Granular Search, Market Structure, and Wages

Gregor Jarosch Jan Sebastian Nimczik Isaac Sorkin*

August 2019

Abstract

We build a model where firm size is a source of labor market power. The key mechanism is that a granular employer can eliminate its own vacancies from a worker's outside option in the wage bargain. Hence, a granular employer does not compete with itself. We show how wages depend on employment concentration and then use the model to quantify the effects of granular market power. In Austrian micro-data, we find that granular market power depresses wages by about ten percent and can explain 40 percent of the observed decline in the labor share from 1997 to 2015. Mergers decrease competition for workers and reduce wages even at non-merging firms.

*Jarosch: Princeton and NBER, gregorjarosch@gmail.com. Nimczik: ESMT Berlin and IZA, jan.nimczik@esmt.org. Sorkin: Stanford and NBER, sorkin@stanford.edu. Thanks to seminar and conference participants at Barcelona GSE: EGF, Berkeley, Calgary, EES Stockholm, Minnesota Workshop in Macroeconomic Theory, NBER SI: Macro Perspectives, NBER SI: Labor Studies, New York Fed, Princeton, STLAR, SED, Stanford, and the West Coast Search and Matching Workshop for helpful comments and conversations. All errors are our own, please let us know about them.

There has been a revival of interest in understanding the effect of market power on many aggregate outcomes, including wages. In this paper, we develop a new model of size-based labor market power. 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. In our setup, the vacancies of a granular employer—a firm controlling a strictly positive fraction of employment—do not compete with the vacancies of the same employer. Specifically, employers exert market power by effectively eliminating their own vacancies from a worker's outside option in the wage bargain. As a consequence, the distribution of employment shares affects wages and we derive a structural mapping from a microfounded concentration index to average wages.

We use our framework to gauge the consequences of levels and trends in labor market concentration for wages (or equivalently, the labor share) in the Austrian labor market from 1997 to 2015, and to study the effects of hypothetical mergers. We have three main findings. First, market power in the labor market depresses wages by about ten percent. Second, changes in market structure can explain over 40 percent of the observed decline in the labor share in Austria from 1997-2015. Finally, merging the two largest employers in each labor market reduces competition for workers and, as a consequence, wages fall even at non-merging firms: in our simulations, market-wide wages decline by on average six percent.

The first key ingredient in our model is that employers each control a strictly positive fraction of vacancies—i.e., they are *granular*. In a frictional market, competition is encoded in workers' outside options. Because a granular employer controls a positive fraction of vacancies, one part of the workers' outside option is the employer it is currently matched with. Thus, in a standard setup, a granular employer would compete with its own future vacancies.

The second key ingredient is that firms can (largely) avoid the competition with their own vacancies. Wages are set through standard Nash-bargaining. But we adopt a matching process where unemployed workers apply to jobs subject to coordination frictions. As a consequence, vacancies frequently have multiple applicants they can choose from. We assume that if the firm and worker fail to reach an agreement and the worker applies to another vacancy at the same firm, then the firm selects another worker from the queue of applicants. Hence, the firm does not compete with its own vacancies.

Our model implies that large firms pay less and wages are lower in more concentrated markets. The intuition for the wage result is that workers' outside options are worse when bargaining with a larger firm. The intuition for the concentration result is that firms in more concentrated markets compete less for workers and thus pay lower wages.

We derive a closed form expression for average wages which shows that market structure is summarized by a particular concentration measure. The measure depends on the sum of squared employment shares (as in the Herfindahl-Hirschman Index (HHI)) as well as on all the higher order terms. Intuitively, the source of the power terms is the possibility of repeated encounters with the same employer who does not compete with itself. The inclusion of the higher order terms means that this measure is distinct from the HHI since it places more weight on large employers. However, it shares the same limits, can be just as easily computed in the data, and empirically it is very similar.

We then extend the model to allow firms to differ in both productivity and size. We obtain the same structural mapping between concentration and wages with an additional term. The additional term measures the gap between productivity-weighted and unweighted employment concentration: for a given distribution of employment shares, competition falls if the larger employers become relatively more productive. Thus, the model allows us to disentangle the effects of pure size-driven market power from the effects of the productivity-size gradient.

Finally, we show that the pass-through of firm-level productivity to wages is decreasing in size. As a consequence, the exercise of market power generates wage dispersion.

We then use our model as an accounting device to measure the wage consequences of market power. Our empirical setting is Austria from 1997-2015. The empirical implementation faces 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 worker flows, where our model of clustering is a stochastic block model. This data-driven notion makes market definition an empirical question, rather than an *a priori* choice such as geography or industry. We view these data-driven boundaries as complementary to standard boundaries and also report some results for the latter.

We present three main empirical exercises. In the first, we make all firms atomistic which eliminates market power and hence allows us to quantify its wage consequences. Our framework thus allows us to translate concentration indices into units of interest. We find that wages would rise by about ten percent. The bulk of these gains can be attributed to pure employment concentration rather than the concentration of productivity.

Nevertheless, we show that the Austrian labor market structure is—in the wage space—far closer to perfect competition than to monopsony: the wage losses from moving the economy to the monopsonistic limit far exceed the wage gains from moving it to the atomistic limit. We also measure the wage losses due to market power in terms of search frictions: removing market power yields the same wage gains as a 40 percent increase in the job finding rate.

Concentration has strong distributional consequences. Across markets, we show that the effects are highly non-linear and largely occur in a few highly concentrated labor markets. Across workers, higher-earners are affected most by concentration.

Our second exercise quantifies the consequences of changing market power over time. We find that changes in concentration have contributed to the observed decline in the Austrian labor share: over the entire sample-period, movements in concentration reduced the Austrian labor share by over one percentage point, which is about forty percent of the observed change. About half of this effect comes from the increasing concentration of productivity over time.

Our third main exercise evaluates the labor market consequences of mergers (Naidu, Posner, and Weyl (2018) and Marinescu and Hovenkamp (2019)). To do so, we simulate the merger of the two largest employers in each labor market and recompute wages at all employers. On average, wages at merging firms decline by eight percent. Crucially, the mergers have large spillovers to all other employers who, recognizing the reduction in competition, reduce their wages by about three percent, leading to market-wide wage reductions on average of about six percent. We again highlight the non-linearity of the effects. Indeed, in our model mergers have particularly large effects in markets that are already highly concentrated.

We conclude by running wage-concentration regressions in data simulated from the model. We find elasticities of wages with respect to concentration that are quite similar to those reported in the literature (e.g., Azar, Marinescu, and Steinbaum (2017) and Rinz (2018)).

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 (e.g., Berger, Herkenhoff, and Mongey (2019), Lamadon, Mogstad, and Setzler (2019), MacKenzie (2018) and Haanwinckel (2018)) build static models of the labor market where workers' labor supply resembles consumer demand in the traditional monopolistic competition setup. These papers thus provide an equilibrium microfoundation for the Robinson-style monopsony wage markdown. The most closely related paper to ours in this vein is Berger, Herkenhoff, and Mongey (2019), which also delivers a microfoundation for a structural relationship between a measure of concentration and wages, and uses the model to assess the impact of changes in concentration on wages in their frictionless setting (among other things). In our frictional labor market, market power arises from size, rather than from the ability to exploit an upward-sloping labor supply curve. Importantly, ours is a model of pure rent extraction where quantities are not directly affected by market structure. In contrast, in the differentiated firms literature, wages and employment are closely tied together. When employers have more market power they reduce both wages and employment.

Our paper is conceptually related to Stole and Zwiebel (1996). In that paper, firms manipulate size to affect workers' wages, and so it is similar to our paper in finding a connection between size and wages. The key difference is that in Stole and Zwiebel (1996) firms manipulate size to affect workers' inside option (the marginal product of the match), whereas in our model, size affects workers' outside options.

Our paper is also related to models of imperfect competition in the posting tradition of Burdett and Mortensen (1998), such as Manning (2003) and Gouin-Bonenfant (2018).[1] The key difference is that these models have a continuum of firms and so do not share the notion of size-driven market power studied in this paper.

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

[1] See also Webber (2015) and Webber (2018).

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. Similarly, Zhu (2012) studies an over-the-counter market where when a seller recontacts a buyer the buyer updates negatively about the quality of the seller's good; this adverse-selection-like channel is not the operative mechanism in our model.

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, and analyzes the implications for wages and pass-through. 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 aggregate trends in labor share, describes the data-driven labor markets, and shows trends in various measures of concentration in local labor markets. Section 5 presents our quantitative results about the role of levels and trends in market structure in explaining levels and trends in wages. Section 6 presents our merger simulations. In Section 7 we compute elasticities of wages with respect to HHI from regressions in simulated data. Section 8 concludes.

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 idea: granular employers exert market power by not competing with themselves.

We characterize the relevant concentration index capturing market structure and the mapping to 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 $0 < \beta < 1$.

A worker 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 $b < 1$.

Firms are granular and control a positive measure of vacancies. Because the model does not feature on-the-job search, the vacancy share also corresponds to the employment share. There are N distinct firms. The probability that a particular job opening is at firm i is given by time-invariant f_i and so $\sum_{i=1}^{N} f_i = 1$. In a slight abuse of language, we often refer to the firm's market share, f_i, as the firm's size.

Matching: For each job opening, firm i 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. Each period, u unemployed workers send one application (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 in a period is exponentially distributed.

If a firm receives multiple applications it follows up on a 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)).

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. 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.

Finally, we assume that firms with multiple job openings treat them in isolation from each other. As a consequence, they cannot consolidate the applications across vacancies, which would give large employers even more market power.

Worker value functions: We let U denote the value of unemployment while W_i denotes the value of a worker employed at firm i. Formally, U satisfies

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

Next period the worker receives an offer with probability λ. This offer is from firm i with probability f_i in which case the worker receives value W_i. If the worker does not receive an offer, then she remains unemployed.

In commonly adopted models of wage setting in frictional labor markets, a key determinant of wages is a worker's outside option, namely the value of unemployment. In markets with intense demand side competition workers find other jobs rapidly which is encoded in the outside option and raises the wage. Suppose employer i and a potential hire were using equation (1) to determine a worker's outside option. Then granular firms would compete with themselves: when bargaining with a particular firm, the worker would effectively claim the same firm's future vacancies as an outside option.

Our key departure is that a firm can remove itself from a worker's outside option, thus preventing competition with itself. To do so, suppose the firm and the worker fail to find an agreement and the worker applies to a job opening controlled by the same employer in the future. In the event that the vacancy received multiple applications, the firm can break the tie by hiring one of the other applicants. This tie-breaking rule allows the employer to (partially) remove its own job openings from the outside option of the worker in the wage bargain. Importantly, this strategy 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. If a deviating worker happens to be the sole applicant to one of the firm's job openings, then the firm rationally hires the worker. It is worth noting that this mechanism operates through off-equilibrium payoffs and the parties never fail to reach agreement.

To make the analysis tractable, we make a particular assumption on the duration of this disagreement "punishment". 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 restriction substantially reduces the state space since it cuts the histories the agents have to keep track of.

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. More broadly, it captures the idea that jobs are imperfect substitutes in the search process.

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

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

This equation states that, after disagreement with employer i, a worker's probability of meeting and subsequently working for any other employer j are unaltered. Moreover, the value of working for employer j does not depend on the worker's history. However, if the worker applies to a vacancy controlled by i, then 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 the outside option when bargaining is worse.

We note that equation (2) does not require commitment power for the firm since it is costless for the firm to select another applicant. It only imposes that the firm "recognizes" a worker under punishment.

Let w_i denote the wage firm i pays under the Nash bargaining solution. The value of working for firm i then satisfies

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

This equation says that the value of being employed at firm i is the wage at firm i plus a continuation payoff, which weights the probability of the job being exogenously destroyed and entering unemployment or remaining employed. Importantly, following an exogenous breakdown of an employment spell a worker is free to return to another vacancy posted by the same employer. Thus, the outside option when bargaining and the value of unemployment following a job spell differ.

Firm value functions: Firm i values the bilateral relationship with each of its workers at J_i satisfying

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

This equation says that the value to firm i of filling the vacancy is the flow output of the match less the wage, and, in the event that the job is not exogenously destroyed, the job continues. Note that this equation reflects the assumption discussed below that the job has no continuation value after an exogenous separation (i.e., $V_i = 0$). In turn, we have that a job opening has value

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

To keep a vacancy open, firm i pays fixed cost c_i. The term in parentheses captures the probability that the job opening receives at least one application this period. 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. Instead, we simply read the f_i off the data and study its consequences for wages. 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 $V_i = 0\,\forall i$. We 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.

Surplus and wage determination: The joint net value of forming a match ("surplus") is given by

$$S_i \equiv W_i - U_i + J_i. \tag{6}$$

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 state U_i while the firm has no continuation value.

We adopt the axiomatic Nash bargaining solution to the bargaining problem. In this case, the wage implements a surplus split such that the net value of forming the match to the worker is

$$\alpha S_i = W_i - U_i, \tag{7}$$

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

$$(1-\alpha)S_i = J_i. \tag{8}$$

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

Summary: Our model combines three conceptual ideas so that concentration of employment affects wages. First, the bargaining implies that competition for workers affects wages. Second, for workers, random search implies that jobs are imperfect substitutes: any job in the offer distribution could be the worker's most preferred job in the future, and so the firm's ability to remove that job from the outside option is a relevant threat. Third, the urn-ball matching function implies that in most situations firms view workers as perfect substitutes. This substitutability introduces the important asymmetry in the model: workers value the possibility of a future encounter the firm, but the firm does not value the possibility of a future encounter with the worker (and, hence, it is costless to the firm to rule out a future match).

The key distinction between our granular search framework and the standard setup with atomistic firms is that with granular firms workers value the possibility of the future relationship with the firm because there is a positive probability of a future re-encounter. Thus, granular employers have the power to shape workers' outside options. We note that both granular and atomistic employers do not value the future relationship with a given worker and so do not compete with themselves. With atomistic employers, however, the mechanism is different: repeated encounters are a zero measure event.

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 and model 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, 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 $\tau \approx \alpha \frac{\beta\lambda}{1-\beta(1-\lambda)}$. This form of τ enables us to obtain simpler 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. This approximation is highly accurate because with a monthly job finding rate of less than 10 percent the implied probability of being the only applicant is remote (for $\lambda = 0.093$ we obtain $\underline{\lambda} = 0.00002$), 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.

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 the relevant concentration index in our environment.

Definition 1. *Define concentration 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 different from—yet very closely related to—the standard HHI. It is closely related to the HHI in that the first element of the sum terms is simply f^2, the HHI. The reason the HHI shows up is the following: it captures the unconditional probability of contacting the same firm twice in a row. And, in our framework, the possibility of these repeated encounters affects wages because vacancies of the same firm do not compete with each other.

Our index also shares the same bounds as the HHI: in the limit with atomistic employers, we have that $\mathcal{C} = 0$, just like the HHI. In the limit of a single monopsonistic employer, we have that $\mathcal{C} = 1$, just like the HHI.[2]

What differs between our index and the HHI is the inclusion of the additional higher order terms, (down)-weighted by τ. The reason for these additional terms is that a worker might, in principle, apply to job openings by the same employer not just twice, but multiple times in a row. These events get mediated by τ and are less relevant if job seekers infrequently contact job openings.

It also differs in that our index places more weight on the size of the largest firms in the labor market than the HHI.[3] 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 different, empirically we find that the HHI and $\mathcal{C}$ are very similar in the Austrian data.

1.3 Concentration, Average Surplus, and Wages

We now derive a structural mapping from $\mathcal{C}$ to wages. Define $\omega_i \equiv \frac{w_i - b}{1-b}$ to be the worker share of flow surplus at firm i (recall that all firms produce flow output of 1), which we refer to as *compensation*. Let $\bar{\omega} \equiv \sum_i f_i \omega_i$ denote mean compensation, which is an affine transformation of mean wages and hence shares its comparative statics. Our first result is the following:

[2]To see these bounds, 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 monopsonist.

[3]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.

Proposition 1. *The equilibrium relationship between (employment-weighted) mean compensation and concentration is:*

$$1 - \bar{\omega} = (1 - \alpha) \frac{1 - \beta (1 - \delta)}{1 - \beta \Big(1 - \lambda \alpha \underbrace{[1 - \mathcal{C}]}_{\textit{wedge 1}} - \delta \underbrace{[1 - \tau \mathcal{C}]}_{\textit{wedge 2}} \Big)}.$$

Proof. See Appendix B.1. □

The denominator in this expression shows that granular market power introduces two wedges into the wage equation, which reflect the two mechanisms by which increases in concentration decrease wages. In a static setting, the worker would receive a share α of net output, and so $1 - \bar{\omega} = 1 - \alpha$. In a dynamic setting, the worker's share is increased through competition for workers: the parties recognize that the worker has other options, which is the $\lambda\alpha$ term.

The reason for the first wedge is that concentration reduces competition: granular employers do not compete with themselves. So a worker's outside option—which encodes competition—is reduced relative to the atomistic benchmark. Hence, as concentration increases, mean wages fall because workers have deflated outside options.

The reason for the second wedge is that part of the value to the worker of forming the match comes once the match ends, rather than in wages (i.e., concentration inflates the inside option). By reaching an agreement, the pair increases the worker's continuation value in unemployment from U_i to U: forming a match increases the inside option because a worker then has the possibility of returning right away to the firm. This mechanism decreases wages because workers take more of the value of the match in the non-wage continuation value, rather than in flow wages. The strength of this mechanism is decreasing in τ: if unemployment spells are long because λ is low, then the worker's return to the firm is further in the future and so is a less important consideration in wage-setting.

