Granular Search, Market Structure, and Wages

Gregor Jarosch Jan Sebastian Nimczik Isaac Sorkin*

March 2019

Abstract

This paper develops an approach to measuring labor market power that builds on the structure of a canonical search model. We relax the common assumption of a continuum of firms and assume that large employers exert market power in the wage bargain by effectively eliminating their own future job openings from a worker's threat point. This *granular* extension yields a micro-founded concentration index similar to the Herfindahl and a structural mapping between the index and worker compensation. We extend the model to allow for firm level productivity differences. We then use the model as an accounting device to measure the contribution of labor market concentration to the level and evolution of mean wages in Austrian labor markets. To do so, we define labor markets endogenously by clustering firms based on worker flows.

*Jarosch: Princeton and NBER, gregorjarosch@gmail.com. Nimczik: Humboldt University Berlin and IZA, jan.nimczik@hu-berlin.de. Sorkin: Stanford and NBER, sorkin@stanford.edu. All errors are our own, please let us know about them.

An intuitively appealing approach to measuring the role of market power in the labor market is to compare measures of labor market concentration to measures of wages.[1] This approach allows researchers to assess the role of employer market power in depressing wages relative to an economy in which employers do not have market power. It also allows researchers to assess how changes in concentration contribute to changes in wages. Notably, an increase in labor market concentration could contribute to the decline in the aggregate labor share (e.g., Karabarbounis and Neiman (2014)).

In this paper, we develop a model that provides a new microfoundation for a structural relationship between concentration and wages. To do so, we build on the structure of a canonical search model in the Diamond-Mortensen-Pissarides tradition, but relax the assumption of a continuum of firms. We further assume that non-atomistic employers exert market power by effectively eliminating their own future job openings from a worker's threat point in the wage bargain. As a consequence, the distribution of employment shares matters and we derive a direct mapping from an empirical concentration measure to average wages.

We then use our framework to gauge the consequences of levels and trends in labor market concentration over time for wages in the Austrian labor market from 1997 to 2015. In so doing, we offer two contributions to the literature: first, we use our model to interpret the underlying patterns in the data and translate these patterns into interpretable economic quantities; second, we define labor markets in a data-driven way using worker flows, rather than relying on ad hoc definitions such as various combinations of industry and geography. We find that eliminating employer market power by moving to the atomistic benchmark would raise wages by about eight percentage points. We also find that over our sample period concentration has decreased, and so changes in market structure have increased wages and the labor share.

The model we build has a couple key ingredients. The first key ingredient is that employers each control a strictly positive fraction of job openings and unemployed workers apply to these in a process that is subject to coordination frictions. As a consequence, vacancies frequently have multiple applicants they can choose from.

The second key ingredient is that wages are set through standard Nash-bargaining with the twist that the firm can manipulate the worker's threat point. We assume that if the firm and worker fail to reach an agreement and the disagreeing worker applies to another job opening at the firm, then the firm would select another worker from the queue of applicants. Thus, the threat allows the firm not to compete with its own job openings. Because most vacancies receive multiple applications, this form of punishment is costless for the firm (and the firm only carries out this punishment when it is costless).

Within the model, we derive a closed form expression for average wages which shows that market

[1] Boal and Ransom (1997) suggest that Bunting (1962) represents the earliest version of this regression. Bunting (1962, Appendix 16) finds a positive relationship between wages and concentration. Recently, there has been a surge of papers computing measures of concentration and relating these measures to wages. A presumably incomplete list includes: Azar, Marinescu, and Steinbaum (2017), Azar et al. (2018), Benmelech, Bergman, and Kim (2018), Hershbein, Macaluso, and Yeh (2018), Lipsius (2018), Qui and Sojourner (2019), and Rinz (2018).

structure is summarized by a particular concentration measure. This measure is distinct from the standard Herfindahl-Hirschman Index (HHI) since it places more weight on large employers, but shares the same limits and can be just as easily computed in the data. This measure of concentration affects two wedges which reduce compensation relative to the atomistic benchmark. These wedges reflect first, the deterioration of the outside option because all employers act in this fashion, and, second, the improvement of the inside option since reaching an agreement frees the applicant from this threat.

To make the model a useful accounting device, we then extend it to allow for firms to differ in both productivity and size. We show that the same structural mapping between concentration and worker outcomes exists with an additional term. The additional term measures productivity-weighted concentration weighted by the covariance between employment and productivity. The model yields a clean decomposition of the effects of changes in concentration and productivity-weighted concentration on wages.

To measure productivity-weighted concentration, we show that the model implies a closed-form and simple-to-compute inversion from wages and size to productivity. In the model, wages depend on the firm's size, the firm's productivity as well as the distribution of productivity and employment shares across firms. Once we measure size and wages, we can back out productivity.

Finally, we show that the pass-through of firm-level productivity to wages is decreasing in size, but independent of market structure. As a consequence, the exercise of market power generates wage dispersion, but in a form that does not per se depend on market-wide concentration.

We then take our model to the data. Our empirical setting is Austria from 1997-2015. We face the basic challenge of market definition: what counts as a labor market? We build on Nimczik (2018) to define labor markets based on worker flows. Formally, we cluster firms on the basis of workers flows, where our model of clustering is a stochastic block model. This endogenous notion makes market definition an empirical question, rather than an *a priori* choice such as geography or industry (though we also report results at the level of geography and industry). We find that the number of labor markets that maximizes the objective function is 369. As discussed in more detail in Nimczik (2018), some of these labor markets span geography and industry, while others capture more conventional measures.

We present four sets of results. Our first set of results concerns trends in concentration. We find that the level and trends in the HHI and our measure of concentration are very similar: both follow a u-shape over the sample period, and the magnitude is around 0.11. The size-productivity concentration wedge, however, moves differently. It is flat and then declines towards the end of the sample period. The feature of the data that drives this decline is that the size-wage gradient decreases over the sample period, which has also been documented in the U.S. (Bloom et al. (2018)). Intuitively, the way the model explains large firms paying higher wages is large firms being more productive. So the declining size-wage gradient contributes to a declining contribution from the productivity-weighted concentration wedge.

Our second set of results concerns counterfactuals in the trends of wages. We find, perhaps counterintuitively, that the change in the productivity-concentration wedge has led to a decline in wages. The reason for this divergence from the trend is that the model implies non-linear effects of our various concentration measures on wages, and we report results averaged over markets. In contrast, the changes in the concentration measure offset the declines in the productivity-concentration wedge so that the changes in market structure led to an increase in wages of about three percentage points from 1997 to 2015.

Our third set of results concerns counterfactuals in the level of wages relative to the atomistic benchmark. That is, we ask the question: what would happen to wages if the fundamentals remained the same, but firms behaved as if they did not have market power? We find that shutting down the the productivity-concentration wedge would raise wages by about two percentage points, while additionally shutting down the direct effects of concentration would lead to a total increase in wages of about eight percentage points. This calculation highlights the value of our structural framework: a calculation based solely on the HHI would lead to the conclusion that the Austrian labor market is not concentrated, whereas our framework delivers a precise quantitative answer to the effect of concentration on wages.

In our final set of results, we document heterogeneity in these effects across labor markets, and consider the sensitivity of our results to different market definitions and parameter values. We find considerable heterogeneity in concentration and the effects of concentration across labor markets. While in the employment-weighted average labor market wages would rise by eight percentage points in the atomistic counterfactual, in the median labor market this increase is only 2%. Hence, the average effect is driven by a few markets that are very concentrated.

We find that the endogenous labor markets are more concentrated than when we use conventional labor market definitions and a comparable number of labor markets. When we define labor markets in terms of 4 digit industry or 2 digit industry times region (approximately, commuting zone)—which have a similar number of labor markets as in our baseline endogenous labor markets—we find effects of concentration that are less than half of what we find in the endogenous labor markets. In contrast, when we use labor market definitions that are more conventional in the literature (e.g., 3- and 4-digit industries times region), then we find much bigger effects of concentration than in our baseline. But this conventional definition also implies well over ten times as many labor markets as our baseline definition, which mechanically raises measured concentration. This combination of results emphasizes that the endogenous labor markets capture different features of the labor market than conventional definitions.

We also find that our results are sensitive to the choice of bargaining parameter. In particular, increasing worker bargaining power dramatically decreases the implied effects of concentration on wages. The intuition is that with higher bargaining power worker wages depend more on the inside option of productivity and less on the outside option that changes in concentration can affect.

Relationship to the literature Our approach is complementary to, but distinct, from papers that build on the "differentiated firms" framework of Card et al. (2018). These papers include

Berger, Herkenhoff, and Mongey (2018), Lamadon, Mogstad, and Setzler (2017), MacKenzie (2018) and Haanwinckel (2018). These papers build static models of the labor market where workers' labor supply to a firm resembles consumer product demand. Coupled with a wage-setting protocol that resembles firms' product price setting decision, these papers deliver wage equations that provide an equilibrium microfoundation for the Robinson-style monopsony markdown. In contrast, this paper builds from the logic of a textbook (labor) search model.[2] The source of market power is distinct: in our model the elasticity of labor supply to all firms is the same (it is zero), but the "markdown" relative to productivity differs across firms and is a function of market power measured through size.

The most closely related paper to ours is Berger, Herkenhoff, and Mongey (2018). This paper also provides a microfoundation for a structural relationship between a measure of concentration and wages, and uses the model to assess changes in concentration on wages (among other things). One important difference is that Berger, Herkenhoff, and Mongey (2018) provide a microfoundation for size, rather than taking size as given and reading it off the data. This ambition allows them to consider richer counterfactuals; it means, however, that the mapping to the data is less straightforward than in our framework.

Our paper joins a literature that emphasizes variation in outside options in generating wage variation. Some examples include Beaudry, Green, and Sand (2012), Caldwell and Danieli (2018) and Arnoud (2018). The key novelty is that we emphasize the role of employer size in affecting outside options.

We are not the first paper to consider the role of finiteness in search models. Menzio and Trachter (2015) consider a large firm and a continuum of small firms in the product market. There is also a literature on market power in the directed search literature, e.g., Galenianos, Kircher, and Virag (2011). In the context of this literature, our mechanism is distinct.

Outline This paper proceeds as follows. Section 1 presents the baseline model and analyzes its implications for wages. Section 2 extends the model to include productivity heterogeneity, analyzes the implications for wages and pass-through, and also shows how to use the structure of the model to infer productivity. Section 3 introduces the matched employer-employee data from Austria that we use, discusses how we define labor markets using worker flows, and finally discusses how we define size and wage, as well as how we measure the parameters of the model. Section 4 presents our quantitative results about the role of levels and trends in market structure in explaining levels and trends in wages.

[2] This distinction also helps explain how our paper relates to the broader literature computing labor supply elasticities to firms and interpreting this through the lense of monopsony models. See, for example, Webber (2015) and Webber (2018).

1 Granular search

In this section, we develop a partial equilibrium random search model in which workers apply to job openings that are distributed across a finite number of firms. Wages are set through Nash-bargaining and we introduce our key assumption: large employers exert market power by threatening to discriminate in the future in the aftermath of disagreement with a worker. We characterize the resulting mapping between market structure and average wages as well as the firm size wage gradient. In section 2 we extend the framework to allow for heterogeneous productivity across firms.

1.1 Set-up

We study a discrete time economy populated by a measure one of infinitely lived homogeneous workers. Workers are either employed, producing a flow output of one unit of the economy's single, homogeneous good, or they are unemployed. The common discount factor is $0 < \beta < 1$.

An agent who is employed experiences a separation shock at rate $\delta > 0$. In this event, the worker flows back into unemployment. An unemployed worker receives flow value $b < 1$.

There are N distinct employers in the market, indexed by i. The probability that a particular job opening is at firm i is given by f_i and so $\sum_{i=1}^{N} f_i = 1$. For each job opening, a firm pays a per period fixed cost c_i. The process which pairs unemployed workers with job openings is governed by an urn-ball matching function. u unemployed workers send one application per period (balls) towards v vacancies (urns) . This matching process is subject to coordination frictions and so some vacancies receive no applications while others may receive multiple ones. Standard arguments imply that the number of applications a vacancy receives is exponentially distributed.

If a firm receives multiple applications it follows up on one randomly chosen one. Subsequently, the firm and the worker bargain over the wage. Specifically, there is continuous Nash bargaining over the wage where $\alpha \in [0, 1]$ denotes the bargaining power of workers. We assume that all job openings have strictly positive surplus so that the job finding rate is given by $\lambda \equiv \frac{v}{u}(1 - e^{-\frac{u}{v}})$ (see, e.g., Shimer (2005)).

The key novelty is that we endow employers with a particular threat which effectively allows them not to compete with their own future job openings. Specifically, if the firm and the worker fail to find an agreement and the worker's next application is to a job opening controlled by the same firm, then the firm can punish the worker: In the event that the vacancy received multiple applications the firm breaks the tie by hiring one of the other applicants. This tie-breaking rule allows the employer to manipulate the disagreement payoff of the worker in the wage bargain. Importantly, this form of "punishment" is costless to the firm since it only applies to situations where workers are rationed and the firm never gives up an opportunity to produce. That is, if a worker under punishment happens to be the sole applicant to one of the firm's job openings, then the firm rationally hires the worker. We also highlight that this mechanism operates through off-equilibrium payoffs and the parties never fail to reach agreement.

While the firm can punish the worker, we limit the duration of the disagreement "punishment." In particular, we assume that as soon as a job opportunity arises at some other employer j, the worker gets released from the punishment state by firm i. This assumption substantially reduces the state space (it effectively cuts histories that we have to keep track of) and makes the analysis tractable.

In order for the punishment to have bite, we assume that workers cannot direct their applications away from firm i. That is, a worker applies to firm i with probability f_i, no matter what the chances are that she will be hired. This assumption is consistent with an interpretation of the search process as one where workers randomly encounter job openings and is a natural benchmark.

We now turn to the value functions of the workers. We let W_i denote the value of a worker employed at firm i while U denotes the value of unemployment for an unemployed worker who is not under punishment by any firm. Formally, U satisfies

$$U = b + \beta[\lambda \sum_i f_i W_i + (1 - \lambda)U]. \tag{1}$$

In a slight abuse of notation, we denote by U_i the continuation value of the worker in the event of a trade breakdown, which satisfies

$$U_i = b + \beta[\lambda \sum_{j \neq i} f_j W_j + \underline{\lambda} f_i W_i + (1 - \lambda(1 - f_i) - \underline{\lambda} f_i)U_i]. \tag{2}$$

This states that, after disagreement with employer i, a worker's chances to meet and subsequently work for any other employer j are unaltered. However, if the worker applies to a vacancy controlled by i she only gets hired if she is the only applicant, which happens at rate $\underline{\lambda} \equiv e^{-\frac{u}{v}}$. With complementary probability $1-\lambda(1-f_i)-\underline{\lambda} f_i$ the worker remains unemployed. Critically, if employer i is larger, then rejecting i's offer leads to a larger reduction in the job finding rate and so a worse outside option.

Let w_i denote the wage firm i pays under the Nash bargaining solution. Because a new contact releases a worker from previous punishments, the wage does not depend on any other state variables, such as the firms that may have previously punished a worker. The value of working for firm i then satisfies

$$W_i = w_i + \beta(\delta U + (1 - \delta)W_i). \tag{3}$$

Following an exogenous breakdown of an employment spell a worker is free to return to another vacancy posted by the same employer. Thus, the disagreement payoff when bargaining and the value of unemployment following a job spell differ.

We now turn to the firm. Firm i values the bilateral relationship with each of its workers at J_i

satisfying

$$J_i = 1 - w_i + \beta(1-\delta)J_i, \tag{4}$$

where we again impose that a job has no continuation value after exogenous separation.

