Labor Markets, Wage Inequality, and Hiring Selection*

Alessandra Pizzo
Université Paris 8, LED

Benjamín Villena-Roldán
Universidad Andres Bello & LM^{2}C^2 & MIPP

September 3, 2025

Abstract

Employers hire more selectively between heterogeneous productivity workers when applicants' queues are longer. Consistently, CPS data reveal a positive and concave relation between unemployment rates and wage inequality. We rationalize intuition and evidence altogether using a nonsequential search model of selective hiring that jointly determines worker flows and the wage distribution. Labor market selectivity on top of search frictions amplifies pre-match inequality at the top of the wage distribution and compresses it at the lower end, generating a typical right skewness of empirical wage distributions. Using GMM-estimated parameters, we show that mean worker productivity distribution shifts are consistent with the evidence. Welfare analysis suggests that progressively taxing better matches may enhance efficiency because it offsets excessive vacancy postings that end up hiring less productive workers. However, under higher screening costs, the optimal taxation is regressive, and top matches generate a trickle-down of jobs for less productive workers.

Keywords: Nonsequential search, Hiring, Inequality, Unemployment, Worker Flows, Efficiency.

*Email: alessandra.pizzo@univ-paris8.fr and benjamin.villena@unab.cl. We are grateful to colleagues at FAU; U. Carlos III; 2024 SAET; 2023 SED; 2022 LACEA-LAMES; 2022 EEA; Le Mans University; Catholic University of Chile; Alberto Hurtado University; 2019 Annual Search and Matching workshop; University of Bristol; and Central Bank of Chile for useful comments and discussions at different stages of this project. Villena-Roldán is grateful for financial support from FONDECYT project 1251885; ANID Milenio Labor Market Mismatch – Causes and Consequences, LM^{2}C^2 (NCS2022_045); and the Institute for Research in Market Imperfections and Public Policy, MIPP (ICS13_002 ANID). All remaining errors are ours.

1 Introduction

Businesses today understand that success importantly hinges on finding the right people. Steve Jobs famous quote "The secret of my success is that we have gone to exceptional lengths to hire the best people in the world" and Jim Collins's influential idea of "getting the right people on the bus" reflect a basic truth of the job market: the most capable candidates often land the jobs, and companies frequently aim to attract top talent from competitors. The amount of resources spent on screening the most suitable candidates show that securing quality talent is vital for employers. Beyond anecdotes, personnel economics shows that employer need screening workers since there are large differences in productivity or output in the same job at the same firm (Hoffman and Stanton 2024, and references therein). Moreover, candidate assessments at the application stage consistently predict worker performance (Sandvik et al. 2024) and retention (Hoffman et al. 2018).

This paper formally explores the implications of recruiting selection in shaping macroeconomic labor outcomes, particularly wage inequality and worker flows. When unemployment is high and more applicants are available, the hiring selection becomes more selective and widens wage gaps, especially at the higher end of the pay scale. We see clear evidence for this in the Current Population Survey (CPS): as unemployment rises, so does wage inequality, often in a concave pattern. In addition, top earners see their wages spread out and the lower earners' wages become more compressed, as in previous findings (Autor et al. 2006; Autor et al. 2008; Bonhomme and Hospido 2017). We also observe that high wage inequality often goes hand-in-hand with lower job finding and job-to-job transitions (for non-college workers).

To capture these intuitions, we introduce hiring selectivity into a standard search framework in which firms simultaneously meet several heterogenous applicants and choose the best candidate, i.e. a nonsequential process with costly employer screening. This ingredient is key to understanding how individuals' productivities affects their wages and labor market transitions. The hiring process endogenously generates job-finding and job-to-job transition probabilities that increase in worker's productivity. The pool of applicants includes both unemployed and employed workers who apply and compete for new jobs, a feature similar to (Eeckhout and Lindenlaub 2019). In the competitive equilibrium, employers do not take into account that they worsen the average composition of applicants when hiring the top candidates and dismissing the less attractive ones. A key model's prediction is that hiring selectivity in equilibrium amplifies inequality at the right tail of the wage distribution and dampens it at the left tail. This occurs because selective hiring allows employers to pick the most productive workers, making the composition of the employed pool better at the top and worsening it at the bottom. This process generates a wage distribution for employed workers that is more skewed to the right than the underlying population's productivity distribution, which would be the wage distribution in a Walrasian frictionless economy. This result that aligns with common empirical observations of right-skewed wage distributions.

We take the model to the data by calibrating the average unemployment rate for the US, as well as the frequencies of job finding, job-to-job transitions and of separations, and the wage distribution during 1994–2019. Since we do not try to match the observed correlations between inequality and worker flow measures per se, we investigate —for instance— what type of changes of the exogenous parameters can replicate a positive comovement between the unemployment rate and wage inequality. We find that shifts in the productivity distribution that increase the mean can replicate the observed empirical patterns. For college workers, these shifts appear to preserve the quantile ratios of the productivity distribution, suggesting a uniform change across the spectrum. In contrast, for non-college workers, shifts affecting the left tail of the distribution while increasing the mean are more promising in offering a potential explanation for observed differences. Therefore, without explicitly targeting it, our model can replicate the relationship between unemployment and inequality overall, at the bottom, and at the top of the US wage distribution (percentile gaps 90/10, 50/10, and 90/50, respectively). Plain average productivity growth predicts increasing inequality, particularly at the right tail of the distribution, facts that have traditionally thought associated to skill biased technological change (Autor et al. 2006; Autor et al. 2008).

What are the key mechanisms at play? The first intuition goes as in standard search and matching models: an increase in the average productivity of workers spurs more creation of vacancies, making it easier to find jobs, and reducing the unemployment rate. A second point is the effect that screening has on labor market composition. As the number of applicants per vacancy declines, screening becomes less selective, and thus the composition of the pool of the unemployed *improves* its average productivity. The employed pool, on the other hand, *worsens*. Hence, the pool of applicants, a mixture of employed and unemployed workers, becomes more similar to the population's distribution of productivities. Therefore, hired applicants tend to be more similar to each other, reducing inequality. A third point is an additional general equilibrium effect: the variance reduction in the distribution of applicants leads to a lower likelihood of hiring top applicants. This offsets the increased average productivity and deters, to some extent, the posting of vacancies. In the end, the general equilibrium adjustment partially undoes the negative initial impact of higher productivity on both unemployment and inequality.

Finally, we explore policy implications of hiring selectivity by studying the social planner's problem in our economy. The first lesson is that a small but positive unemployment is critical for improving welfare. In the model unemployment is essential for improving worker allocations via selection, not just a symptom of unavoidable search frictions dampening efficiency. The optimal allocation takes into account that making the most productive workers employed requires substantial effort in screening applicants. Under the baseline calibration, we show that the optimal unemployment rate for both non-college and college workers is higher than the market solution.

We also examine the profit tax schedule required to implement the social planner's solution under a balanced budget. This schedule must satisfy a coincidence-ranking equilibrium constraint: after-tax profits must be strictly increasing with the worker's type to ensure that employers' preferences for workers within an application pool remain unaffected by taxation. Our analysis finds that the efficiency-restoring tax and transfer schedule is progressive: the social planner disincentivizes the hiring of highly productive workers to curb excessive vacancy creation, thereby increasing unemployment to its efficient level.

However, this result is sensitive to the calibration of recruitment costs—specifically, the balance between vacancy posting and applicant screening costs. When screening costs are high relative to posting costs, long applicant queues become too expensive. Consequently, the socially optimal unemployment rate falls below the market outcome. In this scenario, the optimal policy becomes regressive. The economic intuition is that high-productivity individuals create a positive externality by encouraging firms to post numerous vacancies. This ex-ante incentive ultimately improves job prospects for all applicants ex-post.

Taking stock, our model provides a theoretical framework integrating recruiting selection process, a realistic feature of labor market, into a general equilibrium search and matching framework. The competitive equilibrium economy implies a quantile mapping between ex ante productivities and observed wages. The model explains the commonly observed right-skewness in the wage distribution. Moreover, the calibrated version of the model can replicate the relationship between unemployment rate and inequality in CPS data, as well as the changes in the low and high ends of the wage distribution. From a normative perspective, we learn that some unemployment is essential for selective labor employers to screen the labor market and achieve efficient allocations. Although other sources of wage inequality have been explored in the literature, the predominance of worker-specific components to account for wage variation is close to 50% in the US according to Song et al. (2019) and could be even higher[1]

[1] Bonhomme et al. (2023, Table F, online appendix) survey several studies using the Abowd et al. (1999) frame-

Literature review: Our paper is linked to several distinct, yet interconnected, strands of the literature.

We first contribute to the literature of *nonsequential search*, an approach started by Stigler (1961) seminal paper, to formally model hiring selectivity. Within nonsequential models, Blanchard and Diamond (1994) study an economy firms comparing candidates by unemployment duration, and (Moen 1999) study the link between productivity and education. Some literature considers ranking in a directed search approach (Shimer 2005; Shi 2006; Fernández-Blanco and Preugschat 2018; Cai, Gautier, and Wolthoff 2025). Wolthoff (2017) is the most closely related paper studying recruiting decisions in a directed search model with ex ante homogeneous workers. We introduce ex ante heterogeneous workers and focus on how selective hiring amplifies or attenuates the pre-market productivity inequality along the wage distribution. Some aspects of our paper are similar to Villena-Roldan (2010a), Villena-Roldan (2010b) but they neglect on-the-job search and do not characterize the competitive equilibrium nor the efficient allocation.

Previous work portrays hiring as a selection or nonsequential search process within a pool of workers (van Ours and Ridder 1992; van Ommeren and Russo 2013; Davis and Samaniego de la Parra 2024). Often neglected in search models, hiring selection has been an important theme for human resource management and the personnel economic literature for good reasons. First, screening applicants is highly rewarding in practice: the productivity dispersion for the same position is so substantial that can outweigh the differences in pay for the same job for diverse workers: teachers, physicians, and supermarket employees, (Mas and Moretti 2009; Hoffman and Stanton 2024). Moreover, candidate assessments at the application stage are highly correlated with worker performance and retention (Hoffman et al. 2018; Sandvik et al. 2024).

Second, our work connects to recent papers focusing on the composition of employed and unemployed pools, particularly with on-the-job search. Eeckhout and Lindenlaub (2019) show how vacancy creation and on-the-job search complementarity affect cyclical composition. (Engbom 2021) and Bradley (2020) also consider settings with screening and on-the-job search, highlighting how recruiting costs influence the pool of applicants. Unlike our model, these papers focus on ex ante homogeneous workers facing match-specific shocks. Merkl and van Rens (2019) study an economy with heterogeneous training costs showing its efficiency implications..

Third, we contribute to the established literature assessing the relationship between in-

work reporting that only a cross-study average of 29% of the log wage variance (19% for US studies) is accounted for firm and sorting.

come or labor income inequality and unemployment. For instance, (Jäntti 1994) and (Mocan 1999) found non-linear relationships and confirmed that unemployment disproportionately affects lower income quintiles while benefiting the high end of the wage distribution. In line with our empirical findings, (Saldarriaga 2022) shows countercyclical wage inequality with widening at the top and contraction at the bottom during recessions. His work extends a sequential search framework with endogenous hiring standards to explain these observations, contrasting with our hiring selectivity approach. Our paper also distinguishes itself from papers that show that a sequential search model with an exogenous wage offer distribution (Cysne 2009) and a wage-posting model with on-the-job search (Cysne and Turchick 2012) can generate a positive correlation between unemployment and the Gini index. Their results hold in partial equilibrium: either the wage offer distribution is exogenous, or the job offer arrival rate is fixed, implying an exogenously given unemployment rate. Instead, in our setup, the unemployment rate and the wage distribution are jointly determined by introducing hiring selectivity. Hence, in our model we can assess the effect of changes in productivity, recruting costs, and vacancy-posting costs on unemployment and inequality in equilibrium. Our model also predicts the impact of exogenous factors on any ratios of wage percentiles, with nuanced implications on the overall shape of wage distribution, not just the Gini index..

Finally, our work relates to the broader literature on the evolution of inequality along the whole wage distribution as in (Autor et al. 2006; Autor et al. 2008). While factors like increased education, skill-biased technological change, globalization, and taxation are often cited as principal drivers of rising inequality (e.g., Krusell et al. 2000; Card and DiNardo 2002; Moore and Ranjan 2005; Goldin and Katz 2008; Ravallion 2018.), our theory complements these by explaining how the labor market's inherent frictions and hiring selectivity can generate higher inequality at the top and lower at the bottom of the wage distribution through an even neutral productivity increase. The model shows how the selectivity prevalent in hiring and poaching interacts with pre-market productivity inequality, providing an explanation for empirical correlations between inequality measures and worker flows, and the effect of selective employer activity on the shape of the wage distribution.

2 Motivating Facts

In this section, we document significant relations between unemployment and wage inequality, as well as between worker flows and wage inequality. We are *not* claiming any causal effects. Our preferred interpretation is that worker flows cause changes in wage inequality and we will provide a model that provides a plausible mechanism that explains the facts. In this section, however, our purpose is to document correlations and partial correlations that may be interpreted later by using a structural model.

Using CPS monthly files made available by IPUMS-CPS (Flood et al. 2022), we construct unemployment rates (U) and unemployment-to-employment transitions (UE) for the period 1994–2019 by state and year.[2] We also construct job-to-job (JJ) transition rates using the standard method in Fallick and Fleischman (2004) by state, year, and college status. We use hourly wages from the Outgoing Rotation Group (ORG), also available in IPUMS-CPS and deflate them using national CPI.[3] By state, year, and college cells, we compute measures of wage inequality such as the 90–10 percentile gap for log hourly real wages (G9010), as well as other gaps (75–25, 90–50, 50–10, etc.) We focus on these specific measures because our theoretical model makes precise predictions about them. However, we also estimate other inequality statistics in the data and simulate them computationally in the model, finding results consistent with those obtained for percentile log wage gaps. For the sake of brevity, we report these results in the Online Appendix F.

Our measures have time and geographical cross-sectional variation for two levels of skill or education of the worker: with or without a college degree. Our idea is to group workers who compete for similar jobs in the same market (state) in a specific year. While workers may have also being grouped by occupation, there is substantial occupational mobility in the CPS even for 1-digit codes: Moscarini and Thomsson (2007) report a 3.2% monthly occupational transition rate at the 3-digit level and approximately 2.2% at the 1-digit level for 2004–2005 (4.2% and 3.7% for 1983–1986). Kambourov and Manovskii (2009) obtain comparable results using the PSID. Our split according to college status intends to generate coarse categories in which job transitions between them is likely to be rare. Suggestive evidence for the CPS shows that for a majority of non-college occupations, less than 1% have a college degree (Gottschalk and Hansen 2003).

We mainly present the results for the log of the 90–10 wage percentile ratio, although we do not belittle the relevance of other measures: we will, in fact, show that different percentile gaps can vary in a different way when the the unemployment rate changes.

Figures 1–4 show the fitted values of a quadratic regression between the 90–10 log hourly real wage percentile gap (G9010) by state and year on a number of flow or stock variables, one at a time. The figures also depict the 95% confidence intervals for the fitted values. The subfigures on the left show fitted values with no controls, while the ones on the right include state fixed effects. The specification with no controls portrays a relation between the

[2] We also compute transitions from 1976-2019 and the results do not change qualitatively.

[3] All items in U.S. city average, all urban consumers, not seasonally adjusted, series CUUR0000SA0. Of course, for inequality measures based on log wages, this adjustment is irrelevant.

90–10 percentile log wage gap and the unemployment rate that can be associated with either national-level time-varying factors such as business cycles and technological trends; state time-invariant differences; or state-level time-varying factors. We are agnostic regarding the nature of the underlying source driving the relation. We also include the figures on the right in which we control for state fixed effects. Therefore, we show that the underlying shocks shaping the positive (and concave for non-college) relation between G9010 and the unemployment rate are not solely state-specific: time-varying factors play an important role.

The relation between G9010 and the unemployment rate (U) is mostly increasing but flattens out near 7% for non-college workers in the panels of Figure 1. The pattern seems quite similar when controlling for state fixed effects, suggesting that time variability is driving the pattern to a large extent. For college workers, the relation between G9010 and the unemployment rate is also positive, but mostly in a linear way, although the range of unemployment rates covered is narrower.

Figures 2–4 try to dig deeper into the previous relation, as flows in and out of unemployment could shape the unemployment rate as in a standard two-state search and matching model (Pissarides 2000). The job finding frequency (UE)[4] is mostly negatively correlated with wage dispersion, showing that wage inequality tends to increase in times of long unemployment duration, which remains true after controlling for state fixed effects. The sensitivity of wage inequality has a similar magnitude for both college and non-college workers, although for the latter there is greater concavity.

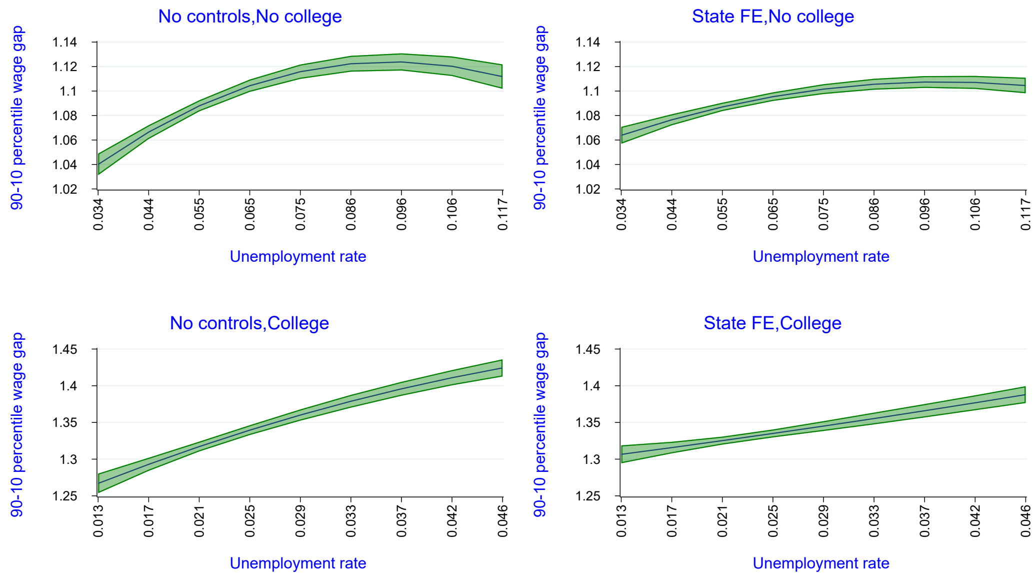
In the case of the separation frequency (EU flow), we observe mostly positive correlations. The sensitivity remains if we control for state fixed effects. Both job finding and separation frequencies are associated to wage dispersion in a way that is consistent with the correlation between wage dispersion and unemployment. Taken together, these pieces suggests a clear linkage between worker flows and inequality.

As occurs with the job finding frequency (UE), for college workers there is a negative relation between job-to-job frequency transitions and inequality, although the slope of the relation is flatter than the one for the UE flow. For non-college workers, the slope is, in contrast, mainly positive.

We perform a number of robustness checks to be sure these relations are meaningful and unrelated to mechanical biases. We refer the reader to the relevant figures in the Online Appendix F. In particular, we checked that

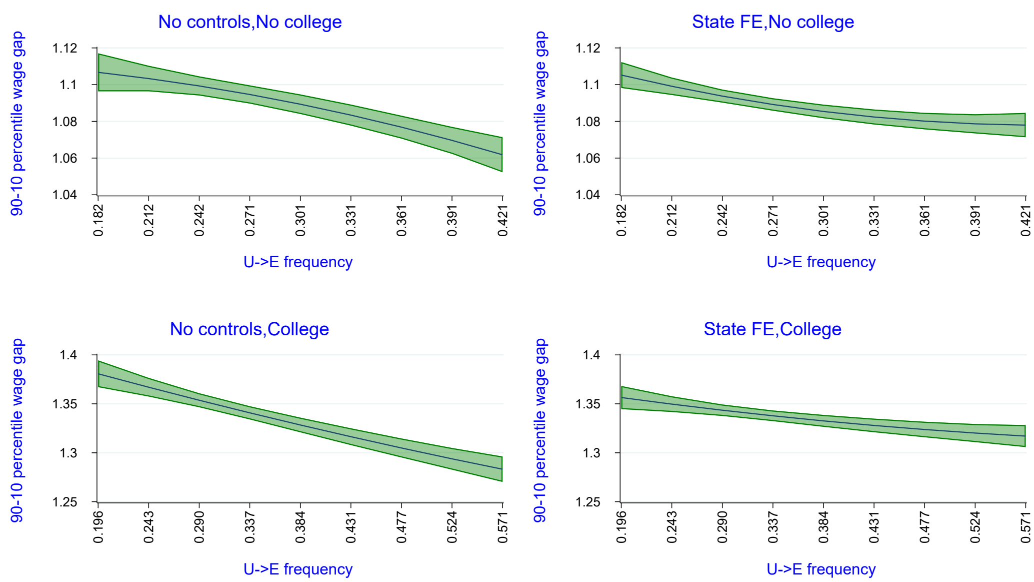
[4]Worker flow transitions cannot be computed for individuals whose month-in-sample is 1 or 5 in the CPS sampling design, which contrasts with the computation of the unemployment rate, which covers all individuals in the sample.

Figure 1: 90–10 log wage percentile gap vs. unemployment rate by state & year



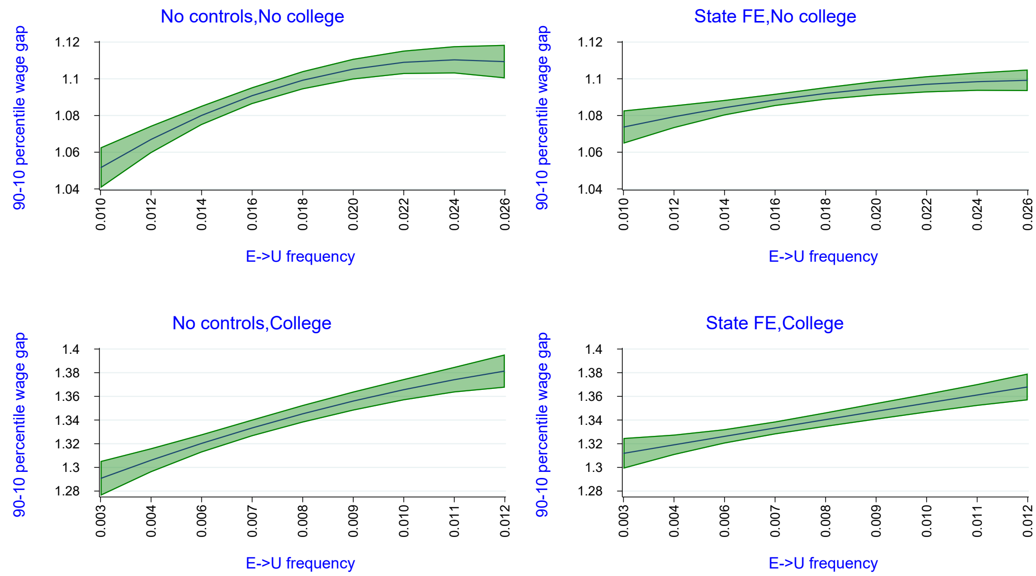
Sample 1994-2019. US real log hourly wages.

Figure 2: 90–10 log wage percentile gap vs. job finding freq. (UE) by state & year



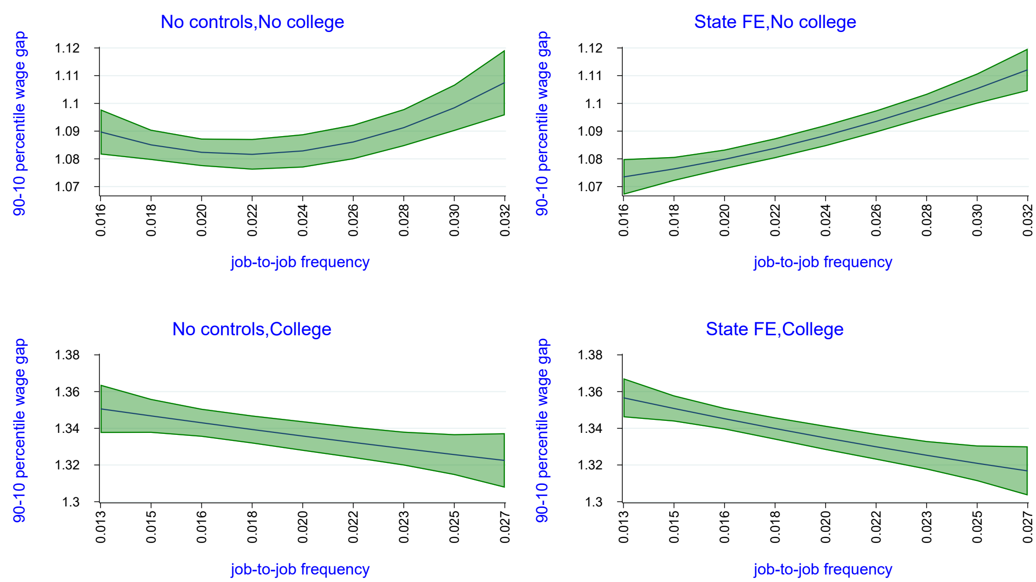
Sample 1994-2019. US real log hourly wages.

