define labor share (LS) as the ratio of "labor cost" to "value added"

[50] We do not exclude women nor part-time workers from our sample: given that these groups of workers are overall paid below average, they therefore constitute a relevant share of the population directly affected by a MW reform.

at the firm-level. To guard against outliers, we winsorize the labor share at 1%. We compute the firm-level market share (ω) in its product market as the ratio of firm-level revenues to sectoral revenues.

D.2 AKM estimation

For each worker, we first select the dominant (= highest-earnings) job spell within each year. We then compute average hourly wages by taking the ratio of yearly earnings to the imputed number of hours worked (obtained by 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:

$$\log w_{n,t} = \alpha_n + \varphi_{f(n,t)} + \epsilon_{n,t},$$

where α_n's are the worker fixed effects and φ_f's are the firm fixed effects. As in Engbom and Moser (2022), we estimate the AKM equation without further controls. This is consistent with our model, where the only source of heterogeneity is a permanent skill component. For the AKM estimation we use the longest possible time span in our data (1990-2018) to minimize bias arising from the effects of temporary shocks to workers' wage trajectories.

D.3 Computing estimation targets

We use our Istat micro-data from 2019 to compute the HHI of each product market k as $HHI_k = \sum_{j=1}^{N_k} \omega_{jk}^2$. 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:

$$LS_j = \delta_{k(j)} + \beta\, \omega_j + \epsilon_j,$$

where LS_j and ω_j denote the labor share and market share of firm j, respectively, and the δ_k's are product-market fixed effects. We normalize the estimated coefficient, $\hat{\beta} = -0.23$, by the average labor share (0.644) to obtain the semi-elasticity of the firm-level labor share with respect to market share, which we use as a target in the structural estimation.[51] The identifying assumption underlying this target is that, for a given distribution of markdowns, firms within a product market operate sufficiently similar technologies. Under this assumption, within-product-market variation in normalized labor shares disciplines the product-market-power gradient across firms.

We partition our sample from INPS in two subsamples, each of which corresponds to half

[51] We target the semi-elasticity of the firm-level labor share with respect to market share, rather than the partial effect in levels, so that the target is independent of the scale of the labor share. Since our model abstracts from production inputs other than labor, the level of the labor share in the model does not directly map to its empirical counterpart.

of the estimated AKM worker fixed effects distribution (low- and high-skilled workers). 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-by-skill level as targets in the structural estimation. Finally, we measure worker transition rates separately for each subgroup. In particular, we measure the EN rate, the job-to-job (J2J) rate, and the NE rate as the Poisson rates consistent with the monthly probability of moving from employment to non-employment, switching to another employer, and of moving from non-employment to employment, respectively.

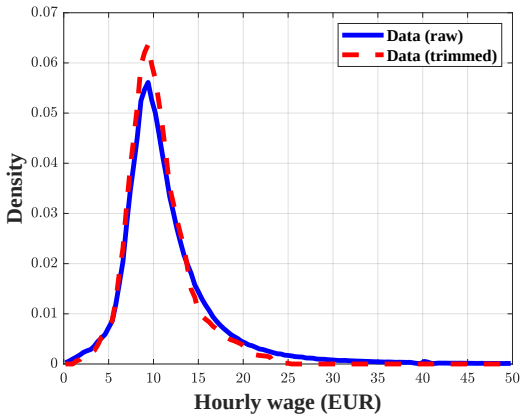
D.4 Estimating productivity from wages

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.

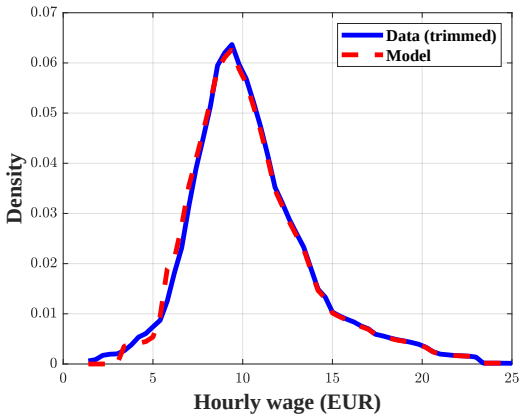
Trimming wage distributions. We trim the lowest 1% and the highest 10% of each wage distribution to guard against outliers. The structure of our model predicts that firms commit to a wage rate for all their employees and that larger firms pay higher wages: it follows that a small share of high observed wages (e.g., managers) would be associated with extremely big firms, whose size would be unrealistic. This observation suggests removing some wages at the top of the distribution. Figure A.2a reports the difference between the raw and the trimmed empirical wage distribution.

Figure A.2: Aggregate wage distribution - data vs model

(a) Trimmed wage distribution



(b) Model-implied wage distribution



Source: INPS data (2016-2018) and model. *Note*: Panel (a) compares the raw and the trimmed economy's employment wage distribution ($G(w)$); Panel (b) compares the trimmed and the model-implied economy's employment wage distribution. The economy's wage distribution in the data aggregates the 16 trimmed labor-market-specific wage distributions.

Cleaning wage distributions. For our model to replicate the trimmed empirical wage distributions, we pass them through the following cleaning algorithm.

Notice that the density of the MRPL distribution is structurally related to firms' wage and vacancy policies by the following relation:

$$\varphi_{a_i}(\tilde{z}) = \frac{\mathcal{H}'_{a_i}(\tilde{z})}{M_{a_i}} \frac{V_{a_i}}{v_{a_i}(\tilde{z})}. \tag{74}$$

Since $\mathcal{H}_{a_i}(\tilde{z}) = F_{a_i}(w_{a_i}(\tilde{z}))$, it follows that $\mathcal{H}'_{a_i}(\tilde{z}) = w'_{a_i}(\tilde{z})\, f_{a_i}(w_{a_i}(\tilde{z}))$. Following Bontemps et al. (2000), one can derive the following expression for $\mathcal{H}'_{a_i}(\tilde{z})$:

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

where the dependence of w on $\tilde{z}$ is left implicit. Turning back to Equation (74), we notice that $\varphi_{a_i}(\tilde{z}) \geq 0 \iff \mathcal{H}'_{a_i}(\tilde{z}) \geq 0$, provided that $v_{a_i}(\tilde{z}) > 0$. 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:

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

Hence, for each labor market, we check this admissibility condition, and marginally shift mass from the (few) points of the empirical wage distributions for which it does not hold. 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 following simple algorithm. Let $G^D_{a_i}(w)$ 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, we then move mass toward $i^Y - 3$;
4. We stop when the first grid point for which the admissibility condition holds is the second one with non-zero mass;
5. Finally, we recheck the condition over the whole distribution. For all points at which the condition is violated, 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. Overall, the distance between the left tails of the empirical (trimmed) wage distributions and those 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. Moreover, we remove grid points where the MRPL density is lower than 10^{-6} to prevent single firms holding approximately unit market share.

Inferring productivity from wages. We use the structure of our model to estimate the underlying productivity distributions from the observed wage distributions in two steps.

First, we leverage the equilibrium markdown function (20) and the mapping between the wage offer distribution and the wage employment distribution to infer the expected 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 expected MRPL distributions, $\tilde{z}_{a_i} \sim \Phi_{a_i}(\psi_{a_i}(w)w)$.

Second, for given CES elasticities of demand, distributions of the number of potential firms, and expected MRPL distributions in each product market, there exists a unique combination of prices, markups, CD industry weights, and physical productivities consistent with our model, which solves the following system of equations:

$$\begin{cases} z(\tilde{z}_{k(a_i)}) = \frac{\mu(\tilde{z}_{k(a_i)})}{\mathbb{E}[p(\tilde{z}_{k(a_i)})]}\tilde{z}_{k(a_i)}, \; p(\tilde{z}_{k(a_i)}) = \left(\frac{y(\tilde{z}_{k(a_i)})}{Y_{k(a_i)}(\tilde{\boldsymbol{z}}_{k(a_i)})}\right)^{-\frac{1}{\sigma}} \left(\frac{Y_{k(a_i)}(\tilde{\boldsymbol{z}}_{k(a_i)})}{Y_i(\tilde{\boldsymbol{z}}_i)/K_i}\right)^{-\frac{1}{\rho}} \left(\frac{Y_i(\tilde{\boldsymbol{z}}_i)}{\alpha_i(\tilde{\boldsymbol{z}})Y(\tilde{\boldsymbol{z}})}\right)^{-1}, \\ \omega(\tilde{\boldsymbol{z}}_{k(a_i)}) = \left(\frac{y(\tilde{z}_{k(a_i)})}{Y_{k(a_i)}(\tilde{\boldsymbol{z}}_{k(a_i)})}\right)^{\frac{\sigma-1}{\sigma}}, \; \mu(\tilde{z}_{k(a_i)}) = \mathbb{E}^p\left[\left(\frac{\sigma}{(\sigma-1)\left[1-\frac{\sigma/\rho-1}{\sigma-1}\omega(\tilde{\boldsymbol{z}}_{k(a_i)})\right]}\right)^{-1}\right]^{-1}, \\ \alpha_i(\tilde{\boldsymbol{z}}) = \frac{P_i(\tilde{\boldsymbol{z}}_i)Y_i(\tilde{\boldsymbol{z}}_i)}{Y(\tilde{\boldsymbol{z}})}, \end{cases} \tag{75}$$

where variables in bold denote vectors at the level of aggregation corresponding to their subscript, and aggregators are defined in (5)-(7). Solving the system of equations (75) allows us to back out the productivity distributions that rationalize the observed wage distributions through the lens of our model, $z_{a_i} \sim \Gamma_{a_i}\left(\frac{\mu_{k(a_i)}(\tilde{z})}{\mathbb{E}[p_{k(a_i)}(\tilde{z})]}\tilde{z}\right)$. Table A.3 shows that the estimated CD industry weights are in line with the empirical revenue shares of the respective industry.