In turn, we have that a job opening has value

$$V_i = -c_i + \beta\left(1 - e^{-\frac{u}{v}}\right)J_i. \tag{5}$$

To keep a vacancy open, firm i pays fixed cost c_i. The term in brackets captures the probability that the job opening receives at least one application this period. Our notation also imposes that, in equilibrium, trade never breaks down and the match is always formed.

We do not take a stance on the details of the job creation process or on what makes a firm large. But we note that since c_i is firm specific, there exist $\{c_i\}_{i=1}^N$ such that, in equilibrium, $V_i = 0 \forall i$ given f_i and $\frac{u}{v}$. Instead of solving for vacancies given a cost function, we simply assume that the cost function satisfies $c_i = \beta\left(1 - e^{-\frac{u}{v}}\right)J_i$ so that $V_i = 0\forall i$. This choice gives us the flexibility to simply read employment shares off the data instead of having to explicitly model them. We also view this choice as natural because it allows us to obtain a normalization akin to a free entry condition without having to explicitly model the details of the entry process. As a consequence, we never actually use equation (5) in what follows and report it solely for expositional purposes.

Finally, we turn to how the surplus is split. The joint net value of forming a match ("surplus") is then given by $S_i \equiv W_i - U_i + J_i$. In words, once the firm has followed up on one of the applications, the pair can form a match or not: if the match forms, then the worker is in state W_i and the firm moves into state J_i. In turn, under disagreement, the worker moves into the punishment state U_i while the firm has no continuation value.

Given that firms sometimes receive multiple applications, one natural question is why the firm cannot have the multiple applicants compete for the job opening. The same issue arises in Blanchard and Diamond (1994, pg. 425). They invoke a standard no commitment assumption to rule out this competition. In particular, the no commitment assumption means that as soon as the other applicants leave the firm, and regardless of the agreement the firm and worker reached, the hired worker would seek to renegotiate the contract, where the bargaining positions would be as stated in the previous paragraph. Similarly, Blanchard and Diamond (1994) also implicitly assume that there are no side payments so that the firm cannot extract the value of the match to the worker in an up-front payment. We follow them here and make both assumptions.

With axiomatic Nash bargaining, the wage then implements a surplus split such that the net value of forming the match to the worker is

$$\alpha S_i = W_i - U_i, \tag{6}$$

while the net value of forming the match to employer i is

$$(1-\alpha)S_i = J_i. \tag{7}$$

Throughout, we already anticipate a result, namely that in equilibrium workers are willing to work for all firms i. That is $S_i \geq 0 \forall i$.

To summarize the set-up, we highlight the key distinction between our granular search framework and the standard setup with atomistic employers: Here, both workers and employers recognize that, with strictly positive probability, they will meet again in finite time which is what gives bite to the firm's threat. The fact that vacancies may receive multiple applications in turn allows the firm to credibly commit to go through with its threat without hurting its own future vacancy yield. As such, the threat is individually rational for the firm.

1.2 A Concentration Index

We are interested in the mapping between market structure—in particular, employment concentration—and equilibrium wages. Concentration is frequently measured via the HHI. But concentration has no inherent cardinality so the right choice of units depends on the question at hand. This subsection presents a particular concentration index that shares many similarities with the HHI and turns out to be the right way to summarize market structure in our model.

To begin with, let $\tau \equiv \alpha \frac{\beta(\lambda-\underline{\lambda})}{1-\beta(1-\lambda)} \in (0, \alpha)$. τ summarizes how costly punishment is for workers: It is increasing in the share of surplus that a worker gives up when under punishment (α), and in the strength of the punishment ($\lambda - \underline{\lambda}$).

Going forward, we use the approximation $\tau \approx \alpha \frac{\beta\lambda}{1-\beta(1-\lambda)}$. The reason is that quantitatively our model implies that $\underline{\lambda}$ is several orders of magnitude smaller than λ (for $\lambda = 0.12$ we obtain $\underline{\lambda} = 0.0002$) and so this approximation is highly accurate. More importantly, this simpler form of τ enables us analytic expressions that cleanly capture the main economic forces at work. Economically, this effectively ignores the possibility that the worker, after disagreement, next applies to a job opening by the same firm and ends up being the only applicant. The reason this approximation is quantitatively inconsequential is that with a monthly job finding rate of around 12% the implied probability of being the only applicant is indeed remote, which is also true in the data (see, e.g., Davis and Samaniego (2019)). We note that we derive our main theoretical results under the exact model and only impose the approximation at the very end of the proofs so the reader can find the exact expressions in the appendix. When we implement our framework quantitatively we work with the exact expressions.

Finally, let $f^k \equiv \sum_i f_i^k$ such that $f^1 = 1$ and f^2 is the HHI index for employment shares in our labor market with $0 \leq f^2 \leq 1$. The following is a useful concentration index.

Definition 1. *Let concentration be measured as*

$$\mathcal{C} \equiv \frac{\sum_{k=2}^{\infty} \tau^{k-2} f^k}{1 + \tau \sum_{k=2}^{\infty} \tau^{k-2} f^k}.$$

This concentration index is distinct from—yet very closely related to—the standard HHI. First, note that the first element of the infinite sums is simply f^2, the HHI. Second, it shares the same bounds: In the limit with atomistic employers, we have that $\mathcal{C} = 0$, just like the HHI. In the limit of a single monopolistic employer, we have that $\mathcal{C} = 1$, just like the HHI.[3]

What differs between our index and the HHI is the inclusion of the additional higher order terms, (down)-weighted by τ. The higher-order terms place more weight on the size of the largest firms in the labor market than the HHI. Clearly, the higher order terms are particularly important if τ is large while $\mathcal{C}$ converges to the HHI as $\tau \to 0$. Despite the theoretical possibility that these two measures could be very different, empirically we find that the HHI and $\mathcal{C}$ strongly comove and yield almost identical quantitative results when using our preferred parameter values.

1.3 Concentration, Average Surplus, and Wages

To see why $\mathcal{C}$ is useful in mapping the connection between changes in market structure and worker compensation, let $\bar{w} \equiv \sum_i f_i w_i$ denote the average wage. All firms produce flow output of 1, so define $\omega \equiv \frac{w_i - b}{1-b}$ to be the fraction of the net flow output produced by a worker-firm pair that goes to the worker. Let $\bar{\omega} \equiv \sum_i f_i \omega_i$. Our first result is the following:

Proposition 1. *The (employment-weighted) mean surplus is*

$$S^1 = \frac{1-b}{1-\beta\left(1-\lambda\alpha\underbrace{\left[1-\mathcal{C}\right]}_{\textit{wedge 1}}-\delta\underbrace{\left[1-\alpha\mathcal{C}\left(\frac{\beta\lambda}{1-\beta\left(1-\lambda\right)}\right)\right]}_{\textit{wedge 2}}\right)}.$$

The equilibrium relationship between compensation and concentration satisfies:

$$1-\bar{\omega} = (1-\alpha)\left(\frac{1-\beta\left(1-\delta\right)}{1-\beta\left(1-\lambda\alpha\left[1-\mathcal{C}\right]-\delta\left[1-\alpha\mathcal{C}\left(\frac{\lambda\beta}{1-\beta(1-\lambda)}\right)\right]\right)}\right)$$

Proof. See Appendix B. □

To interpret the denominator in the surplus and wage expressions, note that there are three terms that multiply β. First, the pure time discounting term which, naturally, is not affected by market structure (this is the 1). Second, discounting due to the foregone option value of search of

[3] To see this, note that $f^k = 0 \ \forall k \geq 2$ in case of perfect competition while $f^k = 1 \ \forall k \geq 2$ in the case of a monopolist. In Appendix A we present an example of two economies where these two measures present different rankings. One economy consists of a monopsonist with a competitive fringe, and another consists of all equal-sized firms. By choosing the relative size of the monopsonist in comparison to the equal-sized firms, we can make these two measures move in opposite direction. The reason is that $\mathcal{C}$ places more weight on the largest firm (the monopsonist) than the HHI.

the worker (this is the $\lambda\alpha$ term). Third, discounting because at some point in the future the match ends (this is the δ term).

Market structure puts wedges into the second and third discounters and reduces them, thereby increasing the value of the bilateral employment relationship. The reason for the first wedge is that the firm and worker recognize that the worker's outside option is reduced because all the other employers that she may potentially match with have market power.

The second wedge reflects the fact that the benefits of forming a match lasts beyond the spell. Why is that? For the worker (this explains the α in the wedge), forming a match has the additional benefit that she then has the possibility of returning right away to the firm, whereas if she fails to form a match, then she will be punished by being unable to return. This additional benefit is captured by the term $\frac{\beta\lambda}{1-\beta(1-\lambda)}$, which is inversely related to the duration of an unemployment spell. That is, the second wedge captures that, by reaching an agreement, the pair increases the worker's outside option from U_i to U. If an unemployment spell is very long because λ is very low, then this benefit goes to zero. If spells are short and the agent is patient, then it goes to 1.

The second part of the proposition relates wages to market structure. In a static setting, the firm would receive a share $(1-\alpha)$. In a dynamic setting, this gets dampened because the parties recognize that the worker has other options, which is the $\lambda\alpha$ term. However, this dampening is weaker in a setting with market power. The second effect which likewise weakens wages is the threat effect, which is particularly large when workers are patient and unemployment spells are short.

It is instructive to consider what happens when the market becomes dominated by one firm (and thus $\mathcal{C} \to 1$). In this case we have that

$$1 - \bar{\omega} = (1-\alpha)\left(\frac{1-\beta\left(1-\delta\right)}{1-\beta\left(1-\delta\left[1-\alpha\left(\frac{\lambda\beta}{1-\beta(1-\lambda)}\right)\right]\right)}\right). \tag{8}$$

It is thus immediate that a large employer is able to extract even more than a fraction $1-\alpha$ of the flow surplus. The latter is an important benchmark since it would be the surplus split that would arise in the absence of any competition from the labor demand side. But gaining market shares in our setup not only directly reduces competition, it also allows the firm to manipulate its worker's outside option more sharply, which further reduces wages. That force is reflected by the remaining wedge in equation 8 (in squared brackets). Again, the threat is particularly strong if unemployment spells are short relative to a worker's rate of time preference.

Most importantly, proposition 1 offers a structural relationship between average wages and market structure. As a consequence, it allows us to directly assess the quantitative contribution of empirically observed changes in employment shares to average wages, given a set of parameters $\{\beta, \delta, \alpha, \lambda, b\}$. Given those parameters, measuring $\mathcal{C}$ empirically does not require any more information than a standard HHI.

An important corollary to Proposition 1 is that:

Corollary 1. *Average wages are monotonically decreasing in concentration $\mathcal{C}$.*

This result provides a theoretical foundation for a negative relationship between concentration—as measured by $\mathcal{C}$— and average wages. The next corollary tells us when we should expect variation in concentration to matter:

Corollary 2. *The elasticity of wages with respect to concentration becomes smaller in magnitude as worker bargaining power (α) increases.*

Proof. See Appendix C. □

This result anticipates an important quantitative feature of our results: variation in concentration matters more when worker bargaining power is low. There are three steps in this logic. First, concentration affects wages through the outside option. Second, when bargaining power is high, wages are determined by the inside option (productivity) whereas when bargaining power is low wages are determined by the outside option. Putting the pieces together, variation in concentration matters more when bargaining power is low.

From the perspective of empirical work that relates variation in concentration into variation in wages, it is useful to note that $\mathcal{C}$ appears in the elasticity. This fact implies that there is a non-constant elasticity and hence a regression is averaging over heterogeneous effects.

1.4 Concentration and Individual Wages

In the previous section, we related market-wide mean pay to concentration. The model also has implications for firm level wages w_i. We are particularly interested in the relationship of w_i with concentration $\mathcal{C}$ and the size of the individual employer i, f_i.

We summarize our key findings in Proposition 2:

Proposition 2. *Firm-specific relative worker compensation is fully characterized by*

$$\frac{\omega_i}{\omega_j} = \frac{1 - \tau f_j}{1 - \tau f_i}$$

Proof. See Appendix D. □

Proposition 1 showed how average wages are governed by concentration $\mathcal{C}$. Proposition 2 in turn shows that relative compensation is independent of concentration. That is, changes in concentration affect compensation at all firms proportionally.

Proposition 1 only proved that average compensation declines. In combination with proposition 2 this further implies that individual wages (and compensation) are monotonically decreasing in $\mathcal{C}$ at all employers i.

Proposition 1 also implies that wages are monotonically decreasing in employer size f_i. That is, naturally, the firm with more market power pays a lower wage.

The proposition also reveals that the effect of size on wages is increasing in τ. That is, the wage-size gradient steepens as τ increases. This is natural since, as discussed above, τ captures the ability of firms to punish. As a consequence, inequality induced by market power in the labor market rises if τ increases. Furthermore, we highlight that the returns to size is independent of the distribution of employment shares across other firms in the market. In other words, the relative returns to size are independent of market structure.

Proposition 2 emphasizes that market power affects wage purely through size, which is a distinct mechanism from the typical “markdown” mechanism embedded in monopsony-style models. In those models, the variation in wages stems from variation in the elasticity of labor supply to the firm (here, the elasticity of labor supply to each firm is 0).

2 Heterogeneous Productivity

The model presented in the previous section has the virtue of simplicity. But it has a pair of stark and counterfactual implications: size perfectly predicts wages, and wages are decreasing in firm size. In this section, we add productivity heterogeneity to the model which allows the model to generate an imperfect relationship between size and wages.

This extension allows us to separate the two distinct ways changing employment shares affects labor market outcomes: First, through the pure size distribution already studied in the previous section. And, second, through how size and productivity comove. Our setup yields a clean decomposition between the two, and hence lets us separately quantify the consequences of empirical movements along both dimensions over time. Underlying this decomposition is the intution that if the firms gaining marke share are low-productivity firms, then increases in concentration matter less than direct measurement suggests.

2.1 Concentration, Average Surplus, and Wages

Let p_i denote output per worker at firm i. As before, let $f^k \equiv \sum_i f_i^k$ and define $p^k \equiv \sum_i p_i f_i^k$ such that p^1 is the employment weighted average output produced by a match. We also define $\tilde{p}_i = p_i - b$ and $\tilde{p}^k \equiv \sum_i (p_i - b) f_i^k$ to be *net* output and the employment weighted average *net* output. For the following theoretical exposition we sometimes normalize, without loss, $\tilde{p}^1 = 1$. When taking the model to the data, we let $\tilde{p}^1$ move over time so as to account for fluctuations in aggregate productivity over time. The definition of $\mathcal{C}$ is unchanged. The following is the productivity counterpart of $\mathcal{C}$, namely a productivity weighted concentration index:

Definition 2. *Measure productivity-weighted concentration as*

$$\mathcal{C}^P \equiv \frac{\sum_{k=2}^{\infty} \tau^{k-2} \tilde{p}^k}{1 + \tau \sum_{k=2}^{\infty} \tau^{k-2} \tilde{p}^k}.$$

This index is identical to $\mathcal{C}$ except the employment weights are productivity weighted. It shares the same properties as $\mathcal{C}$ discussed above. Next, we relate $\mathcal{C}$ and $\mathcal{C}^P$.

Definition 3. *Measure the wedge between concentration and productivity-weighted concentration as*

$$\mathcal{P} \equiv \left[\mathcal{C}^P - \mathcal{C}\right] \left(1 + \frac{\tau \sum_{k=2}^{\infty} \tau^{k-2} \tilde{p}^k}{\tilde{p}^1}\right).$$

This wedge has a couple key properties. First, it is equal to zero if p_i is identical across firms. Second, the wedge is positive when the weighted covariance between size and productivity is positive:

Proposition 3. *The sign of the $\mathcal{P}$ is the same sign as*

$$\sum_i \frac{f_i(\tilde{p}_i - 1)}{1 - \tau f_i}, \tag{9}$$

which is the weighted covariance between size and (normalized) productivity, where the weights are $\frac{1}{1-\tau f_i}$, and so are increasing in size.