Figure 3: 90–10 log wage percentile gap vs. separation freq. (EU) by state & year



Sample 1994-2019. US real log hourly wages.

Figure 4: 90–10 log wage percentile gap vs. job-to-job freq. (JJ) by state & year



Sample 1994-2019. US real log hourly wages.

- The 1994 CPS sampling redesign does not substantially change the observed patterns in the relation between the log wage dispersion and the unemployment rate, nor the relation between the wage dispersion and the EU and UE flow frequencies. These can be seen in Figures 17, 18, and 19.

- The findings generally remain unaltered for alternative measures of wage dispersion, such as the 75–25 percentile wage gaps and the standard deviation of log wage, as can be seen in Figures 20–27. While the former is less sensitive to outliers, the latter is more commonly used.

- We introduced controls for year fixed effects only in Figures 28, 29, 30, and 31, obtaining qualitatively similar results. This shows that the underlying mechanism generating the shapes in the previous evidence may be pertinent to explaining the cross-sectional diversity of inequality across different states.

Our preferred explanation for the relationship between wage inequality and unemployment assumes that employers hire selectively, comparing applicants and offering jobs to the most suitable candidates. When the unemployment rate rises, larger pools of job applicants allow employers to hire better-qualified workers on average, all else being equal. As a result, jobs are allocated to more productive workers or better matches, improving the overall composition of employed workers and widening the gap with previously employed workers. However, as unemployment continues to increase, the employed workforce becomes more productive but less diverse, reducing inequality. This explains why the increase in inequality associated with unemployment diminishes in higher-unemployment markets, resulting in a concave relationship for non-college data. At sufficiently high unemployment rates, selectivity becomes so pronounced that the hired workers are nearly homogeneous, potentially leading to a decrease in wage inequality.

3 The Model

In the model, time is discrete. There is a continuum of homogeneous risk-neutral firms or employers that can post ex ante identical job vacancies. There is also a fixed mass of size 1 of workers, who are characterized by a time-invariant productivity θ according to an exogenous distribution with density $g(\theta)$. Workers can either be employed or unemployed. Employed workers apply to a position with a fixed exogenous probability λ. The mass of applicants $\mathcal{A}$ is given by the sum of the unemployed and the proportion of employed who apply for a job:

$\mathcal{A} = \mathcal{U} + \lambda(1 - \mathcal{U})$. Jobs are destroyed with exogenous probability η. Workers cannot borrow or save.

The general state of the economy is a tuple $\mathcal{X} \equiv (\mathcal{A}, \mathcal{V}, G_A(\theta))$, where $\mathcal{A}$ represents the mass of applicants in the economy, $\mathcal{V}$ is the mass of aggregate vacancies posted, and $G_A(\theta)$ is the endogenous joint distribution of types of the applicants, including unemployed and employed jobseekers. While the setting could be extended to a dynamic framework, in this paper we solely focus on the steady-state symmetric equilibrium of the economy.

3.1 Matching and Job Finding Rate

Since all jobs are ex ante identical, workers randomly apply to vacancies. Thus, the probability that a particular worker applies for a given vacancy is $1/\mathcal{V}$ and the number of applications for a vacancy, K, follows a binomial distribution:

$$Prob(K = k) = \binom{\mathcal{A}}{k} (1/\mathcal{V})^k (1 - 1/\mathcal{V})^{\mathcal{U}-k}$$

As both $\mathcal{A}, \mathcal{V} \to \infty$ with its ratio $q = \mathcal{A}/\mathcal{V}$ constant, the number of applicants per vacancy K converges to a Poisson distribution with mean q, what from now on we refer to as the queue length.

A worker is hired whenever they generate a profit for the employer which is greater than that of the other applicants for the same vacancy. To ease the exposition, we assume for now that the ranking of productivity of the applicants portrayed by the cumulative distribution function of the applicants, $G_A(\theta)$ is the same as the ranking of profits generated in the population of applicants, i.e., a Coincidence Ranking Equilibrium (CRE) holds. This is not guaranteed to actually be the case in any equilibrium, since wage determination may give a high weight to the worker's outside option, making high productivity types less attractive to employers. However, under our simple wage-setting mechanism, to be explained later, the applicant with the highest productivity yields the highest profit in equilibrium.

Under these assumptions, if an employer screens s applicants, the top candidate gets the offer with probability $(\phi G_A(\theta))^{s-1}$, where ϕ is the probability of interviewing an applicant, ex ante decided by the employers in a symmetric equilibrium.

The total number of screened applicants follows a binomial distribution. If k applicants apply for a job, the probability that a worker of type θ will be hired is

$$Prob(\theta \text{ hired}|k \text{ total applicants}) = \sum_{s=1}^{k} \binom{k}{s} (\phi G_A(\theta))^{s-1}(1-\phi)^{k-s+1} = (\phi G_A(\theta) + 1 - \phi)^{k-1}$$

Nevertheless, when workers apply, they ignore how many applicants are competing for the same job they applied to. Assuming that the number of applicants follow a Poisson distribution, the probability of being hired $p(G_A(\theta), q)$ is a Poisson–Binomial mixture.

$$\tilde{p}(G_A(\theta), q) = \sum_{k=1}^{\infty} \frac{e^{-q} q^{k-1}}{(k-1)!} \left(\phi G_A(\theta) + 1 - \phi\right)^{k-1} = e^{-\phi q(1-G_A(\theta))} \tag{1}$$

Henceforth, we take $\phi = 1$ to ease our exposition and defer the explanation of the irrelevance of this assumption. Intuitively, the expected number of interviews ϕq is the relevant variable for both employers and applicants, regardless of the specific values ϕ and q take. Then, we can write the probability of being hired as a function of q and the applicant's ranking $x = G_A(\theta)$:

$$p(x) = e^{-q(1-x)} \tag{2}$$

The average probability of an applicant's being hired, $\mathbb{E}[\tilde{p}(\theta, q)|q]$ is therefore given by

$$\mathbb{E}[p(G_A(\theta), q)|q] = \overline{p}_A = \int \sum_{k=1}^{\infty} \frac{e^{-q} q^{k-1}}{(k-1)!} \left(\phi G_A(\theta) + 1 - \phi\right)^{k-1} dG_A(\theta) = \frac{1 - e^{-q}}{q} \tag{3}$$

3.2 Distributions

The recruiting selection process affects the distribution of unemployed and employed workers. In this section we show how this distribution is endogenously determined in steady state.

First, the exogenous density of types is a weighted average of the densities of the unemployed g_U and of the employed g_E, as follows: $g(\theta) = \mathcal{U} g_U(\theta) + (1 - \mathcal{U}) g_E(\theta)$.

In steady state, for a given type θ and queue length q, the flows in and out of unemployment must be equal, i.e., $\tilde{p}(G_A(\theta), q) g_U(\theta) = \eta^* g_E(\theta)$ where $\eta^* \equiv \eta(1 - \lambda)$ is the effective separation rate once on-the-job searchers skip the separation shock as we assume that an on-the-job application occurs first than a separation shock within a period.

Combining both equations, we obtain that the population density of the type θ is weighted by its steady-state probability of unemployment, and scaled by the mass of the unemployed to ensure the expression integrates to 1.

$$g_U(\theta) = \frac{g(\theta)}{\mathcal{U}} \frac{\eta^*}{\eta^* + \tilde{p}(G_A(\theta), q)} \tag{4}$$

In a similar fashion, we can obtain the density of employed workers as

$$g_E(\theta) = \frac{g(\theta)}{1 - \mathcal{U}} \frac{\tilde{p}(G_A(\theta), q)}{\eta^* + \tilde{p}(G_A(\theta), q)} \tag{5}$$

Since a fraction λ of the employed workers apply for jobs just as the unemployed do, the density of the applicants is $g_A(\theta) = \frac{g_U(\theta)\mathcal{U} + g_E(\theta)\lambda(1-\mathcal{U})}{\mathcal{U}+\lambda(1-\mathcal{U})}$, which turns out to be a separable differential equation

$$\frac{dG_A(\theta)}{d\theta} = \frac{\eta^* + \lambda\tilde{p}(G_A(\theta), q)}{\eta^* + \tilde{p}(G_A(\theta), q)} \frac{1}{\mathcal{U} + \lambda(1-\mathcal{U})} g(\theta) \tag{6}$$

We make a change of variable and define the quantile of the distribution of applicants as $G_A(\theta) = x$. Since $G_A(\theta)$ is a cumulative distribution function, the boundary conditions are $G_A(\infty) = G(\infty) = 1$ and $G_A(0) = G(0) = 0$.

Using the boundary conditions, we can determine the value of the constant of integration and therefore the expression for the mass of the applicants, $\mathcal{A}$, and the unemployment rate, $\mathcal{U}$.[5]

$$\mathcal{A} = \frac{1}{1 + \frac{1-\lambda}{\lambda q} \log\left(\frac{\eta^*+\lambda}{\eta^*+\lambda e^{-q}}\right)} \tag{7}$$

$$\mathcal{U} = \frac{\mathcal{A} - \lambda}{1-\lambda} = 1 - \frac{1}{q}\log\left(\frac{\eta^* + \lambda}{\eta^* + \lambda e^{-q}}\right) \tag{8}$$

Equation (6) shows that there is a closed-form mapping between quantiles of the applicants distribution and the quantiles of the original population, given an equilibrium queue length q.

Using the transformation $x = G_A(\theta) \to G_A^{-1}(x) = \theta$, we therefore obtain

$$G^{-1}(M(x,q)) = G_A^{-1}(x) = \theta \tag{9}$$

with

$$M(x,q) \equiv \frac{m(x,q) - m(0,q)}{m(1,q) - m(0,q)} \text{ with } m(x,q) = x + \left(\frac{1-\lambda}{\lambda q}\right)\log\left(\eta + \lambda e^{-q(1-x)}\right) \tag{10}$$

The result in Equation (9) is key to expressing the equilibrium conditions in a way that they do not depend on the unknown distribution $G_A(\theta)$, but rather on the distribution of the population's productivity $G(\theta)$, which is a primitive of the model.

We can also write down the cumulative distribution functions of the unemployed and the employed by realizing that the population is a weighted average of these two groups: $G(\theta) = \mathcal{U}G_U(\theta) + (1-\mathcal{U})G_E(\theta)$.

[5]Using L'Hôpital's rule, we can show that the mass of the applicants converges to a well-known formula when there is no on-the-job search, i.e., when $\lambda \to 0$

$$\lim_{\lambda \to 0} \mathcal{U} = \frac{\eta^*}{\eta^* + \frac{1-e^{-q}}{q}}$$

where $\frac{1-e^{-q}}{q}$ is the average probability of being hired when there is no on-the-job search.

In addition, the distribution of applicants equals $G_A(\theta) = \frac{\mathcal{U}G_U(\theta)+\lambda(1-\mathcal{U})G_E(\theta)}{\mathcal{A}}$.
Combining these two conditions and the quantile mapping in Equation 9, we obtain

$$G_U(\theta) = \frac{G_A(\theta) - \lambda G(\theta)}{(1-\lambda)\mathcal{U}} = \frac{\mathcal{A}x - \lambda M(x,q)}{(1-\lambda)\mathcal{U}} \tag{11}$$

$$G_E(\theta) = \frac{G(\theta) - \mathcal{A}G_A(\theta)}{(1-\lambda)(1-\mathcal{U})} = \frac{M(x,q) - \mathcal{A}x}{(1-\lambda)(1-\mathcal{U})} \tag{12}$$

We can also determine the average job finding probabilities for applicants, the unemployed and the on-the-job seekers. The last two provide a link between the model and empirical measurements. Thus, the average probability of an applicant's being hired is

$$\overline{p}_A = \int e^{-q(1-G_A(\theta))} dG_A(\theta) = \int_0^1 e^{-q(1-x)} dx = \frac{1-e^{-q}}{q}$$

By integrating over the equilibrium distribution of the unemployed, one can also compute the average probability for an unemployed person to be hired, that is, a UE transition[6]

$$\overline{p}_U = \int e^{-q(1-G_A(\theta))} dG_U(\theta) = \frac{\mathcal{A}\overline{p}_A - \lambda \int_0^1 e^{-q(1-x)} dM(x,q)}{(1-\lambda)\mathcal{U}}$$

The probability for an employed person to be hired is

$$\overline{p}_E = \int e^{-q(1-G_A(\theta))} dG_E(\theta) = \frac{\int_0^1 e^{-q(1-x)} dM(x,q) - \mathcal{A}\overline{p}_A}{(1-\lambda)(1-\mathcal{U})}$$

Therefore, the average job-to-job transition probability is $\lambda\overline{p}_E$.
Some algebra then shows that $\overline{p}_A = \frac{\mathcal{U}\overline{p}_U + \lambda(1-\mathcal{U})\overline{p}_E}{\mathcal{A}}$.
Moreover, the unemployment rate can also be expressed as $\mathcal{U} = \frac{\eta^*}{\eta^* + \overline{p}_U}$.

3.3 Workers

All unemployed workers receive an exogenous income $\rho\theta$ with $0 < \rho < 1$. An applicant randomly chooses a vacancy and faces an *equilibrium* job finding probability $\tilde{p}(\theta, q)$ which depends on the applicant's type and the queue length, as derived above. In case the applicant obtains the job, the worker gets the value of being employed $W(\cdot)$ earning a wage $w(\theta)$. If the applicant comes from unemployment and receives no offers, the applicant remains in this state and applies again the next period.

[6] The integral $\int_0^1 e^{-q(1-x)} dM(x,q)$ equals

$$\mathcal{A}\int_0^1 \frac{\eta^* e^{-q(1-x)} + e^{-2q(1-x)}}{\eta^* + \lambda e^{-q(1-x)}} = \mathcal{A}\left(\frac{1-e^{-q}}{\lambda q} - \frac{\eta^*(1-\lambda)}{q\lambda^2}\left(\log(\eta^* + \lambda) - \log\left(\eta^* + \lambda e^{-q}\right)\right)\right)$$

As previously noted, an employed worker applies to another job with an exogenous probability λ, actually switching with overall probability $\lambda\tilde{p}(\theta, q)$. Employed workers who do not move remain in their current job until the next period. If the workers do not apply to another job, they become unemployed with exogenous probability η. Hence, the effective separation probability is $\eta^* = (1-\lambda)\eta$.

Workers have linear preferences over consumption and have a constant discount factor $\beta \in (0,1)$. Hence, an unemployed worker's lifetime utility is

$$U(\theta) = \rho\theta + \beta[\tilde{p}(\theta, q)W(\theta) + (1-\tilde{p}(\theta, q))U(\theta)] \tag{13}$$

while for the employed agent, the value of being employed W depends on the potential value of a new job $\widetilde{W}$.

$$W(\theta) = w(\theta) + \beta[\lambda\left(\tilde{p}(\theta, q)\widetilde{W}(\theta) + (1-\tilde{p}(\theta, q))W(\theta)\right) + (1-\lambda)\left((1-\eta)W(\theta) + \eta U(\theta)\right)] \tag{14}$$

3.4 Firms

A job filled with a worker of productivity θ generates a value $J(\theta)$ and a profit flow $\theta - w(\theta)$. After production, matched workers apply to another job with probability λ, in which case they are hired by another employer with probability $\tilde{p}(\theta, q)$. In this case, the original employer obtains the value of posting a vacancy again as described later, V.

If the worker does not apply to another job, the match is destroyed with exogenous probability η, in which case the employer obtains the value of posting a vacancy V described below. In case the on-the-job application and the separation shocks do not take place, the match goes on and the employer obtains the discounted profit flow next period. Hence, the value $J(\theta)$ is

$$J(\theta) = \theta - w(\theta) + \beta\left[\lambda\tilde{p}(\theta, q)V + (1-\lambda)\eta V + (1-(1-\lambda)\eta - \lambda\tilde{p}(\theta, q))\, J(\theta)\right] \tag{15}$$

Employers observe the aggregate state of the labor market and optimally create vacancies by paying a per-vacancy flow cost $\kappa + \chi$. In this cost we include the standard flow cost of keeping a job posting open (that will be calibrated), as well as a more general "capital renting cost" that can be associated to the need for the firm to provide the infrastructure for the job post.[7]

Vacancies simultaneously receive K applications drawn from the distribution of applicants $G_A(\theta)$, which will be defined below. This simultaneous hiring is a key departure from

[7] This cost is related to securing financial funds, obtaining capital goods, or access to product markets inline with Smith (1999), Petrosky-Nadeau and Wasmer (2015), Petrosky-Nadeau and Wasmer (2017).

most of sequential search and matching models. In addition, after receiving applications, the employer attaches a probability ϕ of interviewing or screening each applicant with marginal cost $\xi \geq 0$.[8]

After each costly interview, the employer perfectly learns the applicant's type θ. Due to this assumption, we focus on selection issues, leaving aside informational effects. The employer offers the position to the most profitable worker or posts the vacancy again. Thus, the value of posting a vacancy is

$$V = \max\left\{\max_{\phi}\left\{-\kappa - \chi + \beta(Prob(K > 0)H(k) + Prob(K = 0)V)\right\}, 0\right\}$$

where $H(k)$ is the maximum profit obtained from a pool of $k > 0$ screened applicants.

$$H(k) = \mathbb{E}_K\left[-\xi k + \max_j \left\{J(\theta_j)\right\}_{j=1}^{k} | k > 0\right]$$

3.5 Solving the Competitive Equilibrium

3.5.1 Solving the hiring problem

To construct the solution we assume (and discuss later) a Coincidence Ranking Equilibrium (CRE) in which the productivity and profitability rankings coincide, i.e., employers always prefer more productive types. Under CRE, we characterize the problem of (3.4) given a distribution of applicants $G_A(\theta)$ which is exogenous from the viewpoint of an individual employer. We note that employers never choose to reject all applicants and repost a vacancy since no applicant whose value to the firm is lower than V will ever submit a marginally costly application.

In Online Appendix A we show that the free entry condition can be written as

$$\kappa + \chi + \beta\xi\phi q = \beta\phi q\left(\int_0^1 (J(G^{-1}(x, q)))e^{-\phi q(1-x)}dx\right) \tag{16}$$

In general equilibrium, the queue length must satisfy the entry condition $V = 0$, that is, an employer posts vacancies up to the point where the expected value of doing so exactly compensates for the opportunity cost of entry.

[8]This decision could be contingent on the realized number of arrived applicants k, in which case an optimal hiring policy would set a cap on the number of interviewed candidates. This would assuage the hiring advantage high productivity applicants have, as does the ex ante hiring probability ϕ, at the cost of substantially decreasing tractability. Therefore, for the sake of simplicity, we only consider the case in which ϕ is constant and set before learning the realized number of applicants.

3.5.2 The choice of the hiring probability ϕ

The information we can retrieve from the data is the average number of interviews per vacancy $\tilde{q}$. This can be thought of as resulting from a choice of the firm, which decides to interview each applicant with probability ϕ: $\tilde{q} \equiv \phi q$, where q is the number of applications received per vacancy.

When we consider the optimal choice of vacancy opening, we realize that in fact ϕ and q always enter in a multiplicative way in the model, as we can see in Equation 17, where, for the reader's convenience, we show the expressions for $p(G_A(\theta), q)$ and $J(\theta)$.

$$V = \max_{\phi} -\kappa - \chi - \beta\xi\phi q + \beta \int_0^{\infty} J(\theta)\phi q p(G_A(\theta), q) d(G_A(\theta)) \tag{17}$$

$$\begin{aligned} J(\theta) &= \frac{\theta - w(\theta)}{1 - \beta[1 - \eta^* - \lambda p(G_A(\theta), q)]} \\ p(G_A(\theta), q) &= e^{-\phi q(1 - G_A(\theta))} \end{aligned}$$

Since they always enter as a product in the model, we cannot distinguish between ϕq and $\tilde{q}$. Indeed, employers can set an average number of interviews per vacancy by either adjusting vacancy-posting or choosing the interview probability ϕ to achieve the desired level $\tilde{q} = \phi q$. For this reason, without loss of generality we fixe $\phi = 1$ henceforth, so that $\tilde{q} = q$.

3.6 Wages

We consider a bargaining protocol à la Hall and Milgrom (2008), in which the firm and the worker can alternate in propose a wage offer; in this setting, differently from the Nash protocol, the threat points for both parties consist in delay bargaining, instead of walking out of the negotiation. We follow Boitier and Lepetit (2018) in setting the probability that the match is destroyed equal to the probability of exogenous termination of the bargaining round, in order to obtain a closed-form analytical solution; we consider that the cost of delay for the firm is proportional to the productivity level of the worker, i.e. we suppose that the cost of delay for the firm is given by $\gamma\theta$.[9]

We adapt the methodology in Boitier and Lepetit (2018) to our model[10], and we obtain the following wage equation in steady state[11]:

[9]We follow Wang (2020) who stresses the importance of considering a variable cost of delay for the firm in line with productivity. Huckfeldt (2022) too adopts this assumption for the part of his model in which production depends on certain characteristics of the worker.

[10]See Online Appendix B for the detailed equations.

[11]By setting the probability of being able to look for a job while employed (λ) to zero, our steady state wage equation becomes the same as in Boitier and Lepetit (2018).

$$w(\theta) = \frac{1}{2[1-\beta(1-\eta^*)] + \beta\lambda p(G_A(\theta), q)} \theta[1-\beta(1-\eta^*) + \beta(\gamma+\rho)(1-\beta(1-\eta^* - \lambda p(G_A(\theta), q))] \tag{18}$$

3.7 Inequality measures

Remarkably, in our framework we obtain a quasi-closed form for the equilibrium distribution, making the model amenable to studying comparative statics for inequality statistics. In particular, we construct a ratio to compare the outcome of the model in terms of the ranking gap for the employed distribution with the correspondent measure characterizing the population distribution: we thus compare an indicator of *ex post* inequality with an indicator of *ex ante* inequality. If a Walrasian market prevails ex ante inequality would reflect one-to-one in wage inequality. Therefore, this ratio would be interpreted as the amplification or compression of the wage distribution due to search and selection frictions.

We consider two levels of productivity ($\underline{\theta}$ and $\overline{\theta}$) that correspond to two percentiles of the applicants' distribution ($\underline{x}$ and $\overline{x}$), for example the $\overline{x} = 90^{th}$ and the $\underline{x} = 10^{th}$ percentiles: we therefore have that $\overline{\theta} = G_A^{-1}(\overline{x}) = G^{-1}(M(\overline{x}; q))$ and $\underline{\theta} = G_A^{-1}(\underline{x}) = G^{-1}(M(\underline{x}; q))$. We then compute the ratio between the gap of the c.d.f. evaluated at these levels for the employed agents with respect to the gap evaluated for the entire population:

$$\frac{G_E(\overline{\theta}) - G_E(\underline{\theta})}{G(\overline{\theta}) - G(\underline{\theta})} = \frac{G_E(G^{-1}(M(\overline{x}; q))) - G_E(G^{-1}(M(\underline{x}; q))}{M(\overline{x}; q) - M(\underline{x}; q)}$$

If this ratio is bigger than one, then, for the chosen percentiles, and for an equilibrium queue length, the hiring process in the labor market magnifies the pre-existing level of inequality observed in the population expressed as the ranking gap, and attenuates it if the ratio is smaller than one. To characterize this ratio, we apply the Cauchy Mean Value Theorem and obtain the following proposition.[12]

Proposition 1 ***Quantile amplification:*** *The quantiles of the employed distribution and the population distribution are related through*

$$\frac{G_E(\overline{\theta}) - G_E(\underline{\theta})}{G(\overline{\theta}) - G(\underline{\theta})} = \frac{s_E(x_c; q)}{s_E(x^*; q)} \tag{19}$$

in which $s_E(x; q)$ denotes the probability of employment of the type with ranking x given a certain queue length q, i.e.,

$$s_E(x; q) = \frac{p(x; q)}{\eta^* + p(x; q)};$$

[12] The proof can be found in Online Appendix C.

x_c denotes the applicants' ranking associated to the Cauchy Mean Value (CMV) of productivity θ_c, i.e., $x_c = G_A(\theta_c)$, with $\theta_c \in (\underline{\theta}, \overline{\theta})$. Finally, x^* denotes the ranking of the type that defines the average applicants' ratio, i.e., $x^* = G_A(\theta^*)$ is such that

$$\mathcal{A}(q) = \frac{\eta^* + \lambda p(x^*; q)}{\eta^* + p(x^*; q)}$$

Proposition 1 states that the ranking gap for types $(\underline{\theta}, \overline{\theta})$ is amplified by the functioning of the labor market if the average applicant type θ^* is less than θ_c, which obviously can also be stated in terms of the applicants' rankings $x_c > x^*$; in the opposite case, the ranking gap is compressed. In other words, when comparing two types, the hiring selection in the labor market amplifies the ranking gap for individuals whose productivity is higher than the average applicant, and does the opposite for those whose productivity is below that point. Loosely speaking, the hiring selectivity generates an employed distribution with a thickened left tail and a stretched right tail in comparison to the population distribution. Thus, selectivity generates positive skewness, as often occurs in empirical wage distributions.