D.5 Estimation results

Table A.1: Sensitivity matrix of estimated parameters

	Emp share 50+ firms	Vacancy rate	Wgt avg HHI	Wgt avg markup rate	AKM sorting share	Semi-elasticity labor share-market share
ζ	-1.18	0.00	-0.00	0.00	0.00	0.00
χ	0.00	-0.73	-0.00	0.00	0.00	0.00
ϕ	-3.23	-0.00	1.81	-0.01	-0.00	-0.84
σ	-0.07	-0.00	0.60	-1.74	0.00	0.67
ρ	0.38	-0.00	0.15	-0.14	0.00	-0.57
a_H	2.42	0.00	-1.72	0.01	0.08	0.80

Source: Model. *Note*: Let $\hat{\mathbf{m}}(\Theta) \in \mathbb{R}^M$ be the simulated model moments, $\mathbf{m}^D \in \mathbb{R}^M$ the targeted empirical moments, and $\Theta \in \mathbb{R}^P$ the model parameters. Define the vector of relative moment deviations as: $g(\Theta) := \frac{\hat{\mathbf{m}}(\Theta) - \mathbf{m}^D}{\mathbf{m}^D} \in \mathbb{R}^M$. The minimum distance estimator solves: $\Theta^* = \arg\min_\Theta L(\Theta), \quad \text{where} \quad L(\Theta) := g(\Theta)' \mathcal{W} g(\Theta)$, where $\mathcal{W}$ is the identity matrix. We define the Jacobian of the relative moment deviations as: $J(\Theta, \mathbf{m}^D) := \nabla_\Theta g(\Theta) = \text{diag}\left(\frac{1}{\mathbf{m}^D}\right) \nabla_\Theta \hat{\mathbf{m}}(\Theta)$. The sensitivity of Θ^* with respect to $\mathbf{m}^D$ proposed by Andrews et al. (2017) is: $M_1 \equiv \frac{d\Theta^*}{d\mathbf{m}^D} \approx (J'\mathcal{W}J)^{-1} J'\mathcal{W} \cdot \text{diag}\left(\frac{\hat{\mathbf{m}}(\Theta)}{\mathbf{m}^{D2}}\right)$. To facilitate interpretation, we report the elasticity of estimated parameters with respect to moments: $M_1^{\text{elasticity}} = M_1 \frac{\mathbf{m}^D}{\Theta^*}$, which expresses how a 1% change in a moment affects each parameter estimate in percentage terms.

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	0.085	0.089
L_{a_i}	0.150	0.165	0.044	0.044	0.095	0.084	0.049	0.040	0.066	0.053	0.012	0.020	0.019	0.020	0.079	0.059

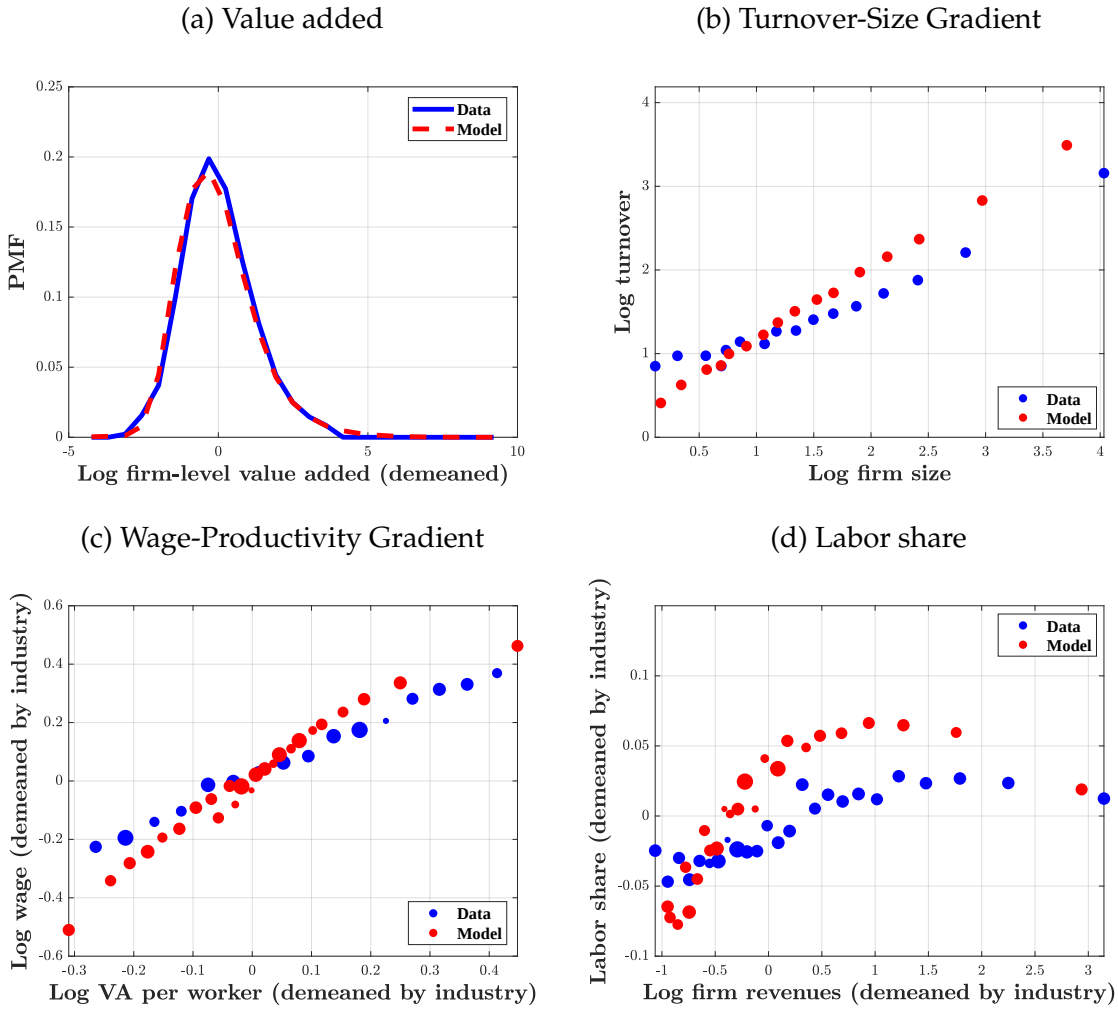
Source: Model.

Table A.3: Industry weights and revenue shares

Industry	CD weight (model)	Revenue share (data)
Manufacturing	0.345	0.349
Constructions	0.088	0.111
Trade	0.181	0.179
Transport	0.090	0.087
Hospitality	0.090	0.046
IT	0.040	0.066
Professional services	0.044	0.062
Services to firms	0.121	0.100

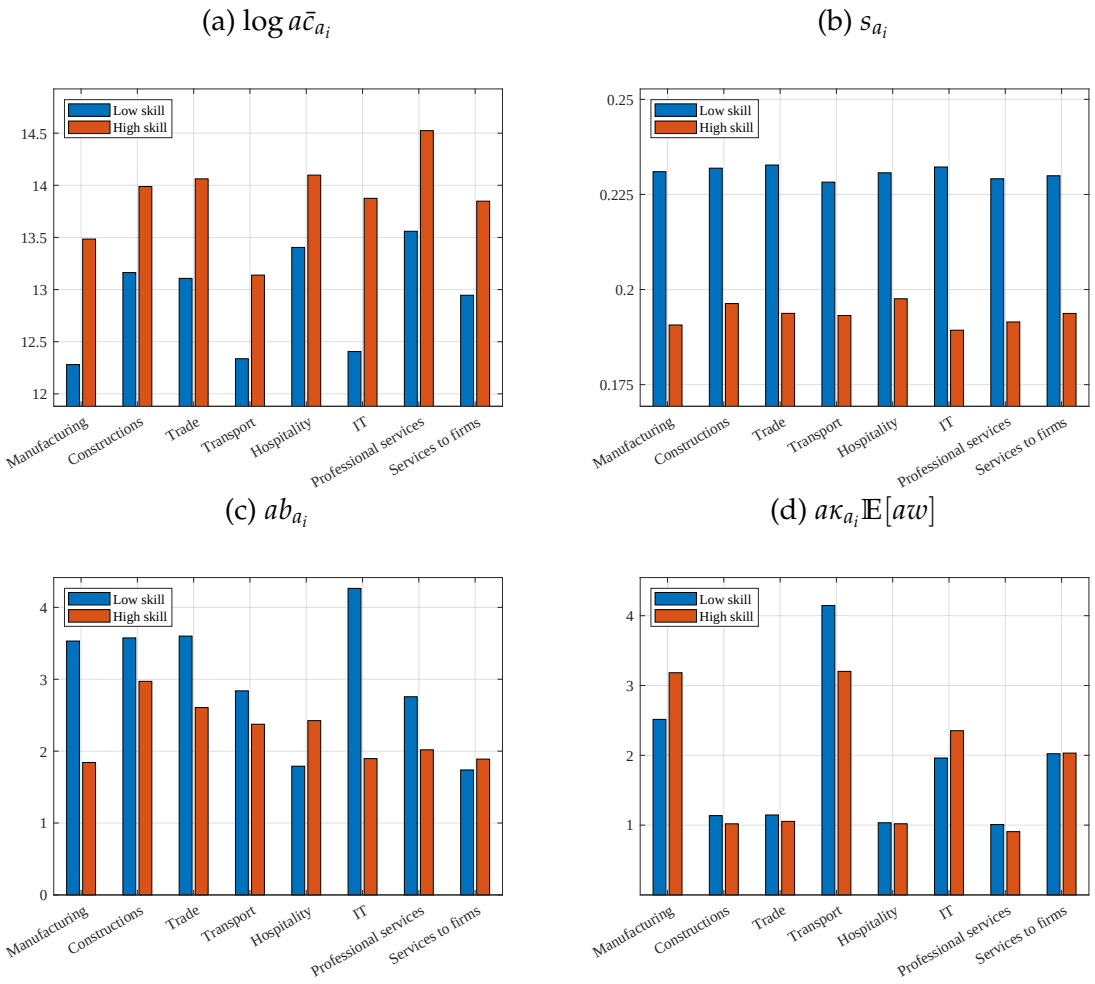
Source: Model and Istat. *Note*: The table reports the model-implied Cobb-Douglas industry weights and the empirical industry share of value added. The empirical values are computed on the Structural Business Statistics dataset of Istat (2019). We map model-based revenue shares into empirical value added shares to net out the effect of goods and services used as intermediate inputs in other industries.