Proof. See Appendix E. □

Denote by S^{1*} and ω^* average surplus and average worker compensation in the homogeneous firms benchmark presented in proposition 1. Similar to before let $\bar{\omega} \equiv \frac{\bar{w}-b}{p_1-b} = \frac{\bar{w}-b}{\tilde{p}^1}$ the fraction of the average net flow output that goes to workers. Let $\hat{\beta} \equiv \beta \frac{1-\beta(1-\delta-\lambda)}{(1-\beta(1-\delta))(1-\beta(1-\lambda))} > 0$. Our key result is summarized in the following proposition:

Proposition 4. *Mean surplus is given by*

$$S^1 = S^{1*}\left(1 + \hat{\beta}\lambda\alpha\mathcal{P}\right).$$

The equilibrium relationship between compensation and concentration satisfies:

$$1 - \bar{\omega} = (1 - \bar{\omega}^*)\left(1 + \hat{\beta}\lambda\alpha\mathcal{P}\right).$$

Proof. See Appendix F. □

Proposition 4 naturally extends the results in Proposition 1 to the heterogeneous firms case. First, it shows that the average joint surplus is given by exactly the same expression as in the baseline case up to a wedge $\mathcal{P}$. This wedge is positive if productivity is more concentrated than employment. The reason is that the net value of an employment relationship is larger if the outside option of the worker deteriorates. One way the worker outside option deteriorates is if, relative to employment, productivity becomes more concentrated as measured by an increase in $\mathcal{P}$.

The expression for compensation clarifies that heterogeneous productivity has a similar effect on wages. That is, compensation equals the homogeneous firm compensation level subject to a wedge which reduces compensation when productivity is positively correlated with size and vice versa. This wedge reflects the fact that when productivity becomes more concentrated this has a direct effect on wages since large firms have a more market power.

We also have that:

Corollary 3. *Average wages are monotonically decreasing in concentration $\mathcal{C}$ and $\mathcal{P}$.*

Proof. Follows from differentiating. $\square$

This result shows that the basic result from the homogeneous firms case extends to the heterogeneous firms case: there is a negative relationship between the concentration of employment shares as measured by $\mathcal{C}$ and wages. What is new is that when productivity concentration as measured by $\mathcal{P}$ becomes more concentrated this also depresses wages. Below, we offer a model-based strategy to systematically decompose the contribution of the two using only data on firm level wages and firm size.

2.2 Concentration, Pass-Through, and Individual Wages

We now extend our previous results on firm-level wages to the heterogeneous productivity case. To that end, it is useful to define $\Pi \equiv \frac{\beta\lambda(1-\alpha)}{1-\beta+\beta(\lambda+\delta)}(b-\bar{w})$. Importantly, Π depends only on the mean wage and primitives. It is linearly decreasing in the mean wage and, as such, an affine transformation of $1-\bar{\omega}$ as defined in 4. As a consequence, it is simply another way of summarizing market power that comes in handy role in the following result:

Proposition 5. *Firm level wages w_i satisfy*

$$(1-\tau f_i)(p_i - w_i) = (1-\alpha)(p_i - b) + \Pi$$

Proof. See Appendix G. $\square$

To interpret this result, note that $p_i - w_i = (1-\alpha)(p_i - b)$ is the solution to static Nash bargain. Suppose market structure changes, leading to a decline in the mean wage $\bar{w}$ and an increase in Π. This affects employers market-wide, even those with unchanged size. In this case, wages at all firms fall as follows from the expression in the proposition. The multiplier $(1-\tau f_i)$ is the size markdown. It again reflects the fact that larger firms have a larger ability to punish off the equilibrium path. The result in proposition 5 is also key in our empirical strategy when mapping the full model to the data as discussed below.[4]

Of course, the proposition also shows that, all else equal, more productive firms pay higher wages. The following corollary records the coefficient that governs the pass-through from a linear projection of productivity levels on wage levels:

Corollary 4. *The firm-level productivity pass-through coefficient is:*

$$\frac{\alpha - \tau f_i}{1-\tau f_i}.$$

[4]We can also use proposition 5 to express relative wages in a form similar to proposition 2,

$$(1-\tau f_i)(p_i - w_i) - (1-\alpha)(p_i - b) = (1-\tau f_j)(p_j - w_j) - (1-\alpha)(p_j - b).$$

This expression nests Proposition 2.

This expression shows that the model generates size-dependent pass through of productivity to wages. We can see that the pass-through coefficient is maximized at α at the smallest firms in the economy. This pass-through reflects the fact that firms and workers divide the surplus, and the worker share is given by α. As the firm's size-based market power increases, the pass-through rate declines. In the monopolist limit, the pass-through coefficient can be arbitrarily close to zero when workers are patient and unemployment spells are short for the same reasons discussed above in the context of proposition 1.

Another important aspect of the corollary is the implication that firm level pass-through in levels is independent of the overall market structure. That is, market level concentration matters for the level of wages, but not for the distribution of relative wages across employers. In addition, the result shows that the the pass-through of size or productivity shocks at the firm level and, as a consequence, the size-productivity gradient, are disconnected from market structure.

2.3 Backing out Firm-Level Productivity

We next show how one can use proposition 5 to invert the model to solve for the market-wide vector of firm-level productivities p_i using information on firm-level wages, w_i, and market shares, f_i.

To do so, we express b as a multiple of mean output such that $b = \tilde{b}p^1$. We further define $c_1 \equiv \frac{\beta\lambda(1-\alpha)}{1-\beta+\beta(\lambda+\delta)}$ and $c_2 \equiv \tilde{b}(c_1 - (1-\alpha))$. Also, define $\mathbf{1}$ to be the $N \times 1$ vector of all ones. Define $\mathbf{F}$ to be the $N \times 1$ vector of f_i. Let $\mathbf{I}$ denote the $N \times N$ identity matrix, and let $\mathbf{D_F}$ denote the $N \times N$ matrix that has the employment shares f_i on its diagonal. Let $\mathbf{P}$ be the $N \times 1$ vector of productivities and $\mathbf{W}$ the $N \times 1$ vector of wages. To ease notation, let $\mathbf{A} \equiv \alpha\mathbf{I} - \tau\mathbf{D_F}$. Then, we can write the expression in Proposition 5 in matrix form as

$$[\alpha\mathbf{I} - \tau\mathbf{D_F}]\mathbf{P} = [\mathbf{I} - \tau\mathbf{D_F}]\mathbf{W} + c_2\mathbf{1}\mathbf{F}'\mathbf{P} - c_1\mathbf{1}\mathbf{F}'\mathbf{W}$$

and so

$$\mathbf{P} = [\mathbf{A} - c_2\mathbf{1}\mathbf{F}']^{-1}[\mathbf{I} - \tau\mathbf{D_F} - c_1\mathbf{1}\mathbf{F}']\mathbf{W}.$$

The following Proposition derives a closed form expression for $\mathbf{P}$:

Proposition 6. *The productivity vector* $\mathbf{P}$ *exists and is given by*

$$\mathbf{P} = \left[\mathbf{A}^{-1} + c_2\frac{\mathbf{A}^{-1}\mathbf{1}\mathbf{F}'\mathbf{A}^{-1}}{1 - c_2\mathbf{F}'\mathbf{A}^{-1}\mathbf{1}}\right][\mathbf{I} - \tau\mathbf{D_F} - c_1\mathbf{1}\mathbf{F}']\mathbf{W}.$$

Proof. See Appendix H.[5] □

We highlight that $\mathbf{A}$ is a diagonal matrix and so there is a closed-form expression for the inverse. The result in proposition 6 is key for the empirical implementation of the full model with heterogeneous productivity. It states that, having taken a stance on the five model parameters, the

[5] Thanks to Mary Wootters for pointing us to the Sherman-Morrison lemma, which is used in the proof.

vector of firm level productivities can be solved for from the firm size and wage distribution in one single step by inverting the model.

3 Data and measurement

In this section we introduce the data that we use and the sample restrictions we impose. We then discuss how we define a labor market, and how we define and measure the variables and parameters that appear in the model.

3.1 Matched employer-employee data

Our analysis is based on the Austrian labor market data base (AMDB) that covers the universe of private sector employment in Austria. For the period from 1997 to 2015, the AMDB provides daily information on employment and unemployment spells, reports annual wages (including base pay and bonus payments) for each worker-firm combination, and contains some worker characteristics (age, gender, nationality) and firm characteristics (industry, geographical location, year of foundation).

We make the following sample restrictions. First, we restrict our sample to regular workers (blue and white collar workers) and exclude marginal workers, short-time workers, and apprentices. Second, we restrict the analysis to firms with 5 or more workers.

3.2 Market definition

We consider a few different notions of market definition. Following the literature, we consider markets based on observable features of firms such as industry and geography (as well as their interaction). In particular, we examine concentration within 4-digit NACE industry codes, within NUTS-3 regions (slightly smaller than commuting zones in the US),[6] and within industry by region cells.

A large share of worker flows, however, occurs across industry and regional boundaries.[7] Predefined categorizations therefore do not necessarily capture the set of reasonable potential employers for a given worker. Likewise, a commensurately long literature discusses whether human capital is industry-, occupation-, or task-specific (e.g., Neal (1995), Kambourov and Manovskii (2009), and Gathmann and Schonberg (2010)).

To address these concerns, we use as our primary definition of a labor market an endogenous notion that clusters firms based on observed worker flows. This definition corresponds to the model in the sense that in the model a labor market is a set of firms where a worker would plausibly go following a spell of unemployment. We follow Nimczik (2018) and estimate a stochastic block model on the network of worker flows. A key tuning parameter is the number of labor markets to

[6]There are on average about 440,000 people per commuting zone in the US; there are on average about 250,000 people per NUTS-3 region.

[7]For the importance of cross-industry flows in the U.S., see Bjelland et al. (2011), especially Figure 7 documenting that over half of employer-to-empoyer flows are across 11 super-sectors (which are coarser than 1 digit NAICS industries).

consider. We pick the number of labor markets to maximize the likelihood of the objective function. This leads us to 369 labor markets. We refer readers to Nimczik (2018) for complete details, but in Appendix I we provide a basic sketch of what we do.

We compute measures market-by-market, and then report results on an employment-weighted basis.

3.3 Measuring variables

The central variables that the model considers are firm size and wages. We now discuss how we define and measure these variables in the data.

Firm size: f_i We employ the following measure of firm size f_i: We count the number of regular employees in a given firm at a reference date (August 10th) each year. This measure has the virtue of simplicity and comparability to previous studies that have computed employment-based HHI (e.g., Azar, Marinescu, and Steinbaum (2017), Benmelech, Bergman, and Kim (2018) and Rinz (2018)). We also report some sensitivity to an alternative definition that tries to capture vacancy shares by measuring the share of new hires at an employer.

Wages: w_i The AMDB provides annual information on gross base wages and bonus payments for each match between a worker and a firm. The wage data are capped at the social security contribution limit. We compute daily salaries by combining annual base and bonus payments and dividing by the number of days employed. We convert daily salaries to real wages using the consumer price index provided by Statistik Austria with 2000 as base year.

The model-relevant notion of a wage is a firm-specific wage in **levels**. The reason it is in levels and not logs is that in the bargaining problem firms and workers value the units equally. Because it is a firm specific wage and the model features homogeneous workers, we would ideally control for compositional differences across firms. But we have limited covariates and it is not clear how well standard wage regression adjustments perform in levels. Moreover, it is not clear whether such adjustments would be model-consistent. Instead, we use the median wage. This has the benefit of addressing the censoring as well as being transparent. We report sensitivity to considering alternative quantiles of the wage distribution.

3.4 Measuring parameters

The model depends on 6 parameters: $\{b, \lambda, \underline{\lambda}, \delta, \alpha, \beta\}$. We now discuss what feature of the data drives our choice of each of the parameters.

Job finding rate: λ We measure λ by calculating the share of workers who are unemployed in month t who are employed in month $t+1$. Across years, this rates ranges from 0.099 to 0.136, with an average of 0.121.

Job finding rate when a sole applicant: $\underline{\lambda}$ The urn-ball matching function implies a unique value of $\frac{v}{u}$ that generates the λ; given this value of $\frac{v}{u}$, we then calculate $\underline{\lambda}$. Given that $\lambda = 0.121$, the implied $\underline{\lambda}$ is extremely small and is 0.0002.

Job destruction rate: δ We measure δ using information in the unemployment rate and λ. In particular, in steady state standard mass balance arguments imply that $u = \frac{\delta}{\lambda+\delta}$. Rearranging, we have that $\delta = \frac{u\lambda}{1-u}$. Given that the average unemployment rate in Austria is 0.047, this gives us $\delta = 0.006$.

Worker bargaining power: α We pick α to match a labor share of 0.66 in an initial time period. To do so, we use the model-implied productivities from Proposition 6. Intuitively, holding everything else constant, increasing α leads to a higher labor share and this relationship is monotonic. Hence, it is straightforward to numerically find the value of α that implies a labor share of $\frac{2}{3}$. We find $\alpha = 0.105$. We also report sensitivity to alternative values of α.

Flow value of unemployment: b We measure b by picking a constant multiple of average productivity in the economy; i.e., $b = \tilde{b}p^1$. The reason is that in this model we treat the mean productivity as a structural (and exogenous) object. Importantly, this normalization means that b is time-varying. In contrast, the typical approach to calibrating b picks it as some fraction of mean (or median wages). In this model, this approach is not model-consistent because we view wages as reflecting both bargaining relative to b, as well as market structure and so b would not be invariant in counterfactuals in which we varied concentration.

To pick a number for $\tilde{b}$, we follow standard calibrations and target $b = 0.4w^1$. We pick α to get labor share $(\frac{w^1}{p^1})$ equal to $\frac{2}{3}$ in an initial period. Combining, this implies that $\tilde{b} = \frac{b}{p^1} = \frac{0.4w^1}{p^1} = 0.4 \times \frac{2}{3} = 0.267$. We also report sensitivity to alternative values of $\tilde{b}$.

Time discount: β There is no information in the data that informs this parameter, and so we follow standard convention and set β so that the annual discount factor is 0.95. On a monthly basis this gives us $\beta = 0.95^{1/12} = 0.9957$.

Summary: Table 1 summarizes our parameter values.

4 Quantifying the effect of market structure on wages

We now use our model to quantify the effect of changes in market structure in Austria from 1997-2014 on wages. We begin by discussing trends in market structure in Austria. We then use our model to quantify the effect of these changes in market structure on the level of wages. We measure how imperfect competition affects the level of wages by considering the effect on wages of shifting from the existing market structure to the atomistic benchmark. We then show how our average results mask considerable heterogeneity across labor markets. Finally, we consider the sensitivity of our results to changes in market definition and some parameter values.

4.1 Trends in market structure

Figure 1 shows that concentration, whether measured by HHI or $\mathcal{C}$ has followed a u-shaped pattern from 1997-2014. The peak of the u-shape is at around 0.120 and the minimum is at around 0.105. The similarity of the trends in the two figures emphasizes that while it is logically possible for the model-based measure to depart in important ways from the HHI that this gap is small in practice. One regularity is that our concentration index is always larger than the HHI.

Our model helps us interpret the magnitudes of the changes. In particular, $\mathcal{C}$ first decreases and then increases by about 0.015, or approximately 10%. Since $\mathcal{C}$ has the same effect on wages as a movement in the job finding parameter or the bargaining parameter, this implies that these changes in concentration are equivalent to about a 10% change in worker bargaining power.