In terms of the consequences for inequality of the hiring process in the labor market, the model is thus able to generate heterogeneous effects, according to the productivity level. This conclusion is valid for a given queue length and therefore unemployment rate. Any changes in the equilibrium value of q imply a change in the value of the ranking gap ratio, for any chosen percentiles.

In Section 4.4 we will provide some numerical illustration of the heterogenous consequences in terms of inequality for different percentiles, considering the steady state equilibrium. In Section 5, we will show numerically how the ranking gap is affected when the unemployment rate changes, as a consequence of some exogenous parameter shift.

We can use the result of Proposition 1 to show that quantile ratio in the wage distribution are also linked to the ranking gap. In particular, using (11) and the inverse quantile mapping in (9), we obtain that $\theta(x; q) = G_A^{-1}(x) = G^{-1}(M(x; q)) = G_E^{-1}\left(\frac{M(x;q) - \mathcal{A}x}{(1-\lambda)(1-\mathcal{U})}\right)$. Using the Cauchy mean value theorem, we can establish that

$$\frac{\log w\left(\theta(\overline{x}; q)\right) - \log w\left(\theta(\underline{x}; q)\right)}{G_E(\theta(\overline{x})) - G_E(\theta(\underline{x}))} = \frac{w'(\theta(\tilde{x}); q)}{\theta(\tilde{x}) g_E(\theta(\tilde{x}); q)} \equiv \Xi(\tilde{x}; q) \text{ with } \tilde{x} \in [\underline{x}, \overline{x}] \tag{20}$$

Combining (19) and (20) we obtain a decomposition of the log quantile ratio as the product of three terms

$$\log \frac{w(\theta(\overline{x}))}{w(\theta(\underline{x}))} = \Xi(\tilde{x}; q) \frac{s_E(x_c; q)}{s_E(x^*; q)} \left(G(\overline{x}) - G(\underline{x})\right) \tag{21}$$

The model predicts that a wage gap depends a wage scale factor $\Xi(\tilde{x}; q)$, a ranking amplification/compression factor $\frac{s_E(x_c;q)}{s_E(x^*;q)}$ and a ranking population gap $G(\overline{x}) - G(\underline{x})$. The last term $(G(\overline{x}) - G(\underline{x}))$ provides the productivity gap regardless any search and selection market structure. The amplification factor tells us how the search and selection frictions in the market amplify or dampen the population inequality ranking gap. Finally, the factor Ξ, since it is always positive, simply scales the ranking gap of the employed distribution into actual wages in the market given the equilibrium distribution of employed types G_E.

4 Calibration and Results

Once we solved the model, we bring it to the data to estimate some of the parameters, and test its empirical performance. Our strategy consists on targeting monthly CPS and CPS-ORG statistics such as the unemployment rate, the job-to-job transition probability (EE), the separation rate (EU), and some of the moments of the wage distribution. We aggregate the data that at the state-year level as in Section 2 and assume that the parameters are time-invariant. Moreover, in our model, agents with different levels of productivity compete for the same type of jobs. As also explained in Section 2, we consider two separate labor markets, one for college and the other for non-college workers. We also assume that the population productivity θ follows a Type I Dagum distribution, i.e. $\theta \sim \text{Dagum}(a, b, p)$.[13] Previous to the estimation, we need to calibrate a set of parameters $\Psi_1 = \{\beta, \gamma, \kappa, \xi\}$; the first represents the discount factor, the second one the employer's cost of delay in the bargaining process; κ represents the part of the vacancy posting costs linked to the online advertisement of job posts, while ξ is the screening cost paid per applicant.

Once we obtain the estimated values, we then use the equilibrium entry condition (Equation 16) to close the model and obtain the total value of the vacancy posting costs.

We set the discount factor to 0.996, implying an annual interest rate of 4.8%. For the worker's delay cost parameter, γ, we are guided by the literature on alternating-offer bargaining models. Following Hall and Milgrom (2008), studies such as Boitier and Lepetit (2018) and Huckfeldt (2022) calibrate this cost to match unemployment moments, yielding values between 18% and 26% of worker productivity. Wang (2020) provides a higher estimate of 30%. We adopt a conservative value and set γ to 15% of productivity for both worker types. This choice is primarily driven by our estimation strategy, which identifies the sum $\gamma + \rho$. By calibrating γ to a conservative lower bound, we establish a credible minimum threshold for estimating the parameter ρ.

Section 4.3 discusses the calbration of the parameters ξ, the screening cost per applicant, and κ, the cost of posting a vacancy (online).

[13] For an overall view of the Dagum distribution and its application in economics, see Kleiber (2008).

Finally, with the calibrated and estimated values of the parameters, we obtain the remaining vacancy posting cost, χ, from the equilibrium entry condition.

4.1 Estimation procedure

We use a Generalized Method of Moments (GMM) approach: we look for those values of the parameters that minimize the difference between the moments of the data and those implied by the model. It is important to note that for our steady state calibration, we also need to estimate the queue length, even if this is not a parameter but an endogenous variable of the model. Once we obtain the estimated values of the parameters of interest, we perform some counter-factual experiments: in these cases, the value of q changes endogenously.

The set of parameters we need to estimate is given by $\Psi_2 = \{\eta, \lambda, a, b, p, \rho, q\}$ where η is the exogenous match destruction rate (remember that the overall separation rate is given by $\eta^* = \eta(1-\lambda)$), λ is the probability of applying for a job while working, while a, b, and p are the parameters characterizing the Dagum productivity distribution of the population; ρ is the proportion of the productivity that the worker receives when unemployed[14], and q is the length of the queue length in steady state equilibrium.

For what regards the wage distributions, we compare the predictions of the model with the data in terms of the level of the wage for a selection of percentiles, in particular the 10th, the 50th, and the 90th. Since we solve the model in terms of the applicants' cumulative distribution function, the predictions of the model are therefore made in terms of the applicant's percentiles. We therefore transform the empirical wage distribution to obtain the percentile of the applicants' distribution that corresponds to the observed wage distribution.

Using the definition of the density of employed, as we did in (12), we can write

$$G_E(G^{-1}(M(x;q,\eta,\lambda))) = \frac{M(x;q,\eta,\lambda) - \mathcal{A}x}{(1-\lambda)(1-\mathcal{U})}$$

Let us write $\tilde{G}_w(.)$ for the empirical cumulative distribution function of observed wages. The cumulative distribution function of the employed $G_E(.)$ is unknown, but we can approximate it using the empirical wage c.d.f.: $G_E(G^{-1}(M(x;q,\eta,\lambda))) \approx \hat{G}_w(\hat{w}(x)) = \frac{M(x;q)-\mathcal{A}x}{(1-\lambda)(1-\mathcal{U})}$. Thus if we want to obtain the level of wages implied by the observed wage distribution for a certain percentile x of the applicants' distribution, we just need to take the inverse function of the expression for $\hat{G}_w$:

[14]Since in our setting we do not explicitly model labor market institutions such as unemployment benefits, the coefficient ρ is an overall measure of the outside option of the workers, including income support such as unemployment insurance.

$$\hat{w}(x) = \widehat{G}_w^{-1}\left(\frac{M(x;q,\eta,\lambda) - \mathcal{A}x}{(1-\lambda)(1-\mathcal{U})}\right) \tag{22}$$

To find the parameters' values, we thus solve the following minimization problem:

$$\min_{\eta,\lambda,\mu,\sigma,\rho,q} Q = \left\{\varphi_1 \sum_{i=1}^{N} \frac{1}{N}(\mathcal{U}(q,\eta,\lambda) - \mathcal{U}_i)^2 + \varphi_2 \sum_{i=1}^{N} \frac{1}{N}(\lambda \overline{p}_E(q,\eta,\lambda) - JJ_i)^2 + \varphi_3 \sum_{i=1}^{N} \frac{1}{N}(\eta(1-\lambda) - EU_i)^2 \right.$$
$$\left. + \varphi_4 \sum_{x \in N_x} \omega_x \left(\frac{1}{2} G^{-1}(M(x;q,\eta,\lambda))[1+\beta(\gamma+\rho)] - \hat{w}(x)\right)^2 \right\}$$

where N represents the number of observations by state and by year in the CPS data[15], while $N_x = \{0.1, 0.5, 0.9\}$; the weights of the 10th, 50th and 90th percentiles are respectively $\omega_{0.1} = 0.33, \omega_{0.5} = 0.34, \omega_{0.9} = 0.33$. The parameters ϕ_1, ϕ_2 and ϕ_3 are all set to one, while the parameter ϕ_4 is set to 0.01 to take into account the different scale of the variable.[16]

4.2 Results

Table 1 shows the results of our estimation of the benchmark model. The results in terms of the exogenous separation rate η and the poaching probability λ are in line with standard values in the literature. The estimated values for the scale parameter b and the shape parameters a and p of the Dagum distribution imply a higher mean productivity for college workers, as expected, together with a higher variance. The parameter ρ is a rough measure that includes the value of home production and unemployment benefits; it represents a proportion of individual productivity, so it is also quite in line with standard calibration values.

Table 1: Parameters' baseline estimation

Group	η	λ	q	a	b	p	ρ
College	0.008	0.044	2.358	2.885	17.314	1.560	0.729
Non-College	0.018	0.052	2.739	3.149	10.605	2.553	0.745

Table 2 compares the informative moments from the data with the ones obtained by simulating the model.

In order to assess the goodness of the fit, we present the mean square errors (MSE) for the four types of moments we used in the estimation. For each variable x, let us define x_i to be

[15] N=1308 for college and N=1326 for non-college.

[16] Hourly wages are expressed in logs, while the first three variables are rates.

Table 2: Data vs. Model generated moments, baseline estimation

	College		Non-College	
Statistic	Data	Model	Data	Model
Labor Market Flows (%)				
Unemployment rate	2.599	2.600	6.412	6.412
Job-to-job transition	1.923	1.922	2.319	2.319
Separation rate	0.742	0.743	1.718	1.718
Wage Levels ($)				
90th percentile	35.251	35.252	22.737	22.736
Median	17.148	17.148	12.135	12.135
10th percentile	9.254	9.255	7.658	7.658
Mean	19.572	20.794	13.495	14.269

the observations by state and year, $\bar{x}$ the simple average, and $\hat{x}$ the simulated value according to the model using the estimated values for the parameters.

For the unemployment rate, separation rates and job-to-job transitions, we compute the MSE in the standard way:

$\bar{Q}_1 = \Sigma_i(\mathcal{U}_i - \overline{\mathcal{U}})^2; \bar{Q}_2 = \Sigma_i(JJ_i - \overline{JJ})^2; \bar{Q}_3 = \Sigma_i(EU_i - \overline{EU})^2$

$\hat{Q}_1 = \Sigma_i(\mathcal{U}_i - \widehat{\mathcal{U}})^2; \hat{Q}_2 = \Sigma_i(JJ_i - \widehat{JJ})^2; \hat{Q}_3 = \Sigma_i(EU_i - \widehat{EU})^2$

For the moments of the wage distribution, it is important to remark that the data are not "model independent," since we used the transformation of equation 22.

Table 3: Mean Square Errors, baseline estimation

	College		Non-College	
MSE Statistic	Data ($\bar{Q}$)	Model ($\hat{Q}$)	Data ($\bar{Q}$)	Model ($\hat{Q}$)
Q_1	0.139	0.140	0.627	0.627
Q_2	0.024	0.024	0.031	0.031
Q_3	0.009	0.009	0.023	0.023
Q_4	0.134	0.137	0.086	0.093

4.3 Closing the model

In the previous section, we estimated the values of the parameters of interest and of the queue length, distinguishing between the two types with different levels of education, $j = \{\text{college, non-college}\}$. We now proceed to recover the fixed entry cost χ, by using the general equilibrium condition given by Equation 16. In order to do so, we need additional evidence on the screening costs ξ and the vacancy posting costs κ.

We follow Villena-Roldan (2010a) and consider the National Employer Survey 1997 (NES97) to compute the average monetary cost incurred in that specific year for recruiting activities. Using the survey information we construct (1) the total annual recruitment and selection cost (RSC) and (2) the total annual number of recruiting interviews (NRI).[17] While this is unusually collected and relevant information, there are two important challenges: First, the RSC does not distinguish between fixed and marginal screening costs, so that the average screening cost RSC/NRI cannot be directly used to measure the marginal screening cost represented by the parameter ξ. A second problem is a substantial non-response rate in the survey. Only 48.2% (1486 out of 3081 respondents) answered the four questions we need to compute RSC and NRI, and just 38.6% report non-zero –credible– recruiting costs. To correct for this non-response bias, we follow an inverse probability weighting (IPW) approach (Fitzgerald, Gottschalk, and Moffitt 1998; Wooldridge 2010). We first construct a probit model with covariates (X) that are observable for the whole sample: categorical variables for size, industry and multi-establishment firm, and the share of four- and two-year college degrees hired in the last two years (CS). We then use NES97 survey sampling weights divided by the predicted response probabilities from the probit model to weigh each observation in the following model:

$$\log RSC_n = \beta_0 + \beta_1 \log NRI_n + \beta_2 \log NRI_n \times CS_n + \beta_3 \log NRI_n \times CS_n^2 + \beta_4 CS_n + \beta_5 CS_n^2 + X_n \gamma + U_n$$

Under this specification we capture a sufficiently flexible approximation to the shape of the conditional expectation of RSC considering non-linear effects of the share of college hirings into their cost structure. Therefore, our measure for the marginal increase of recruiting costs for non-college workers is the marginal effect evaluated in $CS = 0$ since it reflects the cost of a firm only hiring non-college workers keeping constant its size, industry and multi-establishment structure.

$$\frac{\partial E[RSC|NRI, CS = 0, X]}{\partial NRI} \approx \widehat{\beta}_1 \overline{RSC/NRI}$$

[17] We describe in Appendix D details on the construction of these variables using NES97 original variables.

Conversely, our measure for the marginal increase of recruiting costs for non-college workers is therefore the marginal effect evaluated in $CS = 1$

$$\frac{\partial E[RSC|NRI, CS = 1, X]}{\partial NRI} \approx \left(\widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_3\right) \overline{RSC/NRI} + \widehat{\beta}_4 + \widehat{\beta}_5$$

All these computations render values for 1997 only. How can we compute values for ξ for other years? We adapt the idea of Landais, Michaillat, and Saez (2018), who use some NES97 information reported by Villena-Roldan (2010a). We assume that these 1997 costs increase at the rate of the average hourly real wage of workers in areas related to Human Resources Management. The details of the occupations comprised into this category are in Appendix D. The following table summarizes the estimated parameters

Table 4: Measurement of screening cost ξ (NES 1997)

	Non-college	College
Estimated ξ (USD 1997)	209	843
HRM wage adjusted ξ (USD 2010)	320	1288

Therefore, we consider a value for the screening costs per applicant ξ of 1288 USD for college and 320 USD for non-college workers.

The vacancy posting cost κ can be interpreted as the cost of maintaining an online job post for one month.[18] Current costs range from 200 USD to 400 USD, so we use an average value of 300 USD.[19]

Considering the mean screening costs ξ over the sample period and the vacancy posting costs κ, we then apply the free entry condition (16) to estimate the remaining value of vacancy posting costs (the capital renting cost) χ. It is important to emphasize that the time period considered is one month, while wages are expressed on an hourly basis. Therefore, we must assume an average number of hours worked per month. We consider a standard full-time workload of 160 hours per month.

The values for χ we obtain for our benchmark case for college and non-college are, respectively, nearly 8375 USD and 3560 USD.[20] In our model, which excludes modelling capital, investment, and any form of financial or credit friction, the vacancy flow cost χ represents all other expenses for the firm, aside from labor costs: even if the model does not explicitly include capital as a production factor, flow vacancy cost can be interpreted as representing the overall cost of renting capital.

[18]Since our data begin in 1994, we assume that job posting is mainly done through the Internet.

[19]For example, https://www.glassdoor.com/employers/blog/how-much-it-costs-to-post-a-job-online/

[20]We verified that varying assumptions about the number of monthly work hours do not impact our results: while fewer hours would lead to lower fixed entry costs, the overall functioning of the model remains unchanged.

4.4 The ranking gap

Once the baseline steady state of the model has been characterized, in this section we provide a numerical illustration of the ranking gap, i.e., of the amplification/contraction effect of the functioning of the model on inequality that was illustrated in Section 3.7.

The ranking gap can be expressed as

$$\frac{G_E(\overline{\theta}) - G_E(\underline{\theta})}{G(\overline{\theta}) - G(\underline{\theta})} = \frac{s_E(x_c; q)}{s_E(x^*; q)}$$

Table 5 presents the values for the amplification/contraction factor, as well as the corresponding percentiles of the Cauchy Mean Value (x_c) and the percentile corresponding to the average applicant (x^*).

Considering college workers, for applicants' types above the percentile 0.458, the ranking gap is amplified with respect to the population, while for non-college workers, the amplification happens for applicants' types above the percentile 0.495.

Table 5: Ranking gap, baseline steady state

	College			**Non-College**		
Parameters	$\frac{G_E(\overline{\theta})-G_E(\underline{\theta})}{G(\overline{\theta})-G(\underline{\theta})}$	x_c	x^*	$\frac{G_E(\overline{\theta})-G_E(\underline{\theta})}{G(\overline{\theta})-G(\underline{\theta})}$	x_c	x^*
$\overline{x} = 0.9, \underline{x} = 0.1$	1.00047	0.46567	0.45801	0.99917	0.49025	0.49497
$\overline{x} = 0.9, \underline{x} = 0.5$	1.01047	0.67386	0.45801	1.02596	0.67819	0.49497
$\overline{x} = 0.5, \underline{x} = 0.1$	0.98668	0.28051	0.45801	0.95322	0.28748	0.49497

4.5 Coincidence Ranking equilibrium (CRE)

In order for the Coincidence Ranking equilibrium to hold, it is necessary that the value function of a filled job is a positive function of the level of productivity. In terms of our change of variable, it is necessary that the value function of a filled job is increasing in the ranking of the worker in terms of productivity.

We can express the condition in terms of productivity as $\frac{dJ(\theta)}{d\theta} > 0$, and by substituting the expression of $J(\theta)$ and doing some algebra we obtain the following condition[21]

$$1 - \frac{dw(\theta)}{d\theta} > \beta\lambda \frac{dp(G_A(\theta))}{d\theta} J(\theta) \tag{23}$$

The LHS of equation (23) represents the surplus-sharing condition. It requires that for a one-unit increase in a worker's innate productivity, the resulting increase in their bargained

[21] In terms of the productivity percentile we have $\frac{dG^{-1}(M(x;q))}{dx} - \frac{dw(x)}{dx} > \beta\lambda \frac{dp(x)}{dx} J(\theta)$.

wage must be strictly less than one unit. If this condition holds, the firm's instantaneous profit flow, $\theta - w(\theta)$, is strictly increasing in θ. The RHS can be interpreted as the expected marginal capital loss from increased turnover risk associated with hiring a marginally more productive worker.

The inequality states that for the firm's value to increase with productivity, the marginal gain in the flow profit must be greater than the discounted expected marginal capital loss from the shorter expected duration of the match. We checked that this property holds numerically for all our numerical exercises.

5 Counterfactual experiments

In this section, we perform some counterfactual experiments to explore the mechanisms of the model and to explain the empirical findings of Section 2. Modeling productivity with a Dagum distribution enables us to study different types of changes of the productivity distribution, and see how they are translated into the wage distribution.

We consider three types of exogenous changes to the model's productivity distribution to study the impacts on our mesures of inequality, i.e., the quantile ratios.

In the first experiment we focus on varying the median and mean of the productivity distribution, while keeping the quantile ratios as constant. An increase (decrease) in the scale parameter b of the Dagum productivity distribution (while holding the shape parameters a and p constant) results in an increase (decrease) in both the median and mean, without affecting the quantile ratios. This is due to the quantile formula: $Q(u; a, b, p) = b(u^{-1/p} - 1)^{-1/a}$.[22]

In the second experiment, we focus on the left tail of the productivity distribution: we study changes in the shape parameters a and p such that the median and the mean of the distribution increase (decrease) but the 90/50 quantile ratio remains constant.

In the third experiment, we focus on the right tail of the productivity distribution: we study changes in the shape parameters a and p such that the median and the mean of the distribution increase (decrease) but the 50/10 quantile ratio remains constant.

For each counterfactual exercise, we keep all other parameters, except those of interest, fixed and compute the equilibrium queue length (q) that satisfies the entry condition (Equation 16).

Individual wages are, in our model, proportional to the level of productivity of the worker, and they also depend on the probability to find a job, which in turn is due to the position

[22]For a variable following a Dagum distribution of parameters (a, b, p) we also know that $Median = b(-1 + 2^{1/p})^{-1/a}$, $Mean = b\left(\frac{\Gamma(1-1/a)\Gamma(p+1/a)}{\Gamma(p)}\right)$ and $Variance = b^2\left[\frac{\Gamma(1-2/a)\Gamma(p+2/a)}{\Gamma(p)} - \left(\frac{\Gamma(1-1/a)\Gamma(p+1/a)}{\Gamma(p)}\right)^2\right]$.

of the worker in the ranking and the queue length. An exogenous change in the population distribution of productivity affects the distribution of wages (or the distribution of types for employed workers) through the functioning of the labor market. In particular, the hiring process determines the endogenous distribution of productivity of the employed workers and the equilibrium value of the queue length.

We do not claim that only exogenous shifts in the production distribution drive the relation between unemployment and wage inequality, but we are rather interested in proposing potential mechanisms. Our counterfactual exercices thus have to be interpreted as suggesting that the functioning of the labor market interacts with exogenous shifts in the model's parameters to dampen or amplify the effects of those shifts on the economic variables we can observe, such as the unemployment rate and wage inequality.[23]

5.1 The effects of changes in productivity keeping quantile ratios constant

In the first experiment we focus on varying the scale parameter (b) of the Dagum distribution of productivity: as b increases (while the shape parameters a and p remain unchanged), the median and mean increase, while the quantile ratios remain constant.

The increase in median and mean productivity is beneficial for the firm, which therefore opens more vacancies. The length of the queue and therefore the unemployment rate decrease monotonically as median productivity increases. As productivity increases (and therefore as the queue length decreases), both the job-to-job (EE) and the unemployment-to-employment (UE) transition rates increase.

