Federal Reserve Bank of New York
Staff Reports

Replacement Hiring and the Productivity-Wage Gap

Sushant Acharya
Shu Lin Wee

Staff Report No. 860
June 2018

Replacement Hiring and the Productivity-Wage Gap
Sushant Acharya and Shu Lin Wee
Federal Reserve Bank of New York Staff Reports, no. 860
June 2018
JEL classification: E32, J63, J64

Abstract

A large and growing share of hires in the United States are replacement hires. This increase coincides with a growing productivity-wage gap. We connect these trends by building a model where firms post long-lived vacancies and engage in on-the-job search for more productive workers. These features improve a firm's bargaining position while raising workers' job insecurity and the wedge between hiring and meeting rates. All three channels lower wages while raising productivity. Quantitatively, increased replacement hiring explains half the increase in the productivity-wage gap. The socially efficient outcome features fewer low-productivity jobs and a 10 percent narrower productivity-wage gap.

Key words: replacement hiring, productivity-wage gap, unemployment, labor share, efficiency

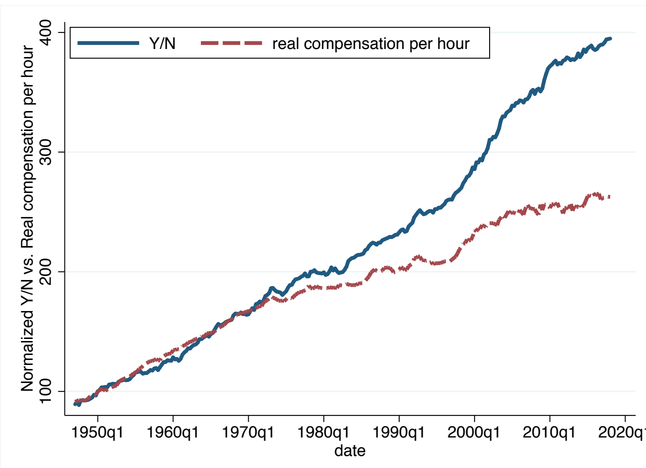
____

Acharya: Federal Reserve Bank of New York (email: sushant.acharya@ny.frb.org). Wee: Carnegie Mellon University (email: shuwee@andrew.cmu.edu). This paper was formerly titled "Replacement Hiring." The authors thank Isabel Cairo, Keshav Dogra, Ross Doppelt, John Haltiwanger, Ryan Michaels, and Laura Pilossoph for useful comments and suggestions. They also thank seminar/conference participants at the Census Bureau, Carnegie Mellon University, the Federal Reserve Bank of Kansas City-CADRE Workshop, the Federal Reserve Bank of St. Louis, and the Fall 2017 Midwest Macro Meetings. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

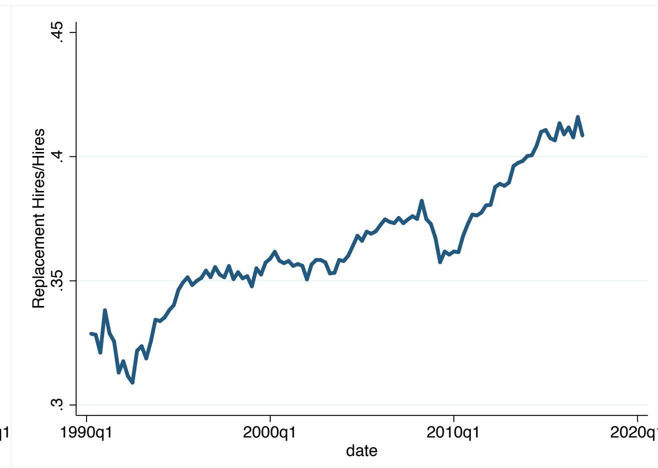
To view the authors' disclosure statements, visit
https://www.newyorkfed.org/research/staff_reports/sr860.html.

1 Introduction

The last three decades have seen an increasing divergence between labor productivity and wages. Figure 1a shows that real compensation per hour has failed to keep up with labor productivity since the 1980s. At the same time, an upward trend in the fraction of total hires that are replacement hires has emerged.[1] Replacement hires are defined by the Census as hires that occur in excess of net employment change.[2] Figure 1b shows that the fraction of total hires that are replacement hires has increased from about 33% in the early 1990s to a high of about 41% in 2017.[3] Given the growing importance of replacement hires in the US economy, we develop a framework that analyzes how increased replacement hiring can contribute to the divergence in labor productivity and wage compensation, or what we subsequently refer to as the productivity-wage gap. In particular, we ask how replacement hiring conducted for the purposes of replacing current workers with better applicants can lead to productivity gains while diminishing the wages of workers.