Figure 2 shows that productivity-weighted concentration wedge, $\mathcal{P}$ was flat and then fell. Over the sample period $\mathcal{P}$ fell by over 50% (from 0.025 to 0.010). Mechanically, for $\mathcal{P}$ to rise means that the productivity-size correlation is decreasing. As with $\mathcal{C}$, $\mathcal{P}$ enters the compensation equation in multiplicative units that are the same as the bargaining parameter. So this says that this change was equivalent to bargaining power increasing by over 50%.

To give some intuition about what in the underlying data drives the change in $\mathcal{P}$, Figure 3 shows that the size wage correlation has changed over time. Intuitively, the model interprets a positive size-wage relationship as telling us that there is a positive size-productivity correlation because larger firms would pay lower wages in the absence of being more productive. The declining size-wage correlation then suggests that the size-productivity correlation has fallen over time (Bloom et al. (2018) document a similar decline in the US).

Both Figures also show parallel results for a single "arbitrary" market. The reason to show results for a single market is that the model contains non-linearities. In the next sections, when we turn to the counterfactuals the employment-weighted counterfactuals are not monotone transformations of the employment weighted trends. In contrast, for the single market, the trends will translate more intuitively into counterfactuals.

4.2 Effects of changes in market structure on compensation

We use the compensation equation in Proposition 4 to quantify the effects of these changes in market structure on compensation. Using the model allows us to isolate the pure role of market structure. Naturally, in the data it is typically difficult to isolate exogenous shifts in market structure that do not have independent effects on wages. To do so, we fix all parameters. We then allow the productivity-weighted concentration wedge ($\mathcal{P}$) to evolve as in the data, while holding concentration ($\mathcal{C}$) fixed. We then additionally allow concentration to vary.

Figure 4 shows the effect of the change in $\mathcal{P}$ and $\mathcal{C}$ on compensation, holding everything else constant. The Figure shows that the time series variation in $\mathcal{P}$ alone leads to a decline in compensation of almost ten percentage points.

Similarly, the Figure shows that when we allow both $\mathcal{C}$ and $\mathcal{P}$ to vary that—relative to just allowing $\mathcal{P}$ to vary—this can also lead to fairly substantial moves in the worker compensation share. In particular, when we allow both $\mathcal{C}$ and $\mathcal{P}$ to vary, the change in market structure leads to an increase in compensation of three percentage points.

We highlight that one appealing feature of our framework is that we can separate the effect of changes in concentration from changes in productivity-weighted concentration. This is important as the productivity weighted concentration reflects the size-wage gradient, which has changed. It is notable that quantitatively the most important source of changes over time in wages due to changes in market structure in fact comes through the changing size-productivity relationship. This effect of changes in market structure on wages is missed in typical analyses that focus on an employment-weighted HHI.

As we discuss more below, this average effect masks substantial heterogeneity across markets. Looking at the bottom Panel, for example, we can see that in this selected market the overall movements in counterfactual wages are much smaller than the average. Moreover, the incremental contribution of changes in $\mathcal{C}$ is substantially smaller. If we consult the bottom Panel of Figure 1, then we can see why: $\mathcal{C}$ did not move by very much in this particular market.

4.3 The level of wages in the atomistic benchmark

We now use the compensation equation in Proposition 4 to quantify the change in wages from moving to the atomistic benchmark over time. This exercise provides a sense of the magnitude of the effects of imperfect competition of the form highlighted in this paper on wages. Specifically, we fix all parameters and then send $\mathcal{P}$ to zero, and then we send $\mathcal{C}$ to zero.

Figure 5 shows that moving to the atomistic benchmark has a large effect on wages. The Figure shows that through the lense of the model the existing productivity-size relationship depresses wages by about two percentage points a year. The Figure shows that if we moved to the atomistic benchmark that the level of wages would rise by about nine percentage points.

This ordering of effects is the reverse of what we found in the previous section. That is, for explaining changes over time the key change in market structure is in the productivity-weighted concentration, and not concentration directly. In contrast, the aspect of market structure that deviates most strongly from imperfect competition is the pure concentration component, and the productivity-weighted concentration contributes less.

One notable feature of our results is that simply reading off the nature of competition from the HHI would suggest that the Austrian labor market is not very concentrated. The threshold for a market to be considered "moderately concentrated" according to US antitrust authorities is 0.15.[8] We find that the maximum value of the HHI over the period we observe is 0.12. Nonetheless, we find that imperfect competition as measured through concentration depresses wages by about eight percentage points per year. This highlights the value of our structural framework which allows us

[8] See `https://www.justice.gov/atr/herfindahl-hirschman-index`.

to translate measures of concentration into wages and so to simply and transparently assess the effects of imperfect competition on wages.

4.4 Heterogeneity across markets

Our main results reported so far reflect employment-weighted averages over 369 distinct labor markets. Table 2 provides some sense of how concentration and our counterfactual results vary across labor markets.

Panel A shows that most labor markets are not very concentrated, but that the distribution of concentration is very skewed so that there are a few labor markets that are very concentrated. Two statistics emphasize this point. First, while the employment-weighted maximum HHI and $\mathcal{C}$ over the sample period are both 0.12, the median of these statistics are a third the size at 0.04. Similarly, the 95th percentile of concentration measures is an order of magnitude larger than the median; for example, while the median of the max of $\mathcal{C}$ is 0.04, the 95th percentile of this measure is 0.63.

Panel B shows that the effects of concentration are also concentrated in a few labor markets. For example, while on average moving to the atomistic benchmark raises wages by 10%, in the median labor market this increase is only 2%. But at the 95th percentile of labor markets, wages would increase by 75% in the atomistic benchmark. Combining the two panels, this emphasizes that our average results mask considerable heterogeneity. Specifically, to the extent that concentration affects wages, these effects are concentrated in a small number of labor markets so that most workers are not affected by concentration, but some workers see their wages dramatically depressed by concentration.

4.5 Sensitivity

Table 3 reports how our main result of the increase in wages when moving to the atomistic benchmark varies as a function of market definition, parameter choices and variable definitions.

Market definition Our baseline results use worker flows to define labor markets. The first few rows of Panel A show that when we define labor markets instead by either region or geography separately that concentration has a small effect on wages: the largest effect comes from 4 digit industries, where the counterfactual indicates a decrease in wages of 3.2 percentage points (as opposed to 8.2 percentage points in our baseline).

We find much larger effects when we follow conventional definitions in the literature and interact region and industry. For example, when we interact region and 3 digit industry (which is approximately comparable to the market definition in Berger, Herkenhoff, and Mongey (2018)) we find that in 2014 moving to the atomistic benchmark increases wages by 18 percentage points. When we interact region and 4 digit industry (which is approximately comparable to the market definition in Rinz (2018), we find that the atomistic benchmark increases wages by 26 percentage points. This last number is over 3 times the size of our baseline.

Holding fixed the number of labor markets, the endogenous labor market definition finds more concentrated labor markets and thus a larger effect of concentration on wages. In the industry $\times$ region labor market definitions when there are larger effects of concentration than in our baseline endogenous labor market definition, there are also over ten times as many labor markets. Naturally, defining more labor markets leads to higher measured concentration and thus larger effects of concentration on wages. In the industry and industry $\times$ region definitions where there are a similar number of labor markets (4 digit industry and 2 digit industry $\times$ region), we find effects of concentration 2.5 to 4 times the size with our endogenous labor markets than in the more conventionally defined labor markets. This emphasizes that the endogenous labor markets capture different features of the data than conventional labor market definitions.

Value of unemployment We chose the value of unemployment to match typical calibrations of b in the literature. Panel B shows how our results vary with the choice of $\tilde{b}$ (recall that $\tilde{b}$ is flow value of unemployment as a share of mean productivity). When we lower $\tilde{b}$ from our benchmark of 0.267 to 0.10, we find larger effects of concentration on wages (0.137 vs. 0.082).

This dimension of sensitivity does not reflect anything about the fundamental economics of our model; instead, it reflects how our approach to measuring productivity depends on parameters. To see this, note that holding observed wages and α constant, if we reduce $\tilde{b}$, then to generate the same level of wages there needs to be a higher level of productivity. Hence, when we move to the atomistic benchmark, there is more productivity for the workers to extract from the employment relationship and so a bigger difference in wages.

When we raise $\tilde{b}$ to 0.4, we find a smaller effect of moving to the atomistic benchmark relative to our baseline value of $\tilde{b}$. The logic is identical to the previous paragraph, but the effect goes in the opposite direction. Thus, as we increase $\tilde{b}$ we will find a smaller effect of changes in concentration on wages.

Worker bargaining power We chose α to match labor share. An alternative way to choose α would be to use Corollary 4 and choose α to match pass-through estimates of productivity to wages at small firms. For example, if we use the average results in Kline et al. (Forthcoming), we would get $\alpha = 0.30$, though if we use the results for workers hired post patent application (which corresponds more closely to the experiment in our model) we would arrive at an α of approximately 0.[9]

Panel B shows that our results are very sensitive to the choice of worker bargaining power (α). In particular, as we increase α from the baseline of 0.1 to 0.4, the implied effects on wages of moving to the atomistic benchmark fall by an order of magnitude. In 2014, the increase in wages falls from an increase of 8.2% to an increase of 0.7%. This quantitative result can be anticipated

[9] Kline et al. (Forthcoming) do not report a pass-through coefficient for new hires. This guess comes from comparing column (2) and column (4) in Table 7, which shows that average stayer earnings rise by on average $7,780, while new hire earning rise by a statistically insignificant $110—or almost two orders of magnitude less. Admittedly, there are challenging issues around the changes in composition of new hires for interpreting the effects on new hires.

from Corollary 2 where we showed that the sensitivity of wages to concentration is decreasing in worker bargaining power. In the discussion of that Corollary, we emphasized the intuition that as α rises, wages are more dependent on the inside option of productivity, instead of the outside option that concentration can move. Hence, with larger α, wages are less sensitive to movements in concentration.

Firm-level wage Our baseline results use the median firm-level wage. Panel C shows that if we instead use the 25th or the 75th percentiles of the firm-level wage distribution that our results are virtually unchanged.

Firm size Our baseline results measure firm size using employment at a point in time. We instead measure firm size as the share of hires in a year from unemployment. The final row of Panel C shows that this leads to smaller wage effects of moving to the atomistic benchmark. Notably, setting $\mathcal{P}$ to zero decreases wages. The reason for this divergence from our baseline results is that this alternative measure produces a negative correlation between wages and size, which translates into a negative wedge between concentration and productivity-weighted concentration. The productivity-weighted wedge therefore increases wages relative to the atomistic benchmark. Finally, additionally setting $\mathcal{C}$ to zero increases wages, but by only half as much as in our baseline results.

5 Discussion

This paper develops a structural model that provides a microfoundation for an equilibrium relationship between market structure—in particular, concentration—and wages. The core idea of the model is that size is a source of market power because firms do not compete with their own future vacancies. This lack of competition means that workers' outside options are worse when bargaining with a large firm. As a result, wages are lower at large firms and in more concentrated markets.

We then extend the model to account for productivity heterogeneity. The model separates the effects of concentration and productivity-weighted concentration on wages.

The model allows us to simply and transparently assess the effects of (changes in) market structure on levels and trends in wages. We implement our framework in Austrian matched employer-employee data. We overcome the arbitrariness of typical definitions of labor markets, and define endogenous labor markets based on worker flows.

While conventional measures of concentration would suggest that the Austrian labor market is quite competitive, we find that wages are eight percent lower than they would be if all firms acted atomistically. The quantitatively more important force in the distance from the atomistic benchmark is concentration by itself, with the productivity-weighted concentration wedge playing a smaller role. In contrast, when we look at trends over time, the key change in market structure that drives changes in wages in our framework is the changing contribution of productivity-weighted concentration wedge, which reflects the changing size-wage premium.

References

Arnoud, Antoine. 2018. "Automation Threat and Wage Bargaining." Working paper.

Azar, Jose, Ioana Marinescu, and Marshall Steinbaum. 2017. "Labor Market Concentration." Working paper.

Azar, Jose, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska. 2018. "Concentration in US Labor Markets: Evidence from Vacancy Data." Working paper.

Beaudry, Paul, David A. Green, and Benjamin Sand. 2012. "Does industry composition matter for wages? A test of search and bargaining theory." *Econometrica* 80 (3):1063–1104.

Benmelech, Efraim, Nittai Bergman, and Hyunseob Kim. 2018. "Strong Employers and Weak Employees: How does employer concentration affect wages?" Working Paper 24307, NBER.

Berger, David W., Kyle Herkenhoff, and Simon Mongey. 2018. "Labor Market Power." Working paper.

Bjelland, Melissa, Bruce Fallick, John Haltiwanger, and Erika McEntarfer. 2011. "Employer-to-Employer Flows in the United States: Estimates Using Linked Employer-Employee Data." *Journal of Business and Economic Statistics* 29 (4):493–505.

Blanchard, Olivier Jean and Peter Diamond. 1994. "Ranking, Unemployment Duration, and Wages." *The Review of Economic Studies* 61 (3):417–434.

Bloom, Nicholas, Fatih Guvenen, Benjamin S. Smith, Jae Song, and Till von Wachter. 2018. "Is the Large Firm Wage Premium Dead or Merely Resting?" *American Economic Association: Papers and Proceedings* 108:317–322.

Boal, William M. and Michael R. Ransom. 1997. "Monopsony in the labor market." *Journal of Economic Literature* 35 (1):86–112.

Bunting, Robert L. 1962. *Employer Concentration in Local Labor Markets.* University of North Carolina Press.

Caldwell, Sydnee and Oren Danieli. 2018. "Outside options in the labor market." Working paper.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36 (S1):S13–S70.

Davis, Steven J. and Brenda Samaniego. 2019. "Application Flows." Working paper.

Galenianos, Manolis, Philipp Kircher, and Gabor Virag. 2011. "Market power and efficiency in a search model." *International Economic Review* 52 (1):188–215.

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

Haanwinckel, Daniel. 2018. "Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution." Working paper.

Hershbein, Brad, Claudia Macaluso, and Chen Yeh. 2018. "Concentration in U.S. local labor markets: evidence from vacancy and employment data." Working paper.

Kambourov, Gueorgui and Iourri Manovskii. 2009. "Occupational Specificity of Human Capital." *International Economic Review* 50 (1):63–115.

Karabarbounis, Loukas and Brent Neiman. 2014. "The Global Decline in Labor Share." *Quarterly Journal of Economics* 129 (1):61–103.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. Forthcoming. "Who Profits from Patents? Rent-Sharing at Innovative Firms." *Quarterly Journal of Economics* .

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2017. "Imperfect Competition and Rent Sharing in the U.S. Labor Market." Working paper.

Lipsius, Ben. 2018. "Labor market concentration does not explain the falling labor share." Working paper.

MacKenzie, Gaelan. 2018. "Trade and Market Power in Product and Labor Markets." Working paper.

Menzio, Guido and Nicholas Trachter. 2015. "Equilibrium Price dispersion with sequential search." *Journal of Economic Theory* 160:188–215.

Neal, Derek. 1995. "Industry-Specific Human Capital: Evidence from Displaced Workers." *Journal of Labor Economics* 13 (4):653–677.

Nimczik, Jan Sebastian. 2018. "Job mobility and endogenous labor markets." Working paper.

Qui, Yue and Aaron Sojourner. 2019. "Labor-Market Concentration and Labor Compensation." Working paper.

Rinz, Kevin. 2018. "Labor Market Concentration, Earnings Inequality, and Earnings Mobility." Working paper.

Shimer, Robert. 2005. "The Assignment of Workers to Jobs in An Economy with Coordination Frictions." *Journal of Political Economy* 113 (5):996–1025.

Webber, Douglas A. 2015. "Firm market power and the earnings distribution." *Labour Economics* 35:123–134.

———. 2018. "Employment Competition over the Business Cycle: The impact of competition in the labor market." Working paper.