Figure A.3: Untargeted moments



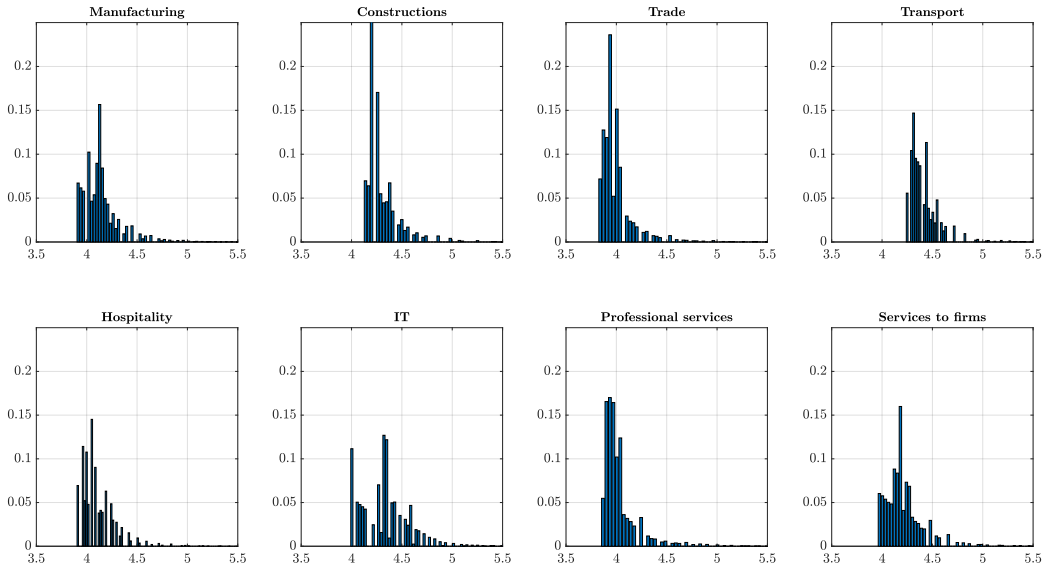
Source: Istat Structural Business Statistics (2019), INPS matched employer-employee data (2016–2018), model. *Note*: All empirical variables are trimmed at 1%. Panel (a) compares the empirical and model-based firm-level distributions of demeaned log value added. Demeaning is performed by subtracting the aggregate mean from each observation. Panel (b) compares the relationship between log turnover and log firm size. Turnover is measured as the average sum of yearly worker inflows to and outflows from a given firm. Dots correspond to equal-mass bins of firm size on the common support of the empirical and model distributions. Panels (c)–(d) compare, respectively, the relationship between demeaned log wages and demeaned log value added per worker, and between demeaned labor share and demeaned log revenues, at the firm level. For these panels, demeaning is performed by subtracting the weighted 1-digit industry mean of each variable from each observation. Data and model observations are restricted to their common support and separately grouped into equal-mass bins. Dot size reflects the relative weight of observations in each bin.

Figure A.4: Labor market parameters, by labor market



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

Figure A.5: Physical productivity distribution, by industry

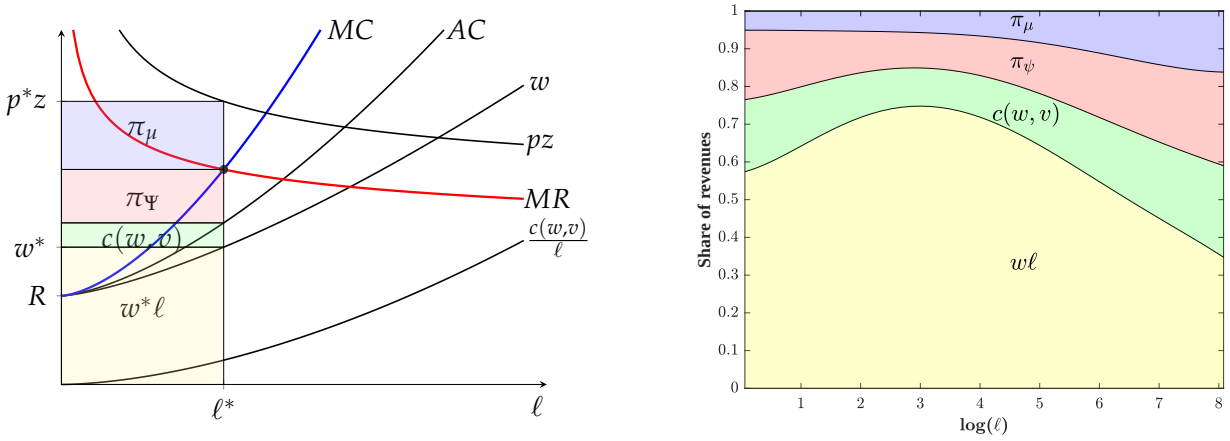


Source: Model. *Note*: The figure plots the model-implied log-physical productivity distributions across 1-digit industries (ATECO classification).

Figure A.6: Revenue shares along the firm size distribution

(a) Revenue decomposition

(b) Model-implied revenue shares



Source: Authors' illustration, model. *Note*: Panel (a) illustrates the theoretical decomposition of firm-level revenues into the wage bill, hiring costs, profits from labor market power, and profits from product market power. Panel (b) reports the average composition of firm-level revenues across log employment in the estimated model. At the firm level, the share of profits from product market power, π_μ, equals $\frac{\mu_{k(a_i)}(z)-1}{\mu_{k(a_i)}(z)}$; the share of profits from labor market power, π_ψ, equals $\left[(\psi_{a_i}(z)-1)-\frac{c_{a_i}(z)}{w_{a_i}\ell_{a_i}(z)}\right]\frac{aw_{a_i}\ell_{a_i}(z)}{\mathbb{E}[p_{k(a_i)}(z)]y_{a_i}(z)}$; the share of hiring costs equals $\frac{ac_{a_i}(z)}{\mathbb{E}[p_{k(a_i)}(z)]y_{a_i}(z)}$; and the share of the wage bill equals $\frac{aw_{a_i}(z)\ell_{a_i}(z)}{\mathbb{E}[p_{k(a_i)}(z)]y_{a_i}(z)}$.

D.6 Local uniqueness of the equilibrium path

In Section B.6, we establish existence and uniqueness of the stationary equilibrium under sufficient regularity conditions. We now complement the theoretical derivation with a numerical check that the counterfactual equilibrium branch followed in the simulations is locally unique. Our estimation procedure recovers the baseline rank-preserving allocation by inverting the observed wage distribution and calibrating hiring-cost scalars to reproduce the estimated job-finding rates. The MW counterfactuals then trace the equilibrium response around this recovered allocation. The numerical validation focuses on local uniqueness of the rank-preserving branch followed by the counterfactuals, conditional on its realized activity vector and smooth equilibrium regime.

Specifically, at each MW value $\underline{w}_b$, we consider the undamped discretized log-expected-MRPL fixed-point residual underlying the numerical solver. Let q_b collect the locally free log-expected-MRPL coordinates after excluding coordinates pinned by inequality constraints. Potential-market-size laws, unconditional productivity probabilities, activity probabilities, and the prevailing binding regime are held fixed. This treatment allows a boundary productivity rank to have fractional activity, but holds its activity probability fixed when constructing the local derivative. Writing the resulting residual map as $\mathcal{F}_b(q_b; \underline{w}_b) = 0$, we numerically evaluate $J_b \equiv D_{q_b}\mathcal{F}_b(q_b; \underline{w}_b)$ using two-sided, regime-preserving finite differences. Conditional on the maintained regime, continuous differentiability of $\mathcal{F}_b$ and nonsingularity of J_b imply, by the implicit function theorem, that the equilibrium root is locally unique and varies locally smoothly with the MW.

Table A.4 reports the diagnostics along the simulated path. Re-entering each stored equilibrium into the undamped fixed-point map produces a maximum residual of 3.766×10^{-3}, – well below a reasonable tolerance threshold. All finite-difference perturbations preserve the binding regime, productivity support, and activity probabilities. The smallest singular value is uniformly bounded away from zero ($\min_b \sigma_{\min}(J_b) = 0.8324$), as the minimum relative singular value ($\min_b \sigma_{\min}(J_b)/\sigma_{\max}(J_b) = 0.1612$). The maximum scaled condition number is 2.88. Moreover, the maximum finite-difference stability error is 2.081×10^{-3}, while the smallest ratio of the singular value to this error is 417.94. The estimated distance from singularity is therefore several hundred times larger than the numerical uncertainty associated with the finite-difference approximation.

Overall, the fixed-activity certificate is obtained at every MW value, including the ten policy points featuring fractional activity. These results establish local uniqueness of the rank-preserving, fixed-activity, fixed-regime MRPL branch followed by the simulations. The check therefore complements, rather than replaces or directly tests, the sufficient conditions for global uniqueness established in Section B.6.

Table A.4: Local isolation of the simulated minimum-wage equilibrium branch

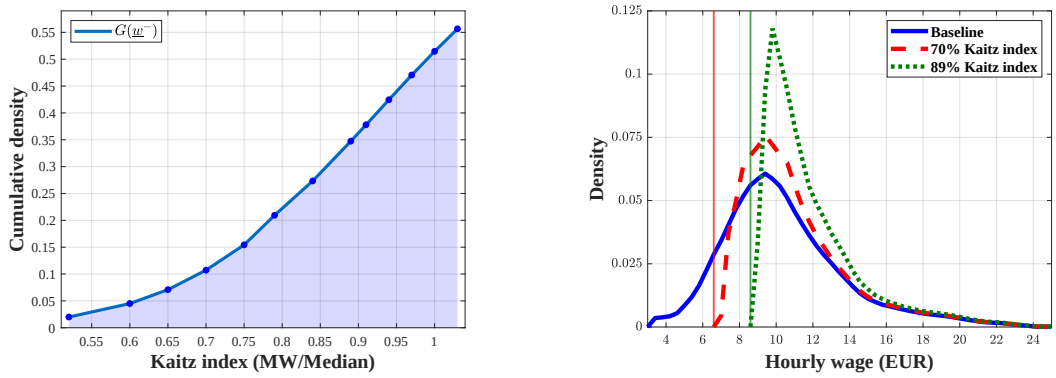
Diagnostic	Value	Interpretation
Minimum $\sigma_{\min}(J)$	$8.324e-01$	Separated from FD error
Minimum relative singular value	$1.612e-01$	Numerically nonzero
Maximum scaled condition number	2.88	Finite
Maximum finite-difference error	$2.081e-03$	Small relative to singular value
Minimum singular-value/FD-error ratio	417.94	Above safety factor
Maximum re-entry residual	$3.766e-03$	Near zero
Fixed regime preserved at every point	Yes	Required
Policy points with fractional activity	10	Allowed and reported
Fixed-activity certificate at every point	Yes	Primary conditional check

Note: J is the Jacobian of the undamped log-expected-MRPL fixed-point residual with respect to locally free coordinates. Potential-market-size laws, unconditional productivity probabilities, activity probabilities, and the binding regime are held fixed. Fractional activity is admissible. The certificate establishes local isolation conditional on that activity vector and regime; it does not test endogenous-activity or global uniqueness.

E Equilibrium Effects of the Minimum Wage

Figure A.7: 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 an 89%-Kaitz-index MW reform (green dotted line).