(a) Labor productivity and real hourly compensation



(b) Replacement hires as a share of total hires

Figure 1: Trends in replacement hiring vs. gap between labor productivity and wages

Intuitively, replacement hiring can occur for two reasons. Firstly, firms may choose to re-fill positions vacated by their worker who has left them. While this channel of replacement hiring has been more commonly studied in the literature, the data suggests that not all of replacement hiring can be accounted for by quits as we show in Section 2. Secondly, firms may choose to replace current workers with better applicants. This latter channel has been less explored and can account for the non-trivial portion of replacement hiring not explained by quits. As such, we ask whether this latter channel of replacement hiring, which can be thought of as on-the-job search by firms, can contribute to the widening

[1] It is important to note that the increased share of replacement hires is not inconsistent with declining labor mobility and declining trends in job creation. In fact, as a fraction of average total employed, the replacement hiring *rate* has been declining over time. Figure 10a in Appendix H shows that the hiring rate has fallen faster than the replacement hiring rate. The sharper decline in total hires relative to replacement hires implies that replacement hiring is increasingly becoming a more important share of total hiring.

[2] See Section 2 for more details. All definitions are taken from `https://lehd.ces.census.gov/doc/QWI_101.pdf`.

[3] While the US Bureau of Labor Statistics (BLS) provides data on Real Output per Hour and Real Compensation per Hour for the non-farm business sector dating back to the early 1950s, information collected on replacement hiring as recorded in the Quarterly Work Force Indicators (QWI) only begins from the 1990s.

productivity-wage gap.

In our model, search is random and firms pay a fixed cost to create a new vacancy. Vacancies are long-lived, allowing firms to replace workers who have quit the firm so long as their vacancies are unexpired. In addition, the longevity of the vacancy in our framework allows firms to continue to meet applicants even after their position has been filled, i.e, firms can conduct on-the-job search. Importantly, extending the random search framework to allow for both long-lived vacancies and on-the-job search by firms allows replacement hiring to occur without the incidence of quits and without firms having to create new vacancies for each replacement hire.[4]

Our simple model uncovers three channels through which higher replacement hiring can depress wages while still raising productivity. First, the fact that vacancies are long-lived imply that there is a positive option value to holding a vacancy. This positive option value or *market power* raises a firm's outside option when bargaining with the worker, allowing the firm to keep wages low and extract a larger share of the surplus. Secondly, the ability of firms to do on-the-job search allows firms to achieve productivity gains whenever they replace their current worker with a higher productivity applicant. However, these productivity gains come at a cost for workers as on-the-job search by firms raises the effective separation rate for workers and consequently the amount of *job insecurity* they face. Increased job insecurity reduces the average employment spell and diminishes workers' outside options, further allowing firms to pay lower wages. Finally, in our model, not all vacancies that unemployed job-seekers encounter are unfilled vacancies. In addition to unfilled vacancies, workers can also meet currently matched firms with unexpired vacancies - which we label as *recruiting matched firms*. In order for an unemployed applicant to be hired at this latter type of vacancy, her productivity must be higher than that of the firm's incumbent worker. Relative to meeting an unfilled vacancy, unemployed job-seekers must pass a higher bar before they are hired by the firm. This generates a larger wedge between hiring and meeting rates - lowering *measured matching efficiency*. This larger wedge again lowers workers' outside option, further reducing wages.