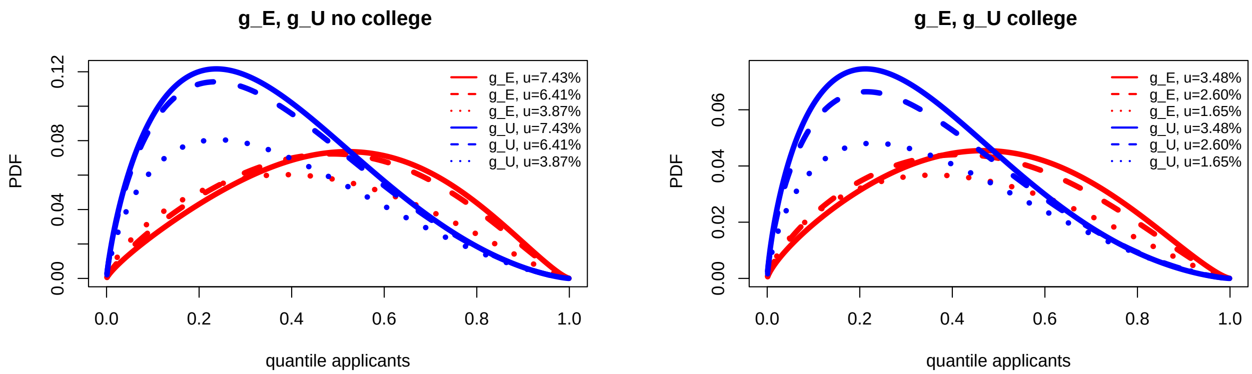
The increase in the mean and median of the productivity distribution caused by the increase in the scale parameter b implies an increase in the mean and median wage too. The change in median and average wage is driven by a composition effect: individual wages in our model reflect the productivity of the employed agents, so the average wage reflects changes in the distribution of the employed agents as well as changes in the wage function.

The higher levels of productivity associated with an increase in the scale parameter b of the productivity distribution are reflected in a higher wage schedule, for every level of productivity.However, when the unemployment rate is very high (because the average productivity is low), the relatively few employed people are concentrated among the most productive. This occurs because employers receive a large number of applicants and are able to hire very selectively. The pool of applicants becomes more homogeneous as the economy becomes less selective, as is clear from comparing the employed and the unemployed productivity distri-

[23] In Online Appendix G, we present the results of an additional experiment, such as an increase in the exogenous probability λ of poaching.

butions at different levels of the unemployment rate in Figure 5. As the effect of the positive shift in productivity distribution gains, not only does the average wage increase but also wage inequality is affected.

Figure 5: The effects of changes in productivity (scale parameter) on mass functions

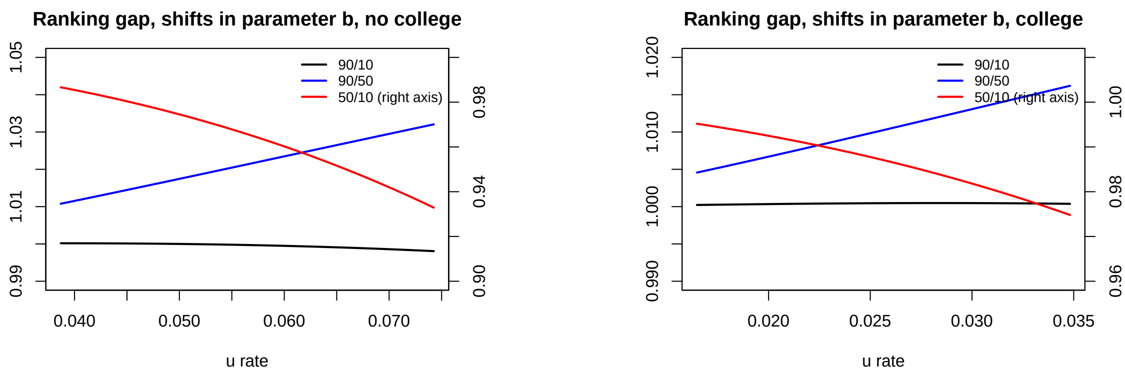


As we saw in Section 3.7, the model predicts a heterogeneous effect of the functioning of the labor market in terms of amplifying or compressing the ex ante ranking gap among different percentiles of the distribution. The ranking gap can be higher or lower than one. In particular, it is lower than one in the left tail of the distribution and higher than one in the right one, as was illustrated in Section 4.4. Here we plot the values for the ranking gaps obtained for three different sets of percentiles of the applicants' distribution (90–10, 90–50 and 50–10), to show the general equilibrium effects of a change in the unemployment rate, driven by shifts of the population distribution generated by the scale parameter b.

Figure 6 shows the heterogenous effects of the unemployment rate on the ranking gap according to the chosen portion of the distribution. The ranking gap, already less than one, further decreases in the left tail of the distribution (i.e., for the 50–10 percentiles), while it increases with the unemployment rate for the right tail of the distribution (the 90–50 percentiles). In other words, as unemployment increases (which corresponds to a decrease of the parameter b and therefore a decrease in median and mean productivity), the left tail of the employed distribution is even more compressed than the population distribution, while the opposite happens in the right tail. The overall quite flat effect for the combination of the 90th and 10th percentiles is thus the result of very different effects on the two tails of the distribution.

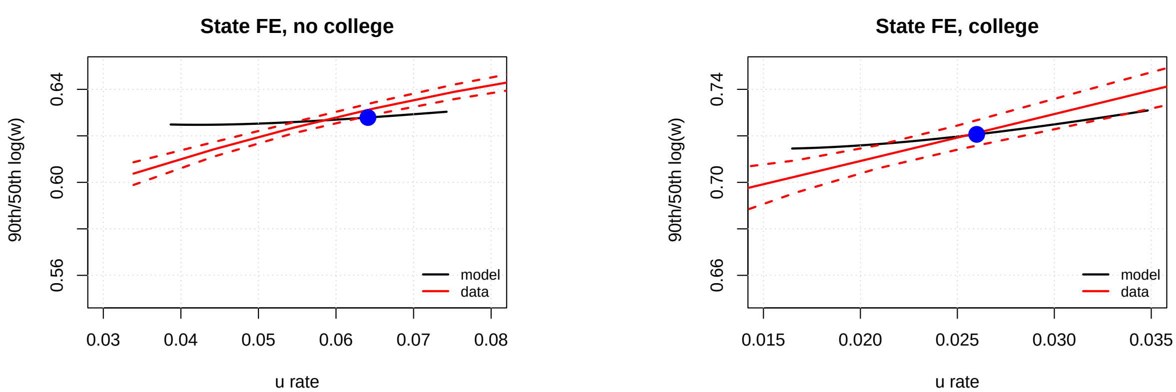
The shift in the productivity distribution also has consequences for the wage distribution's quantile ratios, providing a way to test the model's predictions with data. Equation (21) shows that the wage quantile ratio is affected not only by the productivity distribution and the rank-

Figure 6: The effects of changes in productivity (scale parameter) on the ranking gap



ing gap but also by the scaling factor $\Xi(\tilde{x};q)$. Therefore, the behavior of the wage quantile ratio does not necessarily mirror that of the ranking gap. Figures 7-9 illustrate the overall relationship between the unemployment rate and wage inequality measures as implied by the model.

Figure 7: The effects of changes in productivity (scale parameter) on unemployment and wage inequality (90/50): Model and data



For non-college workers, the 90–10 percentile gap for log hourly real wages increases as unemployment increases, as a result of a decrease in the lower tail inequality and an increase in the upper tail inequality (respectively the 50–10 and the 90–50 percentile gap for log hourly real wages), as it can be seen in the left panels of Figures 7-9.

For college workers, however, the 50–10 percentile gap for log hourly real wages slightly increases too, so that together with an increase in the 90–50 percentile gap it implies an overall increase of the 90–10 percentile gap for log hourly real wages, as it can be seen in the right

Figure 8: The effects of changes in productivity (scale parameter) on unemployment and wage inequality (50/10): Model and data

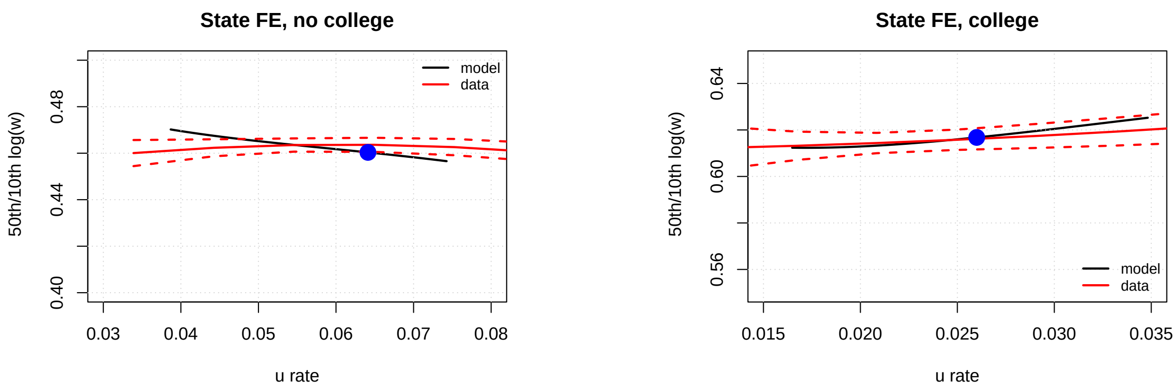
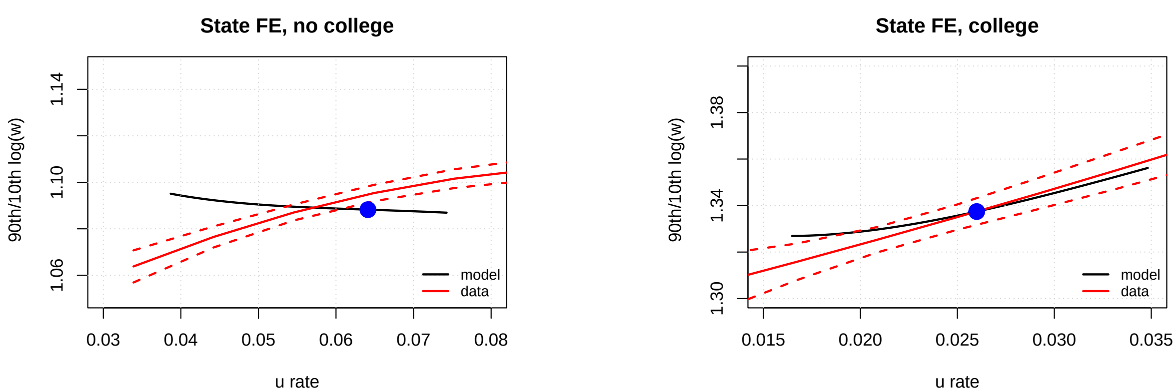


Figure 9: The effects of changes in productivity (scale parameter) on unemployment and wage inequality (90/10): Model and data



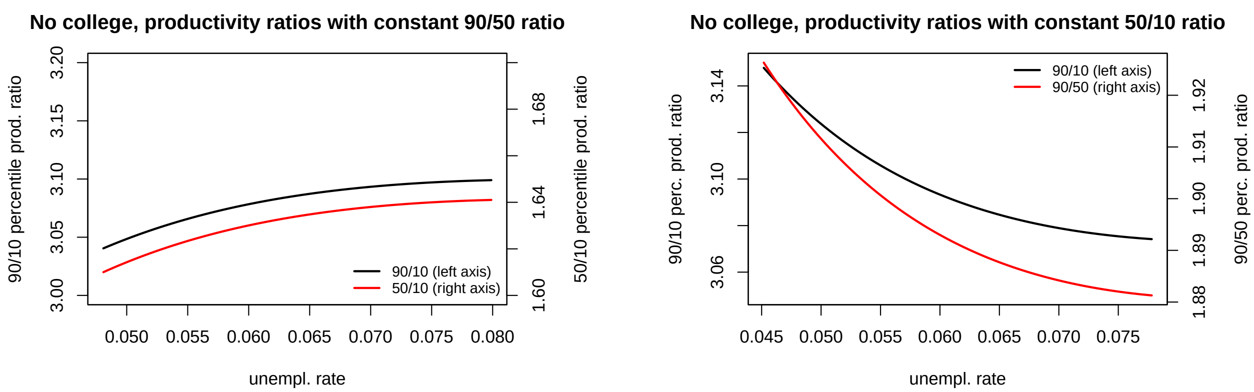
panels of Figures 7-9.

While for college workers the model fits quite will the data, for non-college workers it understates the effects on the right tail of wage distribution (90–50 percentile gap) and predicts a decrease in the inequality measure of the left tail (50–10 percentile gap).

5.2 The effects of changes in productivity keeping 90/50 or 50/10 ratio constant

In our second experiment, we focus on the left tail of the productivity distribution. To do this, we adjust the Dagum distribution's shape parameters—decreasing a while increasing p—to hold the 90/50 quantile ratio constant, keeping the scale parameter b fixed.[24]

Figure 10: The effects of changes in productivity keeping 90/50 or 50/10 percentile ratio constant: other quantile ratios



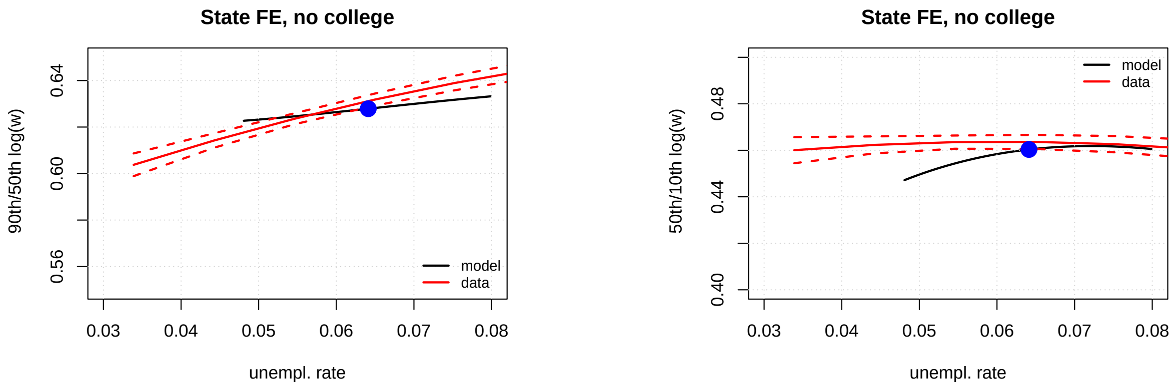
This change shifts the productivity distribution to the right, raising both median and mean productivity. Consistent with the mechanisms previously discussed, the higher productivity leads to a lower unemployment rate. While the upper-tail inequality is fixed by construction, this shift compresses the lower tail of the distribution, causing both the 50/10 and the overall 90/10 quantile ratios to decrease. The key result, illustrated in the left panel of Figure 10, is a positive correlation between the unemployment rate and these measures of lower-tail inequality.

If we analyse the performance of the model with respect to the data, for non-college workers, we can observe in Figure 11 that the model captures the consequences on the 90–50 percentile gap of the wage distribution, but still overstates the changes in the left tail. In our third experiment, we focus on the right tail of productivity distribution: we change the shape

[24]For the sake of brevity we only focus on non-college workers. The results for college workers can be found in Online Appendix H.

parameters a and p of the Dagum distribution of productivity in order to obtain an increase in the median, while keeping the 50/10 percentile ratio constant.

Figure 11: The effects of changes in productivity keeping 90/50 percentile ratio constant on unemployment and wage inequality (90/50 and 50/10): Model and data



These shifts of the productivity distribution imply an increase in the 90/50 as well of the 90/10 percentile ratios. Since those changes are associated with an improvement in the median and mean productivity, and therefore a reduction of the unemployment rate, the overall result is a negative correlation between the unemployment rate and the productivity inequality measures given by the 90/50 and 90/10 percentile ratios, as it is illustrated in the right panel of Figure 10.

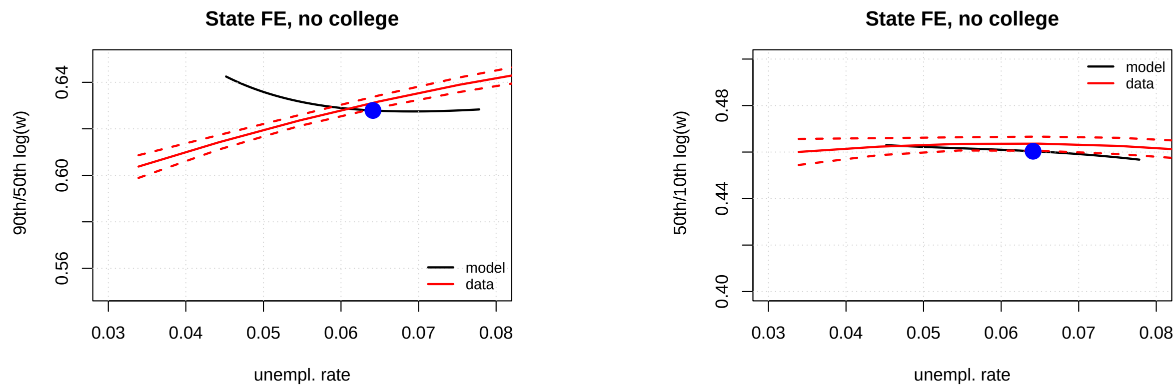
Our model produces a negative correlation between the unemployment rate and the 90–50 percentile gap of the wage distribution for non-college workers, as it can be seen in the left panel of Figure 12, differently from what we can see in the data, as well as a slightly negative correlation with the 50–10 percentile gap (right panel of Figure 12).

By comparing the predictions of the model with the data, we can conclude that for non-college workers it seems that shifts to the right of the productivity distribution driven by the left tail are more likely to be able to explain the observed correlations between the unemployment rate wage inequality, while for college workers shifts in the productivity distribution that did not impact directly the percentile ratios seem a good candidate to explain the observed correlations.

6 Efficiency Analysis

As is common in matching models, there are many Pareto-constrained allocations. Indeed, it is not possible to switch a given employed worker with an unemployed one without hurting

Figure 12: The effects of changes in productivity keeping 50/10 percentile ratio constant on unemployment and wage inequality (90/50 and 50/10): Model and data



the former. We focus on allocations that maximize aggregate welfare or production since workers are risk neutral, following the tradition of Hosios (1990). The Social Planner (SP) faces the same constraints as private agents, in particular the same recruiting and screening technologies. The SP may instruct firms to post vacancies and workers to apply. The control of screening effort is embedded into the decision about the optimal q, because the planner can affect the effective queue length by changing the probability of screening ϕ or by posting vacancies and changing q, i.e., $q = \phi\tilde{q}$.

The matching process faced by the firms instructed by the SP is still characterized by the same risk of the decentralized economy: it is possible that the firm does not receive any applications at all. Since workers are sending their applications to firms, the number of applications received K still follows a Poisson distribution with a mean of $q = \mathcal{A}/\mathcal{V}$. Thus, the Social Planner's objective consists in maximizing total production net of vacancy posting and screening costs.

The instantaneous flow of gross production is given by (abstracting from time subscripts): $Y(\mathcal{U}) = (1-\mathcal{U})\int_0^\infty \theta g_E(\theta)d\theta + \mathcal{U}\int_0^\infty \rho\theta g_U(\theta)d\theta$.

Using the definition of the distribution of applicants $g_A(\theta)$ and the fact that the population distribution always equals $g(\theta) = \mathcal{U}g_U(\theta) + (1-\mathcal{U})g_E(\theta)$, we realize that the distribution of the unemployed is $g_U(\theta) = \frac{\mathcal{A}g_A(\theta)-\lambda g(\theta)}{\mathcal{U}(1-\lambda)}$ while the distribution of the employed can be written as $g_E(\theta) = \frac{g(\theta)-\mathcal{A}g_A(\theta)}{(1-\mathcal{U})(1-\lambda)}$.

Elaborating the previous results, we obtain that $Y(\mathcal{U}) = E[\theta] - (1-\rho)\mathcal{U}\int_0^\infty \theta g_U(\theta)d\theta$.[25]

[25] The detailed derivation is provided in the Online Appendix E.

We can therefore, after some algebra, write the instantaneous production as

$$Y(q) = \frac{1-\lambda\rho}{1-\lambda}E[\theta] - \frac{(1-\rho)}{(1-\lambda)}\mathcal{A}(q)\int_0^1 G^{-1}(M(x;q))dx \tag{24}$$

The first term of Equation (24) is proportional to the output obtained in a frictionless environment, $\overline{Y} = \mathbb{E}[\theta]$. The proportionality factor $\frac{1-\lambda\rho}{1-\lambda}$ indicates there is an obvious loss because unemployed workers only generate a fraction ρ of their productivity. However, the probability of on-the-job search, λ attenuates this effect because a fraction of new hirings do not originate from jobless workers. The second term in (24) refers to the loss due to the negative impact that recruiting activities generate on the quality of the pool of applicants. The more selective the recruiting process is through a higher q, the lower the quality of the average pool of applicants.

The SP problem can therefore be written as

$$\max_{q_t} \sum_{t=0}^{\infty} \beta^t \left\{ \frac{1-\lambda\rho}{1-\lambda}E[\theta] - \frac{(1-\rho)}{(1-\lambda)}\mathcal{A}(q_t)\int_0^1 G^{-1}(M(x;q_t))dx - (\chi+\kappa)\mathcal{V}_t - \xi\mathcal{A}(q_t) \right\} \tag{25}$$

subject to $q_t = \mathcal{A}_t/\mathcal{V}_t$

The third and fourth terms in the maximization problem (25) indicate the recruiting costs, given by the vacancy posting and the screening costs. The number of posted vacancies $\mathcal{V}$ is managed by the Social Planner to control the extensive margin of hiring at the firm level, whereas q controls the screening activity. If the Social Planner is allowed to mandate vacancy-posting and screening, it is clearly constrained by the identity $q = \mathcal{A}_t/\mathcal{V}_t$.

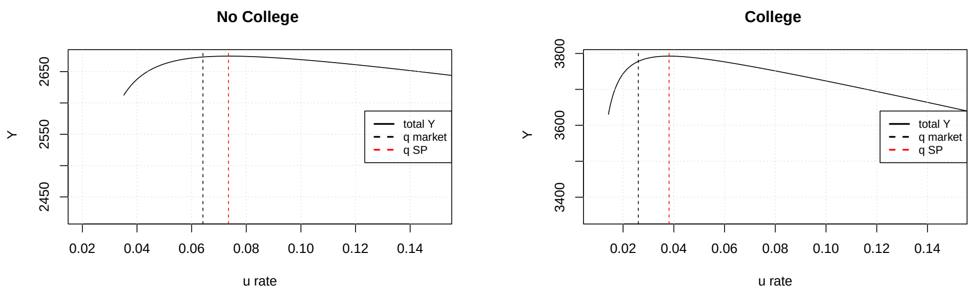
The Social Planner (SP) faces a key trade-off when setting the level of vacancies. On the one hand, more vacancies ($\mathcal{V}$) increase total posting costs, given by $(\chi+\kappa)\mathcal{V}$). On the other hand, for a given number of applicants, more vacancies shorten the average queue length, thereby reducing screening costs, given by $\xi\mathcal{A}$. Consequently, the calibration of these two components of recruiting costs is crucial for determining the optimal unemployment rate, as we demonstrate in Subsection 6.1.

Figure 13 shows that for the calibrated and estimated baseline values of the parameters, the optimal unemployment rate would be higher than the market solution for both non-college (7.35% instead of 6.41%) and college workers (3.81% instead of 2.60%).

In a competitive equilibrium, q is determined by the entry condition (Equation 16). We can therefore look for a tax/subsidy schedule that allows obtaining the queue length that the Social Planner would choose as the outcome of a decentralized market equilibrium.

We adapt a general tax schedule that has been studied at length in the literature on optimal income taxation and apply it to our framework, in which we consider the firm to be the

Figure 13: Queue length and unemployment: Market vs. social planner



subject of taxation. A widely used tax schedule in public finance, as described by Heathcote, Storesletten, and Violante (2017), defines the total tax revenues ($T(y)$) for the level of income y with the two-parameter functional form $T(y) = y - \tau_0 y^{1-\tau_1}$, where the parameter τ_1 is an index of the progressivity of the tax system.[26]

A tax schedule is considered progressive if the Coefficient of Residual Income Progression (CRIP) is less than one and regressive if larger than one; in the case of a flat tax, the CRIP is equal to one. One advantage of this measure of tax progressivity is that it is well defined even when the average tax rate is zero. The CRIP is related to the marginal and average tax rates as follows: $\text{CRIP}(y) = \frac{\partial y^d}{\partial y}\frac{y}{y^d} = \frac{1-T'(y)}{1-\overline{T}(y)}$ where $\overline{T}(y)$ is the average and $T'(y)$ is the marginal tax rate. With the adopted functional form for the tax and transfer schedule, the expression for the CRIP is given by $1 - \tau_1$.

In our framework, we assume that the tax schedule applies to the profit function of the firm, which in itself depends on the level of productivity of the hired worker. The net value of a filled job, $J(\theta)$, is thus given by $(1 - \tilde{t}(\theta))J(\theta)$, where $\tilde{t}(\theta)$ is the average tax rate paid for productivity level θ. Therefore, by using the tax schedule as in Heathcote, Storesletten, and Violante (2017), $\tilde{t}(\theta) = 1 - \tau_0\theta^{-\tau_1}$. By applying the same change of variables as in Equation (9), the tax schedule can be rewritten as $\overline{t}(x) = 1 - \tau_0(G^{-1}(M(x;q)))^{-\tau_1}$.