Table A.5 reports the effects on the main aggregate variables of the two policy experiments considered 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 35^{th} percentile of the initial distribution (89% Kaitz index).

Table A.5: Policy experiments

Variable	Baseline	70% Kaitz index	89% Kaitz index
Panel a. Aggregate statistics			
Aggregate consumption	1.000	1.012	1.023
Aggregate output	1.000	0.998	0.995
Labor share	0.739	0.753	0.763
Aggregate welfare	1.000	1.008	0.992
Unemployment rate	0.135	0.151	0.188
Labor productivity	1.000	1.017	1.061
Average wage	1.000	1.052	1.130
Average firm size	9.313	9.55	11.24
Panel b. Market power statistics			
Aggregate markup	1.090	1.091	1.096
Aggregate markdown	1.466	1.415	1.368
Panel c. Labor market transitions			
Job finding rate	0.161	0.141	0.108
Job separation rate	0.025	0.025	0.025

Source: Model. *Note*: The variables aggregate consumption, aggregate output, aggregate welfare, labor productivity, and average wage are normalized to 1 in the baseline equilibrium.

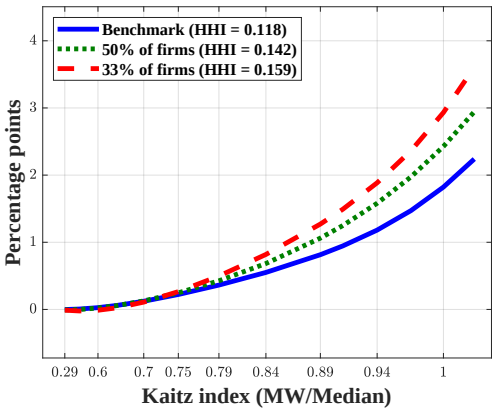
Table A.6: Behavior vs. selection: decomposition of main aggregate effects

Variable	Percent change	Behavioral share	Compositional share
Average wage	12.97%	58.59%	41.41%
Average labor share	3.35%	106.04%	-6.04%
Aggregate markup rate	6.19%	59.71%	40.29%
Aggregate markdown rate	-21.06%	100.03%	-0.03%

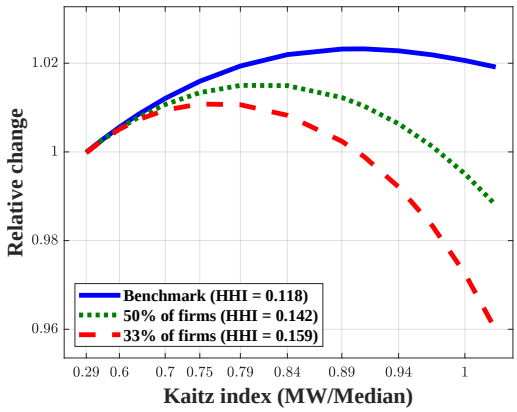
Source: Model. *Note*: The table reports percentage changes between the baseline equilibrium, with no MW, and the consumption-maximizing MW, corresponding to an 89% Kaitz index. Each aggregate variable X corresponds to a weighted average of the firm-specific variable $x(j)$, $X = \int x(j)\,\hat{w}(j)dj$, where j denotes a specific firm and $\hat{w}(j)$ is the firm-specific weight according to the aggregation results reported in Appendix B.7. The change in X is decomposed as $\Delta X = \int \Delta x(j)\,\hat{w}'(j)dj + \int x(j)\,\Delta\hat{w}(j)dj$, where $\hat{w}'(j)$ denotes the weight after the policy change. The first term represents the behavioral effect, while the second represents the compositional effect.

Figure A.8: Heterogeneity by baseline product market concentration

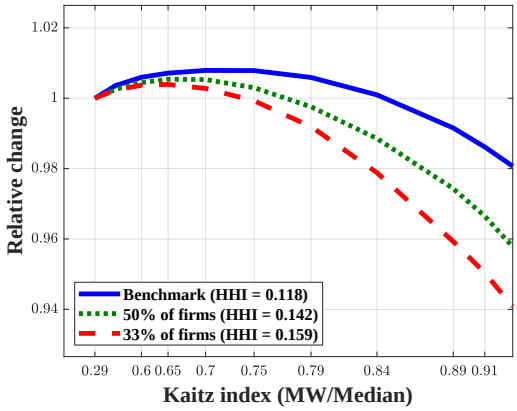
(a) Product market power



(b) Aggregate consumption



(c) Aggregate welfare



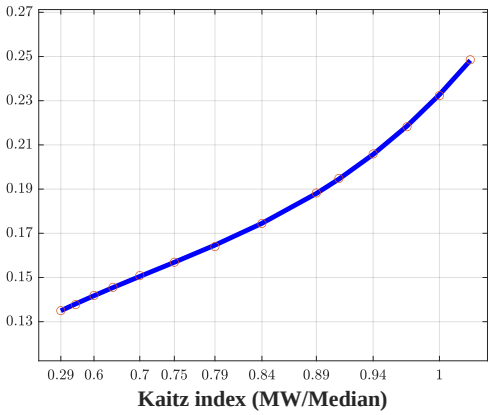
Source: Model. *Note*: The figure compares the economy's response to the MW depending on the number of firms per product market in the baseline equilibrium without a MW. The solid blue line represents our estimated benchmark economy with the empirical distribution of firms per sector; the dotted green line represents an economy with the same productivity distribution but a uniform rescaling of the support of the distribution of firms per sector by one-half; the dot-dashed red line represents an economy with the same productivity distribution but a uniform rescaling of the support of the distribution of firms per sector by one-third. Each economy is denoted by its revenue-weighted average sectoral Herfindahl-Hirschman Index (HHI) absent a MW. The measure of sectors K_{a_i} is adjusted so that the measure of potential firms M_{a_i} remains unchanged across the benchmark economy and the alternative economies.

E.1 Nominal Minimum Wage

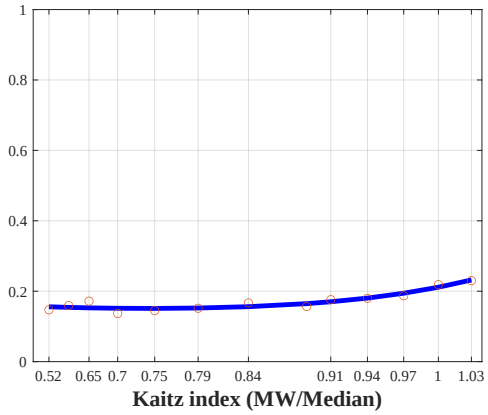
In the main text, the MW is expressed in units of the final good. As a robustness exercise, we instead impose a nominal MW and hold nominal output fixed at its baseline level, $M \equiv P_0 Y_0$. The aggregate price level then adjusts so that $P_t Y_t = M$ in every counterfactual equilibrium.

Figure A.9: Effects of the MW on employment, productivity, and profits

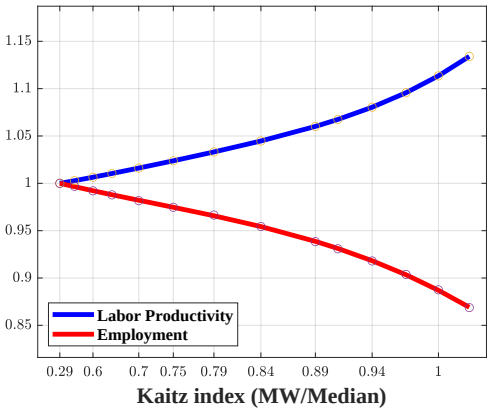
(a) Unemployment rate



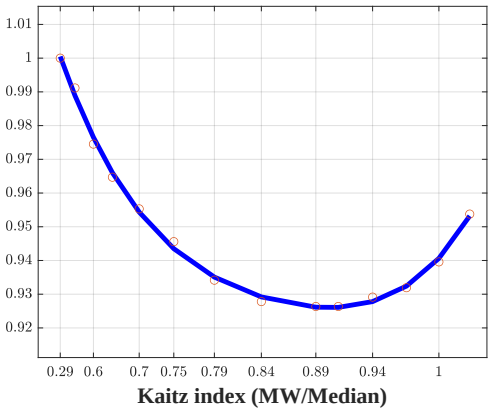
(b) Share of affected jobs destroyed



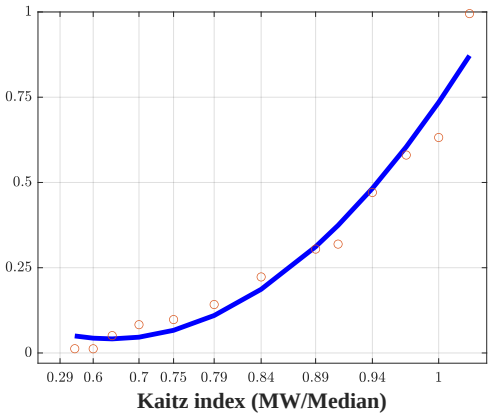
(c) Productivity and employment response



(d) Aggregate profits



(e) Markup share of market power response



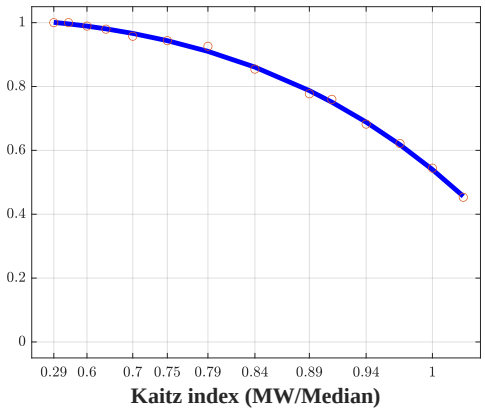
Source: Model. *Note*: Panel (a) plots the aggregate unemployment rate against the Kaitz index of the MW reform. Panel (b) plots the share of affected jobs destroyed against the Kaitz index of the MW reform. The share of affected jobs destroyed is computed as the ratio of the difference in the unemployment rate between the simulated MW reform and the baseline to the share of workers earning less than the MW in the baseline equilibrium. Panel (c) plots the response of aggregate labor productivity (in blue) and headcount employment (in red) to each MW reform. Both variables are normalized to 1 in the baseline equilibrium without a MW. Aggregate labor productivity is defined as $\frac{Y}{E}$, where Y is aggregate output and E is the mass of employed. Panel (d) plots the response of aggregate profits to each MW reform. Panel (e) plots the share of the adjusted market power response due to product market power. See Equation (46) for the formal definition. The red circular markers represent the simulated data, the solid lines their 4^{th}-order polynomial fit.