After demonstrating how an increased replacement hiring share is associated with a wider productivity-wage gap in our framework, we next test the predictions of our model using data. The first testable implication of our model is that the replacement hiring share declines with the vacancy expiration rate. Intuitively, if vacancies last for longer, firms have a longer window to conduct on-the-job search as well as re-fill any position that has been vacated by a worker, increasing the replacement hiring share. While the expiration rate of vacancies is not directly observable, our structural model allows us to identify this expiration rate as the ratio of new-to-old vacancies. The Conference Board Help Wanted Online™ (HWOL) data series provides information on the volume of new and continuing vacancies posted each month, which we use to construct a measure of the vacancy expiration rate. We use cross-sectional variation across industries and states from the QWI on the replacement hiring share. We find strong

[4]In contrast, the standard search model cannot reconcile replacement hiring that occurs for firm on-the-job search reasons. In the standard model with short-lived vacancies and free entry, firms currently attached to a worker never attempt to post another vacancy to replace their current worker or engage in on-the-job search. Intuitively, by free entry, a firm with an unfilled vacancy is indifferent between posting and staying out of the market, because the expected benefit of meeting a worker is exactly balanced by the cost of posting a vacancy. Thus, a firm who is already attached to a worker strictly prefers not to post a vacancy: the cost of posting a vacancy would be the same as that paid by an unattached firm but the net benefit is strictly lower, as firms who meet a new applicant lose their current worker when conducting a replacement hire. Appendix A.1 formalizes this argument.

evidence for this relationship - states and industries with lower vacancy expiration rates observe higher replacement hiring shares, lending credence to our model.

The second testable prediction of our model is that states and industries with higher shares of recruiting matched firms relative to non-recruiting matched firms, i.e firms with expired vacancies, observe lower wage earnings. As aforementioned, recruiting matched firms have higher outside options as well as the opportunity to conduct on-the-job search, allowing them to push wages below what non-recruiting firms offer. Again, while the relative share of recruiting matched firms to non-recruiting firms is not directly observable, our model allows us to relate this to the ratio of worker separation rates to vacancy expiration rates. As such, we can use information from the QWI on the separation rate and average earnings of workers as well as our earlier measure of the expiration rate from the HWOL to test this prediction. We find that the data validates this prediction as well.

Next, we calibrate the model to match labor market flows in the US as well as the replacement hiring share across two time periods: before 2005 and after 2005.[5] In conducting this exercise, we examine if our model, when calibrated to match the rise in the replacement hiring share, can at the same time replicate the observed wider gap between productivity and wages. Across the time periods, our model predicts that the higher replacement hiring share widened the productivity-wage gap by 8%, accounting for half of the total increase in the productivity wage gap over the time periods of interest.[6] Overall, this wider productivity-wage gap would have caused the labor share in our model economy to fall by 5 percentage points, a decline on par with the actual observed fall in the labor share of 6.4 percentage points (Elsby et al. (2013)).

We further use our calibrated model across these two time periods to identify the key factors that contributed towards amplifying the rise in replacement hiring while driving a larger productivity-wage gap. We find that the rise in the value of an unfilled vacancy and the decline in the vacancy expiration rate are key towards driving a rise in the replacement hiring share, which in turn widens the productivity-wage gap. In contrast, we find that a declining quits rate has the opposite impact on the replacement hiring share and the productivity-wage gap. In fact, our model suggests that the declining trend in quits would have moderated the divergence in productivity and wages, and led to a smaller decline in the labor share.

Next, we show that the amount of replacement hiring in the decentralized economy tends to be inefficiently large. The decentralized economy tends to create too many low productivity matches relative to what a social planner would choose. The planner recognizes the accompanying value lost whenever an unemployed applicant displaces a currently matched worker into unemployment and values only the net gain associated with recruiting matched firms re-matching with better applicants. To minimize such losses, the social planner instead only chooses to form matches of higher productivity value to lower the incidence of replacement hiring. Overall, comparing our calibrated economy to the efficient

[5]Notably, there was a significant change in how vacancies were posted with the advent of internet technology and the availability of online job posting. Barnichon (2010) documents that the use of print advertising observes a sharp drop-off in the early 2000s and that online vacancies as documented in the HWOL became the prevalent source of vacancy postings. Since the HWOL dataset which collects information on online vacancies posted is first available in the year 2005, we use 2005 as the cut-off year.

[6]We measure the productivity-wage gap as the ratio of labor productivity to mean compensation. Since the our data for replacement hiring only begins in the 1990s, we only calculate the change in the productivity-wage gap between the periods 1990-2005 and post 2005.

allocations, we observe a welfare loss of between 3-4% in terms of lost consumption. Furthermore, in our calibrated economy, the productivity-wage gap is about 9-10% higher than what would be achieved under the efficient benchmark.