Table 1: Parameter values

Parameter	Value
λ	0.121
$\underline{\lambda}$	0.0002
δ	0.006
α	0.105
$\tilde{b}$	0.267
β	0.9992

Notes: This Table reports monthly parameter values.

Table 2: Heterogeneity of effects of market structure across markets

	Average	Median	5th percentile	95th percentile
Panel A. Concentration measures				
max HHI	0.12	0.04	0.01	0.62
min HHI	0.10	0.03	0.01	0.43
max $\mathcal{C}$	0.12	0.04	0.01	0.63
min $\mathcal{C}$	0.10	0.03	0.01	0.44
max $\mathcal{P}$	0.03	0.01	-0.04	0.20
min $\mathcal{P}$	0.01	0.01	-0.08	0.13
Panel B. Counterfactuals				
max $\Delta\omega$ in time trends with just $\mathcal{P}$	1.00	1.00	1.00	1.18
max $\Delta\omega$ in time trends with $\mathcal{P}$ and $\mathcal{C}$	1.03	1.00	1.00	1.25
max $\Delta\bar{w}$ in atomistic benchmark $\mathcal{P} = 0$	0.03	0.01	-0.02	0.19
max $\Delta\bar{w}$ in atomistic benchmark $\mathcal{P} = \mathcal{C} = 0$	0.10	0.02	-0.00	0.75

Notes: This Table reports how a variety of measures vary across markets. The average column corresponds to the main results in the text and reflects employment-weighted averages. The remaining three columns report results for employment-weighted quantiles of the markets. In Panel A, the maximum of the concentration measures is calculated as follows: for each market, we take the maximum value of the measure over time and then compute the employment-weighted quantile. Similarly, the minimum reflects quantiles of the distribution of the minimum, where there is one observation per market. HHI is the Hirschman-Herfindahl index, $\mathcal{C}$ is our model-based measure of concentration, and $\mathcal{P}$ is the productivity-concentration weighted wedge. In Panel B, we similarly compute the largest change in each measure within market, and then report quantiles of this distribution across markets. The two exercises use Proposition 4 to compute counterfactual wages either using shifts in $\mathcal{C}$ and $\mathcal{P}$ over time, or by sending these quantities to zero.

Table 3: Sensitivity of effects of market structure

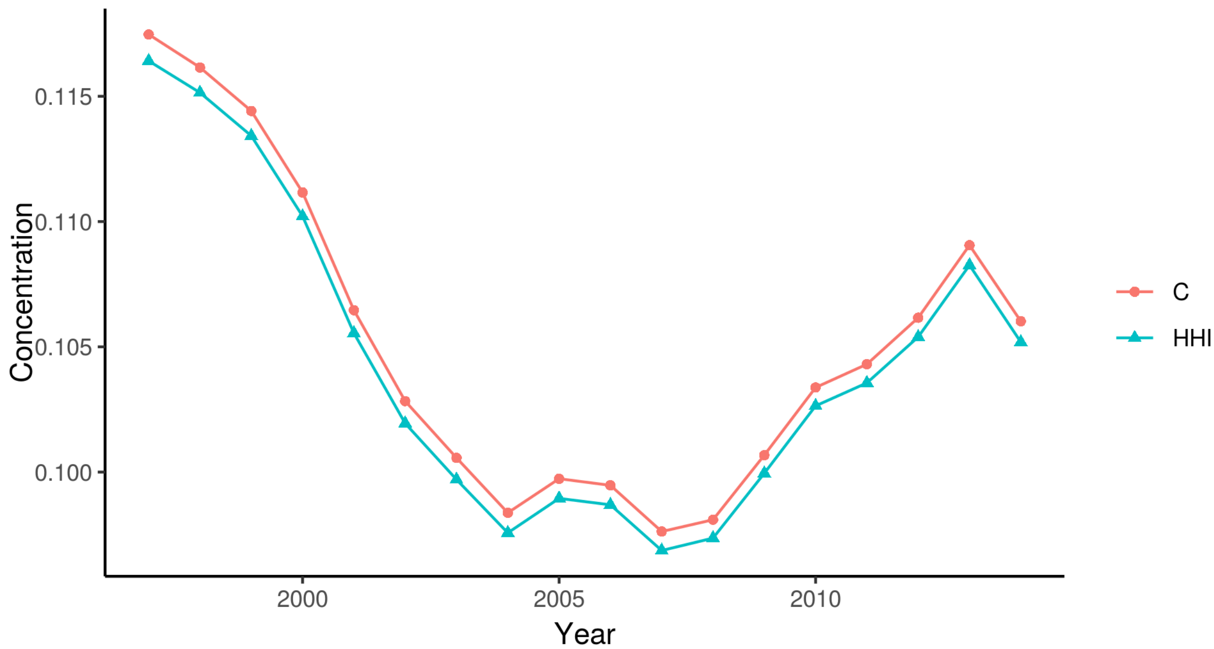
Increase in wages in the atomistic benchmark in 1997 and 2014

	Setting $\mathcal{P}=0$		Setting $\mathcal{C}=\mathcal{P}=0$	
	1997	2014	1997	2014
Baseline	0.025	0.023	0.097	0.082
Panel A. Alternative market definitions				
NUTS-3 regions (35)	0.002	0.003	0.004	0.005
2-digit industries (80)	0.008	0.006	0.014	0.011
3-digit industries (273)	0.013	0.009	0.030	0.024
4-digit industries (599)	0.017	0.012	0.043	0.032
2-digit industry × region (459)	0.018	0.011	0.035	0.023
3-digit industry × region (5410)	0.039	0.031	0.225	0.181
4-digit industry × region (9326)	0.040	0.037	0.315	0.259
Panel B. Alternative parameterizations				
$\tilde{b}=0.1$	0.031	0.028	0.171	0.137
$\tilde{b}=0.4$	0.023	0.021	0.070	0.060
$\alpha=0.2$	0.019	0.017	0.079	0.065
$\alpha=0.3$	0.005	0.005	0.018	0.016
$\alpha=0.4$	0.002	0.002	0.008	0.007
$\alpha=0.5$	0.001	0.001	0.004	0.004
Panel C. Alternative wage and size definitions				
w_i 25th percentile of firm-level wage distribution	0.031	0.024	0.104	0.083
w_i 75th percentile of firm-level wage distribution	0.022	0.023	0.093	0.082
f_i hires from unemployment	-0.003	-0.027	0.033	0.014

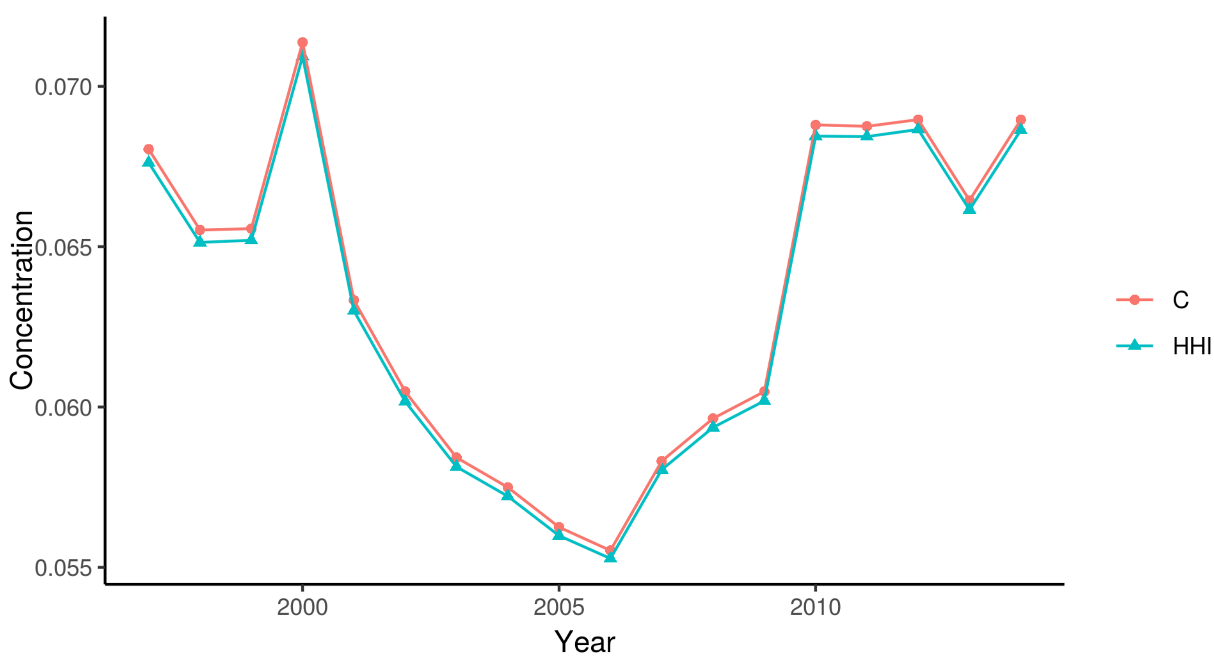
Notes: This Table considers the sensitivity of our benchmark results on the counterfactual of having firms behave atomistically on wages using the results in Proposition 4. The first row shows our baseline results where we use 369 endogenous labor markets and the parameterization reported in Table 1, where $\alpha = 0.105$ and $\tilde{b} = 0.267$. Panel A considers alternative market definitions based just on region, industry and industry × region. The number of markets is in parentheses. Panel B considers alternative choices of α and $\tilde{b}$. Panel C considers alternative definitions of the firm-level wage, where our baseline results use the median firm-level wage. It also considers a definition of f_i based on share of hires from unemployment, where our baseline results instead use the share of employment.

Figure 1: Trends in HHI and $\mathcal{C}$

(a) Employment-weighted average of all markets



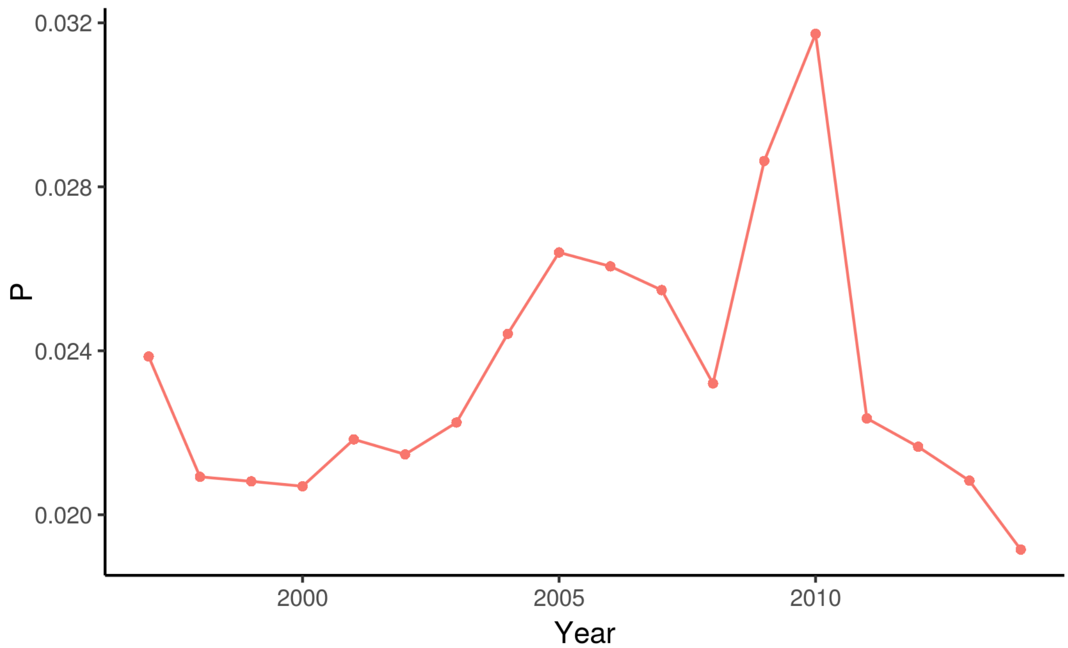
(b) A single market



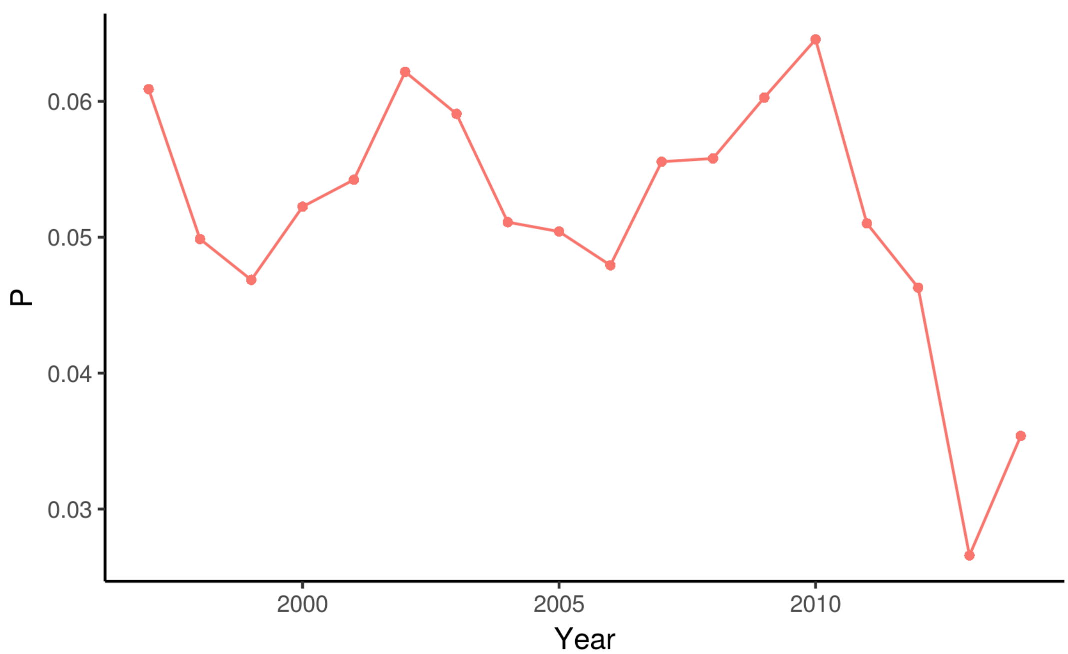
Notes: This figure plots concentration indexes $\mathcal{C}$ and HHI from 1997 - 2015 based on micro data from the Austrian labor market database. The top panel displays the employment-weighted average over all 369 endogenous labor markets, while the bottom panel displays the results for a single market.