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

We conclude with an important corollary to Proposition 1:

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

Proof. Follows from the definition of ω and differentiation. □

This result provides a theoretical foundation for a negative relationship between concentration—as measured by $\mathcal{C}$— and average wages. Furthermore, the strict concavity is a cautionary note on aggregation: the literature on trends in aggregate concentration often aggregates local concentration measures in a weighted linear fashion (e.g., Rinz (2018)). But if the mapping between concentration and the outcome of interest is non-linear at the local level, then the aggregated trends may be misleading. We later show that there are periods where concentration measured as a simple weighted linear aggregate index fell, but using the model we find that concentration changes depressed wages. The non-linearity in the model is the culprit.

1.4 Concentration and Firm-Level 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 between firm-level wages w_i, 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 wages are fully characterized by*

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

Proof. See Appendix B.2. □

Proposition 2 implies that wages are monotonically decreasing in employer size f_i. That is, the firm with more market power pays a lower wage. It also shows that relative wages are independent of the level of concentration.

The combination of Proposition 1 and 2 implies that individual wages are monotonically decreasing in $\mathcal{C}$ at all employers. That is, changes in concentration affect wages at all firms with unchanged market share proportionally.

The proposition also reveals that the profit-size gradient steepens as τ increases (recall that $1 - w_i$ is the flow profit of firm i). The reason is that τ measures the importance of the mechanism: being able to evade competition is particularly powerful if workers find jobs rapidly and competition for them is intense. Furthermore, we highlight that the returns to size as measured by relative flow profits is independent of the distribution of employment shares across other firms in the market (i.e., market structure).

Proposition 2 emphasizes that market power affects wages 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 fundamentally derives from variation in the elasticity of labor supply to the firm (here, in matches with positive surplus, 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. To generate an imperfect relationship between size and wages, in this section we add productivity heterogeneity to the model.

This extension allows us to separate the two ways employment shares affect labor market outcomes: first, through the pure size distribution already studied in the previous section. And, second, through how size and productivity are correlated. Our setup yields a clean decomposition between the two, and hence lets us separately quantify the consequences of each dimension. Underlying this decomposition is the result that the model generates size-dependent pass-through of productivity with higher pass-through at smaller firms. Thus, holding aggregate productivity constant, worker wages are lower when productivity is in larger firms. Hence, if the firms with large market shares are more productive firms, then market structure has larger effects on wages than pure employment concentration 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. The definition of $\mathcal{C}$ is unchanged. We note that, with heterogeneous productivity, not all matches may have positive surplus. Our exposition imposes, however, that all matches are formed.[4]

The following is the productivity counterpart of $\mathcal{C}$, namely a productivity-weighted concentration index:

Definition 2. *Define productivity-weighted concentration as*

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

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

Definition 3. *Define 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 two 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. In particular, we show in Appendix B.3 that the sign of $\mathcal{P}$ is the same sign as $\sum_i \frac{f_i(\tilde{p}_i - 1)}{1 - \tau f_i}$, 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.

The object $\mathcal{P}$ effectively measures to what extent productivity is correlated with size. If size and productivity are positively correlated, then effective concentration is higher than implied through a simple measure of employment concentration. Put differently, market structure can depress wages

[4]This will endogenously be the case in our baseline calibration but not in one of the robustness exercises. We hence revisit this assumption in the robustness section 5.2.

either because employment grows more concentrated or because productivity and size become more correlated (the latter case is the "superstar" firms effect of Autor et al. (2019)). $\mathcal{P}$ separates these forces.

We now relate concentration to wages in this richer environment. Denote by $\bar{\omega}^*$ average worker compensation in the homogeneous firms benchmark presented in Proposition 1. Similar to before let $\bar{\omega} \equiv \frac{\bar{w}-b}{p^1-b} = \frac{\bar{w}-b}{\tilde{p}^1}$ be the fraction of the average net flow output that goes to workers. Let $\hat{\tau} \equiv \tau \left(1 + \frac{\beta\lambda}{(1-\beta(1-\delta))}\right) > 0$. Our key result is summarized in the following proposition:

Proposition 3. *The equilibrium relationship between compensation and concentration satisfies:*

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

Proof. See Appendix B.4. □

Proposition 3 naturally extends the results in Proposition 1 to the heterogeneous firms case. It shows that average compensation is given by exactly the same expression as in the baseline case up to an additional wedge $\hat{\tau}\mathcal{P}$. This wedge is positive if productivity is more concentrated than employment. It reflects the fact that workers' outside options deteriorate if, given a distribution of employment, productivity shares become more concentrated. The reason is that pass-through is lower at larger firms as we discuss further below.

We also have that:

Corollary 2. *Average wages are monotonically decreasing and strictly concave in concentration $\mathcal{C}$ and $\mathcal{P}$.*

Proof. Follows from differentiating. □

This result extends Corollary 1 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 increases in productivity concentration as measured by $\mathcal{P}$ also depress wages.

2.2 Concentration, Pass-Through, and Firm-Level 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 parameters. It is linearly decreasing in the mean wage and, as such, an affine transformation of $1 - \bar{\omega}$ as defined in Proposition 3. As a consequence, it is simply another way of summarizing market power that is useful in the following result:

Proposition 4. *Firm level wages w_i satisfy*

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

Proof. See Appendix B.5. □

To interpret this result, note that $p_i - w_i = (1-\alpha)(p_i - b)$ is the solution to the static Nash bargain. Suppose market structure changes, leading to a decline in the mean wage $\bar{w}$ and an increase in Π. This change affects employers market-wide, even those with unchanged size. In this case, wages at all firms fall. The multiplier $(1-\tau f_i)$ is the size mark-down. It again reflects the fact that larger firms have more power to shape workers' outside options.[5]

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 productivity levels to wage levels:

Corollary 3. *The firm-level productivity pass-through coefficient* ($\frac{\partial w_i}{\partial p_i}$) *is:*

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

Proof. Follows from rearranging and differentiating the equation in Proposition 4. □

This expression shows that the model generates size-dependent pass-through of productivity to wages.[6] 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 monopsonistic 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. Hence, market level concentration matters for the level of wages, but not for relative wages across employers.

This corollary revealed a tight connection between pass-through and worker bargaining power, α. The next corollary shows the relationship between worker bargaining power and the effect of changes in concentration on wages:

Corollary 4. *The elasticity of wages with respect to concentration becomes smaller in magnitude as worker bargaining power (α) increases.*

Proof. See Appendix B.6. □

This corollary shows that, all else equal, variation in concentration matters more when worker bargaining power is low. The reason is that, when bargaining power is low, wages are primarily determined by the outside option which is precisely what concentration affects. As a consequence,

[5] We can also use proposition 4 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.

[6] One can think of this derivative as a cross-sectional wage-productivity gradient. To interpret it literally as a pass-through coefficient one needs to implicitly also change c_i to keep $V_i = 0$.

lower pass-through of productivity shocks to wages suggests a more important role of concentration for wages.

We conclude by noting that Proposition 4 implies that wages decrease when firms increase their size. One way firms can increase their size is to merge, which we discuss quantitatively below. We note that the strength of this relationship is again governed by τ: the more fluid the labor market, the stronger the relationship between size and wages.

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

We use the Austrian labor market data base (AMDB) that covers the universe of private sector employment in Austria. For 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, age). The notion of an employer in the dataset is closer to a firm than an establishment.[7]

We make the following sample restrictions. First, we restrict our sample to regular workers aged 20-60 years and exclude marginal workers, short-time workers, and apprentices. Second, we restrict the analysis to firm-years with 5 or more workers. See Appendix C for further details.

3.2 Market definition

We consider several different market definitions. 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 U.S.),[8] and within industry by region cells. The latter is the definition most commonly employed in the literature (see Lamadon, Mogstad, and Setzler (2019) and Berger, Herkenhoff, and Mongey (2019))

As we document below, a large share of worker flows cross industry and regional boundaries.[9] Pre-defined categorizations therefore do not necessarily capture the set of reasonable potential

[7] Fink et al. (2010, pg. 5) contrast the number of employer IDs in the AMDB with the number of firms in the Austrian firm register. The AMDB has more units than the firm register but the difference is small. As a consequence, the authors conclude that employers in the AMDB are mostly firms.

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

[9] 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).

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 a data-driven 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. The model assumes that worker mobility is driven by unobserved markets and backs out the assignment of each firm to an unobserved market.

To pick the number of markets, we maximize the penalized likelihood of the objective function. Our main choice for regularization is the minimum description length criterion, which penalizes the likelihood with the amount of "information" needed to describe the model (i.e., a particular functional form on the number of parameters). A Bayesian interpretation is that this approach is equivalent to maximizing the posterior probability using uniform priors over the number of markets (i.e. finding the mode of the posterior distribution of the parameters, see, e.g., (Peixoto, 2017)).[10]

This approach leads us to 369 labor markets. For comparison, our 2-digit industry by region definition implies over five times as many distinct labor markets.[11] We refer readers to Nimczik (2018) for complete details, but in Appendix D we provide a basic sketch of what we do.

We compute measures market-by-market, and then report results on an employment-weighted basis. Independent of market definition, a maintained assumption is that labor markets are isolated islands. Furthermore, we keep the boundaries of labor markets fixed throughout our sample.

3.3 Measuring variables and parameters

Here we discuss how we define and measure variables and parameters in the data. The two key variables that we extract from the matched employer-employee data are firm size and wages. We supplement this with information on the aggregate labor share. The model also depends on 6 parameters: $\{b, \lambda, \underline{\lambda}, \delta, \alpha, \beta\}$. We treat our model as a monthly model and take annual averages of the labor market parameters (λ, $\underline{\lambda}$, and δ). We discuss how we combine the variables and parameters to back out firm-level productivity p.

To the extent that the respective empirical moments can be measured at the market level and over time, we choose market-time-specific parameters. We denote a market by m and time by t. We measure year-specific values for these parameters and variables because we want to speak to

[10] To see the equivalence, let A be the observed $N \times N$ matrix of transitions between firms, $z = \{z_i\}$ be the assignment of firms to one of K markets for $i = 1, \ldots, N$, and let the $K \times K$ matrix M denote transitions between markets. The posterior probability of observing the data A given parameters is $P(z, M|A) = \frac{P(A|z,M)P(z,M)}{P(A)}$. The numerator can be expressed as $\exp(-\Sigma)$ where $\Sigma = -\ln P(A|z, M) - \ln P(z, M)$ is the description length. The first term in the description length is the negative log likelihood of the model given parameters z and M. The second term is the penalization term that measures the number of bits necessary to describe the model parameters.

[11] In a robustness check, we use modularity maximization as an alternative choice of regularization which yields a far coarser classification into 9 labor markets. The modularity score measures the share of transitions that are within market relative to the null of random mobility holding the inflow and outflow probabilities at the firm-level constant.

the evolution of concentration and their contribution to wages over time. We recognize that the model imposes a steady state with time-invariant employment and productivity shares. We believe that the resulting discrepancies are small because the model is known to have fast-moving state variables and because of the high-persistence of firm-level employment and wages.

Firm market share f_{it}: We employ the following measure of firm market share f_{it}: We count the number of regular employees in a given firm at a reference date (August 1st) each year. We then divide it by the total number of employees in the relevant market. 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 our main results when we measure f_{it} as the share of new hires and as the share of new hires from unemployment.

Wages w_{it}: Our notion of a wage is a daily wage (or, daily earnings). 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 earnings 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*. Because it is a firm specific wage and the model features homogeneous workers, we control for compositional differences across firms by using a residualized wage measure. To do the residualization, we work in logs. We regress log wages on functions of the observable characteristics in our data: a fourth degree polynomial in age, a second degree polynomial in tenure at the firm, a gender dummy, a dummy for Austrian nationality, and a set of interactions between the dummy variables and the polynomials. We then compute residualized individual wages in levels as the exponential function of the sum of residuals and the predicted value at average characteristics.

We use the median of the residualized wages in a firm as our firm-specific wage measure. This measure has the additional benefit of reducing the role of censoring of wages relative to taking mean (residualized) wages. We report sensitivity to considering alternative quantiles of the wage distribution. See Appendix C for further details.

Productivity p_{it}: We use Proposition 4 to back out firm-level productivity. Given variables that we discussed above and parameters that we discuss below, the Proposition implies one equation in one unknown per firm and so we can use the model to recover the productivities.

Aggregate labor share: We use the 2017 release of the KLEMS data (O'Mahony and Timmer (2009), see also `http://www.euklems.net/`) to measure the trend in the labor share in Austria. The labor share in this data is defined as aggregate payments to labor over aggregate value-added for all industries in Austria.

Job finding rate λ_{mt}: We measure market and time specific parameters λ_{mt} by calculating the share of workers unemployed in market m in month t who are employed in month $t+1$. We classify by "destination", that is we classify a worker as unemployed in market m if her spell ends with a job in market m. Across years, the employment-weighted average of the job finding rate drops from 16% to 10% (see Appendix Figure A1). There is also substantial heterogeneity in job finding rates across markets. In 2015, the 25th to 75th percentile ranges from 7% to 12% on a monthly basis.

Likelihood of being the only applicant $\underline{\lambda}_{mt}$: Let $\theta \equiv \frac{v}{u}$. The urn-ball matching function implies a unique value of θ_{mt} associated with a λ_{mt}, which in turn implies $\underline{\lambda}_{mt}$. That is, the rate at which workers exit unemployment dictates the probability of being the only applicant for a job. Given that workers exit unemployment at a very sluggish pace, the implied median value of $\underline{\lambda}_{mt}$ is 0.00002. There is some heterogeneity in $\underline{\lambda}_{mt}$ across markets with a large right tail so that the average is 0.0006.

Job destruction rate δ_{mt}: We measure market and time specific parameters δ_{mt} by computing the share of workers who are employed in a firm in market m in month t and are unemployed in month $t+1$. In 2015, the monthly unemployment inflow rate measured this way is 0.9%.

Using standard mass balance arguments, the steady state unemployment rate is given by $u = \frac{\delta}{\lambda+\delta}$. If we compute market-month specific steady state unemployment rates associated with the market-month specific flows we measure and then aggregate we match the increase in Austrian unemployment from about 7 to 9% (Statistik Austria) over our sample period quite precisely.

Worker bargaining power and flow value of unemployment α_t and b_{mt}: We jointly calibrate workers' bargaining power and the flow value of unemployment to match two targets: First, we target the time-varying aggregate labor share. Second, we set those parameters such that the least productive firm in a market pays the reservation wage or, equivalently, just breaks even. As a consequence, all matches have (weakly) positive surplus. This strategy gives us a time-varying, country-wide α_t, and a market-time specific b_{mt}.

Intuitively, α governs the "split of the pie" and, as such, the share of income going to workers. In turn, b_{mt} determines the "size of the pie" and, as such, whether there is non-negative surplus in all matches: if we see firms with vastly different pay co-exist in the market, then b_{mt} must be low for all matches to have non-negative surplus.

Importantly, if we calibrate so as to match the aggregate labor share, then the remaining output gets fully absorbed by the c_i. We can thus no longer interpret c_i as a vacancy creation cost in the standard sense. Instead, we view c_i as also capturing the fixed and variable cost of (pre-installed) capital that is complementary to a worker, similar to Acemoglu and Shimer (1999).

This strategy gives us an α of 0.48 in 2015, and a mean value of b of -227, where the units are euros per day (adjusted to the year 2000). This compares to a mean daily wage of 77 euros. Why is our b so negative? Intuitively, we are asking our model to match the empirical extent of residual wage dispersion. As Hornstein, Krusell, and Violante (2011) emphasize, in a benchmark search model, unemployment must be very painful for workers to rationally accept the lowest paying jobs in the economy. How plausible is our α? We use Corollary 3 to convert our α into measures of pass-through and find a mean value (across firms and markets) of 0.45. There is limited evidence on pass-through of productivity shocks to wages in levels; for example, the central estimate in Kline et al. (Forthcoming, Table 8, Panel A, Column 1b) is 0.29, but this masks substantial heterogeneity between incumbents and new hires that our model does not capture (the value for incumbents is 0.61; see Table 8, Panel A, Column 4b). Below, we consider extensive sensitivity for α and 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 provides summary statistics on the main variables and summarizes our parameter values. In terms of summary statistics, the average market has about 5000 workers and around 100 firms. The main idea of the paper is that, with 100 firms in a labor market, their granularity could have important implications for competition for workers.

4 Descriptive facts

We begin by presenting a number of descriptive facts.

4.1 Trends in the labor share and wages

Figure 1 shows that over our sample period the labor share has declined and wage growth has been slow. A similar development in the U.S. has led to speculation that changes in market structure might be the culprit. The top panel shows that the labor share has declined over our sample period by about three percentage points. This overall pattern masks a u-shape, where in the mid-2000s the labor share had declined by over five percentage points from its level at the start of the sample period. The bottom panel shows that real wages rose slowly over the sample period. The annualized real wage growth is under half a percentage point a year.

4.2 Data-driven labor markets

We now provide some descriptive understanding of the labor markets.