Nominal wages and all nominal cost components, including hiring costs, scale with P_t.

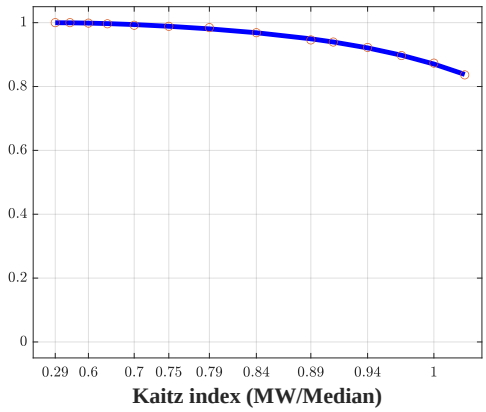
In Figure A.11 we compare the aggregate response of the economy where the MW is set in nominal terms, money is the numeraire, and money supply is fixed, to our benchmark. As

Figure A.10: Firm exit effects of the MW

(a) Share of active firms, headcount



(b) Share of active firms, emp-weighted



Source: Model. *Note*: The *share of active firms, headcount* in Panel (a) is the ratio of the number of active firms after a MW reform relative to the baseline equilibrium. The *share of active firms, employment-weighted* in Panel (b) is the ratio of the baseline employment of active firms after a MW reform relative to total employment in the baseline equilibrium. 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.

apparent, setting the MW in real terms or in nominal terms does not make any significant difference. The reason is that the aggregate price index increases only mildly with the MW.

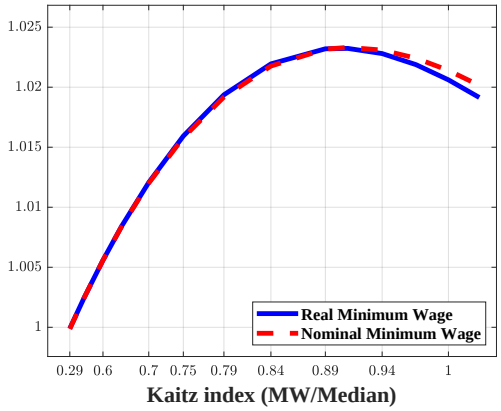
E.2 The Role of the Extensive Margin Adjustment

In the main text, we consider a model in which firms must cover their overhead costs with operating profits in order to remain active. This assumption generates firm exit in response to MW reforms, in line with the empirical evidence. In Section 7, we show that the chosen parametrization of overhead costs matches the empirical evidence on the firm-exit response to the MW from the US. However, the baseline equilibrium is, in principle, consistent with any parametrization of overhead costs that is weakly below the operating profits of the lowest-productivity firms.

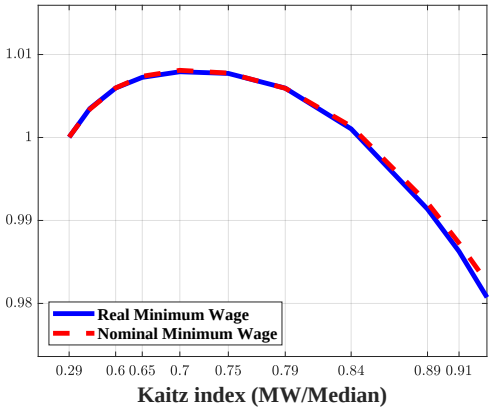
To gauge the role of firm exit in driving the effects of the MW, we repeat our policy experiments in two alternative economies, in which overhead costs are set to 10% and 75% of the operating profits of the lowest-productivity firms in each labor market. Figure A.12 compares the aggregate responses in the benchmark model with those in the two alternative economies, which can be interpreted as lower and upper bounds on the aggregate effects of firm exit.

Figure A.11: Aggregate response, nominal minimum wage

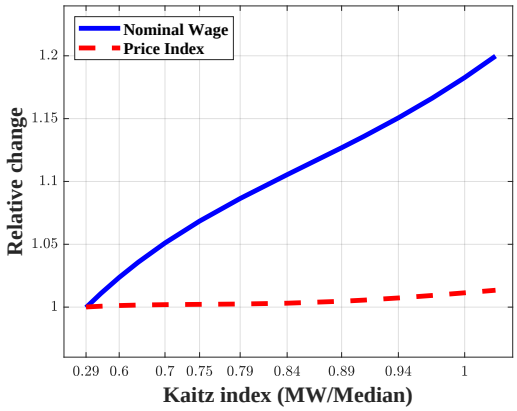
(a) Aggregate consumption



(b) Aggregate welfare



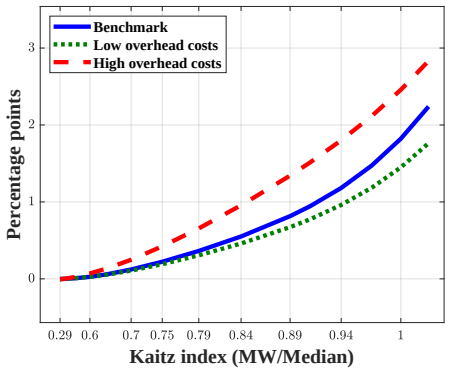
(c) Average nominal wage vs price



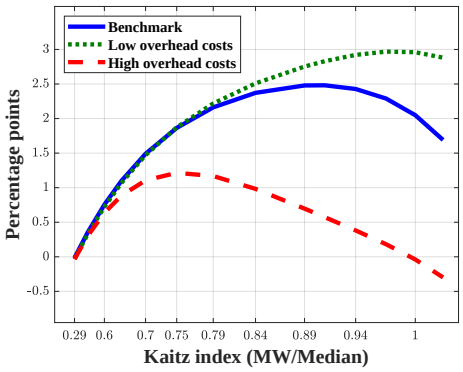
Source: Model. *Note*: In Panels (a)–(b) the solid blue lines represent the response of our benchmark economy to a real minimum wage, while the dashed red line represents the response of our benchmark economy to a nominal minimum wage. In Panel (c) the solid blue line represents the response of the economy-wide average wage $\bar{W}$ and the dashed red line the response of the aggregate price index P. The x-axis is scaled so as to reflect the share of directly affected workers. Each line represents the 4^{th}-order polynomial fit of the simulated data for aggregate consumption, wage, and price, and the 7^{th}-order polynomial fit of the simulated data for welfare.

Figure A.12: Aggregate response, by size of overhead costs

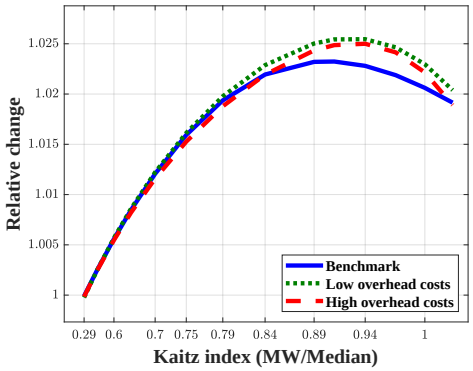
(a) Product market power response



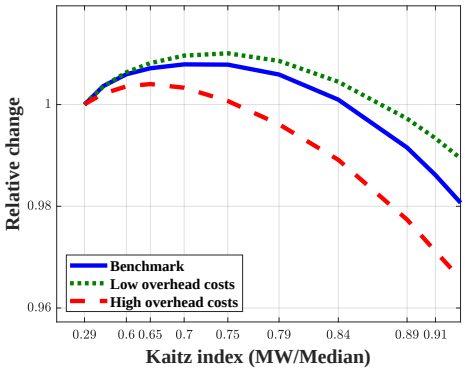
(b) Aggregate labor share response



(c) Aggregate consumption



(d) Aggregate welfare



Source: Model. *Note*: The solid blue lines represent the response of our benchmark economy where overhead costs equal 35% of the operating profits of the lowest-productivity firms in each labor market. The dashed red and the dotted green lines represent the response of alternative economies where overhead costs equal, respectively, 3/4 and 1/10 of the operating profits of the lowest-productivity firms in each labor market. The x-axis is scaled so as to reflect the share of directly affected workers. Each line represents the 4^{th}-order polynomial fit of the simulated data.

F 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,

$$\omega_U = \begin{cases} \left(\frac{\mathcal{W}(\underline{\mathbf{w}}')}{\mathcal{W}(\underline{\mathbf{w}})}\right)^{\frac{1}{1-\vartheta}} - 1 & \vartheta \neq 1, \\ \exp\left\{r\left[\mathcal{W}(\underline{\mathbf{w}}') - \mathcal{W}(\underline{\mathbf{w}})\right]\right\} - 1 & \vartheta = 1, \end{cases}$$

where $\mathcal{W}$ denotes aggregate utilitarian welfare.

Floden (2001) shows that utilitarian welfare admits the following multiplicative decomposition:

$$\omega_U = (1 + \omega_{lev})(1 + \omega_{unc})(1 + \omega_{distr}) - 1. \tag{76}$$

We now derive the three components one at a time. In our context, the *level effect* simply boils down to the aggregate consumption response:

$$\omega_{lev} = \frac{C(\underline{\mathbf{w}}')}{C(\underline{\mathbf{w}})} - 1. \tag{77}$$

Express the value of unemployment and employment at wage piece rate 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 $\mathcal{U}_{a_i}$ is the lifetime utility function associated with our CRRA instantaneous utility function (which turns unemployment benefits and wage piece rates into full consumption). 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,

$$C^e(\underline{\mathbf{w}}) = \sum_{i=1}^{I} L_i \int \left[u_{a_i}(\underline{\mathbf{w}}) C^e\left(U_{a_i}(\underline{\mathbf{w}})\right) + \left(1 - u_{a_i}\left(\underline{\mathbf{w}}\right)\right) \int C^e\left(W_{a_i}(w; \underline{\mathbf{w}})\right)\, dG_{a_i}(w; \underline{\mathbf{w}})\right] d\Xi_i(a).$$