The entry condition including taxes is thus

$$\kappa + \chi + \beta\xi\phi q = \beta \int_0^1 \frac{(1-\rho)G^{-1}(M(x;q))}{1-\beta[1-\eta^* - \lambda e^{-q(1-x)}]} e^{-q(1-x)}(1-\overline{t}(x))dx \tag{26}$$

In addition to the entry condition, we impose a balanced government budget: the net tax revenues that are levied on the work force of the firm have to be null. In fact, the general tax and transfer schedule allows for positive as well as negative taxes, i.e., for subsidies.

[26]We briefly recall that the Coefficient of Residual Income Progression (CRIP) is one of the most used measures of progressivity: it represents the elasticity of post-tax income to pre-tax income.

The balanced budget condition for the government is thus given by

$$(1-\mathcal{U}(q))\int_0^\infty \tilde{t}(\theta)\theta dG_E(\theta)=0$$

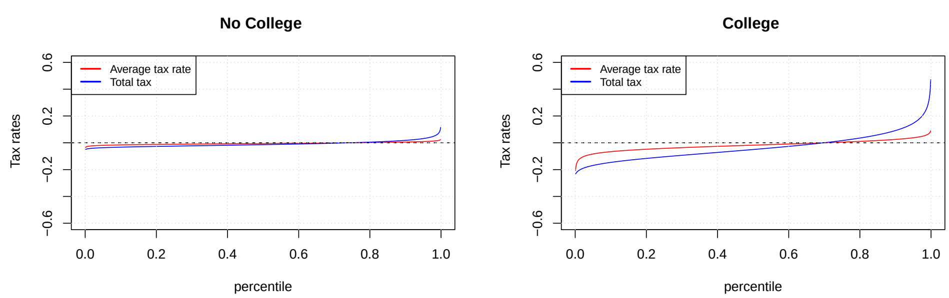
By applying the usual change of variables and re-writing the integral accordingly, after some algebra we obtain the condition

$$\int_0^1 \bar{t}(G^{-1}(M(x;q))M_x(x;q)dx=\mathcal{A}(q)\int_0^1 \bar{t}(G^{-1}(M(x;q))dx \tag{27}$$

Finally, to implement the SP solution as a competitive equilibrium, the tax schedule must also preserve the coincidence ranking equilibrium: the firm has to continue to strictly prefer hiring higher productivity types, even after paying the corresponding tax. We therefore check *ex post* that the coincidence ranking equilibrium is satisfied.

We solve the system of equations (26) and (27) to find the values of the parameters τ_0 and $\tau-1$. The values for the parameters of the tax schedule for non-college types are found to be $\tau_0=1.05$ and $\tau_1=0.05$, while for college $\tau_0=1.23$ and $\tau_1=0.18$; therefore the coefficients of residual income progression $(1-\tau_1)$ indicate that the tax and transfer schedule is progressive for both types (even if for non-college types it is almost proportional). Figure 14 shows that the optimal policy is to tax firms for hiring high-productivity workers and subsidize them for hiring low-productivity ones. This tax and transfer schedule creates a disincentive for firms to open vacancies, which is necessary because the desired unemployment rate is higher than the market-driven outcome.

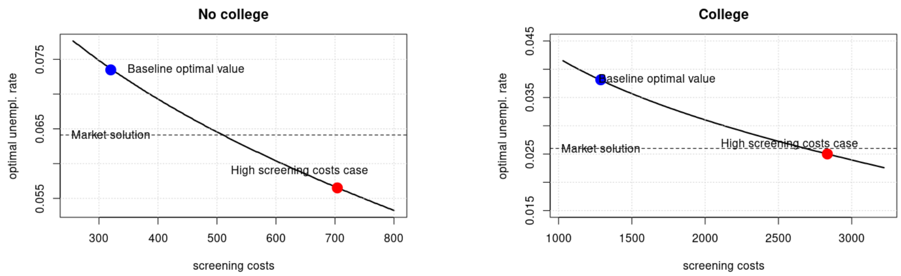
Figure 14: Baseline tax and transfer schedule



6.1 The role of screening and vacancy posting costs

To understand the interplay between the two types of recruiting costs in our model—vacancy posting (χ) and applicant screening (ξ)—we conduct a robustness exercise. In this exercise, we systematically increase the screening cost ξ and allow the posting cost χ to adjust based on the free-entry condition. We then compute the resulting optimal unemployment rate and its associated queue length.

Figure 15: Optimal unemployment rate depending on screening costs



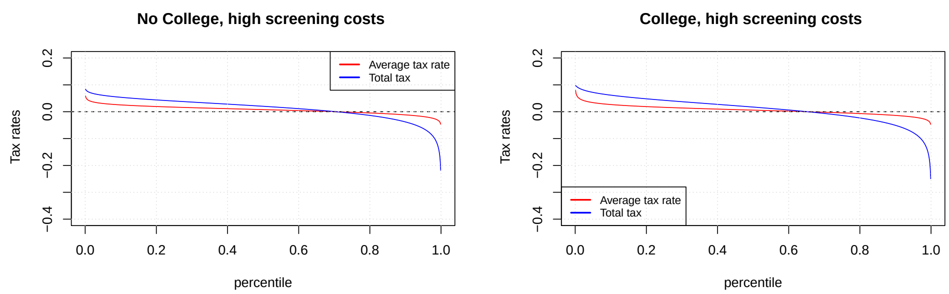
Our central finding, illustrated in Figure 15, is that as screening costs (ξ) rise, the Social Planner (SP) finds it optimal to target a shorter queue length, which corresponds to a lower unemployment rate. To illustrate with a specific case, consider when screening costs are three times their baseline value (the red dot in the figure). This scenario, where screening one college applicant costs nearly $2,900 and a non-college applicant nearly $700, implies a significant drop in vacancy posting costs. Consequently, the SP's optimal unemployment rate falls to just 2.50% for college and 5.66% for non-college workers, both below their respective market equilibrium levels.

Achieving an unemployment rate below the market equilibrium requires a fundamental reversal of the baseline tax policy. As shown in Figure 16, the optimal tax and transfer schedule must now tax firms for hiring low-productivity workers and subsidize them for hiring high-productivity workers. This "regressive" schedule is necessary to create a powerful incentive for firms to open more vacancies, which in turn shortens queues and lowers unemployment to the desired level. This result can be understood by examining the incentives for employers. Top matches chiefly drive firms' interest in the labor market because they can avoid mediocre matches by hiring selectively. An extra incentive for hiring at the top of the distribution is a great incentive to post vacancies. However, good types are scarce, and many employers end up hiring not-so-good applicants, resulting in a trickle-down of job opportunities. Hence, high types generate a positive externality for their lower productivity counterparts by driving more open vacancies.

In addition, there is another composition externality of high types. As their attractiveness spurs more vacancies and shortens queues, subsidies for them reduce hiring selectivity, improve the composition of the unemployed, and increase the share of on-the-job seekers. As employers expect to hire better workers under this scenario, more vacancy posting is reinforced, which translates into higher chances of being hired for less productive workers.

Viewed through this lens, the tax and subsidy scheme is a direct application of the Pigouvian principle. To enhance overall efficiency, the policy aligns private incentives with social benefits. When the planner's goal is to lower unemployment below the market rate, high-productivity workers are subsidized because their recruitment generates positive spillovers for others. Conversely, as seen in our baseline scenario, when the goal is to raise unemployment above the market rate, these workers are taxed. In either case, our balanced budget condition ensures that the costs of the subsidies are covered by taxes on those who benefit from the positive externalities.

Figure 16: Tax and transfer schedule with high screening costs



7 Conclusions

We developed a theoretical model that offers insights into the consequences of selectivity in the hiring process for labor market performance and wage inequality. Within our framework, workers exhibit *ex ante* heterogeneity in their productivity levels. This theoretical model enables us to understand how variations in *ex ante* inequality translate into *ex post* wage inequality, primarily through the endogenous composition of the pool of employed and unemployed workers.

Our analysis began by examining empirical evidence that reveals a positive correlation between the unemployment rate and inequality, specifically as measured by the 90–10 percentile gap in log hourly wages. We also observed a positive correlation when considering job separation probabilities and a negative correlation between job finding probabilities and our measure of inequality. Building on these insights, we constructed a nonsequential search model, where firms can incur screening costs to perfectly discern the productivity type of a worker. Workers can apply for jobs while already employed, resulting in a mix of employed and unemployed job seekers. Labor market transitions significantly impact the composition of the employed workforce, making the productivity distribution of the employed an endogenous variable.

We take our model to the data by estimating a set of parameters using the generalized method of moments and performed various counterfactual experiments. These experiments revealed that the observed correlation between unemployment and wage inequality in the data is consistent with shifts in the distribution of productivity, influencing both its mean and percentile ratios.

Furthermore, our counterfactual analyses unveiled non-linearities in the relation between labor market variables and inequality of wages. Notably, our measure of the ranking gap showed that, within a given steady state equilibrium, the selective hiring process in the labor market results in a productivity distribution of the employed with a skewed left tail and a stretched right tail. This implies that ex ante inequalities in terms of productivity are accentuated for highly productive individuals, while the opposite holds for individuals with relatively lower productivity levels. Moreover, when exogenous parameters change, such as shifts in the distribution of productivities that alter its mean, this has different effects on different parts of the wage distribution. An increase in average productivity, while keeping the percentile ratios constant, leads to a reduction in the ranking gap at the left tail of the distribution (the 50/10 percentiles ratio) but an increase at the right tail (the 90/50 percentiles ratio). Consequently, an overall nearly neutral effect on measures like the 90/10 percentiles ratio masks distinct effects at different locations of the distribution.

Finally, we conducted an efficiency assessment. We considered the scenario of a Social Planner subject to the same technological constraints as the individual agents. Our findings suggest that a small unemployment rate plays an important role for efficiency because it makes it possible hiring selectivity in our model. Beyond a small unemployment rate, the effect of increasing selectivity is relatively less important. In that ballpark, achieving the optimal allocation within our baseline steady-state economy may require implementing either a regressive or a progressive tax and transfer schedule, depending on the objective of attaining a lower or higher unemployment rate. Specifically, if the optimal unemployment rate is lower than the competitive equilibrium market rate, a regressive tax and transfer schedule would be necessary. In such a scenario, the tax and transfer schedule, applied to firms and contingent upon workers' productivity levels (i.e., wages), entails providing subsidies to more productive workers while taxing less productive individuals. The motivation behind this approach lies in maintaining incentives for firms to hire their preferred candidates and the positive externalities derived from encouraging firms to create job opportunities for the most productive individuals.

Taking stock, our model offers a framework to understand how ex ante productivity differences in the labor market map into realized differences in the wage distribution. We highlight hiring selectivity mechanisms that link worker flows, composition, and inequality. By also taking the model to the data and using it for normative economic analysis, we hope to have made a stride in understanding the central role of labor markets in shaping inequality.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999, March). High Wage Workers and High Wage Firms. *Econometrica 2*(67), 251–333.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006). The polarization of the us labor market. *American economic review 96*(2), 189–194.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economics and Statistics 90*(2), 300–323.

Blanchard, O. and P. Diamond (1994). Ranking, unemployment duration, and wages. *Review of Economic Studies 61*(3), 417–434.

Boitier, V. and A. Lepetit (2018). Reduced form wage equations in the credible bargaining model. *Labour Economics 50*, 92–96.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics 41*(2), 291–322.

Bonhomme, S. and L. Hospido (2017, 03). The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data. *The Economic Journal 127*(603), 1244–1278.

Bradley, J. (2020). Worker-Firm Screening and the Business Cycle. Working paper, University of Nottingham.

Cai, X., P. Gautier, and R. Wolthoff (2025). Search, screening, and sorting. *American Economic Journal: Macroeconomics 17*(3), 205–236.

Card, D. and J. E. DiNardo (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of Labor Economics 20*(4), 733–783.

Cysne, R. P. (2009). On the positive correlation between income inequality and unemployment. *The Review of Economics and Statistics 91*(1), 218–226.

Cysne, R. P. and D. Turchick (2012). Equilibrium unemployment-inequality correlation. *Journal of Macroeconomics 34*(2), 454–469.

Davis, S. J. and B. Samaniego de la Parra (2024, April). Application flows. Working Paper 32320, National Bureau of Economic Research.

Eeckhout, J. and I. Lindenlaub (2019). Unemployment cycles. *American Economic Journal: Macroeconomics 11*(4), 175–234.

Engbom, N. (2021). Contagious unemployment. Technical report, National Bureau of Economic Research.

Fallick, B. and C. A. Fleischman (2004). Employer-to-employer flows in the u.s. labor market: the complete picture of gross worker flows. Finance and Economics Discussion Series 2004-34, Board of Governors of the Federal Reserve System (U.S.).

Fernández-Blanco, J. and E. Preugschat (2018). On the effects of ranking by unemployment duration. *European Economic Review 104*(C), 92–110.

Fitzgerald, J., P. Gottschalk, and R. Moffitt (1998). The impact of attrition in the panel study of income dynamics on intergenerational analysis. *Journal of Human Resources 33*, 300–344.

Flood, S., M. King, R. Rodgers, S. Ruggles, J. R. Warren, and M. Westberry (2022). Integrated public use microdata series, current population survey: Version 10.0. dataset.

Goldin, C. and L. F. Katz (2008). *The race between education and technology*. Harvard Univiersity Press.

Gottschalk, P. and M. Hansen (2003). Is the proportion of college workers in noncollege jobs increasing? *Journal of Labor Economics 21*(2), 449–471.

Hall, R. E. and P. R. Milgrom (2008, SEP). The Limited Influence of Unemployment on the Wage Bargain. *American Economic Review 98*(4), 1653–1674.

Heathcote, J., K. Storesletten, and G. L. Violante (2017). Optimal tax progressivity: An analytical framework. *Quarterly Journal of Economics*.

Hoffman, M., L. B. Kahn, and D. Li (2018). Discretion in hiring. *The Quarterly Journal of Economics 133*(2), 765–800.

Hoffman, M. and C. T. Stanton (2024). People, practices, and productivity: A review of new advances in personnel economics.

Hosios, A. J. (1990). On the Efficiency of Matching and Related Models of Search and Unemployment. *The Review of Economic Studies 57*(2), 279–298.

Huckfeldt, C. (2022, April). Understanding the scarring effect of recessions. *American Economic Review 112*(4), 1273–1310.

Jäntti, M. (1994). A more efficient estimate of the effects of macroeconomic activity on the distribution of income. *The Review of Economics and Statistics*, 372–378.

Kambourov, G. and I. Manovskii (2009). Occupational mobility and wage inequality. *Review of Economic Studies 76*(2), 731–759.

Kleiber, C. (2008). *A Guide to the Dagum Distributions*, pp. 97–117. New York, NY: Springer New York.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), pp. 1029–1053.

Landais, C., P. Michaillat, and E. Saez (2018, May). A macroeconomic approach to optimal unemployment insurance: Applications. *American Economic Journal: Economic Policy 10*(2), 182–216.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–145.

Merkl, C. and T. van Rens (2019). Selective hiring and welfare analysis in labor market models. *Labour Economics 57*(C), 117–130.

Mocan, H. N. (1999). Structural unemployment, cyclical unemployment, and income inequality. *Review of Economics and Statistics 81*(1), 122–134.

Moen, E. (1999). Education, Ranking, and Competition for Jobs. *Journal of Labor Economics 17*(4), 694.

Moore, M. P. and P. Ranjan (2005). Globalisation vs Skill-Biased Technological Change: Implications for Unemployment and Wage Inequality. *The Economic Journal 115*(503), 391–422.

Moscarini, G. and K. Thomsson (2007). Occupational and job mobility in the US. *The Scandinavian Journal of Economics 109*(4), 807–836.

Petrosky-Nadeau, N. and E. Wasmer (2015). Macroeconomic dynamics in a model of goods, labor, and credit market frictions. *Journal of Monetary Economics 72*, 97–113.

Petrosky-Nadeau, N. and E. Wasmer (2017). *Labor, Credit, and Goods Markets: The macroeconomics of search and unemployment*. MIT Press.

Pissarides, C. (2000). *Equilibrium Unemployment Theory* (Second ed.). The MIT Press.

Ravallion, M. (2018, June). Inequality and globalization: A review essay. *Journal of Economic Literature 56*(2), 620–42.

Saldarriaga, V. (2022). The distributional dynamics of wages over the business cycle. mimeo.

Sandvik, J., R. Saouma, N. Seegert, and C. T. Stanton (2024). Should workplace programs be voluntary or mandatory? evidence from a field experiment on mentorship. *forthcoming, Management Science*.

Shi, S. (2006, May). Wage differentials, discrimination and efficiency. *European Economic Review 50*(4), 849–875.

Shimer, R. (2005, Oct). The Assignment of Workers to Jobs in an Economy with Coordination Frictions. *Journal of Political Economy 113*(5), 996–1025.

Smith, E. (1999). Search, concave production, and optimal firm size. *Review of Economic Dynamics 2*(2), 456–471.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming Up Inequality. *The Quarterly Journal of Economics 134*(1), 1–50.

Stigler, G. (1961). The Economics of Information. *Journal of Political Economy 3*(69), 213–225.

van Ommeren, J. and G. Russo (2013). Firm Recruitment Behaviour: Sequential or Non-sequential Search? *Oxford Bulletin of Economics and Statistics*, 1–24.

van Ours, J. and G. Ridder (1992). Vacancies and the Recruitment of New Employees. *Journal of Labor Economics 10*(2), 138–155.

Villena-Roldan, B. (2010a). Aggregate Implications of Employer Search and Recruiting Selection. Working Paper 271, Center for Applied Economics, University of Chile.

Villena-Roldan, B. (2010b). Wage Dispersion and Recruiting Selection. Working Paper 270, Center for Applied Economics, University of Chile.

Wang, B. (2020). A note on the unemployment volatility puzzle: Is credible wage bargaining the answer? *Macroeconomic Dynamics*.

Wolthoff, R. (2017, 08). Applications and Interviews: Firms' Recruiting Decisions in a Frictional Labour Market. *The Review of Economic Studies 85*(2), 1314–1351.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (Second ed.). The MIT Press.

Online Appendix

Appendix A Deriving Entry Condition (16)

The entry condition can be written as

$$-\kappa+\chi+\beta\left\{\sum_{k=1}^{\infty}\frac{e^{-\phi q}(\phi q)^k}{k!}\left[-\xi k+\int_0^{\infty}J(v)k(\phi G_A(v)+(1-\phi))^{k-1}g_A(v)dv\right]+e^{-\phi q}V\right\}=V \tag{28}$$

In equilibrium, $V=0$. We also assume that the value of hiring the first worker surpasses the interviewing cost, e.g., $E[J(\theta)]=\int J(\theta)dG_A(\theta)>\xi$. This condition ensures that the employer expects to hire a worker rather than discarding the pool of applicants received in one period, even if the employer gets only one applicant.

Doing some algebra helps us show that

$$\begin{aligned}
\kappa+\chi+\beta\xi\phi q&=\beta\left(\sum_{k=1}^{\infty}\frac{e^{-\phi q}(\phi q)^k}{k!}\int_0^{\infty}J(v)k(\phi G_A(v)+(1-\phi))^{k-1}g_A(v)dv\right)\\
\kappa+\chi+\beta\xi\phi q&=\beta\left(\sum_{k=1}^{\infty}\frac{e^{-\phi q}(\phi q)^{k-1}}{(k-1)!}\int_0^{\infty}J(v)\phi q(\phi G_A(v)+(1-\phi))^{k-1}g_A(v)dv\right)\\
\kappa+\chi+\beta\xi\phi q&=\beta\left(\int_0^{\infty}J(v)\phi q e^{-\phi q(1-G_A(v))}dG_A(v)\right)\\
\kappa+\chi+\beta\xi\phi q&=\beta\phi q\left(\int_0^{1}J(G^{-1}(x;q))e^{-\phi q(1-x)}dx\right)
\end{aligned}$$

For the last step, we use the inverse quantile mapping in (9) to replace the unknown distribution of applicants G_A by the population distribution G, a primitive of the model.

Appendix B Deriving the wage equation (18)

We follow the methodology of Boitier and Lepetit (2018) and their notation: in particular, τ represents the length of the time period in which bargaining takes place (when parties can alternate their wage offers), δ represents the hazard that negotiation breaks down during bargaining, and we consider the approximation $\beta\approx e^{-r}$.

$$
\begin{aligned}
J &= \theta - w + e^{-r}\left[\lambda \tilde{p}(\theta,q)V + (1-\lambda)\eta V + (1-(1-\lambda)\eta - \lambda\tilde{p}(\theta,q))\, J(\theta)\right] && (29)\\
W &= w + e^{-r}[\lambda\left(\tilde{p}(\theta,q)\widetilde{W}(\theta) + (1-\tilde{p}(\theta,q))W(\theta)\right) + (1-\lambda)\left((1-\eta)W(\theta) + \eta U(\theta)\right)] && (30)\\
U &= \rho\theta + e^{-r}[\tilde{p}(\theta,q)W(\theta) + (1-\tilde{p}(\theta,q))U(\theta)] && (31)\\
J^{w} &= \theta - w + e^{-r}\left[\lambda \tilde{p}(\theta,q)V + (1-\lambda)\eta V + (1-(1-\lambda)\eta - \lambda\tilde{p}(\theta,q))\, J(\theta)\right] && (32)\\
J^{w'} &= -\gamma\theta\tau + e^{-(r+\delta)\tau} J^{w} && (33)\\
J^{w'} &= \theta - w' + e^{-r}\left[\lambda \tilde{p}(\theta,q)V + (1-\lambda)\eta V + (1-(1-\lambda)\eta - \lambda\tilde{p}(\theta,q))\, J(\theta)\right] && (34)\\
W^{w} &= \rho\theta\tau + e^{-r\tau}[(1-e^{-\delta\tau})U + e^{-\delta\tau}W^{w'}] && (35)\\
W^{w} &= w + e^{-r}[\lambda\left(\tilde{p}(\theta,q)\widetilde{W}(\theta) + (1-\tilde{p}(\theta,q))W(\theta)\right) + (1-\lambda)\left((1-\eta)W(\theta) + \eta U(\theta)\right)] && (36)\\
W^{w'} &= w' + e^{-r}[\lambda\left(\tilde{p}(\theta,q)\widetilde{W}(\theta) + (1-\tilde{p}(\theta,q))W(\theta)\right) + (1-\lambda)\left((1-\eta)W(\theta) + \eta U(\theta)\right)] && (37)
\end{aligned}
$$

In equilibrium we have that $V = 0$ and that $\widetilde{W}(\theta) = W(\theta)$. To obtain the wage equation, first we combine equations (35) and (36) and use (37) to substitute for $W^{w'}$. Then we combine equations (33), (34) and (32). We use the two expressions found to obtain the sharing rule $W - \frac{\rho\theta}{r+\delta} - \frac{\delta U}{r+\delta} = J + \frac{\gamma\theta}{r+\delta}$ (as in Boitier and Lepetit (2018)) and then we use it to find the wage equation in steady state under the hypothesis that $\delta = \eta^*$.

Appendix C Proofs omitted in the main text

A Proof of Proposition 1

Proof. By the Cauchy Mean Value Theorem, we have

$$\frac{G_E(\overline{\theta}) - G_E(\underline{\theta})}{G(\overline{\theta}) - G(\underline{\theta})} = \frac{g_E(\theta_c)}{g(\theta_c)} \text{ with } \theta_c \in (\underline{\theta}, \overline{\theta})$$

in which θ_c is referred to as the Cauchy Mean Value (CMV).

Using the definitions of the density of the employed, the previous expression can be written as

$$\frac{g_E(\theta)}{g(\theta)} = \frac{p(G_A(\theta_c); q)}{\eta^* + p(G_A(\theta_c); q)} \frac{1}{1 - \mathcal{U}(q)} \text{ with } \theta_c \in (\underline{\theta}, \overline{\theta})$$

where $x_c = G_A(\theta_c)$ is the ranking in the applicants' distribution of CMV type θ_c.