The data-driven labor markets frequently cross the boundaries of regions and (4-digit) industries. For each labor market, we compute the share of employment that is in the "dominant" industry or region, which is the industry or region accounting for the largest share of employment in the labor market. If a data-driven labor market lies completely within an industry or region, then this share is 1. Instead, Figure 2 shows that there are many low values of this measure, indicating that most labor markets contain a substantial fraction in employment outside the dominant industry or region. The top panel shows that the average share of the dominant region is below 0.6 (there are 35 regions compared to 369 data-driven labor markets). The bottom panel shows that labor markets are even less well-described by 4-digit industry.

We now show two senses in which data-driven labor markets are more isolated islands than labor market definitions based on industry or industry-region. The first column of Table 2 shows that relative to defining a labor market by industry or industry-region, a larger share of transitions happens within the data-driven labor markets. That said, the absolute level is fairly low: about 40% of transitions are within the data-driven labor markets. For context, when we look at 2-digit industry $\times$ region, only 25% of transitions are within these labor markets.

A second metric which adjusts for mechanical effects due to the number of markets and more clearly shows that data-driven markets better capture isolated islands than traditional definitions is the modularity score. The modularity score adjusts for the number of distinct labor markets by comparing the number of within market transitions to the random null. Thus, if we had a single labor market, all transitions would be within the labor market, but we would expect this to happen even if all moves were random. The last column of the table shows that the modularity scores for the data-driven markets far exceeds the modularity score obtained when employing standard industry-region labor market definitions.

4.3 Trends in market structure

Figure 3 shows that concentration in Austria is low, has followed a u-shaped pattern from 1997-2015, and that these patterns are not sensitive to the particular concentration measure. The figure shows four different measures of concentration: HHI, wage-bill HHI (emphasized by Berger, Herkenhoff, and Mongey (2019)), our concentration index ($\mathcal{C}$) and our productivity-weighted concentration index ($\mathcal{C}^P$). In terms, of levels, simply reading the nature of competition off of the HHI would suggest that on average the Austrian labor market is not very concentrated. The threshold for a market to be considered "moderately concentrated" according to U.S. antitrust authorities is 0.15 and all the concentration measures are always below this number.[12] While it is logically possible for our model-based concentration index to depart in important ways from the HHI, the gap is small in practice. Reflecting the positive size-wage correlation in our data (see Figure A3), the wage bill HHI is always higher than the HHI, and our productivity-weighted concentration index ($\mathcal{C}^P$) is always higher than our concentration index. All four measures show similar trends of a u-shape.

In the Appendix, we present some alternative ways of summarizing these trends. First, Appendix Figure A4 shows the same figure when we weight markets equally, rather than weighting by employment. Reflecting the fact that smaller markets tend to be more concentrated, the figure shows higher levels than in the weighted version. There is still a u-shape, though the shape of the u differs. Second, Appendix Figure A5 shows the same figure for alternative market definitions with combinations of regions and 2-, 3-, and 4-digit industries. Naturally, the level of concentration

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

is lower for market definitions that generate fewer markets. Similarly, the patterns differ across market definitions.

One tempting inference to draw from comparing the trends in concentration in Figure 3 to the trends in the labor share in Figure 1 is that changes in concentration are not a good candidate explanation for the decline in the labor share because the aggregate patterns do not align temporally. In periods when aggregate concentration was declining—and thus competition for workers and the labor share should have increased—the labor share was decreasing; when aggregate concentration was increasing the labor share was decreasing. We emphasize that because the underlying market-level relationship between concentration and wages is nonlinear (as we emphasized in Corollary 1), one cannot draw inferences from these aggregate trends about the effects of changes in concentration on wages. Thus, we next use the model to measure the consequences of concentration for wages in Austria.

5 The effect of market power on the labor share

We now use our model to quantify the impact of granularity-based market power on the Austrian labor market. Our first exercise considers the effect on wages of shifting from the existing market structure to the atomistic benchmark. Our second exercise uses the model to quantify how the observed evolution of market structure from 1997 to 2015 has affected wages in Austria.

5.1 The nonlinear relationship between the labor share and concentration

We begin by highlighting the nonlinear relationship between labor share and concentration in our model. In particular, all changes in concentration are not the same: a given change in concentration from a high initial value of concentration has a larger effect on wages than from a lower initial value. The reason to emphasize this feature of our model is that the quantitative importance of these nonlinearities shapes our quantitative results. To illustrate this nonlinearity, we consider what happens to the labor share as we move concentration from 0 to 1. We do this exercise in all labor markets to average over the parameter values. Figure 4a shows, consistent with Corollary 1, that wages are decreasing in concentration. But for a given increase in concentration, this decrease is small at low levels of concentration and becomes much more dramatic at high-levels of concentration. Put differently, as Panel B shows, the elasticity of wages to concentration grows (in magnitude) as concentration increases. This figure highlights that what will matter in our quantitative exercises is the small number of markets with concentration at very high levels: above $\mathcal{C} = 0.73$, say, which is the top 5% of the most concentrated markets.

5.2 Effects of levels of concentration: the atomistic benchmark

We use Proposition 3 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.[13] In each year, we consider the change in the labor share from sending first $\mathcal{P}$ to zero, and then sending $\mathcal{C}$ to zero. The first step isolates the role of the size-productivity correlation while the second step isolates the role of pure employment concentration.

Figure 5a shows that moving to the atomistic benchmark would increase the labor share by about nine to 13 percent, depending on the year. The Figure shows that employment concentration accounts for the bulk of this increase: the existing productivity-size relationship depresses wages by merely one to two percent while the remainder is accounted for by employment concentration. The relative magnitude of effects can be anticipated from Figure 3 which shows that $\mathcal{C}$ and $\mathcal{C}^P$ are quantitatively similar, implying a limited role for size-productivity covariance.

Are these effects big or small? We now contextualize and interpret the magnitudes in three ways. First, we note that simply reading the nature of competition off of the HHI would suggest that the Austrian labor market is not very concentrated. Nonetheless, we find that imperfect competition as measured through concentration depresses wages by more than nine percent per year. This highlights the value of our structural framework which allows us to translate measures of concentration into wages.

Second, our setup implies that the labor market is much closer to perfect competition than to a monopsonist. To see this, we compute the labor share in the monopsonistic benchmark. We do this by sending $\mathcal{C}$ to 1 and recomputing wages in each market. We then compare the increase in the labor share from moving the observed economy to one with atomistic firms with the gains from moving the monopsonistic economy to one with atomistic firms. The top row of Table 4 shows our baseline effects are only 4.6 percent of the gains from eliminating market power in a perfectly uncompetitive labor market.[14] Thus, when translated into wage space, the Austrian labor market is far closer to perfect competition than to a perfectly uncompetitive world.

Third, another way to gauge the quantities is to ask what decline in labor market frictions would deliver the same wage gains to workers. We solve for the change in the job finding rate such that the labor share would rise by as much as it does when we move to the atomistic benchmark. We find that this percentage change is about 41.6%. To put this number in perspective, Figure A1 displays the trends over time in the job finding rate and shows that from 1997 to 2015 it declined by almost 40 percent (from about 16 percent to about 10 percent).

We highlight that inference from aggregated statistics about the effects of trends in concentration on wages may be misleading. The comparison between Figure 5a and Figure 3 shows that there exist years where all aggregate concentration indices declined yet market power got worse (since eliminating it gained workers more). The reason again is the underlying non-linearity. If concentration rises in concentrated markets yet declines even more in competitive markets, then one may mistakenly diagnose a reduction in market power.

[13] We view these calculations as "accounting exercises" and not counterfactuals because our model does not provide a theory of the driving forces of levels and trends in concentration.

[14] The reason this number is even smaller than the labor share gains implied by Figure 5a is that wages in the perfectly uncompetitive world would be negative (since $b < 0$).

We conclude this exercise by showing the substantial distributional consequences of granular market power. Figure 6 shows that moving to the atomistic benchmark would increase inequality; equivalently, it is higher-earnings workers who experience the largest losses from employer market power because they tend to be in more concentrated labor markets. The Figure ranks workers with uncensored wages by their earnings. We compute the percent increase in their firm-level earnings in moving to the atomistic benchmark. We then take the euro-weighted average of the increase in earnings in each percentile. The Figure shows that the most highly paid workers benefit almost three times as much from moving to the atomistic benchmark—in terms of percent wage gains—compared with the lowest paid workers.

Heterogeneity across markets

Our results reflect employment-weighted averages over 369 distinct labor markets. Table 3 provides some sense of how concentration and our accounting results of moving to the atomistic benchmark vary across labor markets.

Panel A shows that most labor markets are not very concentrated, but there are a few labor markets that are very concentrated. Two statistics emphasize this point. First, while the employment-weighted HHI and $\mathcal{C}$ are on average 0.12, the median of these statistics are less than half the size. Similarly, the 95th percentile of concentration measures is more than an order of magnitude larger than the median.

Panel B shows that the adverse effects of concentration are also concentrated in a few labor markets. For example, while on average moving to the atomistic benchmark raises wages by 12.6%, in the median labor market this increase is only 2.2%. But at the 95th percentile of labor markets, wages would increase by 23% 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.

Robustness

Table 4 reports how our results vary as a function of market definition, parameter choices and variable definitions.

Market definition: Our baseline results use the data-driven labor market boundaries. We now consider alternative market definitions. For each alternative, we recalibrate the model. The first few rows of Panel A show that concentration has a smaller effect on wages when we define labor markets by either region or geography separately rather than our data-driven markets: the largest effect comes from 4 digit industries, where moving to the atomistic benchmark results in a 6 percent change in the labor share (as opposed to 13 percent in our baseline).

We find fairly similar, if somewhat larger, effects than our baseline when we follow conventional definitions in the literature and interact region and industry. For example, when we interact region and 2 digit industry (comparable to Lamadon, Mogstad, and Setzler (2019, pg. 11)) we find that moving to the atomistic benchmark increases wages by 14 percent. When we interact region and 3 digit industry (comparable to Berger, Herkenhoff, and Mongey (2019)) we find that moving to the atomistic benchmark increases wages by 19 percent. When we interact region and 4 digit industry (comparable to Rinz (2018)), we find that the atomistic benchmark increases wages by 25 percent.

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

Measure of firm size: Our baseline results use employment shares to measure firm size. While employment and hiring shares are the same in our model they are not in a world with on-the-job search. For that reason, we consider two alternative definitions of firm size: the firm-level share of new market-year hires, and the firm-level share of new market-year hires from unemployment. In both of these alternatives we find somewhat smaller effects than in our baseline results.

Value of unemployment and worker bargaining power: Our baseline results use two targets to pick the value of unemployment and worker bargaining power: first, we hit the economy-wide labor share, and second, the least productive active firm pays the reservation wage. This strategy gives rise to very low levels of $\bar{b}$ for reasons explained in Hornstein, Krusell, and Violante (2011). To ensure that our results are insensitive to this choice, we shrink surplus by increasing the flow value of unemployment. Specifically, we continue to match the labor share but increase all b_{mt} in a way that shrinks $\bar{w}_{mt} - b_{mt}$ proportionally. This strategy allows us to generate more conventional flow values of unemployment but comes with a downside: the larger b implies a larger fraction of employers with negative surplus.[15]

Figure 7 shows how our results move as we increase the average b from its baseline level of -227 up to 32.71 (the minimum value of wages in our data). Panel A shows that the increase in b implies that α falls to keep the labor share stable. On the right-hand side, α equals 0.13. Panel B shows that as we shrink flow surpluses we get an increasing share of firms earning negative profits (and, correspondingly, workers in nonviable jobs): on the right hand side, this share approaches 33%. Most importantly, panel C shows that our baseline result—the gains workers derive from eliminating employer market power—are remarkably stable across a very wide range of unemployment flow values.

5.3 Effects of changes in concentration on the labor share

We are next interested in the effects of the observed changes in market structure on wages over time. Naturally, in the data it is typically difficult to isolate exogenous shifts in market structure

[15] Specifically, our derivations did not impose that surplus is positive. They solely imposed that all matches are formed and workers receive a split α. In our baseline calibration all surplus is endogenously positive. This is no longer the case as b rises. In case of negative surplus, the Nash-bargain implements a "pain-split": Firms make negative flow profits, $w_i > pi$ and workers derive negative net value $W_i - U_i = \alpha S_i < 0$ from the employment relationship.

that do not have independent effects on labor share. Here, we use the structure of the model to isolate the role of market structure. In particular, we use Proposition 3 to quantify the effects of these changes in market structure on the labor share, $\frac{\bar{w}}{p^1}$.

Our goal is to capture changes in wages that reflect changes in concentration, and to strip out the variation that comes from changes in the level of productivity or other labor market parameters (such as the job finding rate). To do so, we construct two alternative time series for the aggregate labor share using our model. Recall that our calibrated model exactly fits the evolution of the aggregate labor share in Austria. In the first exercise, we fix $\mathcal{P}$ at its baseline (1997) value and let everything else evolve as before. The difference between the observed and the alternative time series thus isolates the impact of shifts in $\mathcal{P}$ over time. In the second exercise, we fix both $\mathcal{P}$ and $\mathcal{C}$ at their 1997 values. The difference between the observed and the second alternative time series thus isolates the impact of shifts in both $\mathcal{P}$ and $\mathcal{C}$ over time. The difference between the first and second exercises isolates the role of changes in pure size-driven market power over time.

Figure 5b shows that changes in market structure over time substantially contributed to the decline in the Austrian labor share. The first exercise shows that changes in the covariance between size and productivity have reduced the labor share over our sample period by almost one percent. The second exercise shows that, over the entire sample period, shifts in employment concentration reduced the labor share by about as much as shifts in productivity shares. Taken together, market power thus reduced the Austrian labor share by about two percent over our sample period, which can explain over 40 percent of the overall decline depicted in Figure 1a.

There are two things worth emphasizing about the patterns in this figure. First, as before, the time path is not a simple monotone transformation of the path of the concentration measures displayed in Figure 3. This finding emphasizes that in the context of our model it is not sufficient to compute a weighted linear average of local concentration to infer the contribution of trends in local concentration to trends in labor share. Second, one appealing feature of our framework is that we can separate the effect of changes in concentration from changes in productivity-weighted concentration, which turn out to have different temporal patterns. Moreover, this separation allows us to quantify the "superstar" firm effect of Autor et al. (2019), and here it turns out to contribute to about a one percentage point decline in the labor share.[16]

In contrast to the atomistic benchmark results, market definition matters for whether trends in concentration explain trends in the labor share. Table 5 shows that different market definitions generate qualitatively (and quantitatively) different accounts of the implications of changes in market structure for the labor share than what we get in our baseline labor markets. This highlights the importance of carefully drawing the boundaries of labor markets.

[16] In Appendix E, we report a Melitz and Polanec (2015) decomposition of the change in labor share using our model-inferred productivity.

6 Merger simulation

In this section, we use our model to simulate the effects of mergers on labor markets. Naidu, Posner, and Weyl (2018), Marinescu and Hovenkamp (2019), and Shapiro (2019) argue that antitrust authorities should take labor market implications of mergers into account when approving mergers. Our model allows us to quantify how big these effects might be. In particular, our model explicitly takes into account the overall structure of the market under consideration. The impact of mergers depends not only on the merging firms but the remaining firms in the market. Indeed, mergers affect wages at all firms in a market, including the non-merging firms.

We simulate mergers as follows. In each labor market in 2015, we merge the two largest employers. We assume that the combined employer has the employment weighted average productivity of the two constituent firms. Given the new concentration of employment shares $\mathcal{C}$ (and correlation between employment and productivity as captured by $\mathcal{P}$), we simply recompute wages at each firm using the model. This allows us to consider effects on wages at the merging firms as well as the remaining firms in the market.[17]

Panel A of Table 6 shows that these mergers would, on average, increase the HHI by 0.05 compared to an average level of 0.12. This change reduces wages by about 6 percent. As before, the average effect is significantly larger than the median effect again highlighting that the force we model here becomes particularly powerful in markets which are already highly concentrated. The panel also shows that there is very large spillovers to the remaining firms in the markets with wages at non-merging firms declining by 3 percent on average: All market participants recognize the reduction in demand-side competition associated with the merger and consequently lower wages.

The last row of Panel A emphasizes that our model features pure rent extraction: market power shifts the distribution of match surplus. Hence, when firms merge, wages change without employment changing. In contrast, in standard models of monopsony quantities and prices are tightly linked: firms reduce wages by reducing employment.

We now discuss tests proposed by Naidu, Posner, and Weyl (2018), which point to mergers that would generate more scrutiny. The benefit of looking at these statistics in our data and model is that first, we can ask how common is it for hypothetical mergers to pass the relevant thresholds, and second, we can use our model to convert the change in HHI thresholds into wages. First, they emphasize (pg. 577) various thresholds of the change in HHI from the merger that would generate extra scrutiny. One region is a change in HHI of 0.1 to 0.2. For our markets, Panel B shows that this happens in fewer than 10 percent (33 of 356) of the mergers. In our model, the median decrease in market-wide wages in such mergers is about nine percent. A more lenient threshold emphasized by Naidu, Posner, and Weyl (2018) is mergers where HHI increases by more than 0.2. This happens in about five percent of mergers (19 of 356) and the median decrease in wages in our model is almost 50 percent. Second, they suggest (pg. 575) that mergers where merging-firm wages would decline by 5% or more are mergers deserving extra scrutiny. For our market definition, this

[17] We drop the 13 markets where there are fewer than three firms because in these markets it is not possible to compute the wage effects on non-merging firms.

includes almost fourty percent of mergers (in about a quarter of cases would market-wide wages fall by this much).