In turn, let C^d be the aggregate consumption-equivalent utility of certainty-equivalent consumption, that is,

$$C^d(\underline{\mathbf{w}}) = \begin{cases} \left(\sum_{i=1}^{I} L_i \int \left[u_{a_i}(\underline{\mathbf{w}}) C^e\left(U_{a_i}(\underline{\mathbf{w}})\right)^{1-\vartheta} + \left(1-u_{a_i}(\underline{\mathbf{w}})\right) \int C^e\left(W_{a_i}(w;\underline{\mathbf{w}})\right)^{1-\vartheta} \, dG_{a_i}(w;\underline{\mathbf{w}})\right] d\Xi_i(a)\right)^{\frac{1}{1-\vartheta}} \\ \text{if } \vartheta \neq 1, \\ \exp\left\{\sum_{i=1}^{I} L_i \int \left[u_{a_i}(\underline{\mathbf{w}}) \log C^e\left(U_{a_i}(\underline{\mathbf{w}})\right) + \left(1-u_{a_i}(\underline{\mathbf{w}})\right) \int \log C^e\left(W_{a_i}(w;\underline{\mathbf{w}})\right) \, dG_{a_i}(w;\underline{\mathbf{w}})\right] d\Xi_i(a)\right\} \\ \text{if } \vartheta = 1. \end{cases}$$

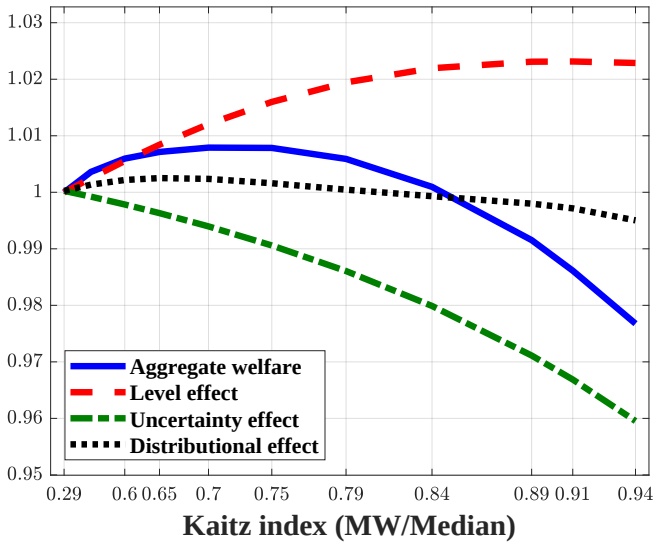
The *uncertainty effect* is the change in the relative aggregate certainty-equivalent consumption with respect to aggregate consumption:

$$\omega_{unc} = \frac{C^e(\underline{\mathbf{w}}')/C(\underline{\mathbf{w}}')}{C^e(\underline{\mathbf{w}})/C(\underline{\mathbf{w}})} - 1. \tag{78}$$

Finally, the *distributional effect* is the change in the aggregate consumption-equivalent utility of certainty-equivalent consumption with respect to certainty-equivalent consumption:

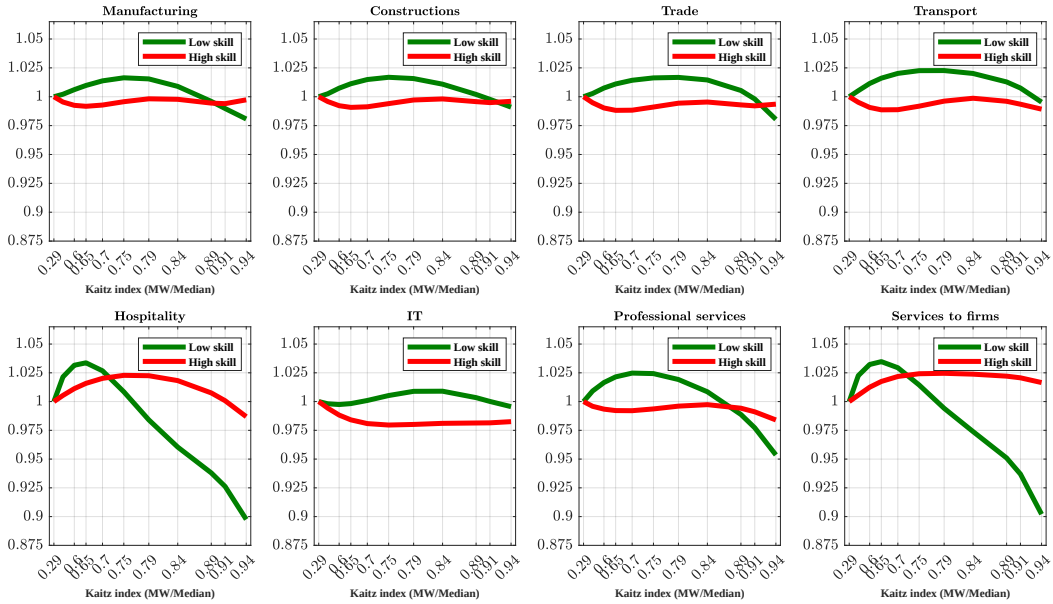
$$\omega_{distr} = \frac{C^d(\underline{\mathbf{w}}')/C^e(\underline{\mathbf{w}}')}{C^d(\underline{\mathbf{w}})/C^e(\underline{\mathbf{w}})} - 1. \tag{79}$$

Figure A.13: Welfare effects, decomposition



Source: Model. *Note*: The figure decomposes the response of aggregate utilitarian welfare into level, uncertainty, and distributional effects as defined in (77)-(79). The x-axis is scaled so as to reflect the share of directly affected workers. Each line represents the 7^{th}-order polynomial fit of the simulated data.

Figure A.14: Breaking down the distributional impact of the MW



Source: Model. *Note*: The graph reports the utilitarian welfare change for workers in each industry. The green lines represent low-skilled workers, the red lines high-skilled workers. The x-axis is scaled so as to reflect the share of directly affected workers. Each line represents the 7^{th}-order polynomial fit of the simulated data.

G Quantitative Validation

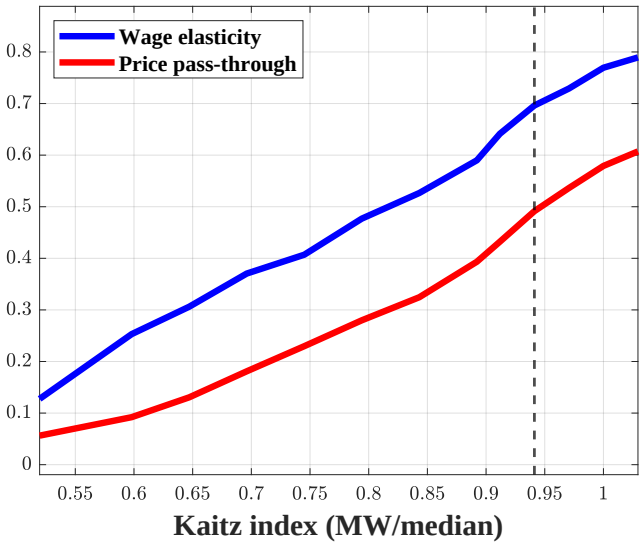
Employment effects. To benchmark our estimated (dis)employment effects of the MW to existing evidence, we compute the elasticity of employment to the MW among the commonly used – and model-consistent – category of *minimum wage workers*, that is, low-skilled workers in hospitality. To ensure comparability with other countries, we consider MW reforms with Kaitz indices from 52 to 65%. According to our estimates, the employment elasticity of low-skilled workers in hospitality with respect to the MW across the reforms considered ranges between -0.06 and -0.04. These figures are slightly smaller in magnitude than the "consensus range" of $[-0.13, -0.07]$ from a meta-study on US data (Belman and Wolfson, 2019). To factor in the wage elasticity to the MW, we complement these statistics with the own-wage employment elasticity, that is the percent change in employment from a percent change in wages induced by a higher MW (Dube and Lindner, 2024). For low-skilled workers in hospitality, we estimate an own-wage employment elasticity of $[-0.33, -0.28]$, which is comparable to the international average of -0.25 reported by Dube and Zipperer (2024). Since our estimated own-wage employment elasticity is increasing with the MW, this divergence might be driven by an over-representation of relatively small MW reforms in the data.

Price effects. By allowing for product market power, our estimated model is well suited to estimate the price pass-through of the MW.[52] Wiltshire et al. (2026) use price and wage data from a large national sample of McDonald's restaurants in the US collected by Ashenfelter and Jurajda (2022) to compute a price pass-through of the MW, that is, the ratio between the price elasticity to the MW and the wage elasticity normalized by the average labor share in the hospitality industry, of about 0.5. To compare our model-based price pass-through to this empirical estimate, we restrict our attention to the labor market of low-skilled workers in hospitality. Figure A.15 shows that both the wage elasticity and the price pass-through implied by the model are increasing with the MW. To ensure comparability, we thus select the MW reform that induces the wage elasticity closest to 0.7 (the elasticity reported in the study). The corresponding price pass-through is 0.50, which matches the empirical evidence.

Pérez Pérez (2026) provides a survey and a meta-analysis of the existing empirical studies estimating the price elasticity to the MW. Even if those studies differ in terms of empirical settings and level of aggregation, the meta-analysis suggests a plausible range for the price elasticity to the MW from 0.03 to 0.11. By following the same strategy as for the employment effects, we compute the own-wage price elasticity in response to MW reforms with Kaitz indices from 52 to 65%. We consider as the relevant wage and price variables the average wage of the low-skilled workers in hospitality and the price index of the respective labor market. According to our estimates, the own-wage price elasticity for the low-skilled hospitality sector to a MW ranges between 0.04 and 0.06. These figures align well with the plausible

[52] For the model-implied price effects to be readily comparable with the data, we also simulate the economy's response to setting a *nominal* MW, using (constant) money balances as numeraire and allowing the aggregate price index to adjust. Results are virtually unchanged by such a numeraire choice. See Appendix E.1 for further details.

Figure A.15: Size-dependent price pass-through



Source: Model. *Note*: The graph reports the elasticity of the average wage (solid blue) and the price pass-through (red line) to a local change in the MW in the labor market for low-skilled workers in hospitality. Both lines are smoothed. The dashed vertical red line represents the local change in the MW giving rise to the closest wage elasticity to 0.7.

range reported by Pérez Pérez (2026).

Reallocation effects (workers). Dustmann et al. (2022) document substantial reallocation effects from the introduction of the 2015 nationwide MW in Germany, which affected 15% of the workforce. Using an individual-level difference-in-differences design, the authors decompose the excess daily-wage growth induced by the reform into wage growth within establishments and wage gains associated with reallocation toward higher-paying establishments. After translating changes in establishment-level average daily wages into individual wage gains, the latter component accounts for 14.1 to 16.8% of the overall increase in individual daily wages.

We replicate the same exercise using the model-generated data from the MW reform whose share of affected workers is closest to 15% of baseline workers—a Kaitz index of 75% on our policy grid. First, we single out workers earning less than the simulated MW reform in the baseline equilibrium (*affected workers*). Let $\mathcal{A}$ be a worker-level dummy variable for being affected by the reform.