Figure 2: Trends in $\mathcal{P}$

(a) Employment-weighted average of all markets



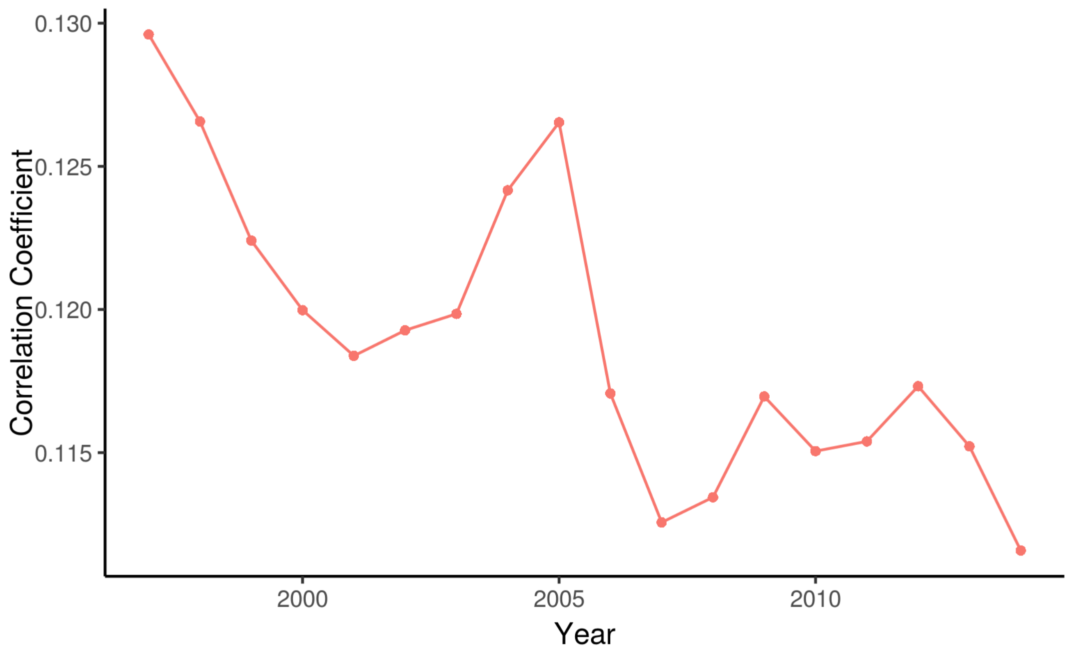
(b) A single market



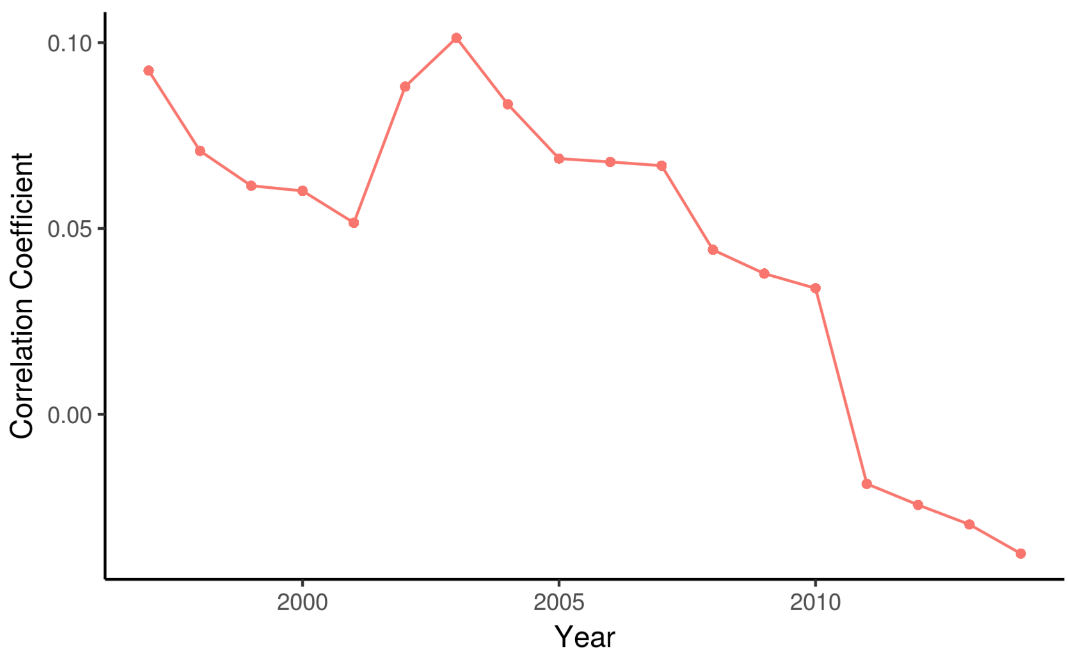
Notes: This figure shows the change over time of the wedge $\mathcal{P}$ between productivity-weighted concentration $\mathcal{C}^{\mathcal{P}}$ and concentration $\mathcal{C}$. The top panel displays the employment-weighted average over all 369 endogenous labor markets, while the bottom panel displays the results for a single market.

Figure 3: Declining size-wage gradient

(a) Employment-weighted average of all markets



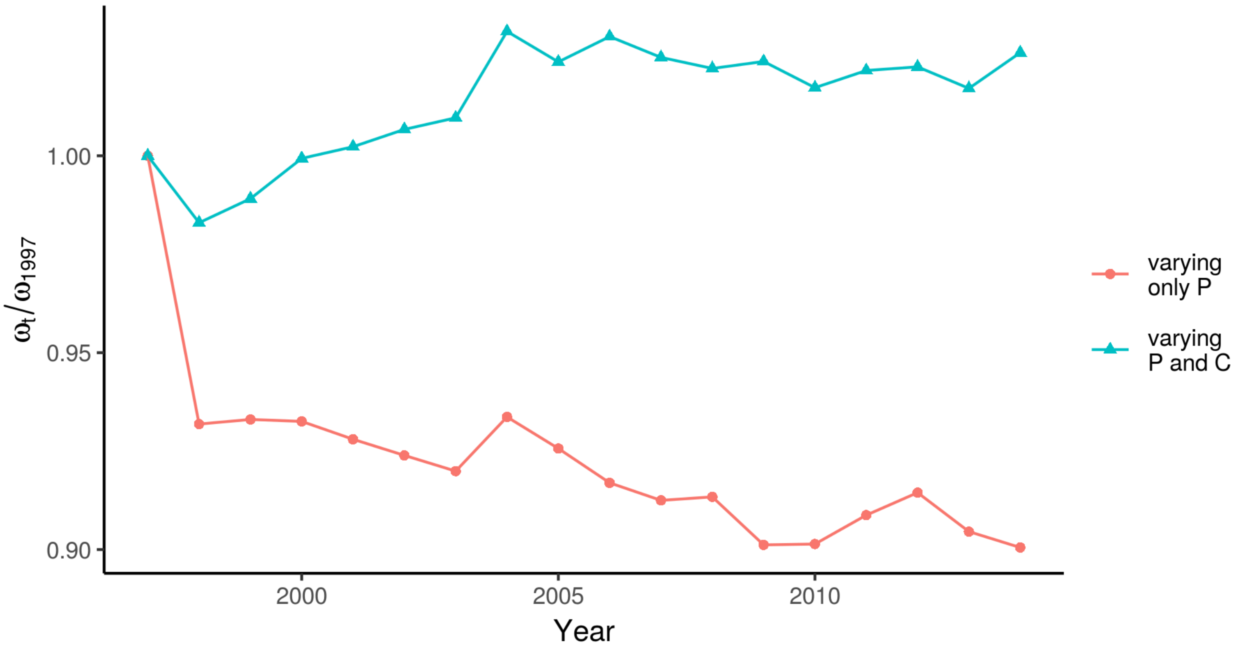
(b) A single market



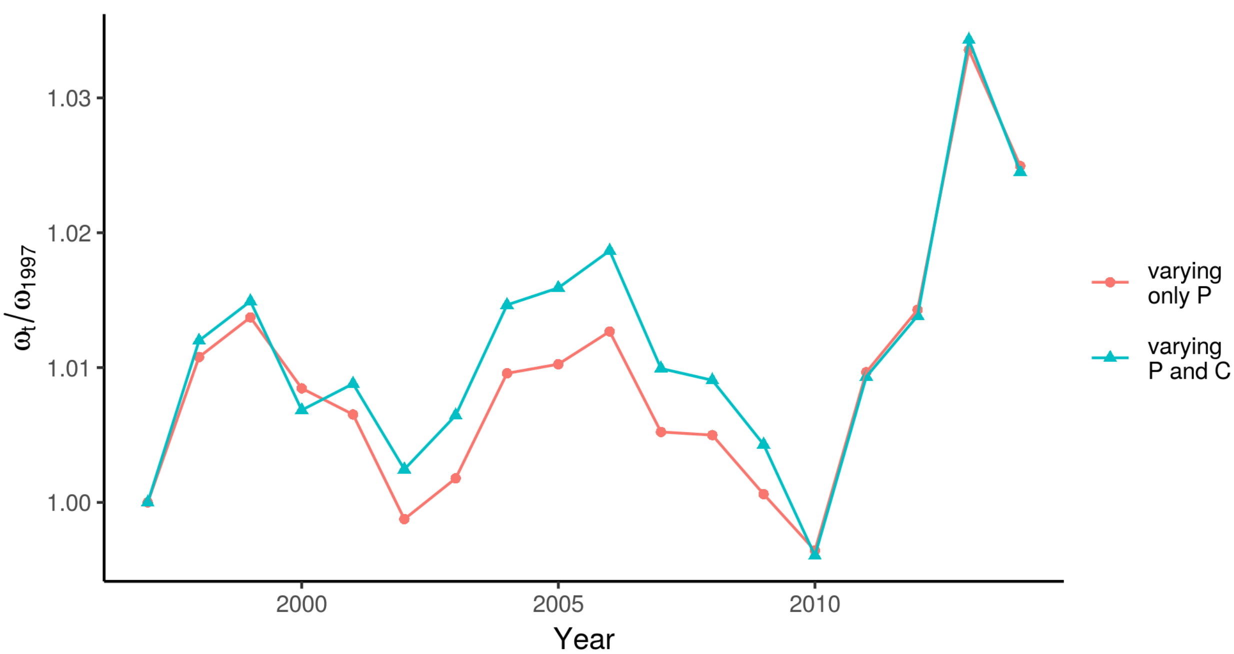
Notes: This figure plots the (employment-weighted average) of the correlation between firm size and firm-level median wages. Firm size is measured at a reference date (August 10th) each year and wages are the median of the firm-level distribution of regular employee wages. The top panel displays the employment-weighted average over all 369 endogenous labor markets, while the bottom panel displays the results for a single market.

Figure 4: Effect of changes over time in $\mathcal{P}$ and $\mathcal{C}$ on compensation (ω)

(a) Employment-weighted average of all markets



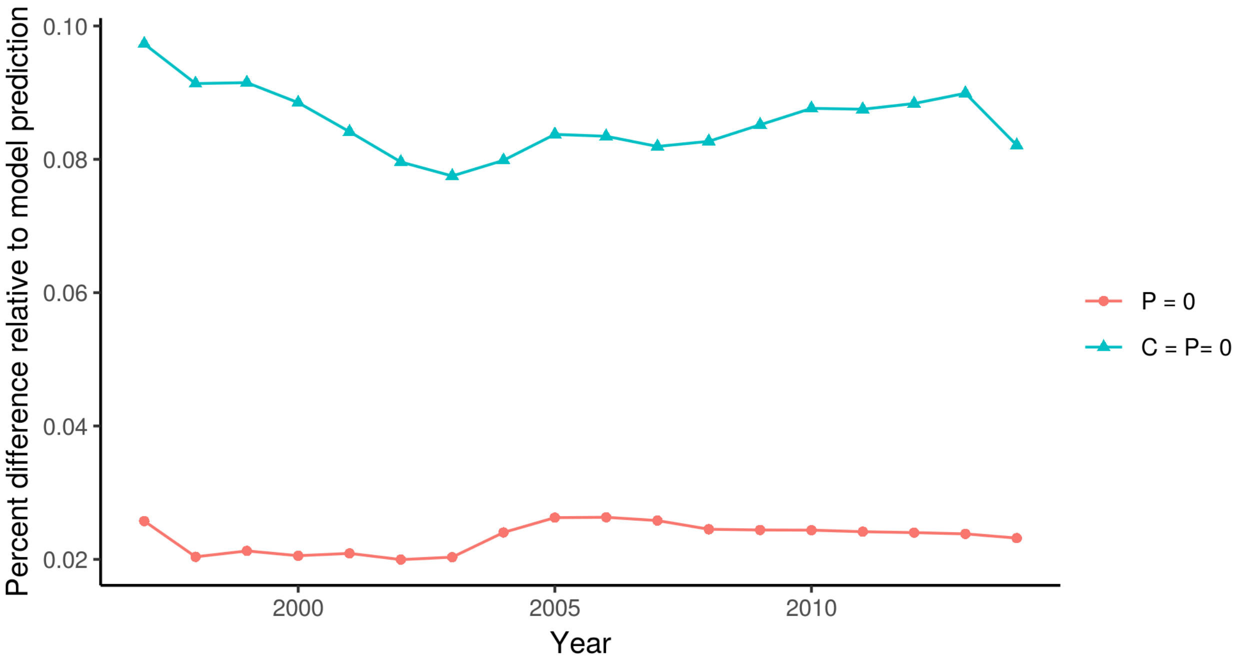
(b) A single market



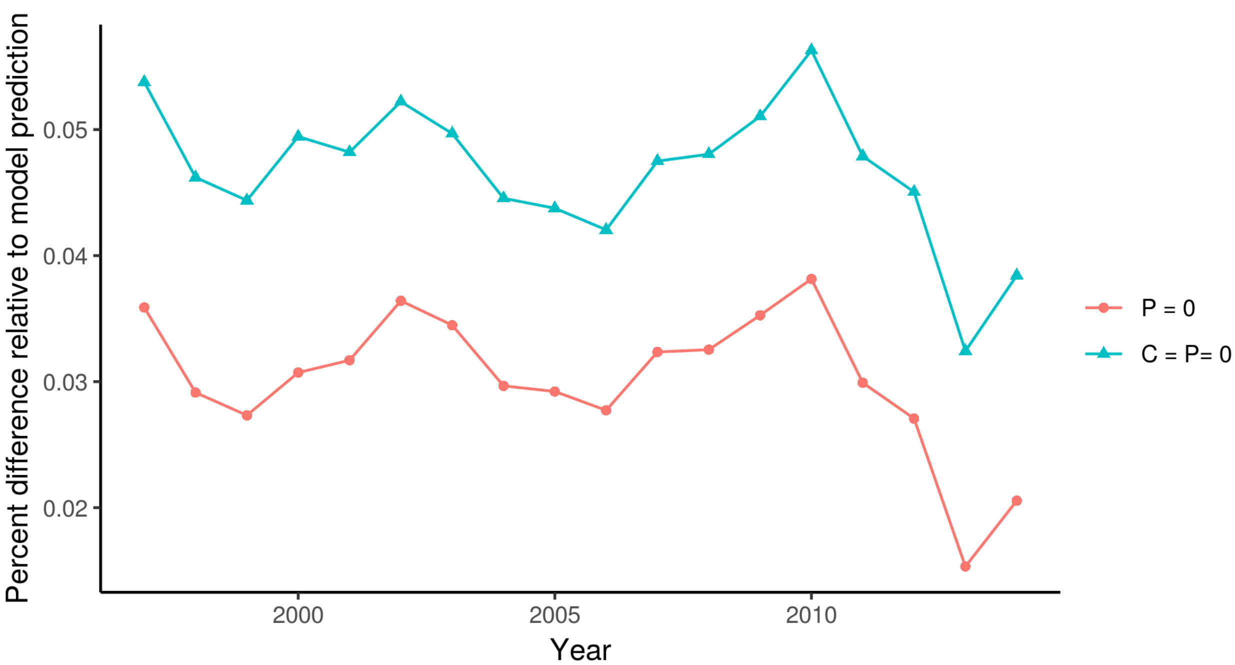
Notes: This Figure uses Proposition 4 to quantify the effect of the change over time in $\mathcal{C}$ and $\mathcal{P}$ documented in Figures 1 and 2 on compensation. The top panel displays the employment-weighted average over all 369 endogenous labor markets, while the bottom panel displays the results for a single market.

Figure 5: Increase in wages by moving to atomistic benchmark (over time)

(a) Employment-weighted average of all markets



(b) A single market



Notes: This Figure uses Proposition 4 to quantify the change in wages implied by moving to the atomistic benchmark over time. The top panel displays the employment-weighted average over all 369 endogenous labor markets, while the bottom panel displays the results for a single market.

A Example where $\mathcal{C}$ and HHI switch positions

In this Appendix, we describe two model economies. The ordering of the concentration of these economies according to $\mathcal{C}$ is different than the ordering according to HHI.