Panel C shows that fewer than half the simulated mergers are in the same region, and only a third are in the same 2-digit industry. These statistics highlight how our definition of labor markets deviates from traditional definitions and emphasize that even mergers of spatially disconnected firms may reduce wages substantially.

Panel D emphasizes the nonlinearity in the model. We show the wage effects of increasing $\mathcal{C}$ by 0.025 (i.e., the median increase in HHI in mergers) at various levels of concentration and averaging over the effects across all markets.[18] As can be anticipated from Figure 4a, the effects of the same change in concentration depend on the initial level of concentration. From the 25th to the 75th percentile of concentration distribution, such a merger would depress market-wide wages by about 1 percent. Increasing the baseline level of concentration by a factor of six (from 0.11 to 0.60), increases the effect of the same increase in concentration by more than a factor of six.

Appendix Table A2 shows how the effects of merging the two largest firms in each market depends on labor market definition. Not surprisingly, the coarser market definitions imply much smaller effects than the finer market definitions.

7 Wage-concentration regressions in the model

So far we have used the structure of our model to quantify the effects of concentration on wages. Here, we compare the model-implied effects of concentration on wages to those found in the literature that relates concentration and wages.[19] We use our model to sidestep the fundamental identification challenge in the literature: we generate data where the only source of variation in wages is concentration. We then run the same regressions that the literature has considered.

We build a dataset where the only variation over time in wages is caused by changes in concentration. We take each labor market in 1997 and compute the path of concentration ($\mathcal{C}$) from 1997 to 2015. We then use Proposition 1 to simulate the effects of these changes in concentration on wages. Thus, we end up with a dataset with 19 years and 369 markets where the only variation over time in wages is caused by variation in concentration.

We then follow the functional form of the regression specification in Azar, Marinescu, and Steinbaum (2017, Table 2) and Rinz (2018, Table 5) and regress log average wages on log HHI with market and time fixed effects:

$$\ln \bar{w}_{mt} = \beta \ln HHI_{mt} + \gamma_m + \gamma_t + \epsilon_{mt}. \tag{9}$$

[18]Here, there are not actual firms in our data that correspond to these changes in concentration and so we only move $\mathcal{C}$, we leave $\mathcal{P}$ unchanged.

[19] 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. A presumably incomplete list of recent papers 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), Rinz (2018), and Schubert, Stansbury, and Taska (2019).

Our coefficient of interest is β: the elasticity of average wages with respect to HHI.

Table 7 shows that the elasticity we estimate is broadly in line with the literature. For our data-driven labor markets, we find an elasticity of -0.097.[20] By way of comparison, Azar, Marinescu, and Steinbaum (2017, Table 2, Panel A, column (6)) estimate an elasticity of -0.127. And, using data from 2005 to 2015, Rinz (2018, Table 5, column (5)) finds an elasticity of -0.161 (for 1976-2015, Rinz (2018, Table 4, column (5)) finds an elasticity of -0.282).

The Table also shows that the elasticity increases as we use finer markets. We perform the same exercise for alternative market definitions and find elasticities ranging from -0.006 when we use regions, to -0.174 when we use 4-digit industry $\times$ region. A general pattern is that we find larger elasticities when we use narrower definitions of labor markets. The reason is the nonlinearity in the model that we highlighted in Figure 4b: narrower definitions of labor markets place us further to the right on that Figure, where there are larger elasticities of wages with respect to concentration.

8 Discussion

This paper develops a new model of size-based market power that provides a microfoundation for an equilibrium relationship between market structure—in particular, concentration—and wages. In our model, the job openings of granular employers do not compete with each other for workers and, as a consequence, the distribution of employment shares matters for wages. Wages are lower at large firms and at firms in more concentrated markets.

The model also provides a natural intuition for why a concentration index includes the sum of squared market shares. Under random search, it captures the probability that an unemployed worker encounters the same firm two times in a row. And it is the possibility of this second encounter that gives rise to market power.

The model allows us to transparently assess the effects of (changes in) market structure on levels and trends in wages, and, similarly, to assess the effects of hypothetical mergers on all workers in a labor market. We implement our framework in Austrian matched employer-employee data. We complement standard definitions of labor markets with data-driven labor markets based on worker flows. We find that granular market power depresses Austrian wages by about ten percent and has contributed substantially to the decline of the Austrian labor share in our sample period. Moreover, the mergers we simulate have large effects: even workers at non-merging suffer substantial wage losses.

This paper offers a new perspective on imperfect competition in labor markets and thus opens up numerous exciting avenues for future research. The model could be extended in a variety of directions. For example, one could add on the job search, or endogenize the size distribution. Similarly, one could analyze various labor market policies such as non-compete clauses and unions, or any shock that affects the distribution of employment and productivity across firms.

[20] To compare to Figure 4b, recall that a regression is a variance-weighted average of effects and so the relevant way to aggregate over the Figure is using the levels of concentration in markets with the largest changes in concentration, rather than the distribution of the level of concentration.

References

Acemoglu, Daron and Robert Shimer. 1999. "Efficient Unemployment Insurance." *Journal of Political Economy* 107 (5):893–928.

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

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2019. "The Fall of the Labor Share and the Rise of Superstar Firms." 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. 2019. "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.

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.

Burdett, Kenneth and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Reviewl* 39 (2):257–273.

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.

Fink, Martina, Esther Segalla, Andrea Weber, and Christine Zulehner. 2010. "Extracting Firm Information from Administrative Records: The ASSD Firm Panel." Working Paper 1004, NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State.

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.

Gouin-Bonenfant, Emilien. 2018. "Productivity Dispersion, Between-Firm Competition, and the Labor Share." Working paper.

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.

Hornstein, Andreas, Per Krusell, and Giovanni L. Violante. 2011. "Frictional Wage Dispersion in Search Models: A Quantitative Assessment." *American Economic Review* 101 (7):2873–2898.

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

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. 2019. "Imperfect Competition, Compensating Differentials, and Rent Sharing in the U.S. Labor Market." Working Paper 25954, NBER.

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.

Manning, Alan. 2003. *Monopsony in Motion*. Princeton University Press.

Marinescu, Ioana Elena and Herbert J. Hovenkamp. 2019. "Anticompetitive Mergers in Labor Markets." *Indiana Law Journal* 94.

Melitz, Marc J. and Saso Polanec. 2015. "Dynamic Olley-Pakes productivity decomposition with entry and exit." *Rand Journal of Economics* 46 (2):362–375.

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

Naidu, Suresh, Eric A. Posner, and Glen Weyl. 2018. "Antitrust remedies for labor market power." *Harvard Law Review* 132 (2):536–601.

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.

O'Mahony, Mary and Marcel P. Timmer. 2009. "Output, Input and Productivity Measures at the Industry Level: The EU KLEMS Database." *The Economic Journal* 119 (538):F374–F403.

Peixoto, Tiago P. 2017. "Nonparametric Bayesian inference of the microcanonical stochastic block model." *Phys. Rev. E* 95:012317.

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.

Schubert, Gregor, Anna Stansbury, and Bledi Taska. 2019. "Getting labor markets right: occupational mobility, outside options, and labor market concentration." Working paper.

Shapiro, Carl. 2019. "Protecting Competition in the American Economy: Merger Control, Tech Titans, Labor Markets." *Journal of Economic Perspectives* 33 (3):69–93.

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

Stole, Lars A. and Jeffrey Zwiebel. 1996. "Intra-firm bargaining under non-binding contracts." *Review of Economic Studies* 63:375–410.

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.

Zhu, Haoxing. 2012. "Finding a good price in opaque over-the-counter markets." *Review of Financial Studies* 25 (4):1255–1285.

Table 1: Summary statistics and parameter values

	Average	Median	25th	75th
Panel A. Summary Statistics				
w_i	77.13	74.99	69.59	84.50
Firm Employment	45.93	13	8	29
Hires	12.24	4	2	9
Hires from u	7.12	2	1	5
Workers per market	4953.27	3303	1995.5	5966.75
Employers per market	111.74	90	44	142
Panel B. Parameters				
λ_m	0.098	0.093	0.070	0.119
$\underline{\lambda}_m$	0.00065	0.00002	0.00000	0.00023
δ_m	0.009	0.008	0.005	0.012
α	0.4845			
b_m	-226.66	-195.37	-319.40	-108.75
β	0.9957			
Panel C. Average pass-through coefficients				
Productivity	0.4475	0.4712	0.4579	0.4773

Notes: All statistics are for 2015 and when there are market-specific parameters these reflect employment-weighted averages. Panel A of this table reports summary statistics on the distribution of firm-level wages and firm size. Wages, w_i are measured as the firm-level median of residualized daily individual wages in Euros (2000) at firm i. Firm size is measured at a reference date (August 1st). We also measure size as the number of total yearly hires or hires from unemployment. Panel B reports parameter values. For market- and time-specific parameters λ, $\underline{\lambda}$, δ, and b, it reports employment-weighted summary statistics. λ, $\underline{\lambda}$ and δ are at a monthly frequency. b is in the same units as wages. Panel C reports the average pass-through of productivity changes to wages, where the distribution is across employment-weighted markets.

Table 2: Share of transitions within markets

	Share of within-market transitions				Modularity score
	Average	Median	25th	75th	
Data-driven Labor Markets (369)	0.41	0.52	0.07	0.62	0.41
Alternative market definitions					
Data-driven Labor Markets (9)	0.82	0.80	0.78	0.86	0.69
States (9)	0.76	0.75	0.75	0.85	0.60
NUTS3-regions (35)	0.60	0.71	0.43	0.77	0.49
2-digit Industries (80)	0.40	0.39	0.25	0.52	0.36
3-digit Industries (255)	0.34	0.35	0.19	0.47	0.32
4-digit Industries (538)	0.30	0.31	0.14	0.42	0.29
2-digit Industries × Regions (1838)	0.25	0.24	0.12	0.34	0.24
3-digit Industries × Regions (3615)	0.21	0.18	0.05	0.30	0.21
4-digit Industries × Regions (5384)	0.18	0.14	0.03	0.28	0.18

Notes: This Table reports summary statistics on the share of within-market transitions among all employer-employer (EE) transitions between the firms in our sample. An EE transition is defined as a change of the firm with at most 30 days of non-employment between the two spells and at least one year of tenure in the old and the new job. The first set of columns shows the share of EE transitions. The last column shows the modularity score, which is the excess share of within-market transitions over a null model of random transitions.

Table 3: Heterogeneity of effects of market structure across markets

	Average	Median	5th	25th	75th	95th
Panel A. Concentration measures						
HHI	0.120	0.053	0.012	0.029	0.102	0.691
$\mathcal{C}$	0.124	0.056	0.012	0.030	0.108	0.729
$\mathcal{C}^P$	0.125	0.055	0.012	0.031	0.112	0.643
$\mathcal{P}$	0.001	0.001	-0.003	-0.000	0.003	0.024
Panel B. %Δ in labor share in the atomistic benchmark						
$\mathcal{P} = 0$	1.8	0.1	-0.3	-0.0	0.6	4.0
$\mathcal{P} = \mathcal{C} = 0$	12.6	2.2	0.5	1.0	5.0	22.7

Notes: This Table reports how a variety of measures vary across markets. All measures are calculated for the year 2015. The average column reflects employment-weighted averages. The remaining columns report results for employment-weighted quantiles of the markets. Panel A shows the distribution of the Hirschman-Herfindahl index (HHI), our concentration index $\mathcal{C}$, our productivity-weighted concentration index $\mathcal{C}^P$, and our productivity-concentration weighted wedge $\mathcal{P}$. In Panel B, we compute the distribution of the change in the labor share due to moving to the atomistic benchmark by setting $\mathcal{P} = 0$ or $\mathcal{P} = \mathcal{C} = 0$, and then report quantiles of this distribution across markets.

Table 4: Sensitivity of increase in labor share in atomistic benchmark in 2015

	Setting $\mathcal{P}=0$			Setting $\mathcal{C}=\mathcal{P}=0$		
	%Δ labor share	% of Max	%Δ in λ	%Δ labor share	% of Max	%Δ in λ
Baseline (369, $\alpha=0.48, \bar{b}=-227$)	1.8	0.7	4.7	12.6	4.6	41.6
Panel A. Alternative market definitions						
Data-driven markets (9, $\alpha=0.50, \bar{b}=-302$)	0.04	0.01	0.08	0.45	0.13	0.98
NUTS-3 regions (35, $\alpha=0.48, \bar{b}=-276$)	0.06	0.02	0.12	0.52	0.16	1.16
2-digit industries (80, $\alpha=0.46, \bar{b}=-262$)	0.16	0.06	0.37	1.31	0.44	2.99
3-digit industries (255, $\alpha=0.43, \bar{b}=-226$)	0.51	0.19	1.19	3.39	1.25	8.27
4-digit industries (538, $\alpha=0.43, \bar{b}=-202$)	0.84	0.34	2.03	5.65	2.25	14.89
2-digit industry × region (1838, $\alpha=0.45, \bar{b}=-186$)	2.11	0.94	5.62	13.80	5.83	46.75
3-digit industry × region (3615, $\alpha=0.42, \bar{b}=-136$)	2.84	1.53	8.30	19.34	9.57	79.28
4-digit industry × region (5384, $\alpha=0.42, \bar{b}=-109$)	3.67	2.30	11.80	25.44	14.01	126.18
Panel B. Alternative wage and size definitions						
w_i 25th percent. of firm-level wage distr. ($\alpha=0.48, \bar{b}=-186$)	1.82	0.70	4.81	12.58	4.64	41.62
w_i 75th percent. of firm-level wage distr. ($\alpha=0.50, \bar{b}=-290$)	1.75	0.63	4.58	12.81	4.46	42.14
f_i share of new hires ($\alpha=0.48, \bar{b}=-187$)	1.14	0.42	2.89	9.80	3.46	29.40
f_i share of new hires from u ($\alpha=0.44, \bar{b}=-158$)	0.60	0.23	1.44	5.66	2.10	14.84

Notes: This Table reports the sensitivity of the effects of moving to the atomistic benchmark to market definition, and alternative definitions of wages and employer size. The first row shows our baseline results where we use 369 data-driven labor markets. We consider two quantitative exercises: setting $\mathcal{P}$ to zero and setting $\mathcal{P}$ and $\mathcal{C}$ to zero. Columns 1 and 4 report the percent increase in the labor share in these exercises. Columns 2 and 5 express those gains relative to the largest possible gains from eliminating concentration (going from full monopsonist to atomistic firms). Columns 3 and 6 report the percent change in the job finding rate that would deliver the same gains to workers as moving to the atomistic benchmark. In each row we recalibrate the model. In parentheses, we report the number of markets, α, and $\bar{b}$ (in units of euros per day). Panel A considers alternative market definitions. Panel B considers alternative definitions of the firm-level wage, where our baseline results use the median firm-level wage. It also considers alternative definition of f_i based on share of new hires in year t and the share of new hires from unemployment in year t.

Table 5: Effects of changes in market structure on the labor share

	Contribution of	
	$\mathcal{P}$ and $\mathcal{C}$	$\mathcal{P}$
	(1)	(2)
Baseline (369)	-1.19	-0.65
Alternative market definitions		
Data-driven markets (9)	0.07	-0.01
NUTS-3 regions (35)	0.02	-0.02
2-digit industries (80)	0.17	0.02
3-digit industries (255)	0.28	-0.03
4-digit industries (538)	0.48	-0.06
2-digit industry × region (1838)	0.26	-0.14
3-digit industry × region (3615)	0.27	0.67
4-digit industry × region (5384)	-0.57	0.98

Notes: This Table reports the percentage point change in the labor share explained by changes in $\mathcal{P}$ and $\mathcal{C}$ from 1997 to 2015 in our baseline market definition, as well as various alternative market definitions. The numbers in the baseline row correspond to the last point (scaled by the level of the labor share to convert to percentage points) in Figure 5b. The remaining rows report the parallel exercise for other market definitions.

Table 6: Merger simulation

	Average	Median	25th	75th
Panel A. Distribution of effects				
Δ HHI	0.046	0.025	0.013	0.058
%Δ wages at merging firms	-7.9	-3.5	-6.2	-2.0
%Δ wages at non-merging firms	-3.0	-0.6	-1.7	-0.2
%Δ market-wide wages	-5.6	-1.4	-3.6	-0.6
%Δ employment	0	0	0	0
Panel B. %Δ market-wide wages given certain HHI changes				
$\Delta HHI \in [0.1, 0.2]$ (33)	-13.7	-9.2	-16.2	-8.4
$\Delta HHI > 0.2$ (19)	-64.3	-45.6	-67.3	-27.7
Panel C. Percent of mergers satisfying various criteria				
Market-wide wages $\downarrow > 5\%$	23.0			
Merging firm wages $\downarrow > 5\%$	38.2			
Same region	44.4			
Same 2-digit industry	33.4			
Same 3-digit industry	28.4			
Same 4-digit industry	24.7			
Panel D. %Δ market-wide wages of a merger that increases $\mathcal{C}$ by 0.025				
From $\mathcal{C} = 0.030$ (25th)	-1.0			
From $\mathcal{C} = 0.108$ (75th)	-1.2			
From $\mathcal{C} = 0.250$	-1.7			
From $\mathcal{C} = 0.500$	-4.4			
From $\mathcal{C} = 0.600$	-8.0			
From $\mathcal{C} = 0.650$	-12.6			
From $\mathcal{C} = 0.729$ (95th)	-66.2			

Notes: Panels A through of C of this Table report the effects of combining the two largest employers in each data-driven labor market in 2015. We report results for the 356 markets where there are more than two firms. Panel A and B report employment-weighted statistics, while Panel C reports unweighted statistics across markets. Panel D reports the effects of increasing $\mathcal{C}$ from various levels by 0.025. We average over markets and leave the size-productivity correlation unchanged.