We then construct the monthly worker-transition processes under the baseline and reform equilibria using the corresponding job-contact rates and wage-offer distributions. These processes allow for job-to-job transitions, movements through unemployment, and multiple transitions over the horizon. Following Dustmann et al. (2022), we track workers for 24 months and condition the endpoint wage and establishment-quality outcomes on employment at the end of the period.

Formally, let $r \in \{0, 1\}$ index the no-reform and reform transition processes, z_0 denote a worker's baseline employer productivity type, z_{24}^r denote her employer productivity type after 24 months, and E_{24}^r indicate employment at that date. The model-implied reallocation

share is:

$$\mathcal{S} \equiv \frac{\mathbb{E}_1\left[\log w_0(z_{24}^1) - \log w_0(z_0) \mid \mathcal{A}, E_{24}^1 = 1\right] - \mathbb{E}_0\left[\log w_0(z_{24}^0) - \log w_0(z_0) \mid \mathcal{A}, E_{24}^0 = 1\right]}{\mathbb{E}_1\left[\log w_1(z_{24}^1) - \log w_0(z_0) \mid \mathcal{A}, E_{24}^1 = 1\right] - \mathbb{E}_0\left[\log w_0(z_{24}^0) - \log w_0(z_0) \mid \mathcal{A}, E_{24}^0 = 1\right]}.$$

Because employer wage policy is held at its baseline value, the numerator of the reallocation share changes only when a worker moves to a different employer productivity type. Stayers therefore contribute zero to the numerator.[53] According to our estimates, reallocation accounts for 17.8% of the overall individual wage gains, slightly exceeding the share reported by Dustmann et al. (2022).

Reallocation effects (firms). Aaronson et al. (2018) document a spike in firm exit in the hospitality sector (restaurants and hotels/motels) in response to state-level MW hikes enacted in the US between 2001 and 2006. The MW hikes considered range from 6.2% to 26.2%, with an average increase of 17%. According to the authors, just over half of workers in the restaurant industry were paid within 150% of the MW before the reforms. Let $x_s^{(2)}$ denote the cumulative two-year exit rate in sector s, defined as the share of establishments operating before the reform that exit within the following two years. The corresponding firm exit elasticity to the MW is $\varepsilon_s^x \equiv \partial \log x_s^{(2)} / \partial \log \underline{w}$. Following a difference-in-differences strategy near state borders, the authors estimate $\varepsilon_s^x = 2.40$ (0.86) for limited-service restaurants and $\varepsilon_s^x = 8.00$ (2.11) for hotels/motels, starting from a mean exit rate of $x_s^{(2)} = 5.7\%$.

Since our model does not feature firm flows in equilibrium, we focus on the *excess* firm exit rate induced by the MW, defined as the exit-rate semi-elasticity $\chi_s^x \equiv \partial x_s^{(2)} / \partial \log \underline{w} = \varepsilon_s^x x_s^{(2)}$. Its model analogue is

$$\widehat{\chi}^{x,M} \equiv \frac{\left[N^A(\underline{w}_0) - N^A(\underline{w}_1)\right] / N^A(\underline{w}_0)}{\log(\underline{w}_1 / \underline{w}_0)},$$

where $N^A(\underline{w})$ is the number of active firms under MW $\underline{w}$. To ensure comparability of MW bites, we consider a baseline MW equal to a 60% Kaitz index, such that 54% of hospitality workers are paid within 150% of the MW. We then study the firm exit response to increasing the MW to Kaitz indices of 65% and 75%, corresponding to MW increases of 8.2% and 24.6%, respectively. According to our estimates, the model-implied exit-rate semi-elasticity ranges between 0.13 and 0.64. The corresponding empirical semi-elasticities are 0.137 (0.049) for limited-service restaurants and 0.456 (0.120) for hotels/motels.

[53] Unlike in the empirical calculation, the model requires no establishment-to-worker wage loading because firms do not wage discriminate within employer types: the baseline employer wage is also the wage received by the worker.

H The Role of Endogenous Markups

To isolate the role of endogenous markups in driving our aggregate results, we compare our benchmark economy to two similar economies featuring exogenous markups.

H.1 Observationally Equivalent Economy with Exogenous Markups

The first alternative economy features the same demand elasticities (σ, ρ) as our benchmark model, but assumes that all firms charge a constant markup μ equal to the aggregate markup of the benchmark economy. For the baseline equilibrium in this alternative economy to be observationally equivalent to our benchmark economy, we re-estimate the four remaining internally estimated parameters (a_H, χ, ζ, ϕ) according to the same strategy outlined in Section 4.

Table A.7: Internally estimated parameters: endogenous vs exogenous markups

Parameter	Description	Value (End. markups)	Value (Exog. markups)
a_H	Skill parameter, high skills	1.340	1.365
χ	Meeting efficiency	1.187	1.188
ζ	Vacancy convexity hiring costs	1.032	0.767
ϕ	Hiring rate elasticity hiring costs	1.972	0.799
σ	Elast. of subst. within sectors	20.29	20.29
ρ	Elast. of subst. across sectors	2.500	2.500
	Estimation targets		
	Employment share 50+ firms	0.403	0.401
	AKM sorting share of wage variance	-0.066	-0.066
	Weighted avg HHI, across product markets	0.118	0.127
	Semi-elasticity labor share-market share	-0.365	-0.315

Table A.7 compares the differences in the estimated parameters across the two economies. The re-estimated hiring cost convexity (ζ) and hiring rate elasticity of hiring costs (ϕ) are lower than in the benchmark model. This is intuitive: since product market power is constrained to be constant in the alternative model, large firms need to enjoy more labor market power to replicate the same gradient between market share and labor share. Moreover, fewer hiring frictions are needed to replicate the empirical firm size distribution. Importantly, the baseline wage distribution is exactly the same as in our benchmark model. This means that the bite (share of affected workers) of each MW reform is also the same. On the other hand, the productivity distribution differs across the two economies, reflecting different product- and labor-market wedges. Figure 8 in the main text compares the aggregate response to the MW across our benchmark economy and this alternative economy.

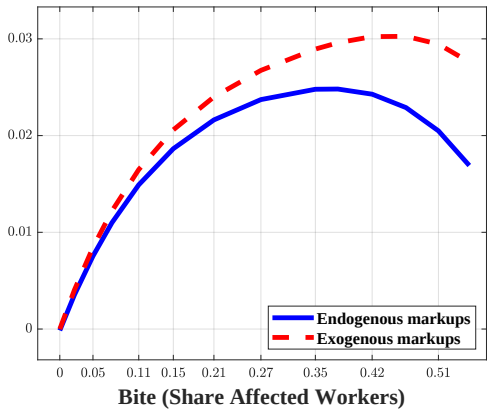
Overall, this alternative economy with exogenous markups is observationally equivalent to our benchmark with endogenous markups absent a MW. It follows that differential responses to the MW across the two economies can be interpreted as a structural counterfactual, i.e., what would happen by ignoring endogenous markups when estimating the equilibrium effects of the MW.

H.2 Structurally Equivalent Economy with Exogenous Markups

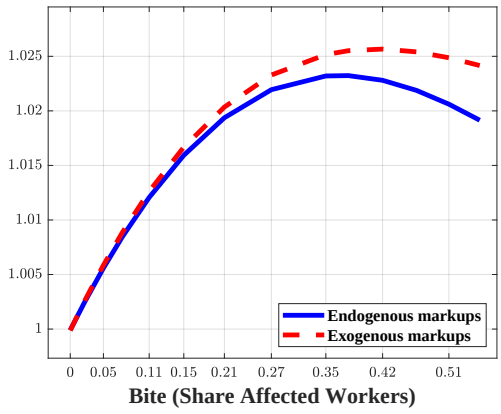
The second alternative economy features the same structural parameters and primitive distributions as our benchmark model, but assumes that all the firms charge the monopolistically-competitive markup $\mu = \frac{\sigma}{\sigma-1}$. This implies that markups are exogenous and constant across all firms. Unlike our first alternative economy, the baseline equilibrium is *not* observationally equivalent to our benchmark model: neither the targeted moments nor the baseline wage distribution are the same across the two economies. For this reason, we compare the aggregate effects of MW reforms affecting the same share of workers. Moreover, we re-estimate the overhead costs in the alternative economy, so that they equal the same share of operating profits of the lowest-productivity firms in each labor market. Figure A.16 compares the aggregate response to the MW across our benchmark economy and this alternative economy.

Figure A.16: Aggregate response: Endogenous vs exogenous markups – Same productivity distribution

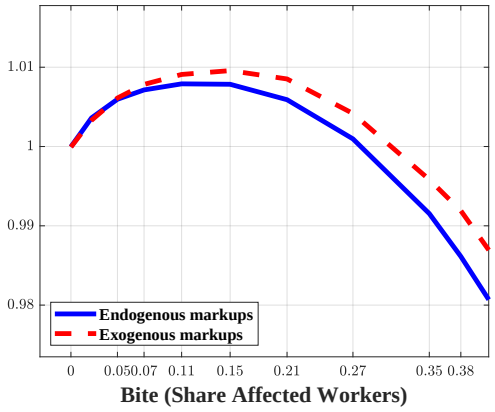
(a) Aggregate labor share response



(b) Aggregate consumption



(c) Aggregate welfare



Source: Model. *Note*: Panel (a) compares the changes in the aggregate labor shares, while Panels (b) and (c) compare the relative change of aggregate consumption and aggregate welfare, normalized to 1 in the baseline equilibrium without a MW. The *endogenous markup* economy in solid blue is our benchmark economy with oligopolistic competition in product markets, while the *exogenous markup* economy in dashed red is a structurally identical economy to our benchmark but with fixed markups. The x-axis reflects the share of directly affected workers.

Overall, this alternative economy with exogenous markups is structurally equivalent to our baseline with endogenous markups absent a MW. It follows that differential responses to the MW across the two economies can be interpreted as a behavioral counterfactual, i.e.,

what would happen to the equilibrium effects of the MW if firms started charging constant markups.

H.3 Empirical Validation from Contractual Wage Floors in Italy

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 three-year duration (2 years prior to 2009). Importantly, these contracts have 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, noncompliance 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.[54] For a more detailed description of the institutional framework, see D'Amuri and Nizzi (2018).