Relationship between the two economies: Choose c_1 such that $c_1 = \sqrt{c_2} - \epsilon$.

Economy 1: monopsonist with a competitive fringe:

- c_1 share of employment at the first firm;
- $\frac{1-c_1}{n-1}$ of employment at the remaining $n-1$ firms, where we let $n \to \infty$.

Economy 2: equally-sized, but finite number of firms:

- c_2 share of employment at each of the $\frac{1}{c_2}$ firms.

HHI in these two economies: For the first one:

$$c_1^2 + \frac{(1-c_1)^2}{n-1} \approx c_1^2,$$

where the $\approx$ relies on $n \to \infty$.

For the second one:

$$\frac{1}{c_2} c_2^2 = c_2.$$

Now $c_1^2 = (\sqrt{c_2} - \epsilon)^2 \approx c_2 - \epsilon < c_2$, so the second economy is more concentrated when measured using HHI.

$\mathcal{C}$ in these two economies: We now consider the $k > 2$ terms.

For the first economy:

$$c_1^k + (n-1)(\frac{1-c_1}{n-1})^k = c_1^k + \frac{(1-c_1)^k}{(n-1)^2} \approx c_1^k,$$

where the $\approx$ relies on taking $n \to \infty$.

For the second economy:

$$\frac{1}{c_2} c_2^k = c_2^{k-1}.$$

For $k > 2$ the first economy is now more concentrated. To see this note that

$$c_1^k = (\sqrt{c_2} - \epsilon)^k \approx c_2^{k/2} - \epsilon^k.$$

Because for $k > 2$ we have $\frac{k}{2} < k-1$, $c_2 < 1$ and ϵ is small,

$$c_2^{k/2} - \epsilon^k > c_2^{k-1}.$$

Hence, for small enough ϵ the first economy will be more concentrated according to $\mathcal{C}$. Intuitively, $\mathcal{C}$ places more weight on the largest firm than HHI (in the limit, only the largest share), and so the monopsonist with the competitive fringe is more concentrated according to $\mathcal{C}$ than HHI.

B Proof of Proposition 1

Proof. Now:

$$\begin{aligned} U_i &= b + \beta[\lambda \sum_{j \neq i} f_j W_j + \underline{\lambda} f_i W_i + (\lambda - \underline{\lambda}) f_i U_i + (1-\lambda) U_i] \\ U_i &= b + \beta[U_i + \lambda \sum_{j \neq i} f_j (W_j - U_i) + \underline{\lambda} f_i (W_i - U_i)]. \end{aligned} \tag{A1}$$

From equations (6), (3), and (2)

$$\begin{aligned} \alpha S_i = (W_i - U_i) &= w_i + \beta[\delta U + (1-\delta) W_i] - b - \beta[U_i + \lambda \sum_{j \neq i} f_j (W_j - U_i) + \underline{\lambda} f_i (W_i - U_i)] \\ &= w_i + \beta \alpha S_i - \beta[\delta \alpha S_i)] - b - \beta[\lambda \alpha S^1 - (\lambda - \underline{\lambda}) f_i \alpha S_i + \lambda \sum_j f_j (U_j - U_i)] + \beta \delta (U - U_i) \\ (1 - \beta(1-\delta)) \alpha S_i &= w_i - b + \beta(\lambda - \underline{\lambda}) f_i \alpha S_i - \beta \lambda [\alpha S^1 + \sum_j f_j (U_j - U_i)] + \beta \delta (U - U_i), \end{aligned} \tag{A2}$$

where we define $S^1 \equiv \sum_i f_i S_i$ and we used the fact that:

$$\begin{aligned} \sum_{j \neq i} f_j (f_i S_i - f_j S_j) &= \sum_{j \neq i} f_j (f_i S_i - f_j S_j) + f_i (f_i S_i - f_i S_i) \\ &= \sum_j f_j (f_i S_i - f_j S_j). \end{aligned}$$

Note that

$$\begin{aligned} U_k &= b + \beta[U_k + \lambda \sum_{j \neq k} f_j (W_j - U_k) + \underline{\lambda} f_k (W_k - U_k)] \\ &= b + \beta[U_k + \lambda W^1 - \lambda f_k W_k - \lambda(1 - f_k) U_k + \underline{\lambda} f_k (W_k - U_k)] \\ (1 - \beta(1-\lambda)) U_k &= b + \beta[\lambda W^1 - \lambda f_k W_k + \lambda f_k U_k + \underline{\lambda} f_k (W_k - U_k)] \\ (1 - \beta(1-\lambda)) U_k &= b + \beta[\lambda W^1 - (\lambda - \underline{\lambda}) f_k \alpha S_k] \end{aligned}$$

where $W^1 \equiv \sum_j f_j W_j$. Hence,

$$(U_j - U_i) = \frac{\beta(\lambda - \underline{\lambda})\alpha}{(1 - \beta(1-\lambda))}[f_i S_i - f_j S_j]. \tag{A3}$$

Note that:

$$\begin{aligned} U - U_i &= \beta[\lambda W + (1-\lambda) U] - \beta[U_i + \lambda \sum_{j \neq i} f_j (W_j - U_i) + \underline{\lambda} f_i (W_i - U_i)] \\ (1 - \beta(1-\lambda))(U - U_i) &= \beta(\lambda - \underline{\lambda}) f_i \alpha S_i \\ \beta \delta (U - U_i) &= \beta \delta \frac{\beta(\lambda - \underline{\lambda})}{(1 - \beta(1-\lambda))} f_i \alpha S_i. \end{aligned} \tag{A4}$$

Plug(A4) and (A3) into (A2) to get

$$
\begin{aligned}
&(1-\beta(1-\delta))\alpha S_i \\
&= w_i - b + \beta(\lambda-\underline{\lambda})f_i\alpha S_i - \beta\lambda[\alpha S^1 + \frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}\sum_j f_j[f_i S_i - f_j S_j]] + \beta\delta\frac{\beta(\lambda-\underline{\lambda})}{(1-\beta(1-\lambda))}f_i\alpha S_i
\end{aligned}
$$

$$
(1-\beta(1-\delta))\alpha S_i = w_i - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2 + \frac{1-\beta(1-\delta)}{1-\beta(1-\lambda)}\beta(\lambda-\underline{\lambda})f_i\alpha S_i. \quad \text{(A5)}
$$

Combine (5), (4), and the normalization that $V_i = 0$ to get that:

$$
w_i = 1 - (1-\beta(1-\delta))(1-\alpha)S_i. \quad \text{(A6)}
$$

Hence, combine (A6) and (A5)

$$
(1-\beta(1-\delta))S_i = 1 - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2 + \frac{1-\beta(1-\delta)}{1-\beta(1-\lambda)}\beta(\lambda-\underline{\lambda})f_i\alpha S_i. \quad \text{(A7)}
$$

Define $S^k \equiv \sum_i f_i^k S_i$, recall that $\tau = \frac{\beta(\lambda-\underline{\lambda})\alpha}{1-\beta(1-\lambda)}$ and that $f^k \equiv \sum_i f_i^k$, to rewrite (A7) as

$$
(1-\beta(1-\delta))S^k = f^k\left[1 - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2\right] + (1-\beta(1-\delta))\tau S^{k+1}. \quad \text{(A8)}
$$

Evaluate (A8) at $k = 1, 2, 3, ...$ and to get

$$
\begin{aligned}
(1-\beta(1-\delta))S^1 &= f^1\left[1 - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2\right] + (1-\beta(1-\delta))\tau S^2 \\
(1-\beta(1-\delta))S^2 &= f^2\left[1 - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2\right] + (1-\beta(1-\delta))\tau S^3 \\
(1-\beta(1-\delta))S^3 &= f^3\left[1 - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2\right] + (1-\beta(1-\delta))\tau S^4.
\end{aligned}
$$

Note that, for $k = 1$, we can also write

$$
(1-\beta(1-\delta))S^1 = 1 - b - \beta\lambda\alpha S^1 + \beta(\lambda-\underline{\lambda})\alpha\frac{1-\beta+\beta(\delta+\lambda)}{(1-\beta(1-\lambda))}S^2. \quad \text{(A9)}
$$

Hence:

$$
(1-\beta(1-\delta))S^1 = \left[1 - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2\right]\left[f^1 + \tau f^2 + \tau^2 f^3....\right] \quad \text{(A10)}
$$

$$
(1-\beta(1-\delta))S^2 = \left[1 - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2\right]\left[f^2 + \tau f^3 + \tau^2 f^4....\right]. \quad \text{(A11)}
$$

Define

$$
F \equiv \left(f^2 + \left(\frac{\lambda}{\lambda+r}\right)f^3 + \left(\frac{\lambda}{\lambda+r}\right)^2 f^4 + ...\right) = \sum_{k=2}^{\infty}\tau^{k-2}f^k \quad \text{(A12)}
$$

to get that, directly from equations (A10) and (A11)

$$S^2 = S^1 \frac{F}{1+\tau F} = S^1 \mathcal{C}. \tag{A13}$$

Plug this into equation (A9) to get that mean surplus is given by

$$S^1 = \frac{1-b}{\Big[(1-\beta(1-\delta)) + \beta\lambda\alpha\Big] - \tau\Big[1-\beta(1-\delta)+\beta\lambda\Big]\mathcal{C}}. \tag{A14}$$

This is where we use the approximation that $\underline{\lambda} \approx 0$. As a consequence,

$$\tau \approx \alpha \frac{\beta\lambda}{1-\beta(1-\lambda)}$$

and so

$$\frac{1-b}{\Big[(1-\beta(1-\delta)) + \beta\lambda\alpha\Big] - \Big[\lambda + \frac{\beta\lambda\delta}{1-\beta(1-\lambda)}\Big]\alpha\beta\mathcal{C}}$$

or

$$\frac{1-b}{1-\beta\left(1-\lambda\alpha\underbrace{[1-C]}_{\text{wedge 1}} - \underbrace{\delta\left[1-\alpha\mathcal{C}\left(\frac{\beta\lambda}{1-\beta\left(1-\lambda\right)}\right)\right]}_{\text{wedge 2}}\right)}. \tag{A15}$$

Integrate across equation (A6) to get

$$\begin{aligned} w^1 &= 1-(1-\beta(1-\delta))(1-\alpha)S^1 \\ \frac{w^1-b}{1-b} &= \frac{(1-\beta(1-\delta)+\beta\lambda)(\alpha-\tau\mathcal{C})}{\Big[(1-\beta(1-\delta)) + \beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}}. \end{aligned} \tag{A16}$$

For the expression for $1-\bar{\omega}$ reported in the proposition, again integrate across (A6) and plug in (A15). □

C Proof of Corollary 2

Proof. We want to show that the elasticity is increasing in α. Because the elasticity is always negative, increasing means that it becomes smaller in magnitude.

To keep notation more compact, it is helpful to note that:

$$\bar{w} = \bar{\omega}(1-b) + b \tag{A17}$$

Then:

$$\frac{\partial \bar{w}}{\partial \mathcal{C}} = (1-b)\frac{\partial \bar{\omega}}{\partial \mathcal{C}}. \tag{A18}$$

And:

$$\frac{\partial \bar{w}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{w}} = \frac{\partial \bar{\omega}}{\partial \mathcal{C}}\frac{(1-b)\mathcal{C}}{\bar{\omega}(1-b)+b}$$
$$= \frac{\partial \bar{\omega}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{\omega}+\frac{b}{1-b}}. \tag{A19}$$

Then:

$$\frac{\partial \frac{\partial \bar{w}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{w}}}{\partial \alpha} = \frac{\partial \frac{\partial \bar{\omega}}{\partial \mathcal{C}}}{\partial \alpha}\frac{\mathcal{C}}{\bar{\omega}+\frac{b}{1-b}} - \frac{\partial \bar{\omega}}{\partial \mathcal{C}}\frac{\partial \bar{\omega}}{\partial \alpha}\frac{\mathcal{C}}{(\bar{\omega}+\frac{b}{1-b})^2}$$
$$= \frac{\mathcal{C}}{\bar{\omega}+\frac{b}{1-b}}\left(\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C}\partial \alpha} - \frac{\partial \bar{\omega}}{\partial \mathcal{C}}\frac{\partial \bar{\omega}}{\partial \alpha}\frac{1}{(\bar{\omega}+\frac{b}{1-b})}\right). \tag{A20}$$

Note that:

$$\frac{\partial \bar{\omega}}{\partial \mathcal{C}} = \frac{\left((1-\beta(1-\delta)+\beta\lambda)(\alpha-\tau\mathcal{C}) - \Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big] + \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}\right)\Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau}{\left(\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}\right)^2}$$
$$= \frac{-(1-\alpha)\tau[1-\beta(1-\delta)][1-\beta(1-\delta)+\beta\lambda]}{\left(\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}\right)^2} < 0. \tag{A21}$$

Note that:

$$\frac{\partial \bar{\omega}}{\partial \alpha} = \frac{\left(\Big[1-\beta(1-\delta))+\beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C} - (\alpha-\tau\mathcal{C})\beta\lambda\right)\Big[1-\beta(1-\delta)+\beta\lambda\Big]}{\left(\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}\right)^2}$$
$$= \frac{(1-\tau\mathcal{C})[1-\beta(1-\delta)][1-\beta(1-\delta)+\beta\lambda]}{\left(\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}\right)^2} > 0 \tag{A22}$$

Now we can consider the cross-partial:

$$\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C}\partial \alpha}$$
$$= \frac{\left(\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C} + 2\beta\lambda(1-\alpha)\right)\tau[1-\beta(1-\delta)][1-\beta(1-\delta)+\beta\lambda]}{\left(\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big] - \Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}\right)^3} > 0. \tag{A23}$$

Combining pieces, we can see that the relevant elasticity is positive, which is what we wanted to show.