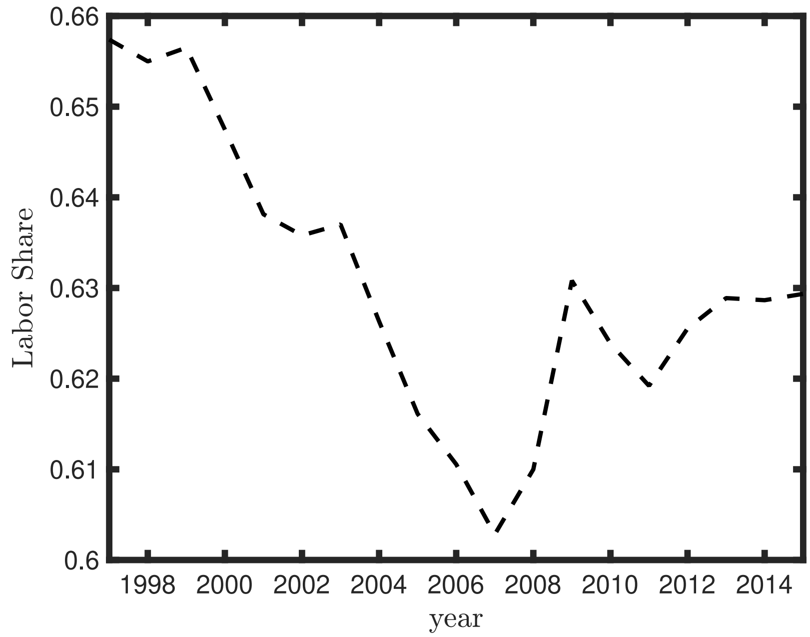
Table 7: Wage-concentration regressions in the model

	Elasticity of wages to HHI
Baseline (369)	-0.097
	(0.004)
Alternative market definitions	
Data-driven markets (9)	-0.003
	(0.000)
NUTS-3 regions (35)	-0.006
	(0.000)
2-digit industries (80)	-0.015
	(0.001)
3-digit industries (255)	-0.040
	(0.002)
4-digit industries (538)	-0.070
	(0.002)
2-digit industry × region (1838)	-0.097
	(0.002)
3-digit industry × region (3615)	-0.140
	(0.002)
4-digit industry × region (5384)	-0.174
	(0.002)

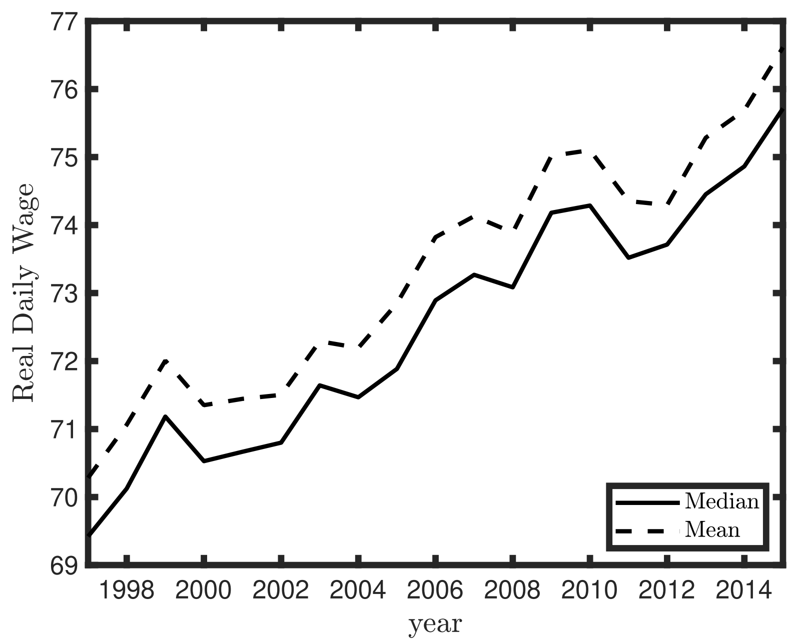
Notes: This Table reports the elasticity of wages with respect to the HHI from a regression estimated on data simulated from the model. Simulated wages for year t are the actual wage in the initial year (1997) plus the variation in wages that derives from changes in $\mathcal{C}$ (holding productivity and parameters fixed at their initial value). The Table reports regression coefficients and standard errors (in parenthesis). We regress simulated log wages on observed log HHI on the market-year level. All regressions include market and year fixed effects.

Figure 1: Trend in labor market aggregates in Austria

(a) Labor Share



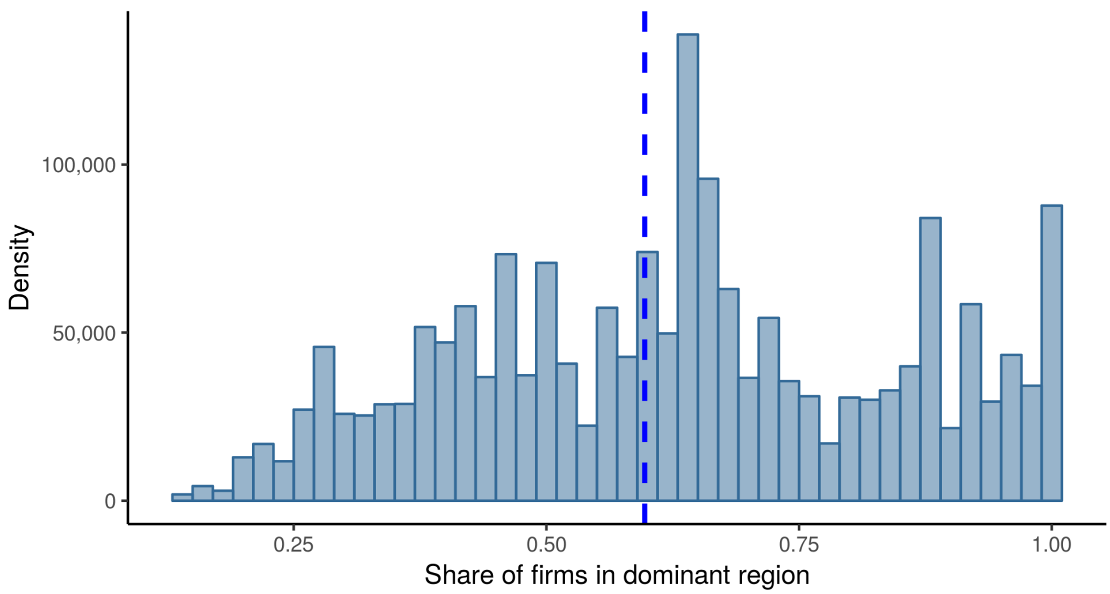
(b) Real Daily Wages



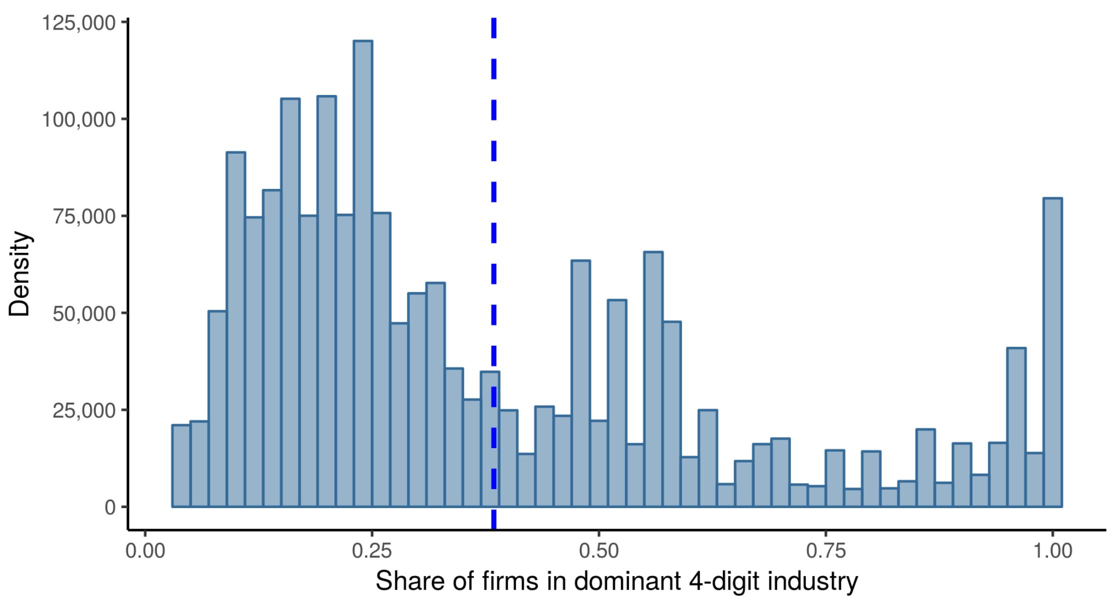
Notes: Panel A of this Figure plots the labor share in Austria based on KLEMS data for the sample period from 1997 to 2015. The labor share is defined as aggregate compensation over aggregate value added for all industries in Austria. Panel B plots employment-weighted median and mean of real daily earnings in our sample using the CPI from Statistic Austria with base year 2000 as the deflator.

Figure 2: Data-driven markets are not the same as region or industry

(a) Regions

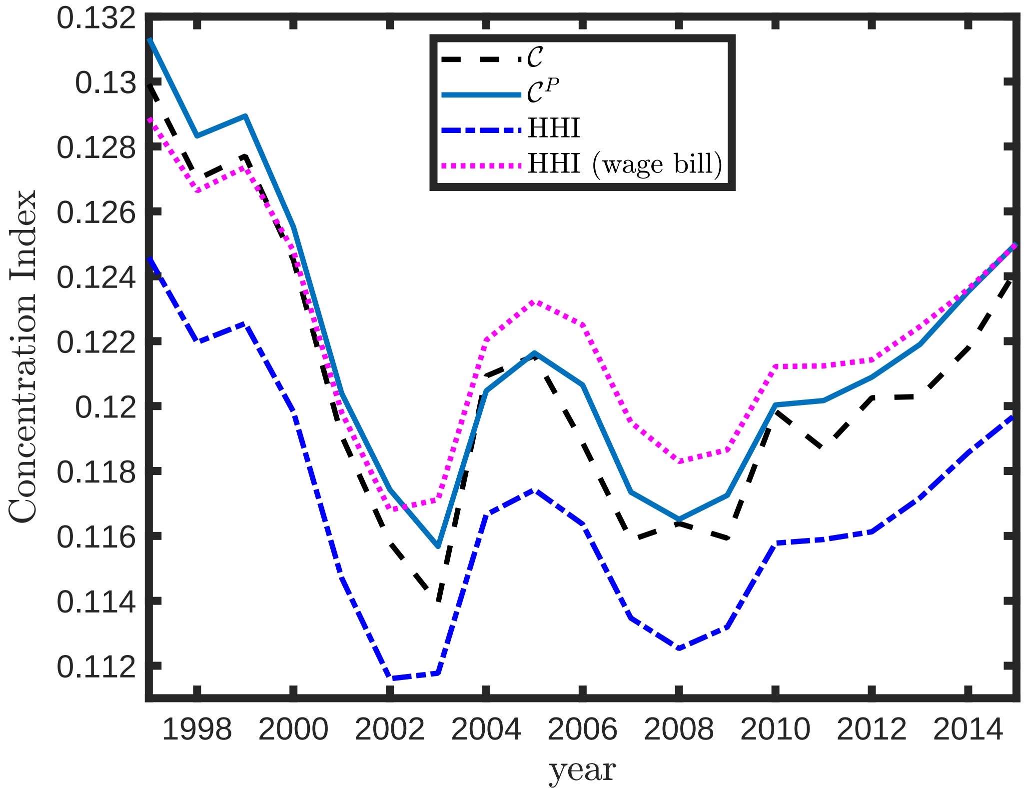


(b) Industries



Notes: This Figure shows a sense in which the data-driven labor markets capture industry or geographic boundaries. For each market, we classify its "dominant" region or industry as the region or industry with the largest share of employment. The figures then show the distribution of the share of employment contained in the dominant region or industry. A value of 1 says that all of the employment is in a single region or industry. The figure displays employment-weighted averages over all 369 data-driven labor markets for the year 2015. The dashed line shows the average value.

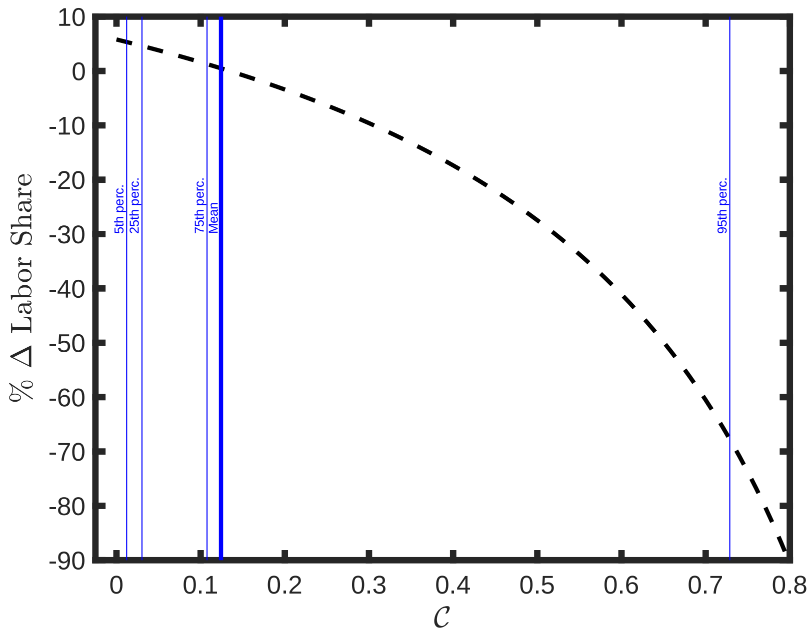
Figure 3: Trends in labor market concentration



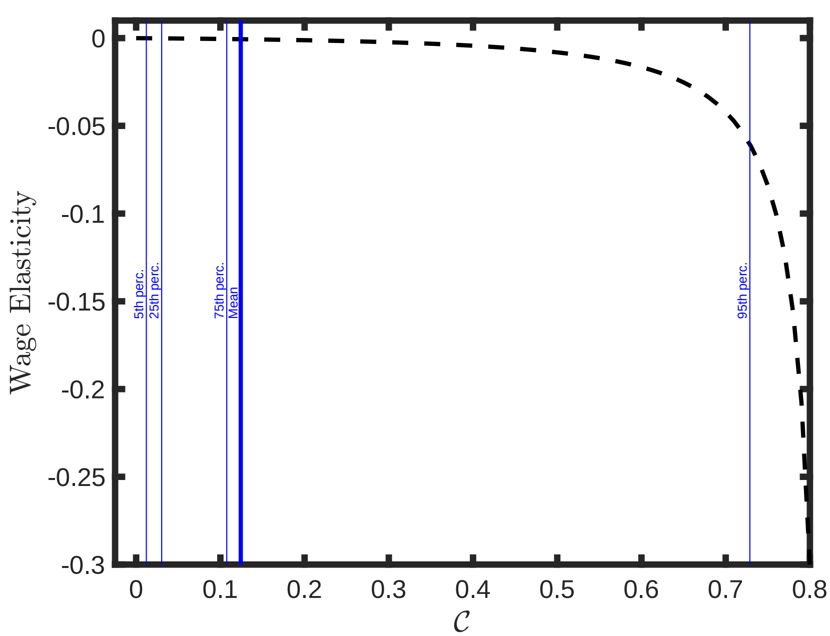
Notes: This Figure plots concentration indexes $\mathcal{C}$, $\mathcal{C}^P$, HHI and wage-bill HHI from 1997 - 2015. The figure displays employment-weighted averages over all 369 data-driven labor markets.