Balance sheet data. We leverage administrative data on the balance sheets for the universe of incorporated Italian firms covering the period 2005-2020 from CERVED. For each firm-year we match information on the number of employees and on the total wage bill coming from INPS, 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. The Italian collective bargaining system is highly centralized. Collective agreements are signed at the nationwide industry level, with little room for further adjustments at the local level. Contracts envisage wage floors for different worker categories. In this paper, we use data on contractual wage-floor indices at 3-digit industry level that are made publicly available by Istat for the period 2005-2020, computed as a weighted average of the floors set by the detailed collective contracts within each industry.

Product-market-level analysis. We use our micro-data to measure concentration by computing the revenue-based HHI for each product market, and to compute the ratio of the product-market-level value added to employment as a proxy of product-market-level productivity. As shown in Table A.8, higher industry-specific wage floors increase both concentration and productivity, consistent with worker reallocating towards more productive

[54] 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 demonstrate the important role of collective bargaining for downward wage rigidity (Devicienti et al., 2007), wage inequality (Devicienti et al., 2019; Leonardi et al., 2019), and regional differences in employment (Boeri et al., 2021) in Italy.

firms. On average, a 10% increase in the wage floor leads to an average increase in product-market HHI of 0.3 and a rise in measured labor productivity slightly below 10 percent.

Table A.8: Product-market response to contractual wage floors

	(1) HHI (product market)	(2) Log labor productivity
Log wage floor	0.029	0.942***
	(0.072)	(0.194)
Time FE	Yes	Yes
Product-market FE	Yes	Yes
N	52,007	52,007
R^2	0.790	0.785

Clustered standard errors at product-market level in parentheses.

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

Source: CERVED (2005-2020) and Istat data. As in the quantitative model, a product market corresponds to a 4-digit ATECO sector for national industries (manufacturing, IT, and transport), and to a 4-digit-ATECO-sector times region pair for local industries (trade, constructions, hospitality, professional services, and services to firms).

Product or Labor Market Concentration? In Section 8 we have established empirically that product-market concentration is an important predictor of the labor share response to the MW. According to our theory, higher product-market concentration implies greater product market power. Mapping product markets into detailed sectors (4-digit for national industries and 4-digit $\times$ region for local industries, in our case) is a rather standard practice in the literature. However, one may still worry that the degree of sectoral concentration also affects *labor* market competition. 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 (Berger et al., 2022; Azar et al., 2024), we work with two definitions of labor market, embracing its geographical dimension: province (administrative geographical units) and province-industry (1-digit).[55]

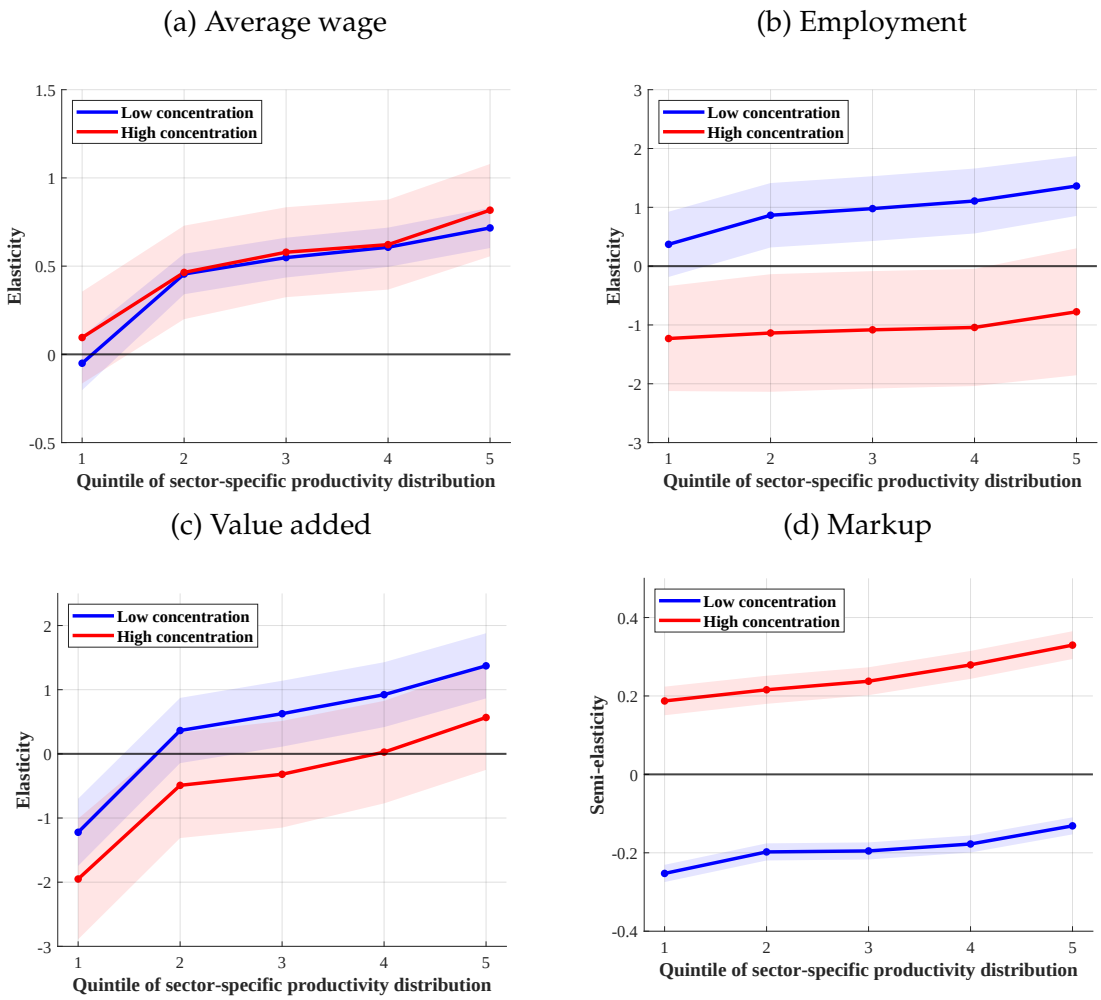
We focus on the following regression model of the heterogeneous firm-level labor share response to a rise in the contractual wage floor by product market concentration:

$$LS_{i,t} = \beta_1 \log MW_{s_3(i),t} + \beta_2 \log MW_{s_3(i),t} \times HHI_{k(i)} + \phi_{t(i)} + \alpha_i + \epsilon_{i,t}. \tag{80}$$

The coefficients of interest are reported in Column 1 of Table A.9: β_1 captures the effect of the wage floor in highly competitive product markets ($HHI_{k(i)} \approx 0$), while β_2 captures the gradient of the effect with respect to product market concentration. We start by adding province-time or province-industry-time effects to the regression model (80), 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 Columns 2 and 3 of Table A.9, the estimates of our coefficients of interest are remarkably robust to the

[55] 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 on the headquarters, which introduces some measurement error for multi-establishment firms.

Figure A.17: Firm-level response, by productivity and product market concentration



Source: CERVED (2005-2020), INPS and Istat data. *Note*: The panels plot the estimated coefficients of the interaction terms between the quintile of the sector-specific productivity distributions by product market concentration and the natural logarithm of wage floor, computed from regression (22). The outcome variables are the natural logarithms of the firm-level average wage in Panel (a), employment in Panel (b), value added in Panel (c), and the firm-level markup in Panel (d). Firm-level markups are estimated on revenue data using a translog production function with time-varying sectoral output elasticities (De Loecker et al., 2020). Observations are weighted by firm employment. The outcome variables are measured at the end of the calendar year corresponding to the first year in which the wage-floor increase is in effect. Time effects are interacted with 2-digit industry and an indicator for product-market concentration being above the economy-wide value-added-weighted average. Standard errors are clustered at the firm-level. Shaded areas represent 90% confidence intervals.

inclusion of these additional fixed effects. Next, we augment our estimating regression 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. The coefficients of interest are robust to the inclusion of the additional regressors (Columns 4 and 5). Crucially, our estimates show that the labor share response is higher in *highly* concentrated labor markets than in less-concentrated labor markets across the entire productivity distribution (see Figure A.18). 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.

Table A.9: Robustness Checks: Controlling for labor market competition

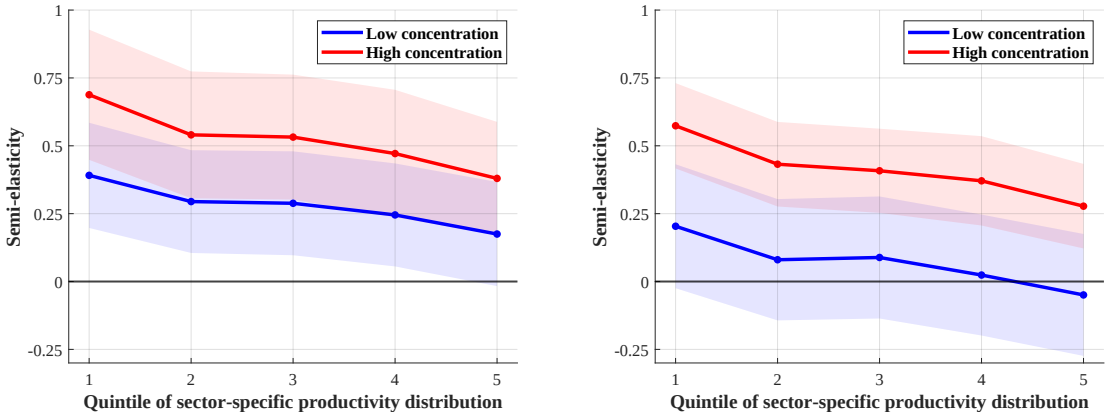
	(1) Labor share	(2) Labor share	(3) Labor share	(4) Labor share	(5) Labor share
Log wage floor	0.401**	0.389***	0.297***	0.348***	0.272**
	(0.128)	(0.114)	(0.084)	(0.087)	(0.084)
Log wage floor × HHI revenues (product market)	-0.749*	-0.695**	-0.493**	-0.685**	-0.478**
	(0.311)	(0.263)	(0.167)	(0.264)	(0.166)
Log wage floor × HHI wage bill (prov)				3.863	
				(6.156)	
Log wage floor × HHI wage bill (prov × 1-digit)					0.265
					(0.935)
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	6,347,463	6,347,461	6,346,704	6,260,229	6,259,354
R^2	0.728	0.730	0.738	0.731	0.738

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*: The labor share is defined as the ratio of total labor costs to value added. Observations are weighted by firm employment.

Figure A.18: Firm-level labor share response, by productivity and labor market concentration

(a) Local concentration (b) Local concentration × 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 industry-specific wage floor by productivity bins, computed from regression (22). 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. Observations are weighted by firm employment. Standard errors are clustered at the firm-level. Shaded areas represent 90% confidence intervals.