□

D Proof of Proposition 2

Proof. Start with (A5)

$$(1-\beta(1-\delta))\alpha S_i = w_i - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2 + \frac{1-\beta(1-\delta)}{1-\beta(1-\lambda)}\beta(\lambda-\underline{\lambda})f_i\alpha S_i$$
$$(\alpha-\tau f_i)S_i = \frac{w_i-b}{1-\beta(1-\delta)} + \frac{1}{1-\beta(1-\delta)}\Big(-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big). \tag{A24}$$

Add $(1-\alpha)S_i = \frac{1-w_i}{1-\beta(1-\delta)}$ on both sides to get

$$(1-\tau f_i)S_i = \frac{1-b}{1-\beta(1-\delta)} + \frac{1}{1-\beta(1-\delta)}\Big(-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big).$$

Plug in for S_i using $S_i = \frac{1-w_i}{(1-\alpha)(1-\beta(1-\delta))}$ and observe that the right hand side is a constant to get that

$$\frac{1-w_i}{1-w_j} = \frac{1-\tau f_j}{1-\tau f_i}. \tag{A25}$$

□

E Proof of Proposition 3

Proof. Note that:

$$\mathcal{C}^P = \frac{\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k}{\tilde{p}^1 + \tau\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k} \times \frac{\tilde{p}^1}{\tilde{p}^1}$$
$$= \frac{\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1}{1+\tau\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1}. \tag{A26}$$

We have that:

$$\mathcal{C}^P - \mathcal{C} = \frac{\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1}{1+\tau\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1} - \frac{\sum_{k=2}^{\infty}\tau^{k-2}f^k}{1+\tau\sum_{k=2}^{\infty}\tau^{k-2}f^k}. \tag{A27}$$

Forming a common denominator, the sign of $\mathcal{C}^p - \mathcal{C}$ depends on the sign of $\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1 - \sum_{k=2}^{\infty}\tau^{k-2}f^k$. So now let us sign this component:

$$\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1 - \sum_{k=2}^{\infty}\tau^{k-2}f^k = \sum_{k=2}^{\infty}\tau^{k-2}(\tilde{p}_i/\tilde{p}^1 - f^k)$$
$$= \sum_i\sum_{k=2}^{\infty}\tau^{k-2}f_i^k(\tilde{p}_i/\tilde{p}^1 - 1)$$
$$= \frac{1}{\tau^2}\sum_i\sum_{k=1}^{\infty}\tau^k f_i^k(\tilde{p}_i/\tilde{p}^1 - 1) - \frac{1}{\tau^2}\sum_i\tau f_i(\tilde{p}_i/\tilde{p}^1 - 1)$$
$$= \frac{1}{\tau^2}\sum_i(\tilde{p}_i/\tilde{p}^1 - 1)\frac{\tau f_i}{1-\tau f_i} - \frac{1}{\tau}(\tilde{p}^1/\tilde{p}^1 - 1). \tag{A28}$$

Note that ${}^{\tilde{p}^1}\!/\!_{\tilde{p}^1} - 1 = 0$. So we have:

$$\begin{aligned}\sum_{k=2}^{\infty} \tau^{k-2} {}^{\tilde{p}^k}\!/\!_{\tilde{p}^1} - \sum_{k=2}^{\infty} \tau^{k-2} f^k &= \frac{1}{\tau^2} \sum_i ({}^{\tilde{p}_i}\!/\!_{\tilde{p}^1} - 1) \frac{\tau f_i}{1 - \tau f_i} \\ &= \frac{1}{\tau} \sum_i \frac{f_i ({}^{\tilde{p}_i}\!/\!_{\tilde{p}^1} - 1)}{1 - \tau f_i}. \end{aligned} \tag{A29}$$

Since $\sum_i f_i {}^{\tilde{p}_i}\!/\!_{\tilde{p}^1} = 1$, the numerator is the *weighted* empirical covariance between f_i and ${}^{\tilde{p}_i}\!/\!_{\tilde{p}^1}$ (note that $\sum_i f_i({}^{\tilde{p}_i}\!/\!_{\tilde{p}^1} - 1) = \sum_i (f_i - \bar{f})({}^{\tilde{p}_i}\!/\!_{\tilde{p}^1} - 1)$), where the weights are $\frac{1}{1-\tau f_i}$, so we place more weight on the larger firms. □

F Proof of Proposition 4

Proof. Now we have:

$$w_i = p_i - (1 - \beta(1-\delta))(1-\alpha)S_i. \tag{A30}$$

We proceed in exactly the same fashion as in the proof of proposition 1. The proof is unaltered up to equation (A5).

$$(1 - \beta(1-\delta))S_i = p_i - b - \beta\lambda\alpha S^1 + \beta\lambda \frac{\beta(\lambda - \underline{\lambda})\alpha}{(1 - \beta(1-\lambda))} S^2 + \frac{1 - \beta(1-\delta)}{1 - \beta(1-\lambda)} \beta(\lambda - \underline{\lambda}) f_i \alpha S_i. \tag{A31}$$

Thus, proceeding identically to the proof of proposition 1, the counterpart to equation (A8) is

$$(1 - \beta(1-\delta))S^k = \tilde{p}^k + f^k \left[-\beta\lambda\alpha S^1 + \beta\lambda \frac{\beta(\lambda - \underline{\lambda})\alpha}{(1 - \beta(1-\lambda))} S^2 \right] + \frac{1 - \beta(1-\delta)}{1 - \beta(1-\lambda)} \beta(\lambda - \underline{\lambda}) \alpha S^{k+1} \tag{A32}$$

where $\tilde{p}^k \equiv \sum_i f_i^k (p_i - b)$ is the employment weighted average (net) productivity.

Evaluate (A32) at $k = 1, 2, 3, ...$ to get

$$\begin{aligned}(1 - \beta(1-\delta))S^1 &= \tilde{p}^1 + f^1 \left[-\beta\lambda\alpha S^1 + \beta\lambda \frac{\beta(\lambda - \underline{\lambda})\alpha}{(1 - \beta(1-\lambda))} S^2 \right] + \frac{1 - \beta(1-\delta)}{1 - \beta(1-\lambda)} \beta(\lambda - \underline{\lambda}) \alpha S^2 \\ (1 - \beta(1-\delta))S^2 &= \tilde{p}^2 + f^2 \left[-\beta\lambda\alpha S^1 + \beta\lambda \frac{\beta(\lambda - \underline{\lambda})\alpha}{(1 - \beta(1-\lambda))} S^2 \right] + \frac{1 - \beta(1-\delta)}{1 - \beta(1-\lambda)} \beta(\lambda - \underline{\lambda}) \alpha S^3 \\ (1 - \beta(1-\delta))S^3 &= \tilde{p}^3 + f^3 \left[-\beta\lambda\alpha S^1 + \beta\lambda \frac{\beta(\lambda - \underline{\lambda})\alpha}{(1 - \beta(1-\lambda))} S^2 \right] + \frac{1 - \beta(1-\delta)}{1 - \beta(1-\lambda)} \beta(\lambda - \underline{\lambda}) \alpha S^4. \end{aligned}$$

Importantly, for $k = 1$, we can also write

$$(1 - \beta(1-\delta))S^1 = \tilde{p}^1 - \beta\lambda\alpha S^1 + \beta(\lambda - \underline{\lambda})\alpha \frac{1 - \beta + \beta(\delta + \lambda)}{(1 - \beta(1-\lambda))} S^2. \tag{A33}$$

Now start the substitution

$$(1-\beta(1-\delta))S^1 = \tilde{p}^1 + f^1\Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big] + \tau\Big(\tilde{p}^2 + f^2\Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big] + (1-\beta(1-\delta))\tau S^3\Big). \tag{A34}$$

If we keep substituting, then we get:

$$(1-\beta(1-\delta))S^1 = \Big(\tilde{p}^1 + \tau\tilde{p}^2 + \tau^2\tilde{p}^3 + ...\Big) + \Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big]\Big(f^1 + \tau f^2 + \tau^2 f^3 + ...\Big). \tag{A35}$$

Proceeding identically for S^2 gives

$$(1-\beta(1-\delta))S^2 = \Big(\tilde{p}^2 + \tau\tilde{p}^3 + \tau^2\tilde{p}^4 + ...\Big) + \Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big]\Big(f^2 + \tau f^3 + \tau^2 f^4 + ...\Big). \tag{A36}$$

Define

$$P \equiv \Big(\tilde{p}^2 + \tau\tilde{p}^3 + \tau^2\tilde{p}^4 + ...) = \sum_{k=2}^{\infty} \tau^{k-2}\tilde{p}^k \tag{A37}$$

and let, as previously, $F \equiv \sum_{k=2}^{\infty} \tau^{k-2} f^k$. This gives, directly from equations (A35) and (A36)

$$S^2 = S^1\frac{F}{1+\tau F} - \frac{1}{1-\beta(1-\delta)}\left[\left(\tilde{p}^1 + \tau P\right)\frac{F}{1+\tau F} - P\right] = S^1\mathcal{C} - \frac{1}{1-\beta(1-\delta)}\left[\left(\tilde{p}^1 + \tau P\right)\mathcal{C} - P\right]. \tag{A38}$$

Note that:

$$\begin{aligned}
\left(\tilde{p}^1 + \tau P\right)\mathcal{C} - P &= \left(\tilde{p}^1 + \tau P\right)\mathcal{C} - P\frac{\tilde{p}^1 + \tau P}{\tilde{p}^1 + \tau P} \\
&= \left(\tilde{p}^1 + \tau P\right)\left(\mathcal{C} - \mathcal{C}^P\right) \\
&= \tilde{p}^1\left(1 + \frac{\tau P}{\tilde{p}^1}\right)\left(\mathcal{C} - \mathcal{C}^P\right) \\
&= -\tilde{p}^1\mathcal{P}.
\end{aligned}$$

Plug this into equation (A33) to get

$$S^1 = \frac{\tilde{p}^1\left(1 + \tau\frac{1-\beta+\beta(\delta+\lambda)}{1-\beta(1-\delta)}\mathcal{P}\right)}{1-\beta+\beta(\delta+\lambda\alpha) - \tau\mathcal{C}(1-\beta+\beta(\delta+\lambda))}. \tag{A39}$$

Integrate across equation (A30) to get

$$\frac{\bar{w}-b}{\tilde{p}^1} = 1 - (1-\beta(1-\delta))(1-\alpha)\frac{S^1}{\tilde{p}^1}$$

and thus, plugging in (A39),

$$\frac{\bar{w}-b}{\tilde{p}^1} = \frac{(1-\beta(1-\delta)+\beta\lambda)(\alpha-\tau\mathcal{C})}{\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big]-\Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}} - \frac{\tau(1-\beta+\beta(\delta+\lambda))(1-\alpha)\mathcal{P}}{\Big[(1-\beta(1-\delta))+\beta\lambda\alpha\Big]-\Big[1-\beta(1-\delta)+\beta\lambda\Big]\tau\mathcal{C}}.$$

□

G Proof of Proposition 5

Proof. Equation (A5) also holds in the extension with heterogeneous productivity:

$$\begin{aligned}(1-\beta(1-\delta))\alpha S_i &= w_i - b - \beta\lambda\alpha S^1 + \beta\lambda\frac{\beta(\lambda-\underline{\lambda})\alpha}{(1-\beta(1-\lambda))}S^2 + \frac{1-\beta(1-\delta)}{1-\beta(1-\lambda)}\beta(\lambda-\underline{\lambda})f_i\alpha S_i \\ (\alpha-\tau f_i)S_i &= \frac{w_i-b}{1-\beta(1-\delta)} + \frac{1}{1-\beta(1-\delta)}\Big(-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big). \qquad \text{(A40)}\end{aligned}$$

Add $(1-\alpha)S_i = \frac{p_i-w_i}{1-\beta(1-\delta)}$ on both sides to get

$$(1-\tau f_i)S_i = \frac{p_i-b}{1-\beta(1-\delta)} + \frac{1}{1-\beta(1-\delta)}\Big(-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big).$$

Plug in for S_i once more to get

$$(1-\tau f_i)(p_i-w_i) = (1-\alpha)(p_i-b) + (1-\alpha)\Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big]. \qquad \text{(A41)}$$

To characterize the term in squared brackets use, from equation (A33),

$$(1-\beta(1-\delta)+\beta\lambda\alpha)S^1 = \tilde{p}^1 + \tau(1-\beta+\beta(\delta+\lambda))S^2.$$

Rewrite as

$$\beta\lambda\Big[\tau S^2 - \frac{1-\beta((1-\delta)+\beta\lambda\alpha}{1-\beta+\beta(\delta+\lambda)}S^1\Big] = -\tilde{p}^1\frac{\beta\lambda}{1-\beta+\beta(\delta+\lambda)}$$

and so

$$\begin{aligned}\beta\lambda\Big[\tau S^2 - \alpha S^1 - \frac{(1-\alpha)[1-\beta(1-\delta)]}{1-\beta+\beta(\delta+\lambda)}S^1\Big] &= -\tilde{p}^1\frac{\beta\lambda}{1-\beta+\beta(\delta+\lambda)} \\ \beta\lambda\Big[\tau S^2 - \alpha S^1\Big] &= -\tilde{p}^1\frac{\beta\lambda}{1-\beta+\beta(\delta+\lambda)} + \beta\lambda\frac{(1-\alpha)[1-\beta(1-\delta)]}{1-\beta+\beta(\delta+\lambda)}S^1 \\ \beta\lambda\Big[\tau S^2 - \alpha S^1\Big] &= -\frac{\beta\lambda}{1-\beta+\beta(\delta+\lambda)}\left[\tilde{p}^1 - S^1(1-\beta(1-\delta))(1-\alpha)\right].\end{aligned}$$

Use this to replace the term in squared brackets in (A41) and plug in for $S^1 = (p^1-\bar{w})\frac{1}{(1-\beta(1-\delta))(1-\alpha)}$

to get that

$$(1-\tau f_i)(p_i - w_i) = (1-\alpha)(p_i - b) - \frac{\beta\lambda(1-\alpha)}{1-\beta+\beta(\delta+\lambda)}\left[\tilde{p}^1 - (p^1 - \bar{w})\right]$$
$$(1-\tau f_i)(p_i - w_i) = (1-\alpha)(p_i - b) - \frac{\beta\lambda(1-\alpha)}{1-\beta+\beta(\delta+\lambda)}(\bar{w} - b),$$

which completes the proof.

□

H Proof of Proposition 6

Proof. $\mathbf{P}$ satisfies

$$\mathbf{P} = [\mathbf{A} - c_2\mathbf{1}\mathbf{F}']^{-1}[\mathbf{I} - \tau\mathbf{D}_{\mathbf{F}} - c_1\mathbf{1}\mathbf{F}']\mathbf{W}.$$

The Sherman-Morrison lemma states that a matrix of the form $\mathbf{A} - c_2\mathbf{1}\mathbf{F}'$ is invertible if and only if $1 - c_2\mathbf{F}'\mathbf{A}^{-1}\mathbf{1} \neq 0$. To verify this, note that

$$1 - c_2\mathbf{F}'\mathbf{A}^{-1}\mathbf{1} = 1 - c_2\sum_i \frac{f_i}{\alpha - \tau f_i} \tag{A42}$$

$$= 1 + \tilde{b}(1-\alpha)\frac{r+\delta}{r+\delta+\lambda}\frac{1}{\alpha}\sum_i \frac{f_i}{1 - \frac{\lambda}{\lambda+r}f_i} > 1. \tag{A43}$$

The first line uses that $\mathbf{A}$ is a diagonal matrix whose i^{th} diagonal element is given by $\alpha - \tau f_i$. The second line uses the definitions of c_2 and τ, and that $f_i \leq 1 \forall i$. Thus, a solution to the inverse exists. The Sherman-Morrison lemma states that it is given by

$$[\mathbf{A} - c_2\mathbf{1}\mathbf{F}']^{-1} = \left[\mathbf{A}^{-1} + c_2\frac{\mathbf{A}^{-1}\mathbf{1}\mathbf{F}'\mathbf{A}^{-1}}{1 - c_2\mathbf{F}'\mathbf{A}^{-1}\mathbf{1}}\right]$$

which completes the proof.

□

I Flow-based Labor Markets

We assume that each firm $i = 1, \ldots, N$ in the economy is associated with one out of a finite number K of labor markets. An $N \times 1$ vector z denotes the assignment of firms to markets with $z_i \in \{1, \ldots, K\}$. Rather than being determined by regional or industry classifications, however, we assume that the boundaries of labor markets are unobserved. We assume that worker flows between firms are driven by the latent markets and hence are able to recover the unobserved classification from observed worker flows. In particular, a $K \times K$ matrix M summarizes transition probabilities between labor markets where the typical element M_{uv} indicates how likely a firm in market u experiences a job-to-job transition of one of its workers to a firm in market v.

The dependence of worker flows between firms i and j on market assignments is then expressed as

$$E[A_{ij}] = M_{z_i z_j}\gamma_j^+\gamma_i^-, \tag{A44}$$

where the number of worker transitions from i to j , A_{ij}, depends on the market assignments of firms i and j, z_i and z_j, the transition probability between these markets, and firm-level parameters

γ_j^+ and γ_i^- that capture the individual propensity of firm j to attract workers from other firms and the individual propensity of firm i to release workers. Based on the observed $N \times N$ matrix A of worker transitions between firms, we estimate the parameters of equation A44 by a computational approximation to maximum likelihood. An important tuning parameter is the number of markets to consider, K. In order to assure comparability to industry or geography groups, we choose various values of K that are equal to the number of NUTS-3 regions in Austria, to 1-digit and 4-digit industries, and to the interaction of industries and regions.

The model imposes no restrictions on the structure of the transition matrix M. The resulting labor markets, however, are much more self-contained than standard labor markets defined by geographical regions or industry classifications.