Figure 4: Nonlinear effects of concentration on labor share and wage elasticity

(a) Effect of moving to atomistic benchmark



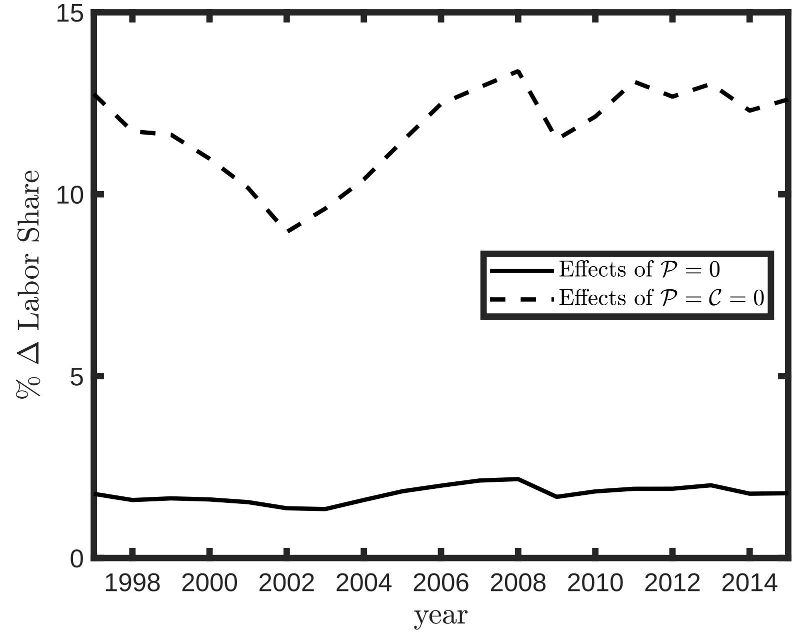
(b) Elasticity of wages with respect to $\mathcal{C}$



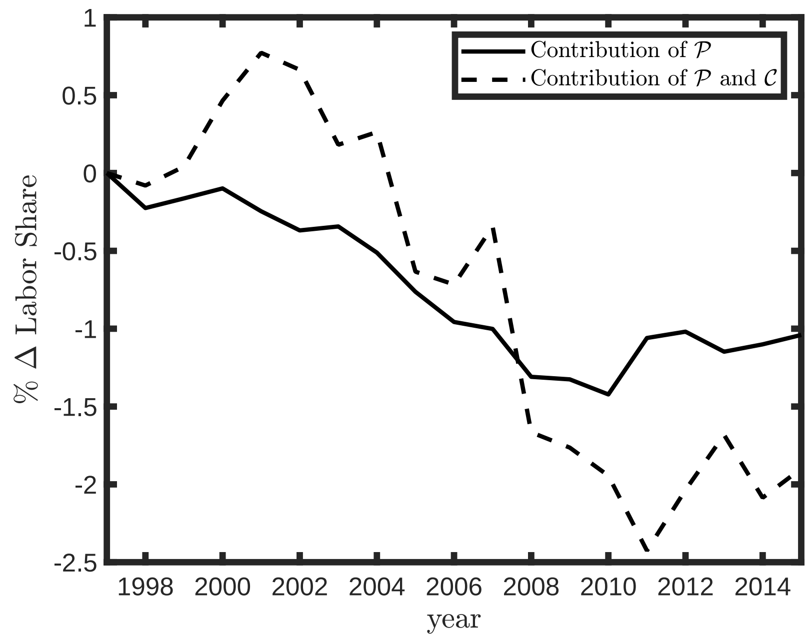
Notes:. This Figure documents nonlinearities in the effect of concentration on the labor share and wages. Panel (a) reports the effect on the aggregate labor share when moving each market from the observed level of concentration in the data over the support of $\mathcal{C}$. The thick dashed lines show the average value of $\mathcal{C}$ in our data in 2015, and the thin dashed lines show the 5th, 25th, 75th and 95th percentiles. Panel (b) shows that the elasticity of wages with respect to $\mathcal{C}$ varies with $\mathcal{C}$. We compute the arc-elasticity of wages with respect to concentration using a one-percent change in $\mathcal{C}$ at different levels of $\mathcal{C}$.

Figure 5: Labor share accounting exercises

(a) Change in labor share in atomistic benchmark (over time)

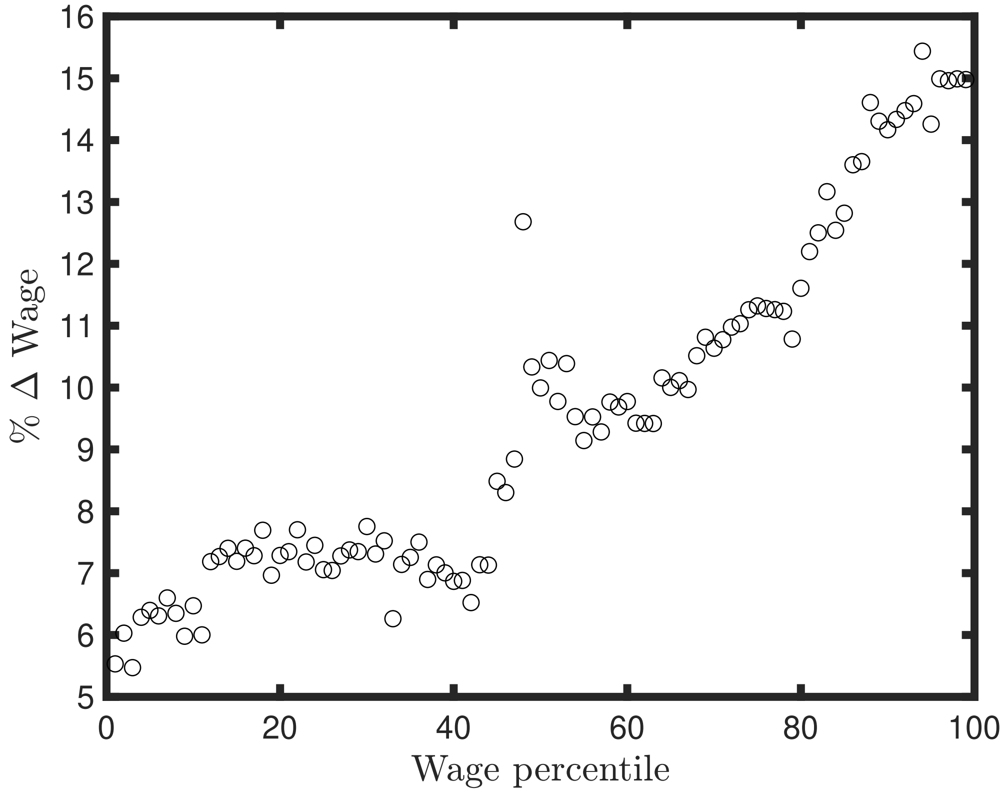


(b) Change in labor share over time



Notes: This Figure uses Proposition 3 to compute changes in labor share due to changes in concentration. In each it reports the change relative to the actual evolution of the labor share $(\frac{\bar{w}}{p^1})$ in percent. We always let λ, δ, b, and α evolve as in the calibration and report changes in the employment-weighted averages across the 369 data-driven labor markets. The top panel moves the economy to atomistic benchmark in two steps: First, it sets $\mathcal{P} = 0$ in all market-years; then it also sets $\mathcal{C} = 0$ in all market-years. The bottom panel first fixes $\mathcal{P}$ at initial 1997 values to isolate the contribution of changes in $\mathcal{P}$ over time. It then fixes both $\mathcal{P}$ and $\mathcal{C}$ at initial 1997 values.

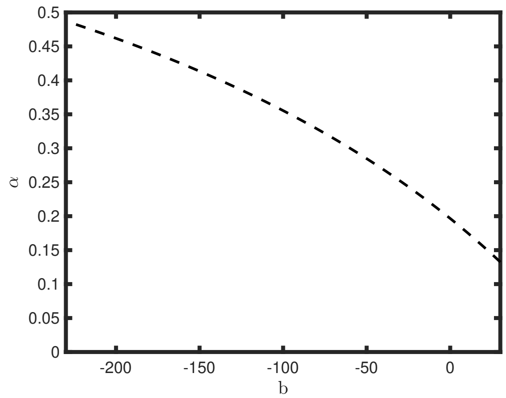
Figure 6: Distributional impact of moving to the atomistic benchmark



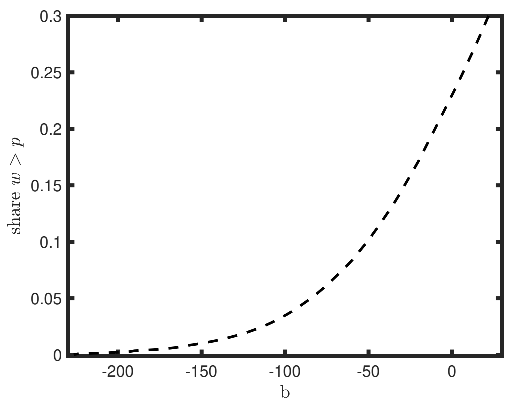
Notes: This Figure shows how the percent change in wages implied by moving to the atomistic benchmark varies over the distribution of individual-level wages. We bin raw wages from 2015 (below the social security contribution cap) into percentiles and compute average wage changes within each bin. We compute the percent wage change as the wage change at the worker's employer.

Figure 7: Alternate values of worker bargaining power and the flow value of unemployment

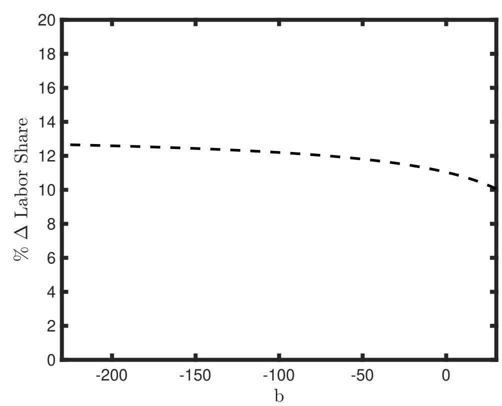
(a) α



(b) Share $w > p$



(c) $\mathcal{C} = \mathcal{P} = 0$



Notes: This Figure shows the effect of increasing b. In all cases, the x-axis show the market-weighted b which we increase in constant proportion across markets (we take the market-specific b and $\bar{w}$ and compute $b' = \bar{w} - x * (\bar{w} - b)$ for various values of x; the x-axis shows employment-weighted values of b). Panel (a) plots values of α that yield an employment-weighted average labor share of 0.629 using data from 2015. Panel (b) shows the share of firms that pay wages above productivity (earn negative profits) given the combination of α and b. Panel (c) shows how the results of our benchmark exercise (increase in the labor share by setting $\mathcal{C} = \mathcal{P} = 0$ in 2015) vary as we shrink b.

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 Omitted proofs

B.1 Proof of Proposition 1

Proof. Now:

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

From equations (7), (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 $S^k \equiv \sum_i f_i S_i^k$ so that $S^1 \equiv \sum_i f_i S_i$ and we used the fact that:

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

Now, we obtain two expressions for $(U_j - U_i)$ and $(U - U_i)$ in order to re-write equation (A2) above. We start with $(U_j - U_i)$. 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 to go from the first line to the second line we use $\sum_i f_i = 1$ and define $W^1 \equiv \sum_j f_j W_j$. Hence,

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

Now, recall, from (1), that $U = b + \beta[\lambda \sum_i f_i W_i + (1-\lambda)U]$. Then, note that:

$$U - U_i = \beta[\lambda W^1 + (1-\lambda)U] - \beta[U_i + \lambda \sum_{j \neq i} f_j(W_j - U_i) + \underline{\lambda} f_i(W_i - U_i)]$$

$$(1-\beta(1-\lambda))(U - U_i) = \beta(\lambda - \underline{\lambda}) f_i \alpha S_i$$

$$\beta\delta(U - U_i) = \beta\delta \frac{\beta(\lambda - \underline{\lambda})}{(1-\beta(1-\lambda))} f_i \alpha S_i. \quad \text{(A4)}$$

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

$$(1-\beta(1-\delta))\alpha S_i$$

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

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

where we used $S^k \equiv \sum_i f_i^k S_i$.

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)

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

Recall again that $\tau = \frac{\beta(\lambda - \underline{\lambda})\alpha}{1-\beta(1-\lambda)}$, $S^k \equiv \sum_i f_i^k S_i$, 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

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

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

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

Note that, for $k = 1$, we can use $f^1 = 1$ and the definition of τ to 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:

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

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

Define

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

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

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

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

$$S^1 = \frac{1-b}{\left[(1-\beta(1-\delta)) + \beta\lambda\alpha\right] - \left[\lambda + \frac{\beta\lambda\delta}{1-\beta(1-\lambda)}\right]\alpha\beta\mathcal{C}}$$

or

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

Sum across i in equation (A6) to get

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

where the second line uses (A15) and the third line divides by $1-b$ and uses the definition of $\bar{\omega}$. $\square$

B.2 Proof of Proposition 2

Proof. Start with (A5) and use the exact definition of $\tau = \frac{\beta\lambda - \underline{\lambda}\alpha}{1-\beta(1-\lambda)}$ to get

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

$$(\alpha - \tau f_i)S_i = \frac{w_i - b}{1-\beta(1-\delta)} + \frac{1}{1-\beta(1-\delta)}\Big(-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big). \tag{A17}$$

Use equation (A6) to 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

$$\frac{1-w_i}{1-w_j} = \frac{1-\tau f_j}{1-\tau f_i}. \tag{A18}$$

□

B.3 Properties of $\mathcal{P}$

Proof. Note that:

$$\begin{aligned}\mathcal{C}^P &= \frac{\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k}{\tilde{p}^1 + \tau\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k} \times \frac{\tilde{p}^1}{\tilde{p}^1} \\ &= \frac{\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1}{1+\tau\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1}.\end{aligned} \tag{A19}$$

We have that:

$$\mathcal{C}^P - \mathcal{C} = \frac{\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1}{1+\tau\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1} - \frac{\sum_{k=2}^{\infty}\tau^{k-2}f^k}{1+\tau\sum_{k=2}^{\infty}\tau^{k-2}f^k}. \tag{A20}$$

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:

$$\begin{aligned}\sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k/\tilde{p}^1 - \sum_{k=2}^{\infty}\tau^{k-2}f^k &= \sum_{k=2}^{\infty}\tau^{k-2}(\tilde{p}^k/\tilde{p}^1 - f^k) \\ &= \sum_i\sum_{k=2}^{\infty}\tau^{k-2}f_i^k(\tilde{p}_i/\tilde{p}^1 - 1) \\ &= \frac{1}{\tau^2}\sum_i\sum_{k=1}^{\infty}\tau^k f_i^k(\tilde{p}_i/\tilde{p}^1 - 1) - \frac{1}{\tau^2}\sum_i \tau f_i(\tilde{p}_i/\tilde{p}^1 - 1) \\ &= \frac{1}{\tau^2}\sum_i(\tilde{p}_i/\tilde{p}^1 - 1)\frac{\tau f_i}{1-\tau f_i} - \frac{1}{\tau}(\tilde{p}^1/\tilde{p}^1 - 1).\end{aligned} \tag{A21}$$

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

$$\sum_{k=2}^{\infty} \tau^{k-2} {}^{\tilde{p}^k}/_{\tilde{p}^1} - \sum_{k=2}^{\infty} \tau^{k-2} f^k = \frac{1}{\tau^2} \sum_i ({}^{\tilde{p}_i}/_{\tilde{p}^1} - 1) \frac{\tau f_i}{1 - \tau f_i}$$
$$= \frac{1}{\tau} \sum_i \frac{f_i ({}^{\tilde{p}_i}/_{\tilde{p}^1} - 1)}{1 - \tau f_i}. \tag{A22}$$

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. □

B.4 Proof of Proposition 3

Proof. Recall that, under heterogeneous productivity, the output per worker at firm i is given by p_i. Hence, the equivalent of equation (4) is

$$J_i = p_i - w_i + \beta(1 - \delta) J_i.$$

This equation, together with (8), gives us the equivalent of (A6) under heterogeneous productivity:

$$w_i = p_i - (1 - \beta(1 - \delta))(1 - \alpha) S_i. \tag{A23}$$

Now, we proceed in exactly the same fashion as in the proof of Proposition 1. The proof is unaltered up to equation (A5).

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

Thus, proceeding identically to the proof of Proposition 1, we combine (A23) and (A24) to obtain the counterpart to equation (A8):

$$(1 - \beta(1 - \delta)) S^k = \tilde{p}^k + f^k \left[- \beta\lambda\alpha S^1 + \beta\lambda \frac{\beta(\lambda - \underline{\lambda})\alpha}{(1 - \beta(1 - \lambda))} S^2 \right] + \frac{1 - \beta(1 - \delta)}{1 - \beta(1 - \lambda)} \beta(\lambda - \underline{\lambda}) \alpha S^{k+1} \tag{A25}$$

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

Evaluate (A25) at $k = 1, 2, 3, ...$ to get

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

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

$$(1-\beta(1-\delta))S^1 = \tilde{p}^1 - \beta\lambda\alpha S^1 + \beta(\lambda - \underline{\lambda})\alpha\frac{1-\beta+\beta(\delta+\lambda)}{(1-\beta(1-\lambda))}S^2. \tag{A26}$$

Now start the substitution

$$(1-\beta(1-\delta))S^1 = \tilde{p}^1 + f^1\Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big] + \tau\Big(\tilde{p}^2 + f^2\Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big] + (1-\beta(1-\delta))\tau S^3\Big). \tag{A27}$$

If we keep substituting, then we get:

$$(1-\beta(1-\delta))S^1 = \Big(\tilde{p}^1 + \tau\tilde{p}^2 + \tau^2\tilde{p}^3 + ...\Big) + \Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big]\Big(f^1 + \tau f^2 + \tau^2 f^3 + ...\Big). \tag{A28}$$

Proceeding identically for S^2 gives

$$(1-\beta(1-\delta))S^2 = \Big(\tilde{p}^2 + \tau\tilde{p}^3 + \tau^2\tilde{p}^4 + ...\Big) + \Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big]\Big(f^2 + \tau f^3 + \tau^2 f^4 + ...\Big). \tag{A29}$$

Define

$$P \equiv \Big(\tilde{p}^2 + \tau\tilde{p}^3 + \tau^2\tilde{p}^4 + ...) = \sum_{k=2}^{\infty}\tau^{k-2}\tilde{p}^k \tag{A30}$$

and let, as previously, $F \equiv \sum_{k=2}^{\infty}\tau^{k-2}f^k$. This gives, directly from equations (A28) and (A29)

$$S^2 = S^1\frac{F}{1+\tau F} - \frac{1}{1-\beta(1-\delta)}\left[\left(\tilde{p}^1 + \tau P\right)\frac{F}{1+\tau F} - P\right] = S^1\mathcal{C} - \frac{1}{1-\beta(1-\delta)}\left[\left(\tilde{p}^1 + \tau P\right)\mathcal{C} - P\right]. \tag{A31}$$

Note that

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

and so A31 becomes

$$S^2 = S^1\mathcal{C} + \frac{1}{1-\beta(1-\delta)}\tilde{p}^1\mathcal{P}. \tag{A32}$$

Plug this into equation (A26) to get

$$S^1 = \frac{\tilde{p}^1\left(1 + \tau\frac{1-\beta(1-(\delta+\lambda))}{1-\beta(1-\delta)}\mathcal{P}\right)}{1-\beta(1-(\delta+\lambda\alpha)) - \tau\mathcal{C}(1-\beta(1-(\delta+\lambda)))}. \tag{A33}$$

Define S^{1*} to be the employment weighted mean surplus from the homogeneous firm case given in (A15). Use the definition of $\hat{\tau}$ and the steps leading from (A14) to (A15) to get that

$$S^1 = S^{1*}\frac{\tilde{p}^1}{1-b}\left(1+\hat{\tau}\mathcal{P}\right). \tag{A34}$$

Integrate across equation (A23) to get

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

and thus, plugging in the previous expression

$$(1-\beta(1-\delta))(1-\alpha)\frac{S^{1*}}{1-b}\left(1+\hat{\tau}\mathcal{P}\right) = 1-\bar{\omega}$$

and so the result is immediate from comparison with (A16). □

B.5 Proof of Proposition 4

Proof. Equations (A5) and (A17) also hold in the extension with heterogeneous productivity.