On the other hand, the aggregate level of employment $1 - \mathcal{U}$ can be written in terms of the mass of applicants $\mathcal{A}$ as $1 - \mathcal{U}(q) = \frac{1-\mathcal{A}(q)}{1-\lambda}$

We can express the density function of the applicants' distribution as the sum of the unemployed distribution and a share λ of the employed. Hence,

$$g_A(\theta) = \frac{g(\theta)}{\mathcal{U}(q) + \lambda(1 - \mathcal{U}(q))} \frac{\eta^* + \lambda p(G_A(\theta); q)}{\eta^* + p(G_A(\theta); q)}$$

where we define $s_A(G_A(\theta)) \equiv \frac{\eta^*+\lambda p(G_A(\theta);q)}{\eta^*+p(G_A(\theta);q)}$.

By integrating the density of the applicants, we obtain

$$\mathcal{A}(q) = \int_0^1 s_A(G_A(\theta);q)dG(\theta)$$

By the mean value theorem for integrals, there exists a value $\theta^* \in (0,\infty)$ such that $\mathcal{A}(q) = s_A(G_A(\theta^*))$

Remembering that $x \equiv G_A(\theta)$, we can therefore write $\mathcal{A}(q) = s_A(G_A(\theta^*)) = s_A(x^*)$ where x^* is the average applicant type.

We then substitute the expression for the applicants' mass in the definition of the employment rate:

$$\begin{aligned} 1-\mathcal{U}(q) &= \frac{1-\mathcal{A}(q)}{1-\lambda} \\ &= \frac{1-s_A(G_A(\theta^*))}{1-\lambda} = \frac{1-\frac{\eta^*+\lambda p(G_A(\theta^*);q)}{\eta^*+p(G_A(\theta^*);q)}}{1-\lambda} = \frac{p(G_A(\theta^*);q)}{\eta^*+p(G_A(\theta^*);q)} = s_E(x^*;q) \end{aligned}$$

Using the latter, we obtain the result stated in Equation (19). ■

B Unemployment rate and queue length

Proposition 2 ***Increasing unemployment rate*** *The unemployment rate is increasing in q.*

Proof.

$$\mathcal{A} = \frac{1}{m(1,q)-m(0,q)}$$

with

$$m(x,q) \equiv x + \frac{1-\lambda}{\lambda q}\log\left(\eta + \lambda e^{-q(1-x)}\right)$$

We have to compute the derivative

$$\frac{\partial(1/\mathcal{A})}{\partial q} = -\frac{1-\lambda}{\lambda q^2}\log\left(\frac{\eta+\lambda}{\eta+\lambda e^{-q}}\right) + \frac{1-\lambda}{\lambda q}\frac{\lambda e^{-q}}{\eta+\lambda e^{-q}}$$

Using the Mean Value Theorem we realize that the term with a logarithm in the previous expression is a difference evaluated at 1 and 0, so it can be written as

$$\log(\eta+\lambda e^{-q(1-1)}) - \log(\eta+\lambda e^{-q(1-0)}) = \frac{q\lambda e^{-q(1-\bar{x})}}{\eta+\lambda e^{-q(1-\bar{x})}}(1-0) \text{ with } \bar{x}\in(0,1)$$

Hence, the original derivative can be expressed as

$$-\frac{1-\lambda}{\lambda q}\left(\frac{\lambda e^{-q}}{\eta+\lambda e^{-q}} - \frac{\lambda e^{-q(1-\bar{x})}}{\eta+\lambda e^{-q(1-\bar{x})}}\right)$$

$$-\frac{1-\lambda}{q}\left(\frac{1}{\eta e^{q}+\lambda}-\frac{1}{\eta e^{q(1-\bar{x})}+\lambda}\right)<0$$

where the last expression holds because $\frac{1}{\eta e^{q}+\lambda}$ is strictly decreasing in q.

Since $\frac{\partial(1/\mathcal{A})}{\partial q}<0$, it follows that $\frac{\partial\mathcal{A}}{\partial q}>0$ and also $\frac{\partial\mathcal{U}}{\partial q}>0$ because $\mathcal{U}=\frac{\mathcal{A}-\lambda}{1-\lambda}$.

■

If on-the-job search is prevalent in the labor market, i.e., λ is high, hiring selectivity q has less of an effect on the unemployment rate. In the extreme scenario of $\lambda=1$, i.e., all workers apply, selectivity does not matter as it does not affect the composition of the unemployed pool.

Appendix D Calibrating screening costs

To construct the annual annual recruitment and selection cost (RSC) and the annual number of interviews over the last two years (NRI) we use the following survey questions:

Q29: What percent of total labor costs is spent annually on the recruitment and selection of employees?

Q3: What was the total labor cost used in the production of your 1996 sales?

Q30A: How many people have you hired in the past two years?

Q41: How many candidates do you interview for each [JOB TITLE] opening?

We define $RSC=Q29/100\times Q3$ and $NRI=Q30A/2\times Q41$, assuming each position is filled after interviewing $Q41$ applicants on average.

The occupation codes used to compute the adjustment factor are:

Occupation codes 1992–2002:

- 8: Personnel and labor relations workers.
- 27: Personal, training and labor relation specialists.

Occupation codes 2003–2010:

- 130: Human Resources Managers.
- 620: Human resources, training, and labor relations specialists.
- 5360: Human resources assistants, except payroll and timekeeping.

Occupation codes since 2011:

- 136: Human Resources Managers.

- 630: Human resource workers.

- 5360: Human resources assistants, except payroll and timekeeping.

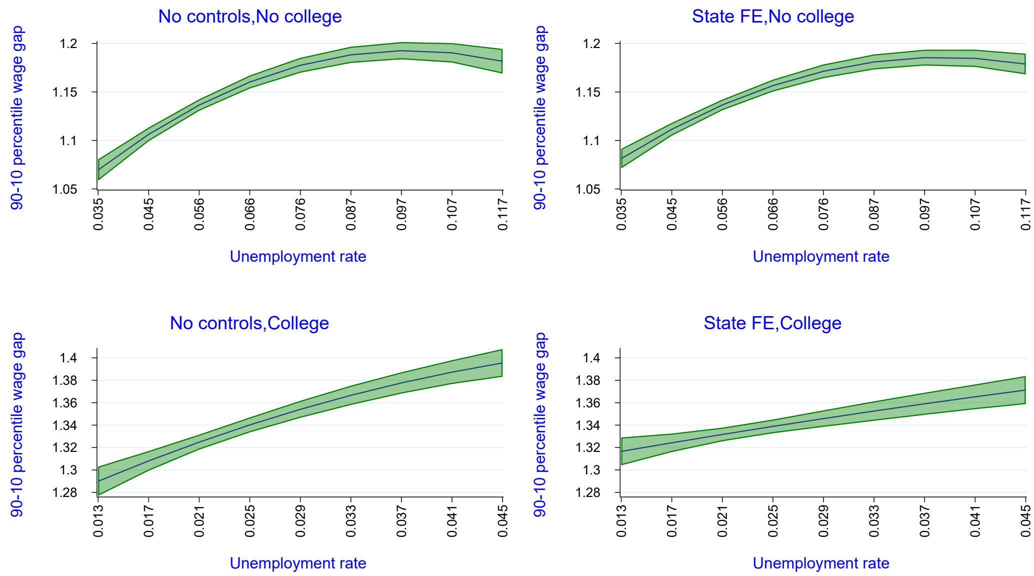
Appendix E Social planner problem

$$
\begin{aligned}
Y(\mathcal{U}) &= \int_0^\infty \theta((1-\mathcal{U})g_E(\theta) + \rho\mathcal{U}g_U(\theta))d\theta \\
&= \int_0^\infty \theta \left((1-\mathcal{U}))\frac{g(\theta)p(G_A(\theta))}{(1-\mathcal{U})(\eta(1-\lambda)+p(G_A(\theta)))} + \rho\mathcal{U}\frac{g(\theta)\eta(1-\lambda)}{\mathcal{U}(\eta(1-\lambda)+p(G_A(\theta)))}\right) d\theta \\
&= \int_0^\infty \theta\frac{g(\theta)(p(G_A(\theta)) + \rho\eta(1-\lambda))}{\eta(1-\lambda)+p(G_A(\theta))}d\theta \\
&= \int_0^\infty \theta g(\theta)\left(1 - \frac{(1-\rho)\eta(1-\lambda)}{p(G_A(\theta))+\eta(1-\lambda)}\right)d\theta \\
&= E[\theta] - (1-\rho)\int_0^\infty \theta g(\theta)\frac{\eta(1-\lambda)}{p(G_A(\theta))+\eta(1-\lambda)} \\
&= E[\theta] - (1-\rho)\int_0^\infty \theta g_U(\theta)\mathcal{U}d\theta \\
&= E[\theta] - (1-\rho)\mathcal{U}\int_0^\infty \theta g_U(\theta)d\theta
\end{aligned}
$$

Appendix F Robustness of empirical facts

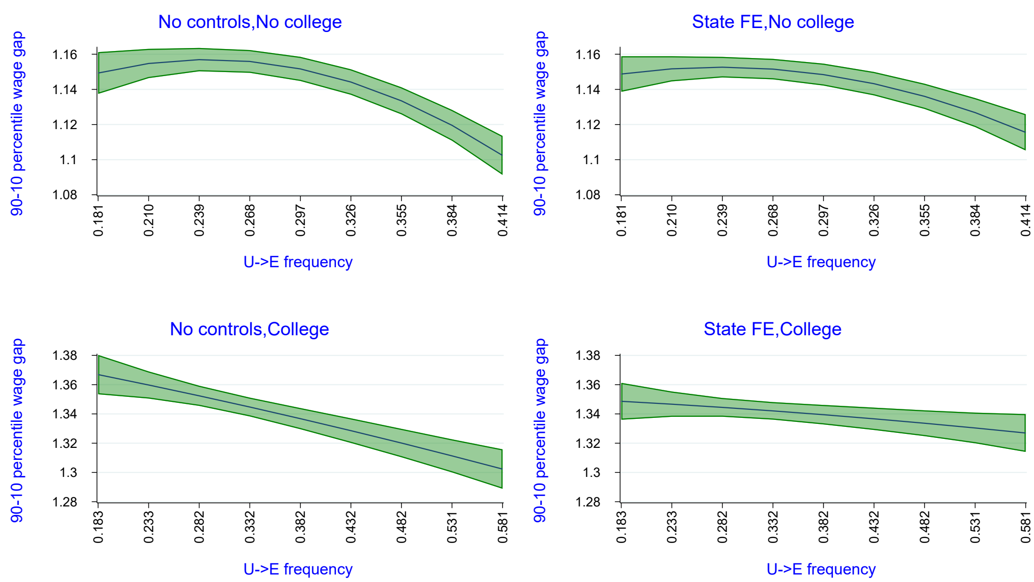
A Robustness to 1994 CPS sampling redesign

Figure 17: 90–10 log wage percentile gap vs. unemployment rate by state & year



Sample 1976-2019. US real log hourly wages.

Figure 18: 90–10 log wage percentile gap vs. UE frequency by state & year



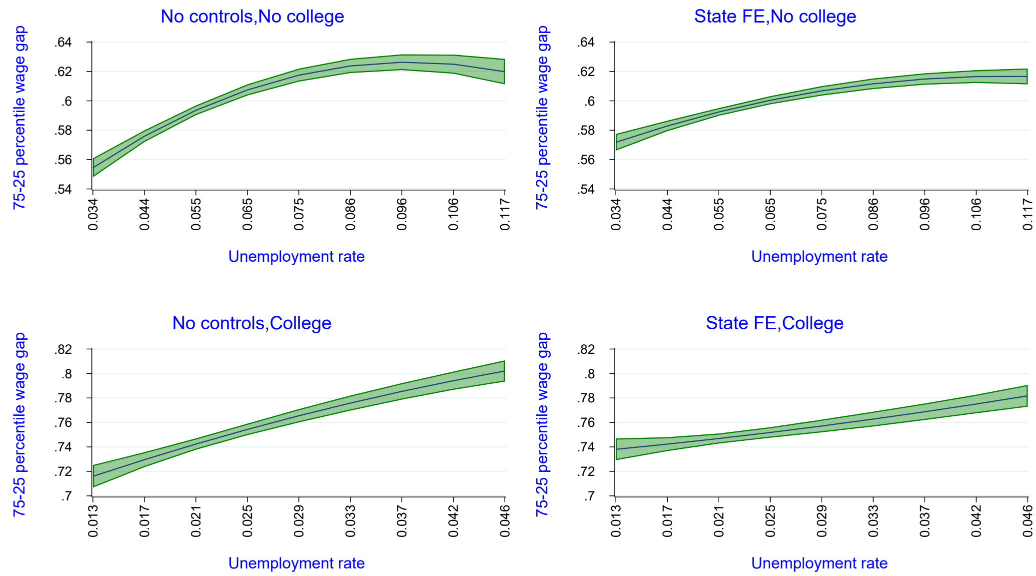
Sample 1976-2019. US real log hourly wages.

Figure 19: 90–10 log wage percentile gap vs. EU frequency by state & year

Sample 1976-2019. US real log hourly wages.

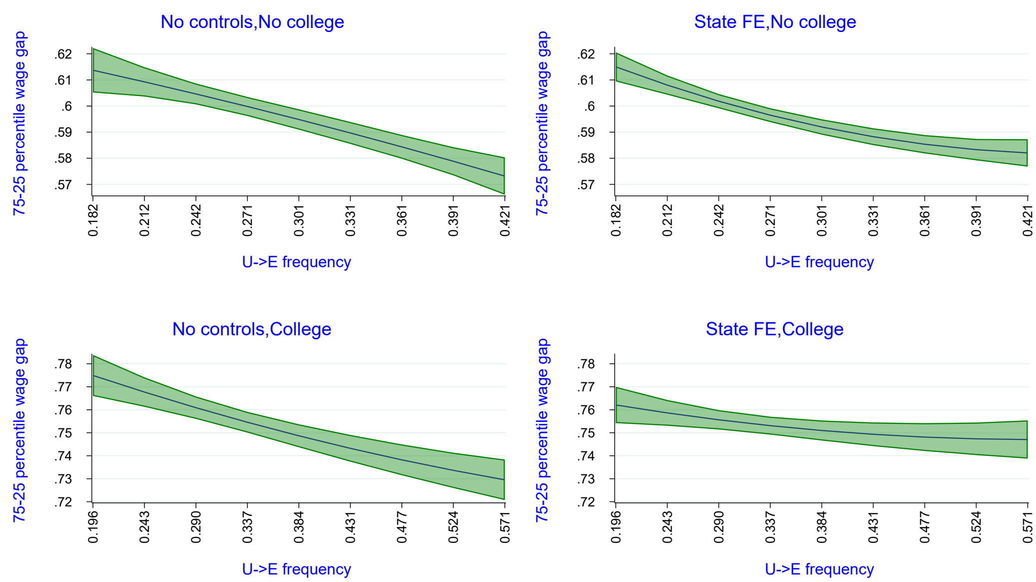
B Alternative measures of dispersion

Figure 20: 75–25 percentile log wage gap vs. unemployment rate by state & year



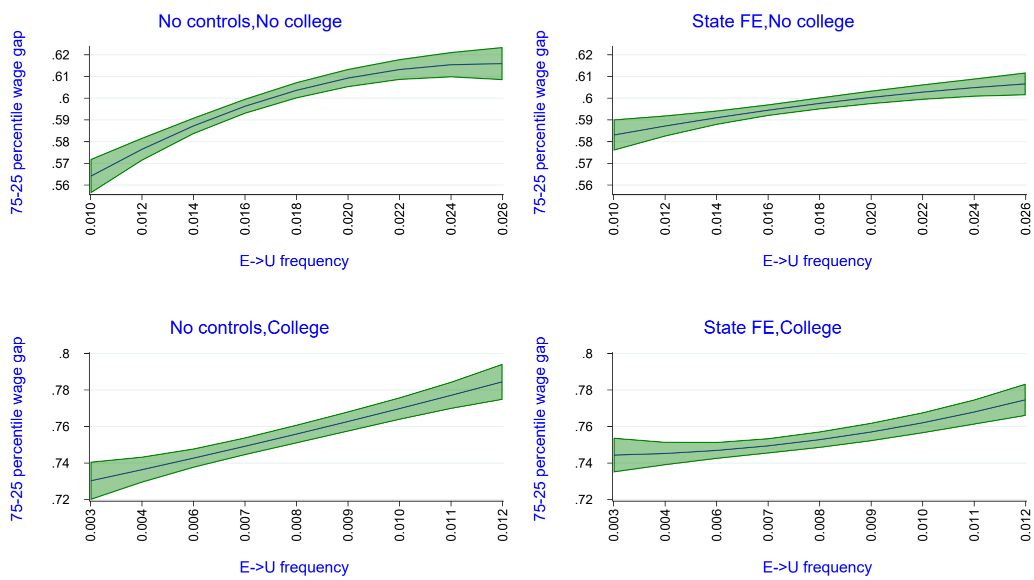
Sample 1994-2019. US real log hourly wages.

Figure 21: 75–25 percentile log wage gap vs. UE frequency by state & year



Sample 1994-2019. US real log hourly wages.

Figure 22: 75–25 percentile log wage gap vs. EU frequency by state & year

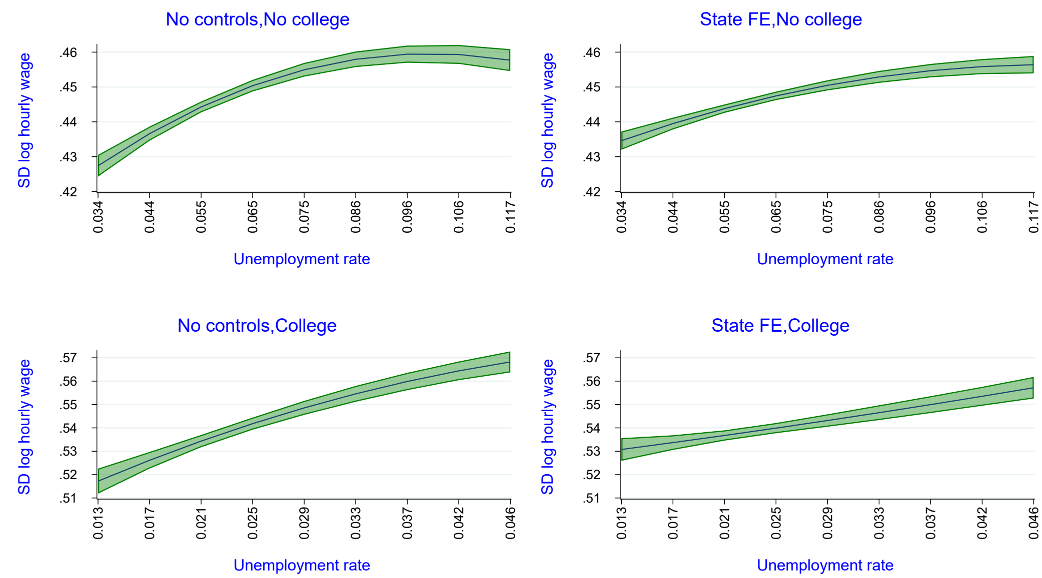


Sample 1994-2019. US real log hourly wages.

Figure 23: 75–25 percentile log wage gap vs. JJ frequency by state & year

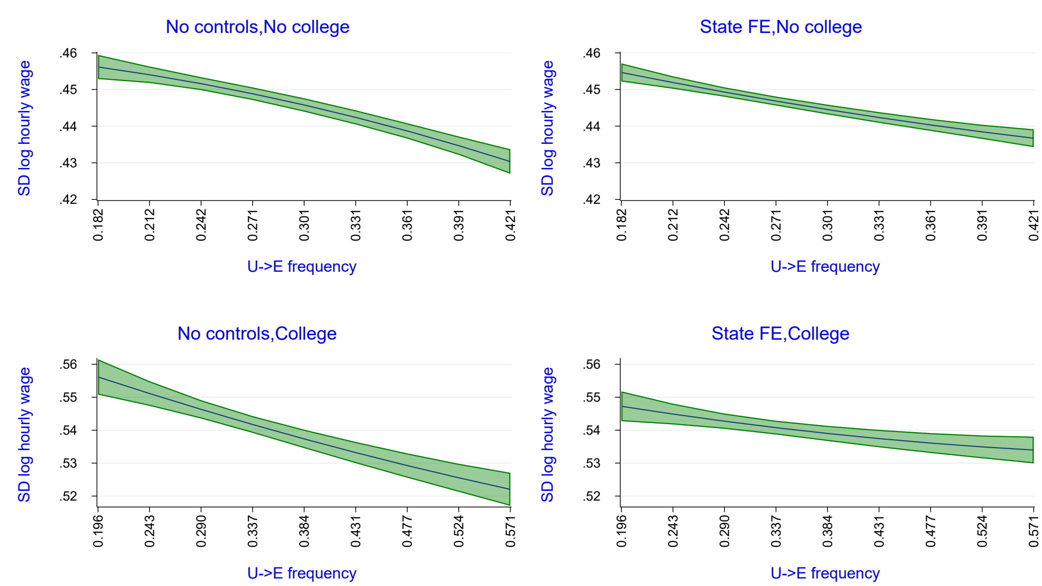
Sample 1994-2019. US real log hourly wages.

Figure 24: Standard deviation log wage vs. unemployment rate by state & year



Sample 1994-2019. US real log hourly wages.

Figure 25: Standard deviation log wage vs. UE frequency by state & year



Sample 1994-2019. US real log hourly wages.

Figure 26: Standard deviation log wage vs. EU frequency by state & year

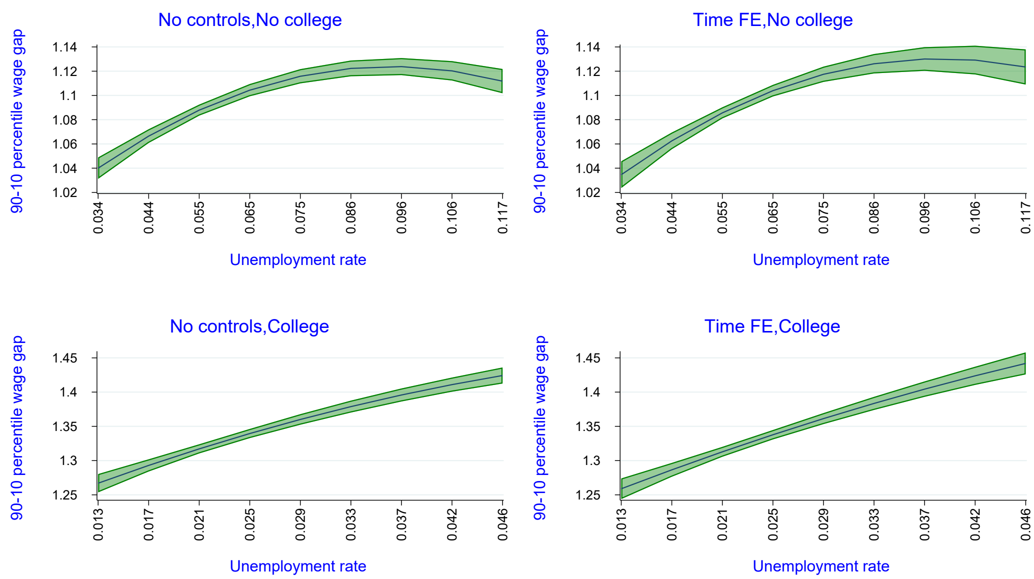
Sample 1994-2019. US real log hourly wages.

Figure 27: Standard deviation log wage vs. JJ frequency by state & year

Sample 1994-2019. US real log hourly wages.

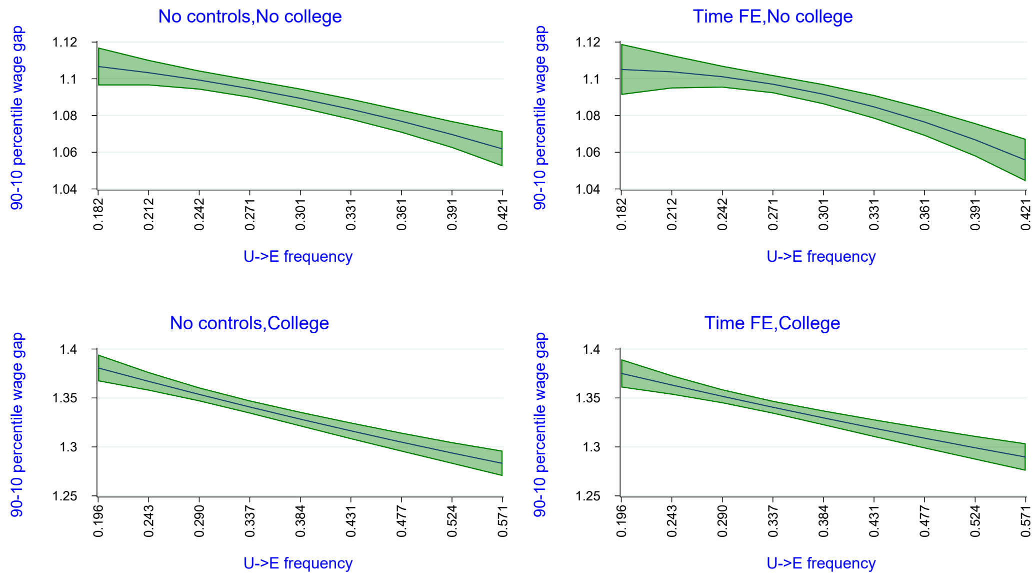
C Time fixed effects

Figure 28: 90–10 percentile wage gap vs. unemployment rate by state & year



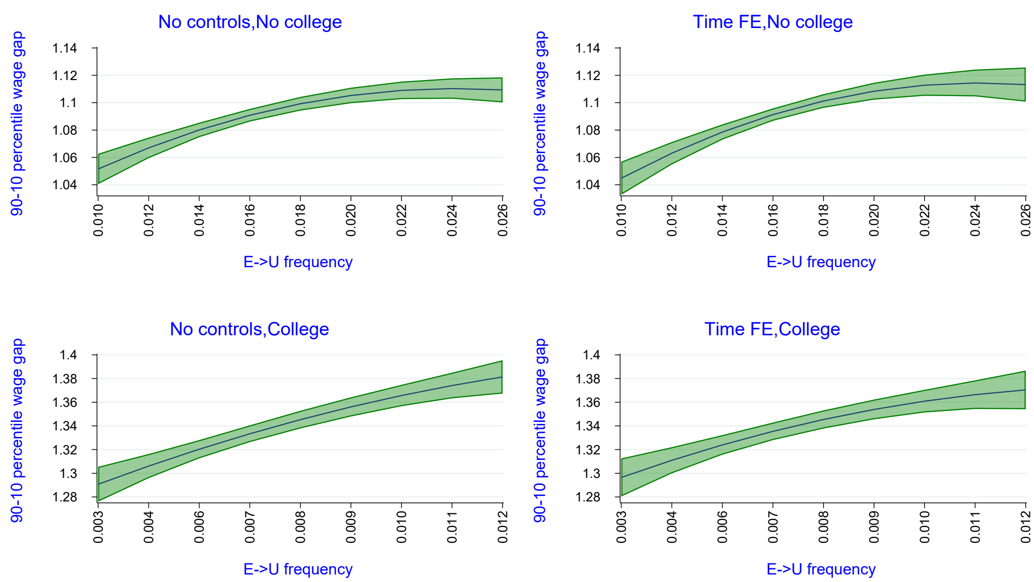
Sample 1994-2019. US real log hourly wages.

Figure 29: 90–10 percentile wage gap vs. UE frequency by state & year



Sample 1994-2019. US real log hourly wages.

Figure 30: 90–10 percentile wage gap vs. EU frequency by state & year



Sample 1994-2019. US real log hourly wages.

Figure 31: 90–10 percentile wage gap vs. JJ frequency by state & year

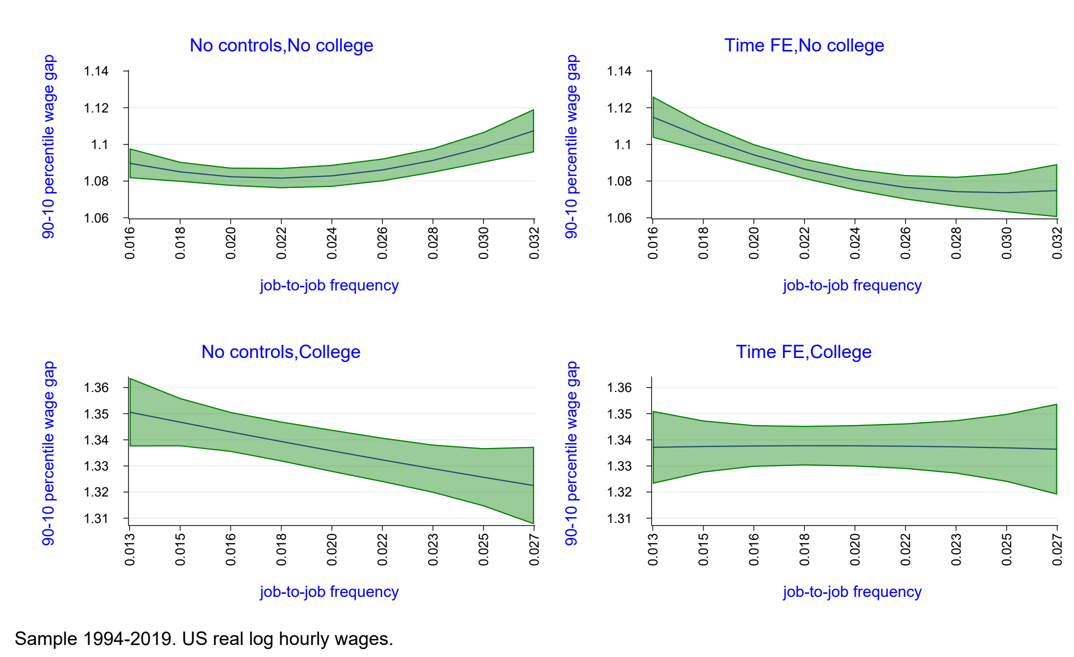


Figure 32: 75–25 percentile wage gap vs. unemployment rate by state & year

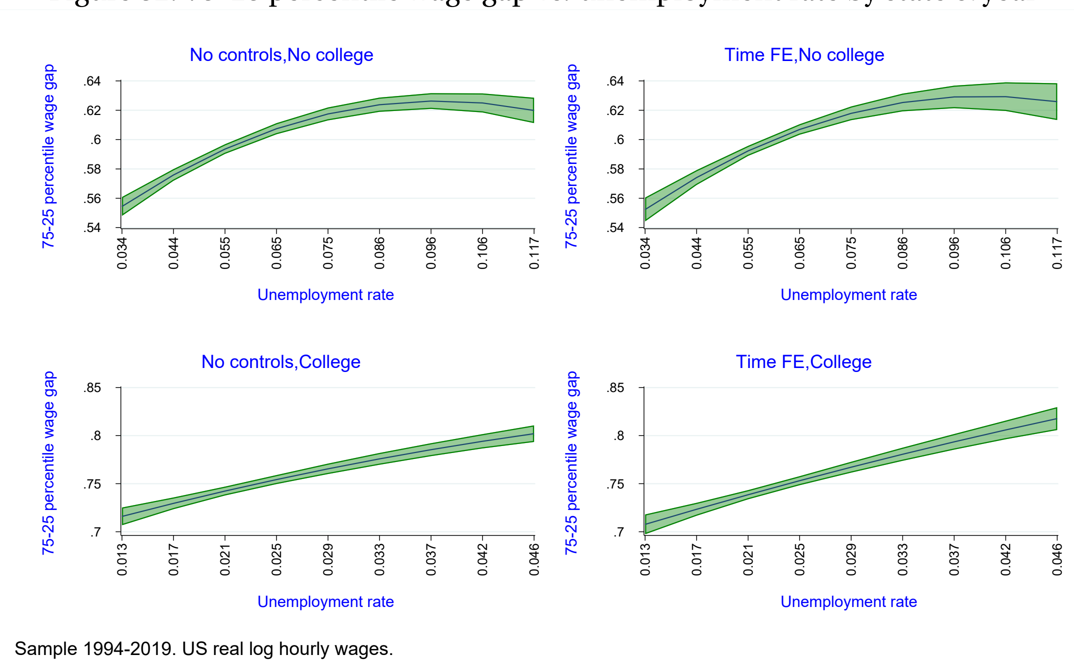
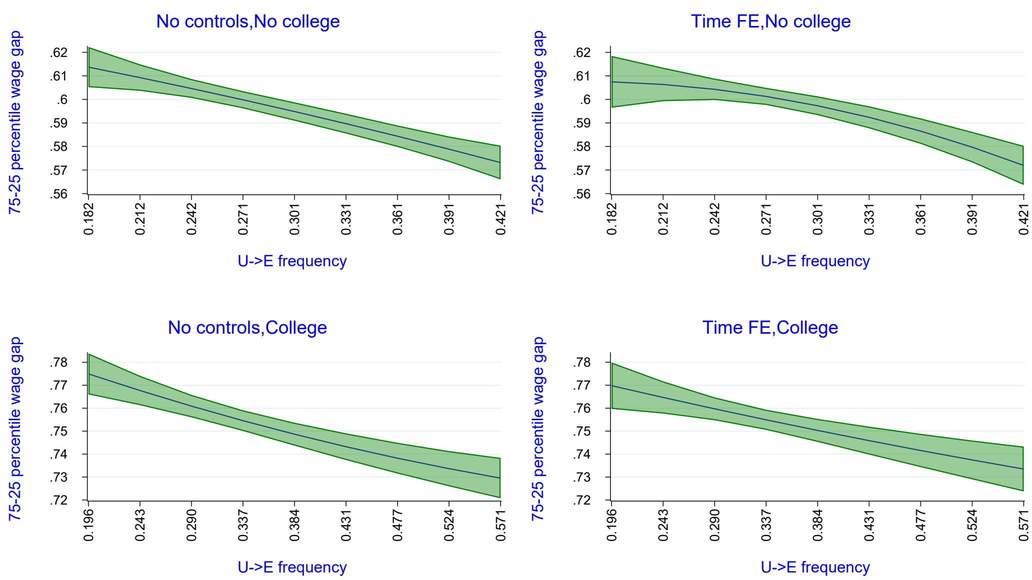
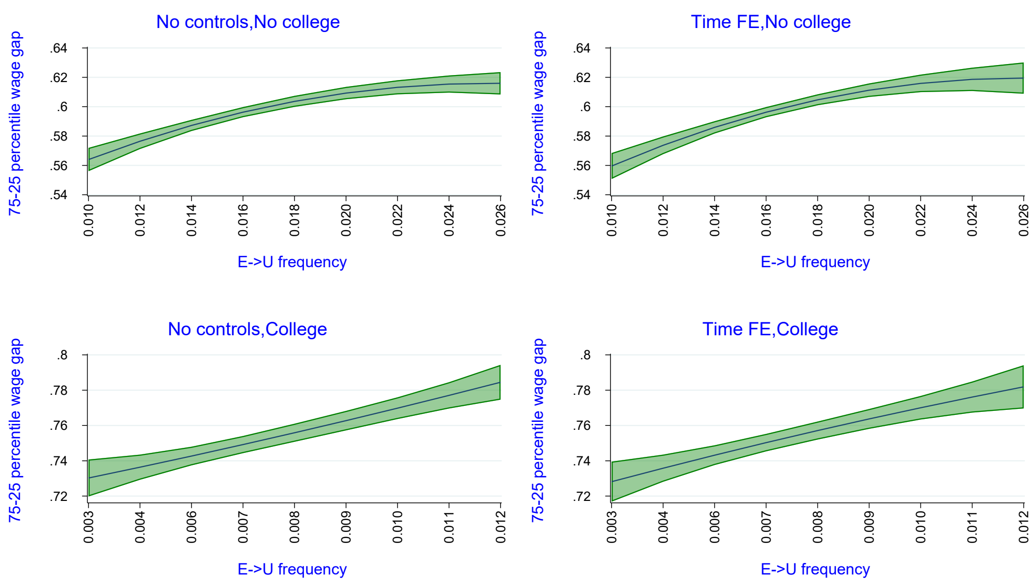


Figure 33: 75–25 percentile wage gap vs. UE frequency by state & year



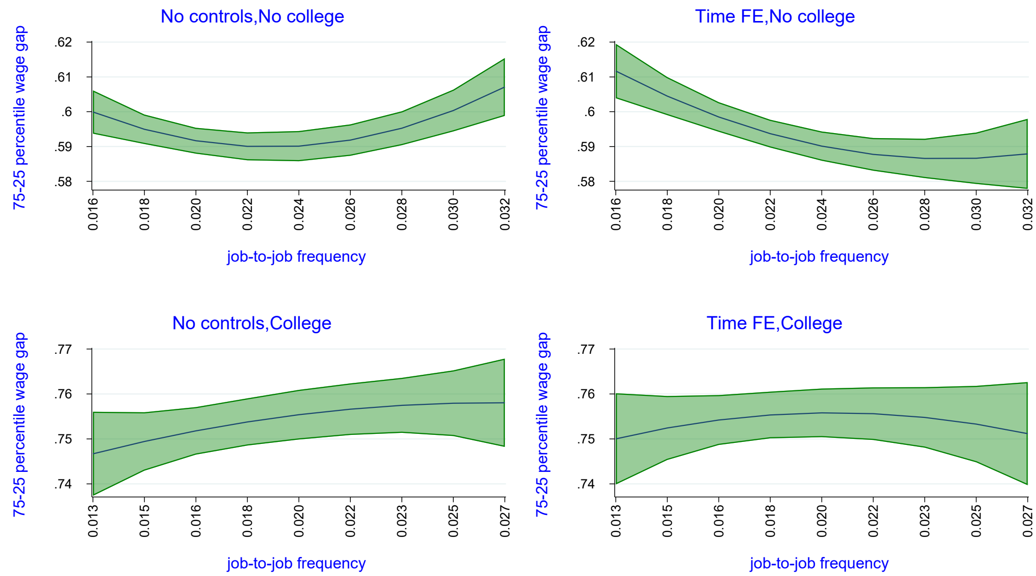
Sample 1994-2019. US real log hourly wages.

Figure 34: 75–25 percentile wage gap vs. EU frequency by state & year



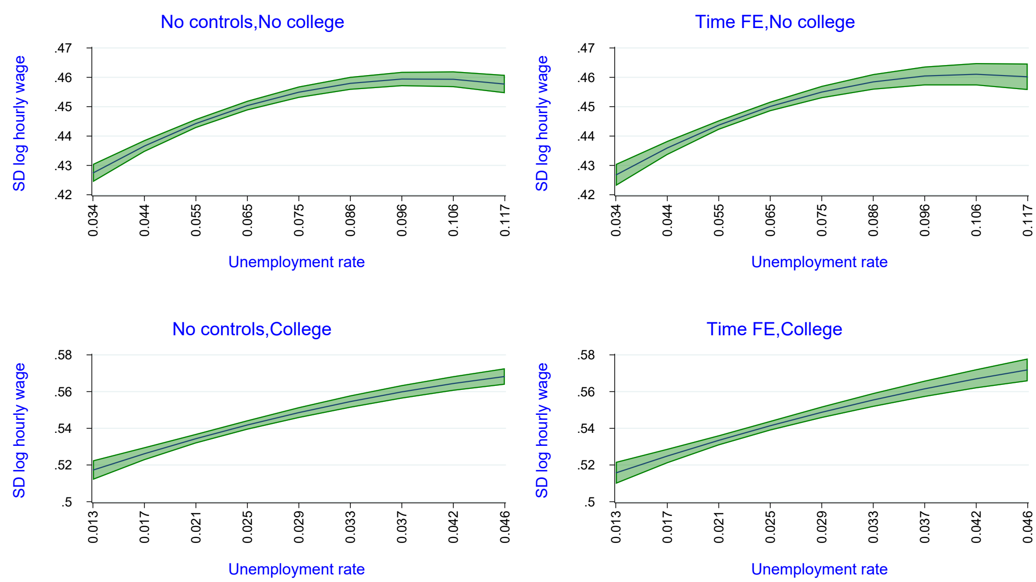
Sample 1994-2019. US real log hourly wages.

Figure 35: 75–25 percentile wage gap vs. JJ frequency by state & year



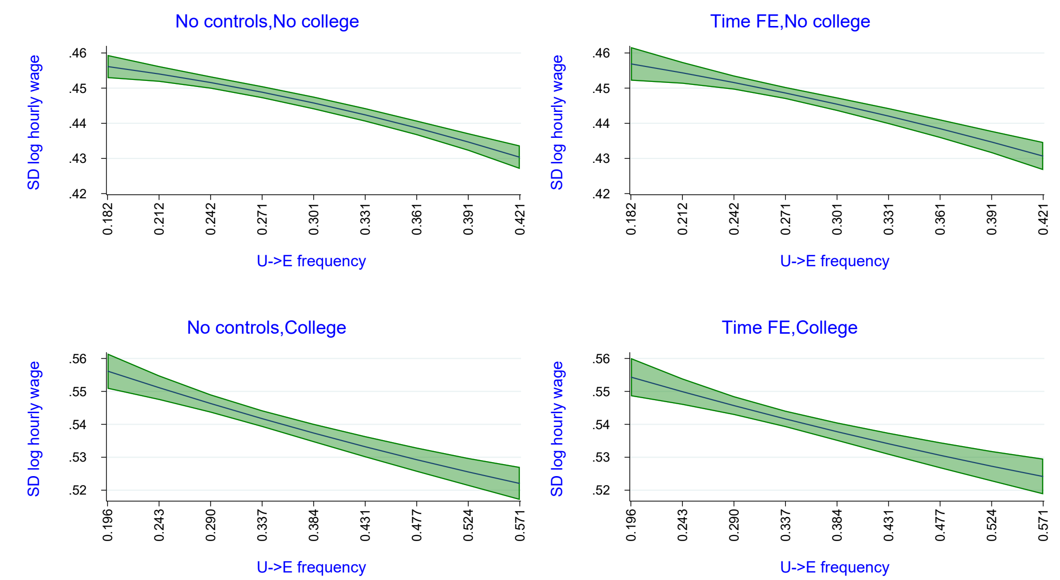
Sample 1994-2019. US real log hourly wages.

Figure 36: Standard deviation log wage vs. unemployment rate by state & year



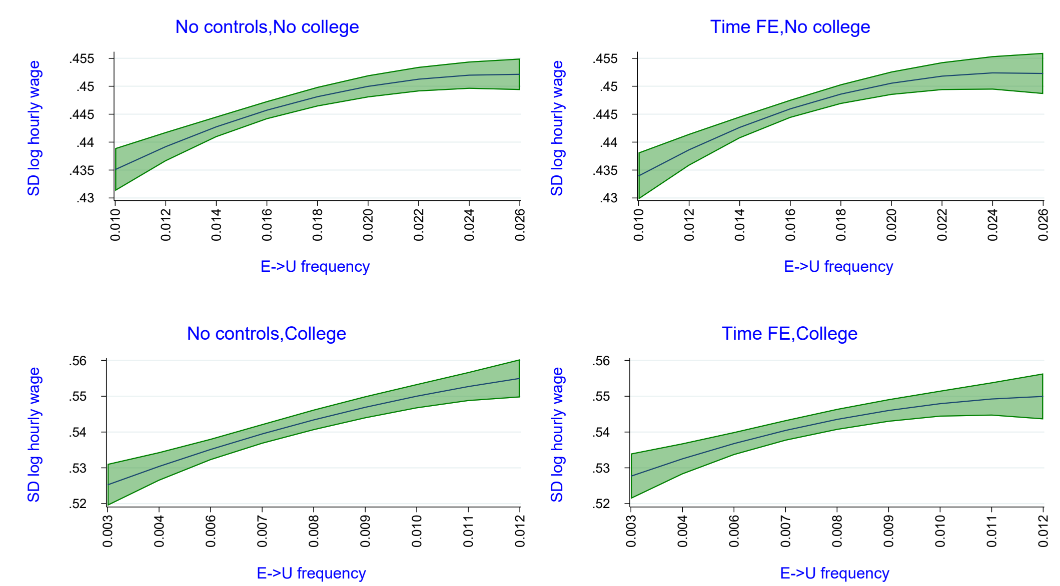
Sample 1994-2019. US real log hourly wages.

Figure 37: Standard deviation log wage vs. UE frequency by state & year



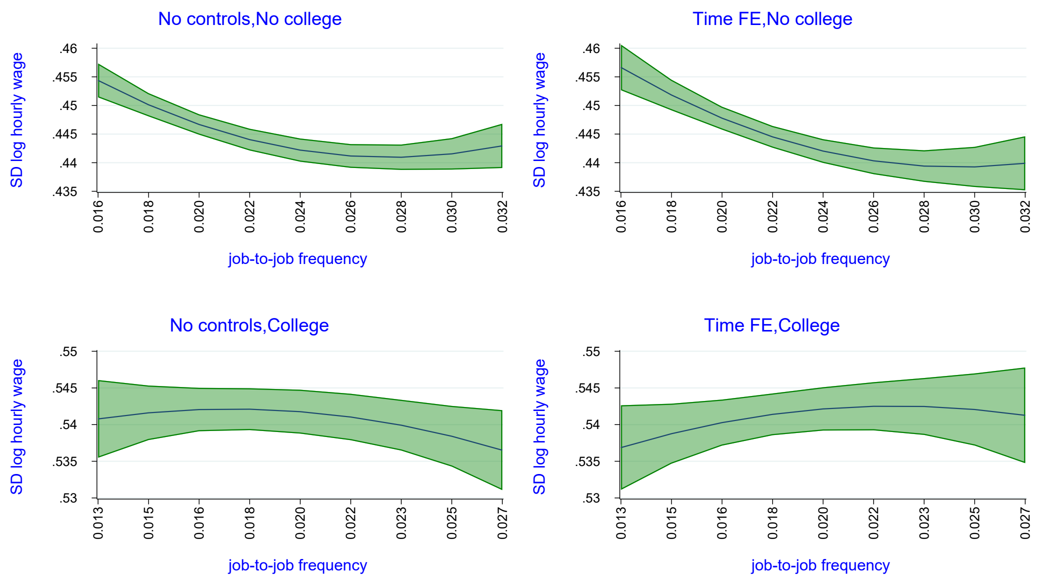
Sample 1994-2019. US real log hourly wages.

Figure 38: Standard deviation log wage vs. EU frequency by state & year



Sample 1994-2019. US real log hourly wages.

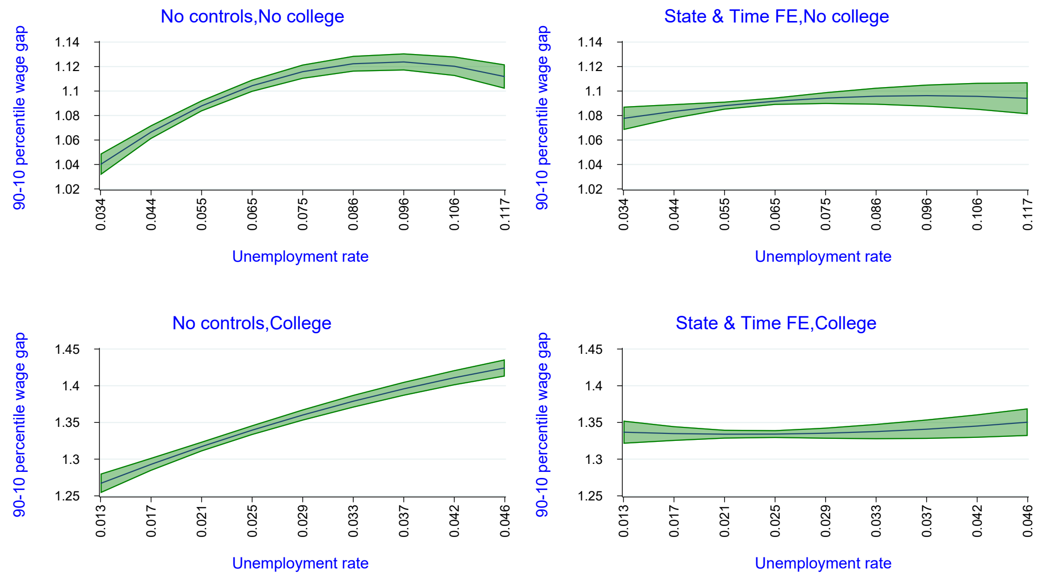
Figure 39: Standard deviation log wage vs. JJ frequency by state & year



Sample 1994-2019. US real log hourly wages.