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

$$(\alpha-\tau f_i)S_i = \frac{w_i-b}{1-\beta(1-\delta)} + \frac{1}{1-\beta(1-\delta)}\Big(-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big). \tag{A35}$$

Use equation (A23) to add $(1-\alpha)S_i = \frac{p_i-w_i}{1-\beta(1-\delta)}$ on both sides of (A17) 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

$$(1-\tau f_i)\left(p_i-w_i\right) = (1-\alpha)\left(p_i-b\right) + (1-\alpha)\Big[-\beta\lambda\alpha S^1 + \beta\lambda\tau S^2\Big]. \tag{A36}$$

To characterize the term in squared brackets use, from equation (A26),

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

Rewrite as

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

and so

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

Use this to replace the term in squared brackets in (A36) 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.

□

B.6 Proof of Corollary 4

Proof. The proof proceeds in a few steps:

1. First, establish that $\frac{\partial \bar{w}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{w}} < 0$.
2. Second, establish that $\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C}\partial\alpha}\frac{\mathcal{C}}{\bar{w}} > 0$.
3. Combined, these steps imply that $\frac{\partial \bar{w}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{w}}$ becomes smaller in magnitude when α increases.

Establish that $\frac{\partial \bar{w}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{w}} < 0$ To keep notation more compact, it is helpful to note that:

$$\bar{w} = \bar{\omega}(1-b) + b \tag{A37}$$

Then:

$$\frac{\partial \bar{w}}{\partial \mathcal{C}} = (1-b)\frac{\partial \bar{\omega}}{\partial \mathcal{C}}. \tag{A38}$$

And:

$$\frac{\partial \bar{w}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{w}} = \frac{\partial \bar{\omega}}{\partial \mathcal{C}}\frac{(1-b)\mathcal{C}}{\bar{\omega}(1-b)+b}$$
$$= \frac{\partial \bar{\omega}}{\partial \mathcal{C}}\frac{\mathcal{C}}{\bar{\omega}+\frac{b}{1-b}}. \tag{A39}$$

Note that:

$$\frac{\partial \bar{\omega}}{\partial \mathcal{C}} = \frac{-(1-\alpha)\tau[1-\beta(1-\delta)][1-\beta(1-\delta)+\beta\lambda]}{\left(\left[(1-\beta(1-\delta))+\beta\lambda\alpha\right] - \left[1-\beta(1-\delta)+\beta\lambda\right]\tau\mathcal{C}\right)^2} < 0. \tag{A40}$$

Hence, combining (A39), (A40) and the fact that $\frac{\mathcal{C}}{\bar{w}} > 0$, we have

$$\frac{\partial \bar{w}}{\partial \mathcal{C}} \frac{\mathcal{C}}{\bar{w}} < 0. \tag{A41}$$

Establish that $\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C} \partial \alpha} \frac{\mathcal{C}}{\bar{w}} > 0$ Now:

$$\begin{aligned}\frac{\partial \frac{\partial \bar{w}}{\partial \mathcal{C}} \frac{\mathcal{C}}{\bar{w}}}{\partial \alpha} &= \frac{\partial \frac{\partial \bar{\omega}}{\partial \mathcal{C}}}{\partial \alpha} \frac{\mathcal{C}}{\bar{\omega} + \frac{b}{1-b}} - \frac{\partial \bar{\omega}}{\partial \mathcal{C}} \frac{\partial \bar{\omega}}{\partial \alpha} \frac{\mathcal{C}}{(\bar{\omega} + \frac{b}{1-b})^2} \\ &= \frac{\mathcal{C}}{\bar{\omega} + \frac{b}{1-b}} \left(\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C} \partial \alpha} - \frac{\partial \bar{\omega}}{\partial \mathcal{C}} \frac{\partial \bar{\omega}}{\partial \alpha} \frac{1}{(\bar{\omega} + \frac{b}{1-b})} \right).\end{aligned} \tag{A42}$$

Note that:

$$\begin{aligned}\frac{\partial \bar{\omega}}{\partial \alpha} &= \frac{\left(\left[1 - \beta(1-\delta)) + \beta\lambda\alpha \right] - \left[1 - \beta(1-\delta) + \beta\lambda \right] \tau\mathcal{C} - (\alpha - \tau\mathcal{C})\beta\lambda \right) \left[1 - \beta(1-\delta) + \beta\lambda \right]}{\left(\left[(1 - \beta(1-\delta)) + \beta\lambda\alpha \right] - \left[1 - \beta(1-\delta) + \beta\lambda \right] \tau\mathcal{C} \right)^2} \\ &= \frac{(1 - \tau\mathcal{C})[1 - \beta(1-\delta)][1 - \beta(1-\delta) + \beta\lambda]}{\left(\left[(1 - \beta(1-\delta)) + \beta\lambda\alpha \right] - \left[1 - \beta(1-\delta) + \beta\lambda \right] \tau\mathcal{C} \right)^2} > 0.\end{aligned} \tag{A43}$$

Now we can consider the cross-partial:

$$\begin{aligned}\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C} \partial \alpha} =& \left(\left[(1 - \beta(1-\delta)) + \beta\lambda\alpha \right] - \left[1 - \beta(1-\delta) + \beta\lambda \right] \tau\mathcal{C} + 2\beta\lambda(1-\alpha) \right) \\ &\times \frac{\tau[1 - \beta(1-\delta)][1 - \beta(1-\delta) + \beta\lambda]}{\left(\left[(1 - \beta(1-\delta)) + \beta\lambda\alpha \right] - \left[1 - \beta(1-\delta) + \beta\lambda \right] \tau\mathcal{C} \right)^3} > 0.\end{aligned} \tag{A44}$$

Plugging (A40), (A43), and (A44) into (A42), we have that $\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C} \partial \alpha} \frac{\mathcal{C}}{\bar{w}} > 0$.

Conclude Since $\frac{\partial \bar{w}}{\partial \mathcal{C}} \frac{\mathcal{C}}{\bar{w}} < 0$ and $\frac{\partial^2 \bar{\omega}}{\partial \mathcal{C} \partial \alpha} \frac{\mathcal{C}}{\bar{w}} > 0$, we have that when α increases the elasticity decreases in magnitude.

□

C Sample Construction

Data and main sample: Our data source is the Austrian labor market data base (AMDB). Our main sample consists of all workers aged 20-60 that have a regular job in a firm at a reference date (August 1st) in a particular year. Regular jobs are defined as blue- and white-collar jobs that last for at least 30 days and exclude marginal work, apprenticeships, or subsidized work. The sample period is from 1997 to 2015.

Wage data and sample selection based on wages: The wage data contains annual earnings (regular pay plus bonus pay) of each employer-employee relation as well as the number of days in that person-firm-year record. Wages are censored at the social security contribution limit which varies by year. We compute average daily salaries by dividing annual earnings by the number of days worked and convert these to real wages using the consumer price index provided by Statistik Austria with 2000 as base year. The Austrian data does not provide any information on working hours and whether workers are part-time or full-time employed. In order to restrict the analysis to likely full-time workers, we drop all observations with earnings below a minimum daily wage of 32.71 Euros.[21]

Residualizing wages: We regress log wages on functions of the observable characteristics in our data: a fourth degree polynomial in age, a second degree polynomial in tenure at the firm, a gender dummy, and a dummy for Austrian nationality and a set of interactions beween the dummy variables with the polynomials in age and tenure. We then compute residualized wages as the exponential function of the sum of residuals and the predicted value at average characteristics.

In equations, let w_j be the daily wage in levels of worker j. We estimate: $\ln w_j = \beta X_j + \epsilon_j$, where X_j is the vector of observed characteristics. Let $\bar{X}$ be the mean of X_j across the population, $\hat{\beta}$ be the estimated value of β, and $\hat{\epsilon}_j = \ln w_j - \hat{\beta} X_j$. Define $\ln \hat{w}_j^r = \hat{\beta}\bar{X} + \hat{\epsilon}_j$. Then our residualized wage in levels is: $\hat{w}_j^r = exp(\hat{\beta}\bar{X} + \hat{\epsilon}_j.)$.

Defining firm level wages: We use the median of residualized wages in a firm as our main firm-specific wage measure. In equations, we take the median of the $\hat{w}_j^r$ over the j that are employed at firm i. Figure A6 displays the employment-weighted distribution of firm-level wages before (Panel a) and after (Panel b) residualizing.

Additional sample restrictions: In the next Appendix, we describe how we estimate data-driven markets. These can be estimated in the connected set. Table A1 summarizes the order in which we impose sample restrictions and the effect these restrictions have on our sample sizes.

[21] Austria has no universal minimum wage. The vast majority of employers and employees however are covered by collective bargaining contracts, which introduced a monthly minimum wage of 1167 Euros in 2009, equivalent to a daily wage of 32.71 Euro in 2000.

D Data-driven labor markets

We assume that each firm $i = 1, \ldots, N$ in the economy is in one of K labor markets. An $N \times 1$ vector z denotes the assignment of firms to markets with $z_i \in \{1, \ldots, K\}$. We assume that worker flows between firms are driven by the latent markets. In particular, a $K \times K$ matrix M summarizes transition probabilities between labor markets where the typical element $M_{mm'}$ indicates how likely a firm in market m experiences a transition of one of its workers to a firm in market m'.

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

$$E[A_{ij}] = M_{z_i z_j} \gamma_j^+ \gamma_i^-, \tag{A45}$$

where the number of worker transitions from i to j , A_{ij}, depends on the markets of firms i and j, z_i and z_j, the transition probability between these markets, and the firm-level parameters γ_j^+ and γ_i^- which measures the propensity of firm j to hire workers and the propensity of workers to leave firm i.

Based on the observed $N \times N$ matrix of worker transitions between firms, we estimate the parameters of equation (A45) by a computational approximation to maximum likelihood. An important tuning parameter is the number of markets to consider, K. A higher number of labor markets increases the flexibility of the stochastic block model to describe the data where in the limit of $K = N$ each firm represents its own market. This additional flexibility comes with the threat of overfitting.

To guide the trade-off between model complexity and flexibility, we rely on a regularization approach where we pick the number of labor markets to maximize the penalized likelihood of the objective function. In our baseline, we choose parameters by minimizing the description length of the model. The description length is given by the difference between the log-likelihood and the information (entropy) of the model. The log-likelihood of the stochastic block model can be written $\log \mathcal{L} = \sum_{m,m'} E_{mm'} \log \frac{E_{mm'}}{d_m^+ d_{m'}^-}$, where $E_{mm'}$ denotes the number of transitions between markets m and m' and d_m^+ and $d_{m'}^-$ denote the number of incoming links in market m and outgoing links in market m', respectively.[22] The information can be written $\frac{K(K+1)}{2} \log E + N \log K$, where E denotes the total number of worker flows.

Minimizing the description length leads us to 369 labor markets. In a robustness check, we use modularity maximization as an alternative regularization approach, which yields a coarser classification into 9 labor markets. Fixing K, we estimate the partition that maximizes the log-likelihood and then evaluate the different variants according to the modularity score. The modularity score, $Q = \frac{1}{2E} \sum_{ij} (A_{ij} - \frac{d_i d_j}{2E}) \mathbf{1}\{z_i = z_j\}$, compares the share of transitions *within* a market to the share of expected within-market transitions in a null model that keeps the number of links constant for each firm but generates links uniformly at random (ignoring the market structure).

For the purposes of estimating the model, we only use employment-to-employment transitions. An EE transition is defined as a change of the firm with at most 30 days of non-employment between the two spells and at least one year of tenure in the old and the new job. Spurious transitions due to firm renamings, mergers or spin-offs are excluded using cutoffs on worker flows.

[22] For a derivation of this result see, e.g., Nimczik (2018).

E Melitz and Polanec (2015) decomposition of the change in the labor share

We use our model-based measure of productivity and report a Melitz and Polanec (2015) decomposition of the change in the labor share. We report a single decomposition for the whole economy from 1997 to 2015. Hence, in terms of time horizon, our numbers are most closely comparable to Autor et al. (2019, Table A1, Panel C). Specifically, we decompose the change in the aggregate labor share into four terms:

$$\Delta S = \underbrace{\Delta \bar{S}_S}_{\text{unweighted mean of survivors}} + \underbrace{\Delta \left[\sum (f_i^p - \bar{f}^p)(S_i - \bar{S}) \right]_S}_{\text{incumbent reallocation}} + \underbrace{f^p_{X,1997}(S_{S,1997} - S_{X,1997})}_{\text{exit}} + \underbrace{f^p_{E,2015}(S_{E,2015} - S_{S,2015})}_{\text{entry}}, \tag{A46}$$

where Δ represents the change between two time periods (1997 to 2015), $S = \frac{\bar{w}}{p^1}$ is the aggregate labor share, $\bar{w}$ and p^1 are employment-weighted averages of wages and productivity across the whole economy, $\bar{S}$ is the unweighted mean of labor share across firms (i.e., $\bar{S} \equiv \frac{1}{N}\sum_i \frac{w_i}{p_i}$, where N is the number of firms in the economy), f_i^p is the productivity share of the firm in the economy as opposed to market ($f_i^p \equiv \frac{f_i p_i}{\sum_j f_j p_j}$), $\bar{f}^p$ is the unweighed mean of the productivity share (i.e., $\bar{f}^p \equiv \frac{1}{N}\sum_i f_i^p$), S are the survivors (i.e., firms that exist in 1997 and 2015), $X, 1997$ are the exiters (i.e., firms that exist in 1997 and not in 2015), and $E, 2015$ are the entrants (i.e., firms that exist in 2015 and not in 1997). The first term captures the within-firm change in the labor share and the second term captures reallocation among the survivors. The third and fourth terms account for the contribution from exiting and entering firms. Table A3 shows the results.

F Additional Tables and Figures

Table A1: Sample Size and Construction

	Person/year	Firm/year
Total number of observations	2,857,835	236,142
Impose daily wage threshold (32.71 Euro)	2,525,519	192,952
Impose firm-size threshold ($\geq$ 5 employees)	2,290,285	65,285
Restrict to largest connected set	1,858,871	39,827

Notes: This Table reports sample sizes for the year 2015. The total number of observations includes all employment spells of workers aged 20-60 that are present at August 1st 2015 and last for at least 30 days. In the second row, we subtract all spells where the average daily wage for the spell is below a minimum daily wage of 32.71 Euros. In the third row, we subtract spells in firms that employ fewer than 5 employees on August 1st. The fourth row shows the number of observations that are in the largest connected set based on employer-to-employer transitions.

Table A2: Effects of mergers

	Δ HHI		%Δ wages at merging firms		%Δ wages at non-merging firms	
	Average	Median	Average	Median	Average	Median
Baseline (356)	0.046	0.025	-7.9	-3.5	-3.0	-0.6
Alternative market definitions						
Data-driven markets (9)	0.004	0.001	-1.1	-0.6	-0.1	-0.0
NUTS-3 regions (35)	0.004	0.003	-1.4	-1.2	-0.1	-0.1
2-digit industries (79)	0.010	0.005	-2.0	-1.7	-0.4	-0.2
3-digit industries (238)	0.021	0.010	-3.2	-2.1	-0.9	-0.3
4-digit industries (490)	0.034	0.016	-4.9	-2.9	-1.7	-0.5
2-digit industry × region (1539)	0.078	0.050	-10.6	-5.1	-4.6	-1.4
3-digit industry × region (2450)	0.106	0.077	-14.2	-5.9	-7.0	-1.9
4-digit industry × region (3028)	0.123	0.092	-16.4	-7.1	-8.4	-2.6

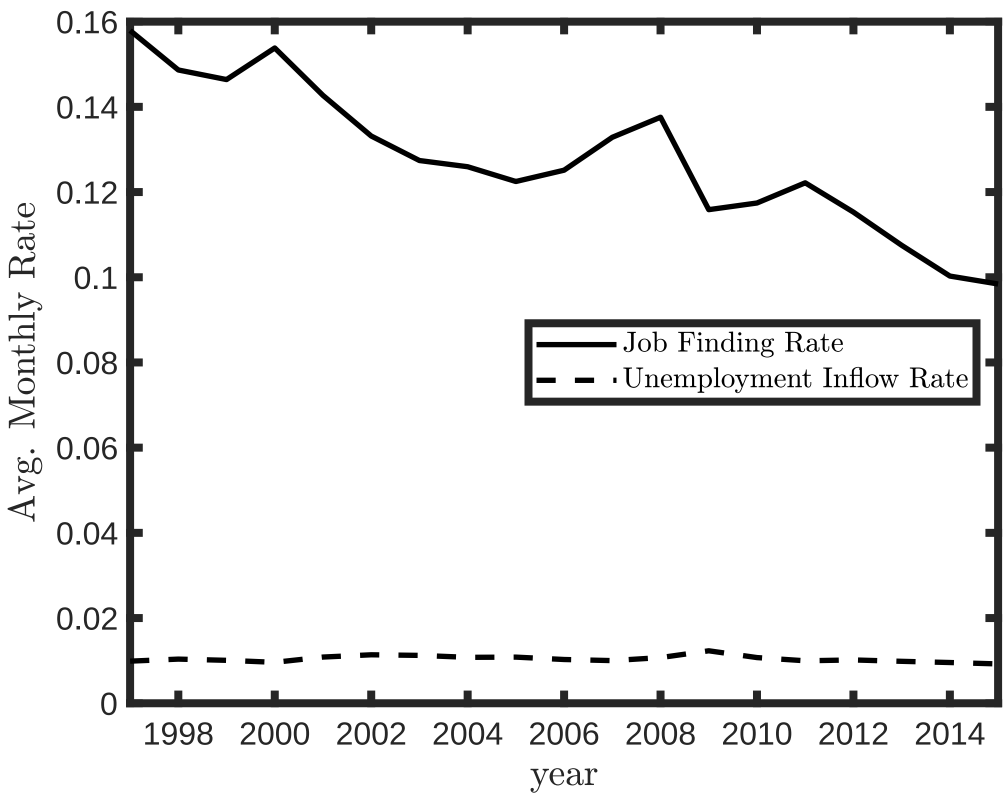
Notes: This table reports sensitivity of the effects of merging the two largest firms in each market to various labor market definitions. We report results for those markets where there are more than two firms and the number of such markets in parentheses. All numbers are employment-weighted statistics across markets.

Table A3: Decomposition of the change in the labor share

	Δ unweighted mean of survivors	Incumbent Reallocation	Exit	Entry	Contribution of	
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline (369)	-2.70	3.49	-2.32	-1.27		
Alternative market definitions						
Data-driven markets (9)	-2.48	-0.04	-0.41	0.13		
NUTS-3 regions (35)	-2.66	0.24	-0.53	0.23		
2-digit industries (80)	-2.53	0.61	-0.77	-0.11		
3-digit industries (255)	-2.07	-0.17	-0.60	0.03		
4-digit industries (538)	-2.61	0.13	-0.82	0.49		
2-digit industry × region (1838)	-3.47	2.27	-2.99	1.37		
3-digit industry × region (3615)	-3.04	1.87	-2.55	0.85		
4-digit industry × region (5384)	-2.58	1.78	-2.66	0.69		

Notes: This Table reports results of decomposing the total change in the labor share between 1997 and 2015 into components using Equation A46. The first column is the unweighted average of labor share changes within surviving firms (that exist both in 1997 and 2015). The second column is the reallocation component that measures the covariance between productivity shares and labor share among survivors. The third and fourth column measure the productivity-weighted contribution of exiting and entering firms. All parts are measured in percentage points.

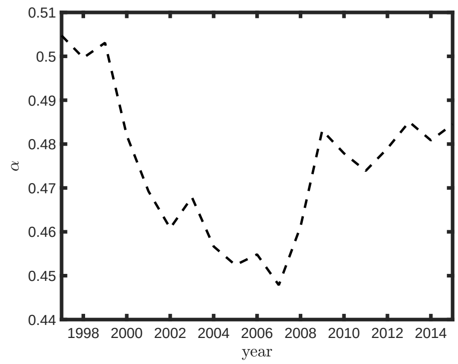
Figure A1: Job finding rate and job destruction rate over time (employment-weighted averages)



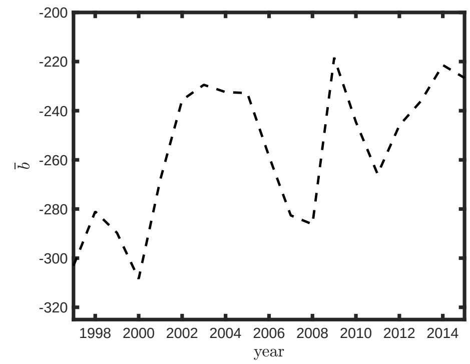
Notes: This figure plots the employment-weighted average of market-specific job finding and job destruction rates over time. For each market m, the yearly rate is an average of twelve monthly rates and the monthly rate is the probability that a worker who is unemployed (employed) on the 1st of a specific month will have a job in market m (be unemployed) on the 1st of the next month. Workers are attached to the market in which they work. Unemployed are attached to the market in which they will eventually find a job.

Figure A2: Worker bargaining power and value of unemployment and over time

(a) Worker bargaining power

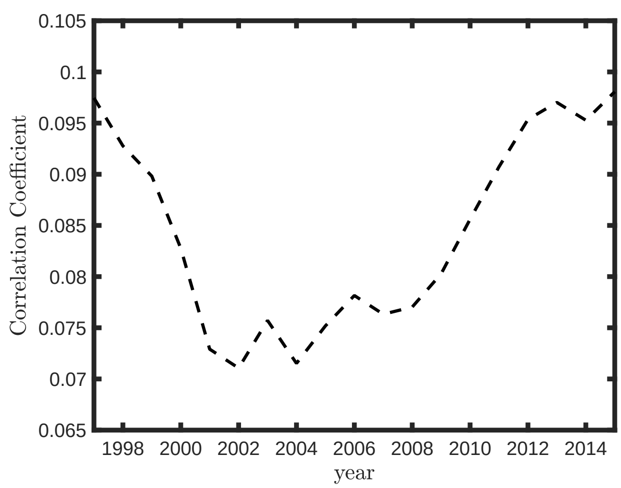


(b) Value of unemployment



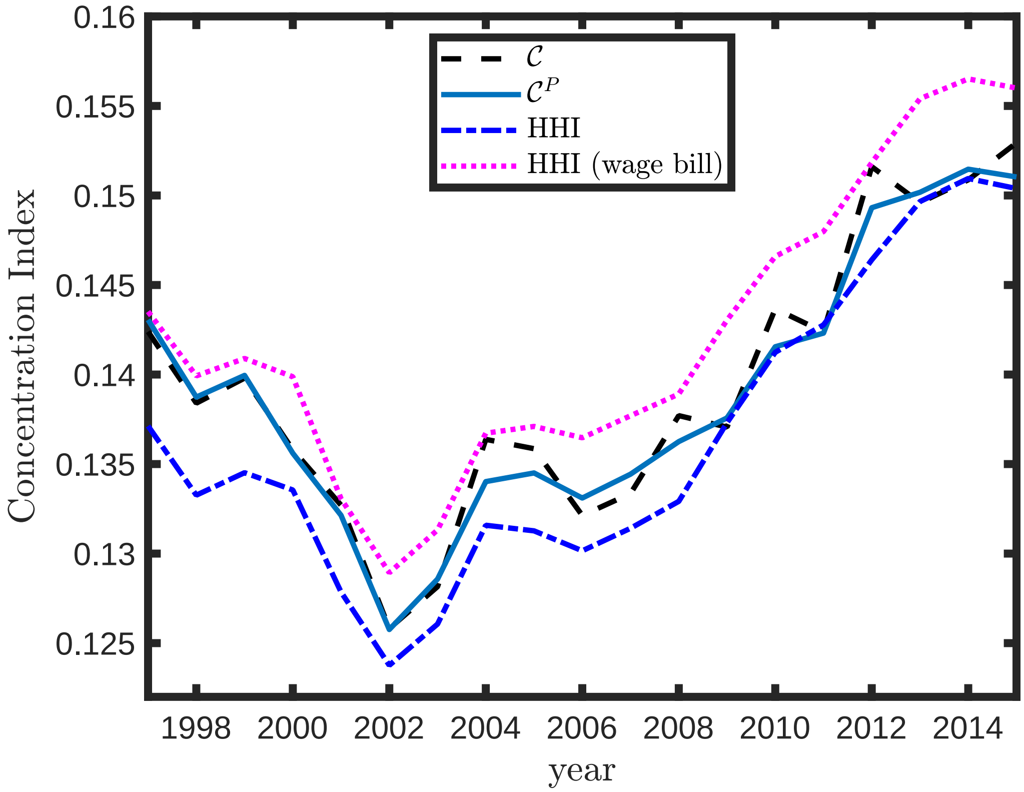
Notes: Panel A of this figure the year-specific values of α that target the labor share from the KLEMS data over time. Panel B plots the employment-weighted average of market specific parameters b over time. For each market m, the parameter b is chosen such that the lowest observed wage in the market equals to the reservation wage.

Figure A3: Size-wage gradient



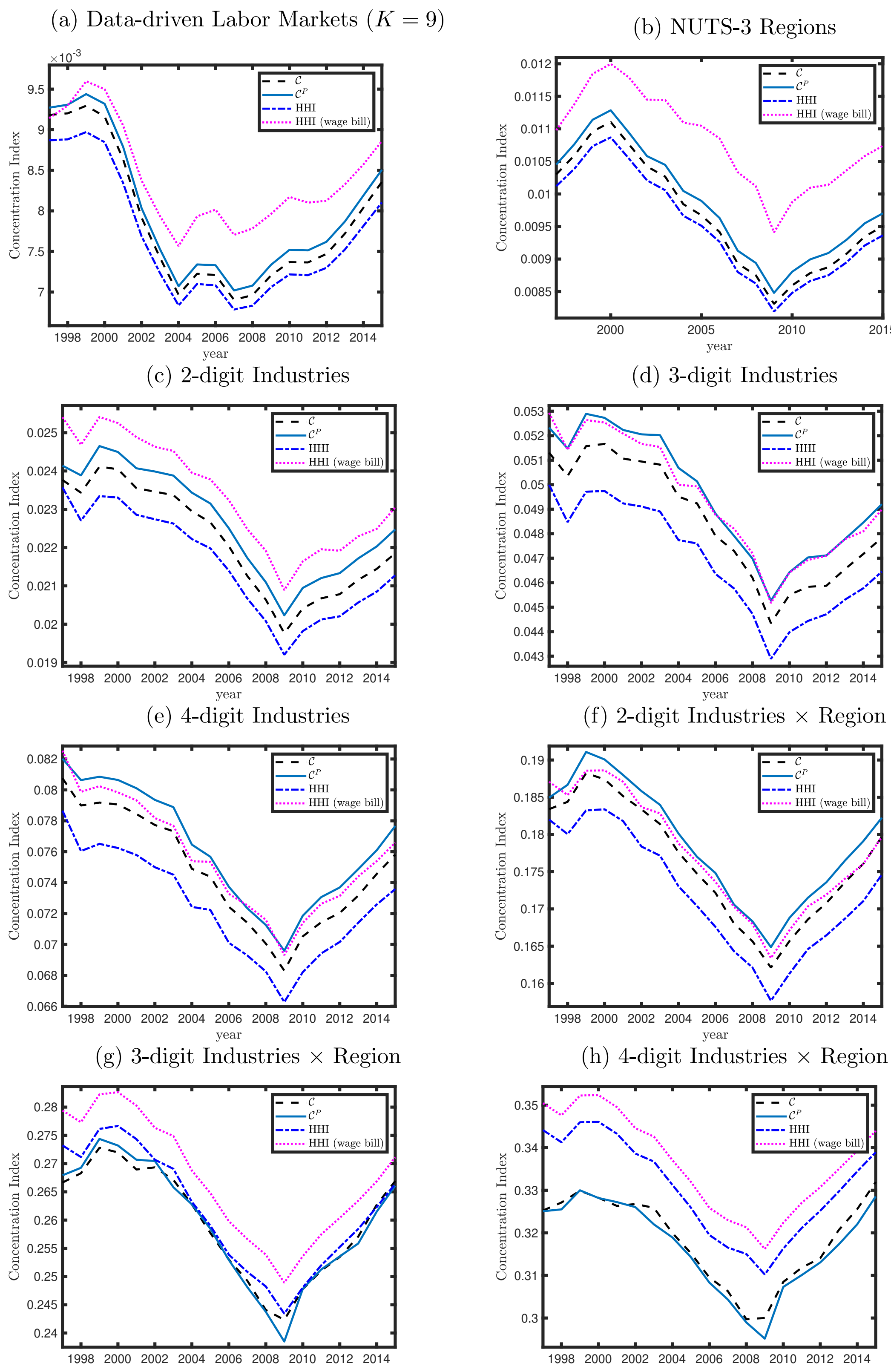
Notes: This figure plots the (employment-weighted average) of the correlation between firm size and firm-level median wages. Firm size is measured at a reference date (August 1st) each year and wages are the median of the firm-level distribution of regular employee wages. The figure displays employment-weighted averages over all 369 data-driven labor markets.

Figure A4: Trends in Labor Market Concentration (unweighted)



Notes: This figure plots concentration indexes $\mathcal{C}$, $\mathcal{C}^P$, HHI and wage-bill HHI from 1997 - 2015 based on micro data from the Austrian labor market database. The figure displays raw averages over all 369 data-driven labor markets.

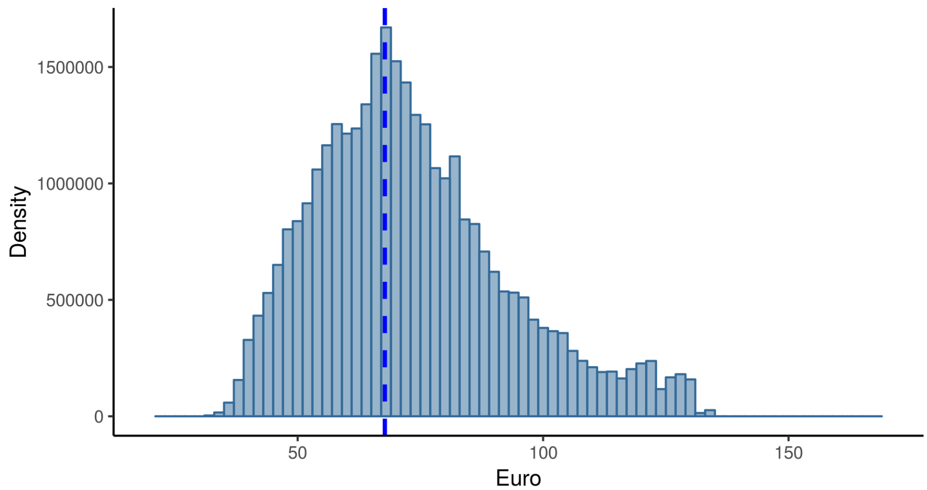
Figure A5: Trends in Labor Market Concentration – Different Labor Market Definitions



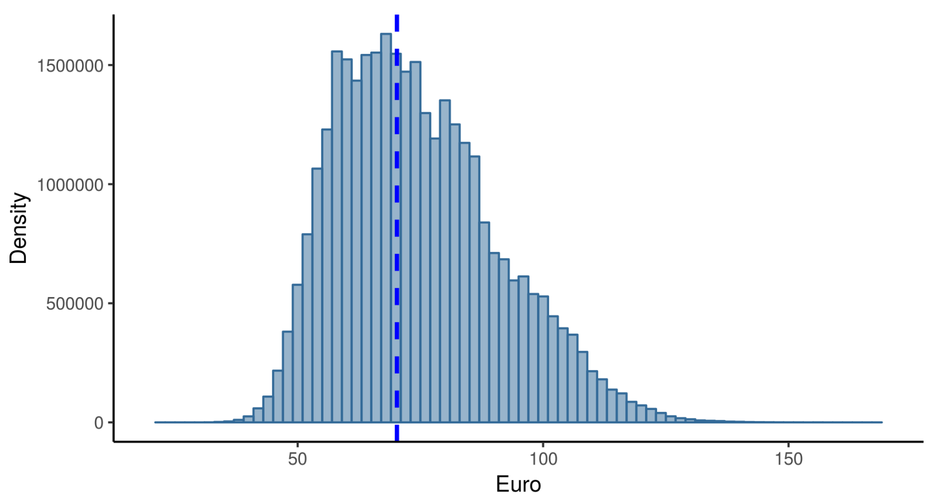
Notes: This figure plots concentration indexes $\mathcal{C}$, $\mathcal{C}^P$, HHI and wage-bill HHI from 1997 - 2015 based on micro data from the Austrian labor market database. The figure displays employment-weighted averages over all markets for various labor market definitions.

Figure A6: Distribution of Firm-level Median Wages

(a) Real Wages



(b) Residualized Real Wages



Notes: This figure plots the (employment-weighted) distribution and mean of firm-level median wages in real Euros where the base year for the Austrian CPI is 2000 and where we pool over all years in the sample period. Panel (a) shows actual median wages while Panel (b) shows wages after residualizing.