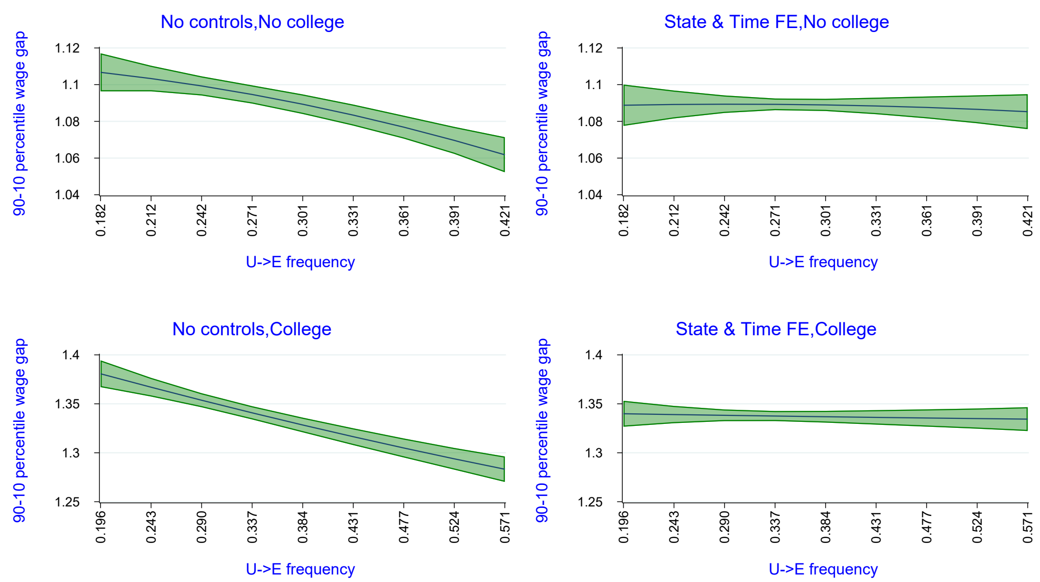
D Time and State fixed effects

Figure 40: 90–10 percentile wage gap vs. unemployment rate by state & year



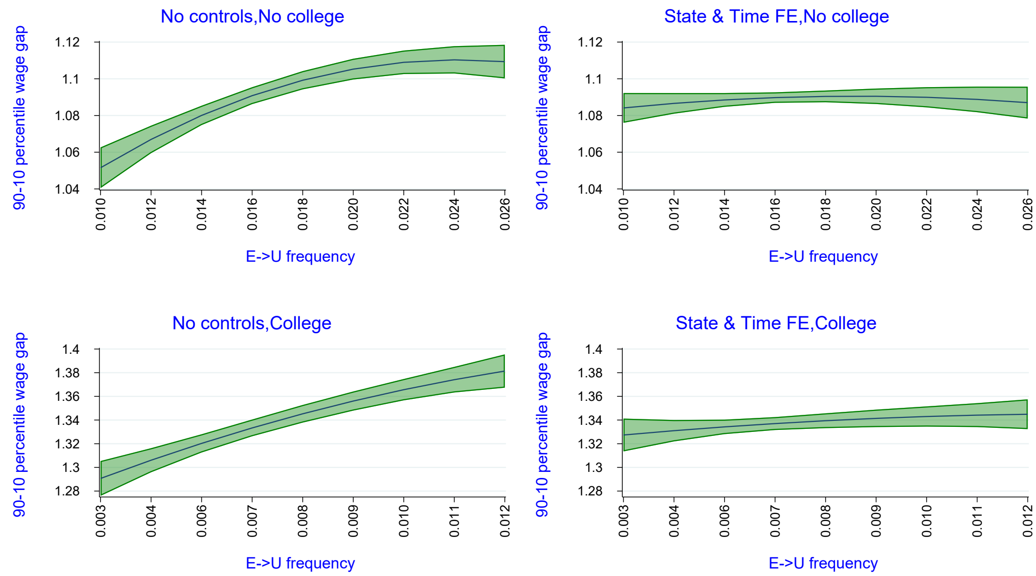
Sample 1994-2019. US real log hourly wages.

Figure 41: 90–10 percentile wage gap vs. UE frequency by state & year



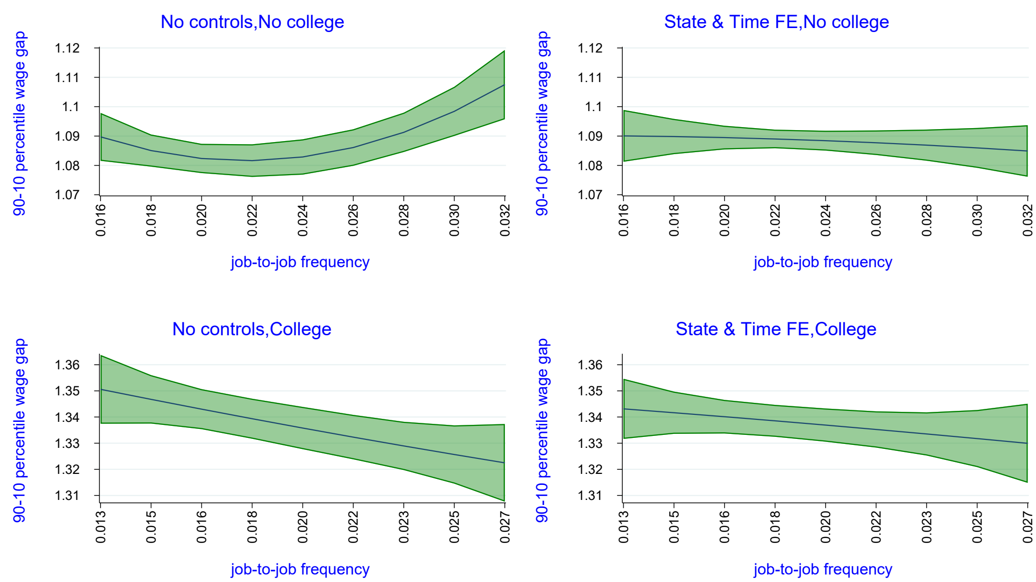
Sample 1994-2019. US real log hourly wages.

Figure 42: 90–10 percentile wage gap vs. EU frequency by state & year



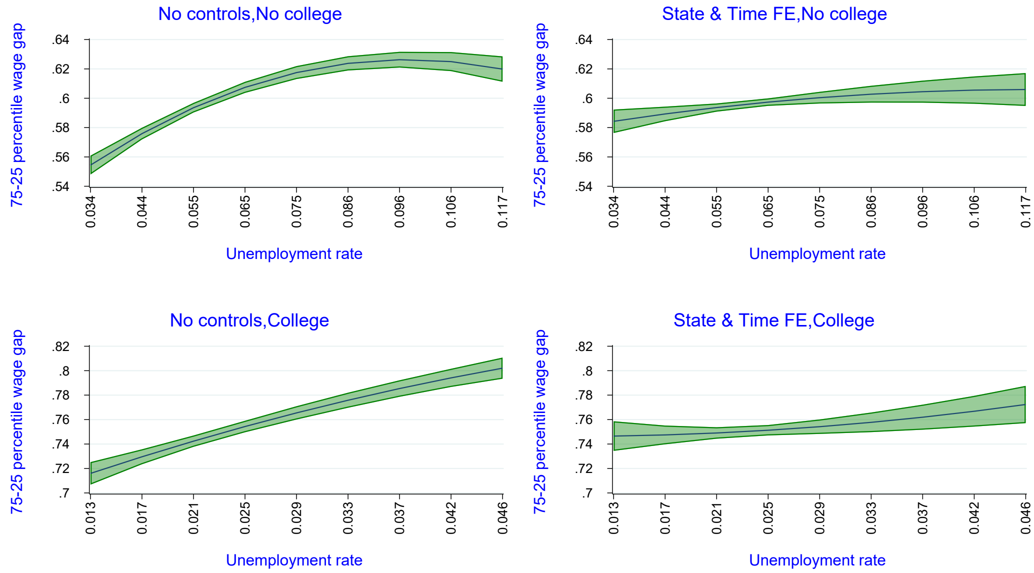
Sample 1994-2019. US real log hourly wages.

Figure 43: 90–10 percentile wage gap vs. JJ frequency by state & year



Sample 1994-2019. US real log hourly wages.

Figure 44: 75–25 percentile wage gap vs. unemployment rate by state & year

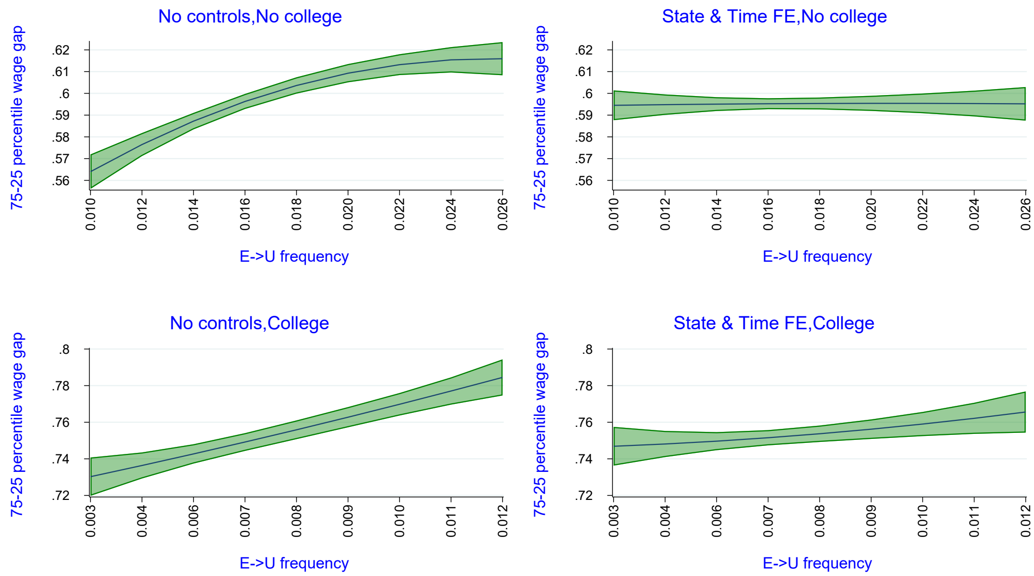


Sample 1994-2019. US real log hourly wages.

Figure 45: 75–25 percentile wage gap vs. UE frequency by state & year

Sample 1994-2019. US real log hourly wages.

Figure 46: 75–25 percentile wage gap vs. EU frequency by state & year

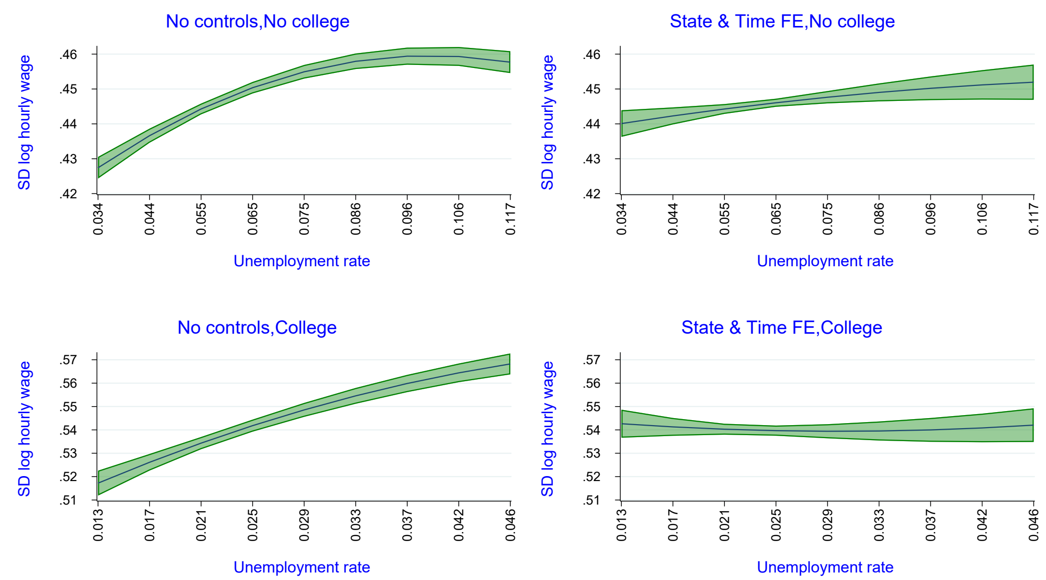


Sample 1994-2019. US real log hourly wages.

Figure 47: 75–25 percentile wage gap vs. JJ frequency by state & year

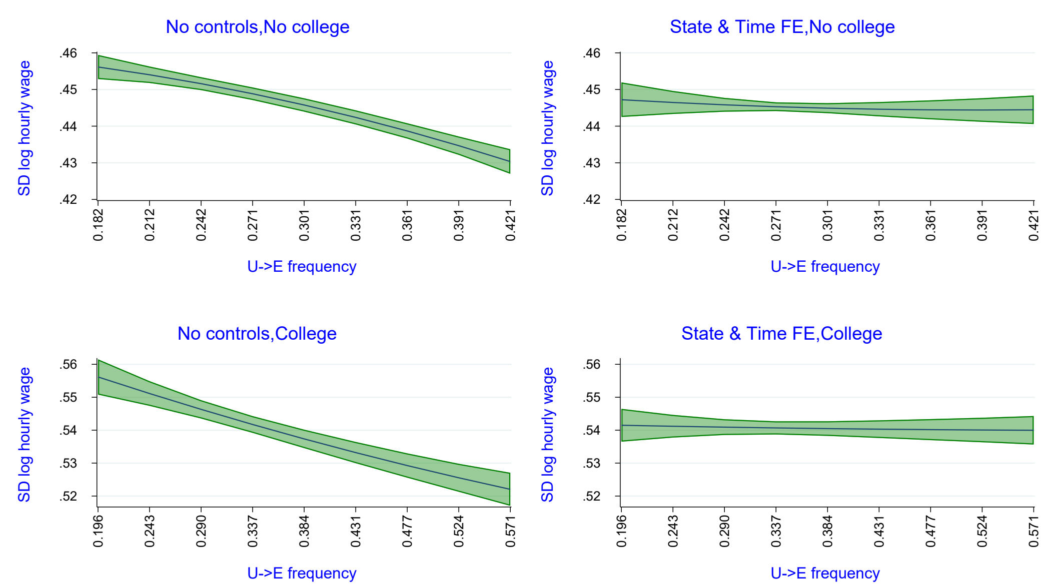
Sample 1994-2019. US real log hourly wages.

Figure 48: Standard deviation log wage vs. unemployment rate by state & year



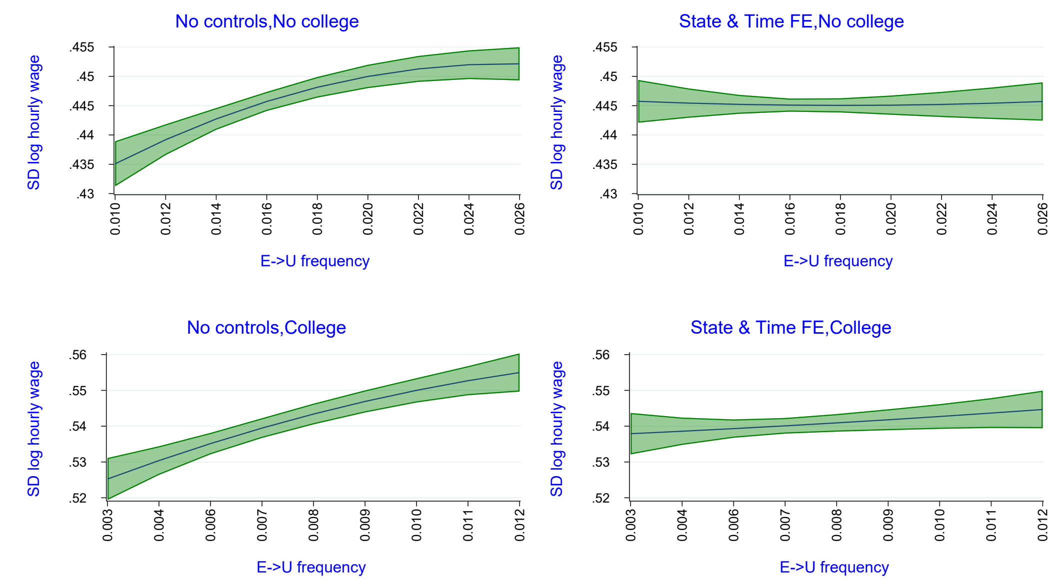
Sample 1994-2019. US real log hourly wages.

Figure 49: Standard deviation log wage vs. UE frequency by state & year



Sample 1994-2019. US real log hourly wages.

Figure 50: Standard deviation log wage vs. EU frequency by state & year



Sample 1994-2019. US real log hourly wages.

Figure 51: Standard deviation log wage vs. JJ frequency by state & year

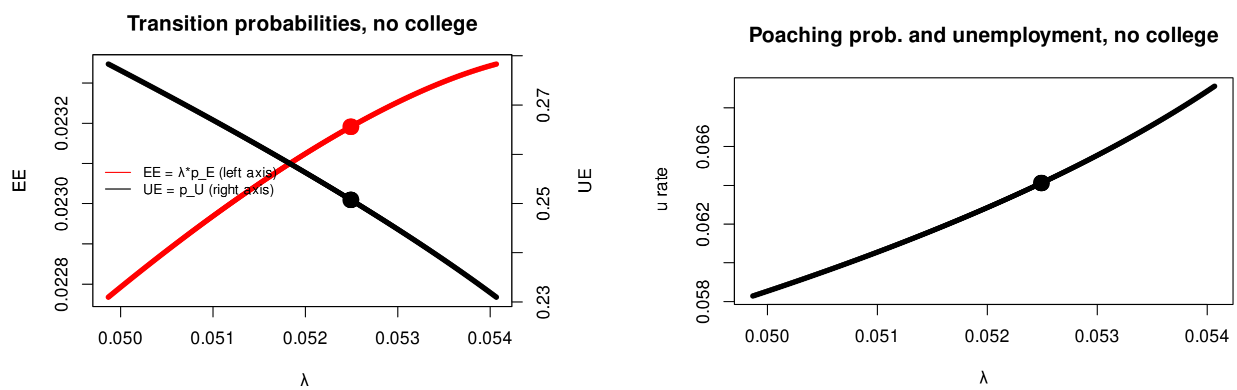
Sample 1994-2019. US real log hourly wages.

Appendix G Additional counterfactual experiments

The effects of an increase in poaching probability λ

If the probability of applying to a job while working (represented by the parameter λ) exogenously increases, there is more competition for jobs, since more employed agents enter the pool of applicants. The average job finding probability of the unemployed (in our notation, $\bar{p}_U$) decreases, while the job-to-job transition probability, $\lambda\bar{p}_E$, increases, as can be seen in the left panel of Figure 52. The average length of the queue, and therefore the unemployment rate, increases with λ, as is shown in the right panel of Figure 52.

Figure 52: The effects of poaching probability on transitions and unemployment



In terms of the wage distribution, this additional competition and therefore selectivity implies that there are fewer employed (a higher unemployment rate), but those who have a job earn more, because they are the most productive: the average log-wage increases as the parameter λ increases, as we can see in Figure 53.

The predictions of the model in terms of the consequences for wage inequality of an increase in the poaching probability depend on the chosen measure of inequality: while the model predicts a slight decrease of the 50/10 percentiles of log wages as λ increases, the consequences in terms of the ratio of the 90th over 10th percentile are the opposite, as we can see in Figure 53.

Appendix H Additional counterfactuals: college workers

In this section we present the results of the counterfactual experiments consisting in changes in productivity keeping 90/50 or 50/10 ratio constant for college workers.

Figure 53: The effects of poaching probability on wages

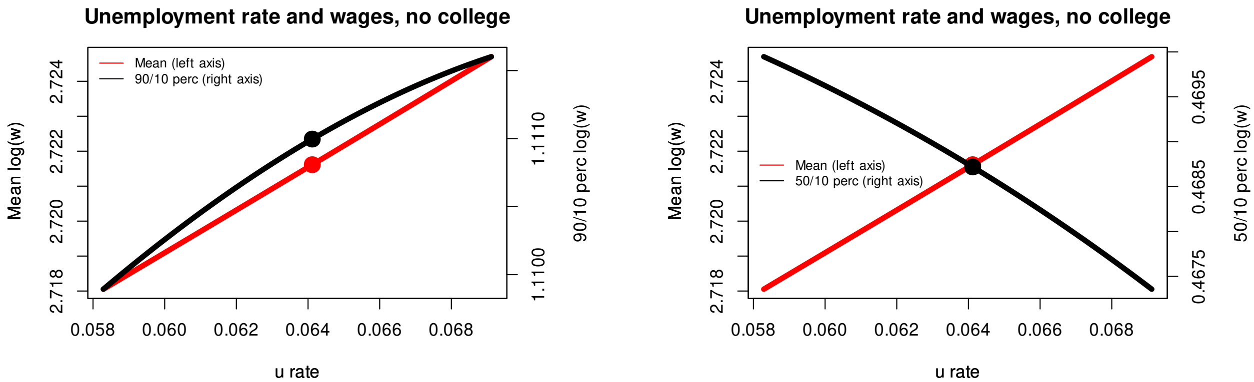


Figure 54: The effects of changes in productivity keeping 90/50 or 50/10 percentile ratio constant: other quantile ratios

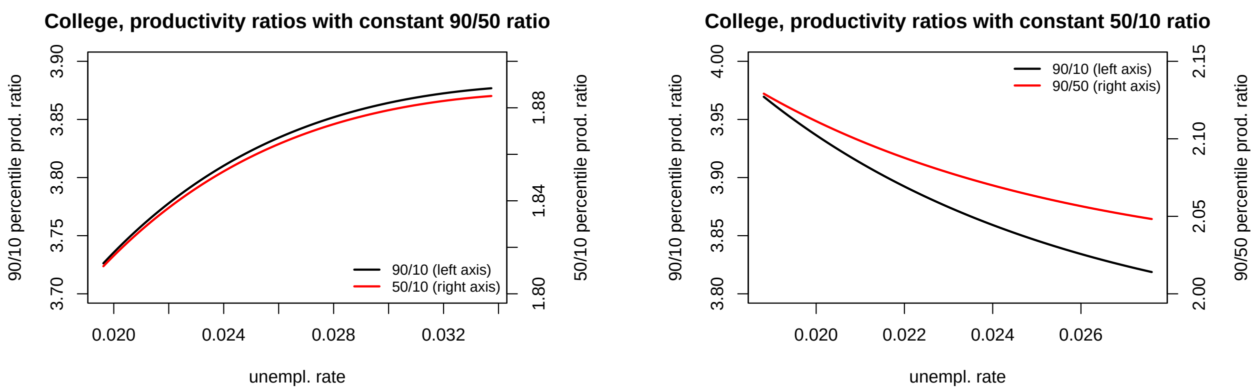


Figure 55: The effects of changes in productivity keeping 90/50 percentile ratio constant on unemployment and wage inequality (90/50 and 50/10): Model and data

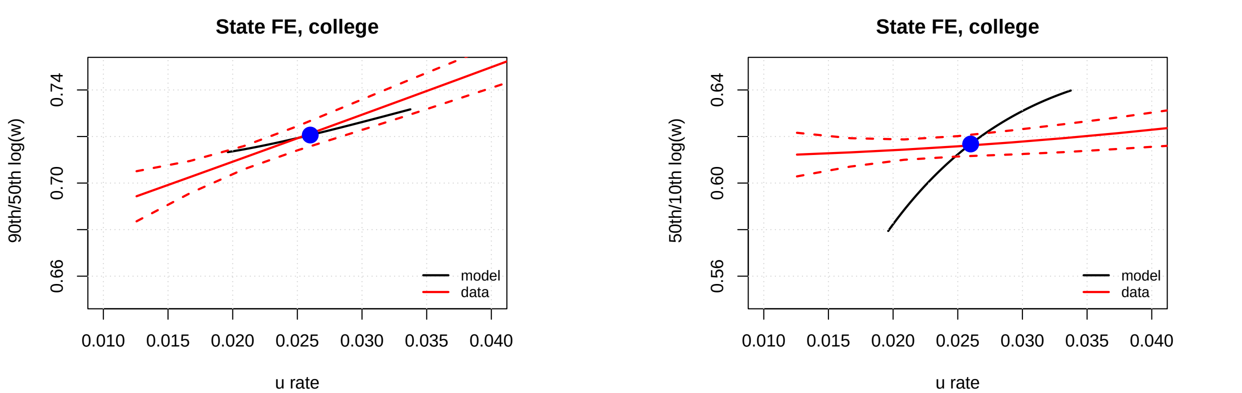


Figure 56: The effects of changes in productivity keeping 50/10 percentile ratio constant on unemployment and wage inequality (90/50 and 50/10): Model and data