Related Literature Our paper contributes to the growing literature on the impact of replacement hiring. In a paper closely related to ours, Elsby et al. (2016) examine replacement hiring and focus on how a quit by a worker necessitates a firm to create a new vacancy in order to re-fill the position. While the authors focus on how replacement hiring generates volatility in vacancy posting over the business cycle, we instead focus on the long run implications of increased replacement hiring. Separately, Menzio and Moen (2010) examine replacement hiring in the context that firms would like to insure workers against income fluctuations but cannot commit to not replacing current workers in a downturn with cheaper new hires. While their paper is concerned with characterizing the efficient wage contract, we examine how under a steady state setting, the decentralized economy tends to create too many low productivity matches relative to what a planner would choose.

Our paper is also related to the literature on long-lived vacancies. Fujita and Ramey (2007) and Haefke and Reiter (2017) consider models where job positions are long-lived and firms do not shut down immediately upon worker separation. Both of these papers demonstrate that the inclusion of long-lived vacancies in a labor search model can better replicate labor market flows in the data. While replacement hiring is not the focus of both the aforementioned papers, it should be noted that firms with unexpired job positions in these models only re-hire new workers when they are separated from their current worker. As such, these models are not equipped to address the issue of firm on-the-job search and its ramifications for the productivity-wage gap.

Although we study how replacement hiring can affect the productivity-wage gap, our paper also has implications for the labor share. Intuitively, the divergence in labor productivity and compensation implies that a smaller share of total output accrues to labor. Karabarbounis and Neiman (2013) document that the labor share has declined across countries and argue that capital deepening is the primary factor behind this decline. Elsby et al. (2013) conduct a comprehensive study and find a strong negative relationship between import exposure and the labor share at the industry level. We add to this debate on the labor share and show how increased replacement hiring can lead to a smaller labor share by raising productivity while keeping wages low. Recent work by Autor et al. (2017) and Azar et al. (2017) suggest that product market concentration is associated with labor market concentration. These papers empirically show that wages are lower when firms observe increased market power. By focusing on long-lived vacancies, our paper offers an alternative view of firm market power and shows that the positive outside option of firms from having an unexpired vacancy allows them to offer lower wages.

Lastly, our paper is also related to the recent literature on phantom vacancies. Cheron and Decreuse (2017) and Albrecht et al. (2017) argue that phantoms are vacancies that have already found a match and as such cannot generate any more new hires. The existence of phantoms lowers matching efficiecy as unemployed jobseekers cannot convert a meeting with a phantom into a hire. We take an alternative view: matched firms with unexpired vacancies can still generate new hires. An unemployed job applicant who contacts a recruiting matched firm, however, must surpass the productivity of the incumbent worker before she is hired. As such, these long-lived vacancies which allow firms to continue to re-match with better applicants also lowers measured matching efficiency in the labor market. Recent work using online vacancy job board data by Davis and de la Parra (2017) suggests that a non-trivial portion of job postings are "long-duration" postings which are continuously on the look-out for new applicants, giving support to our supposition that vacancies are long-lived and can re-match with multiple workers.

The rest of this paper is organized as follows. Section 2 discusses the data on replacement hiring. Section 3 introduces the model. Section 4 examines the testable implications of our model and examines how large a productivity-wage gap our model would predict given the rise in replacement hiring. Section 5 explores the efficiency implications of our model. Section 6 contains a brief discussion about some assumptions of our model while section 7 concludes.

2 Data

Building on the underlying Longitudinal Employer Household Dynamics (LEHD) linked employee-employer database, the Quarterly Workforce Indicators (QWI) provides information on labor market outcomes by state, industry, worker characteristics, employer age and size. In particular, the QWI provides information on the number of total hires at a firm, the total number of separations, the number of job gains and losses as well as earnings. The QWI defines job gains at a firm as the change in employment within a quarter, which can be formally written as:

$$\text{Job Gain/Loss} = Emp_t^{end} - Emp_t^{beginning}$$

In contrast, the number of hires at a firm in quarter t is defined at the total number of new employees at a firm that did not have earning in period $t-1$ but that reported earnings at that firm in period t. The total number of hires measures the gross inflows into a firm, while the measure of job gains measures the net employment change at the firm. Replacement hires are defined as the hires that continue into the next quarter *in excess* of job gains at a firm. Using data from the QWI, we calculate the replacement hiring share as the fraction of total hires that are replacement hires, i.e.

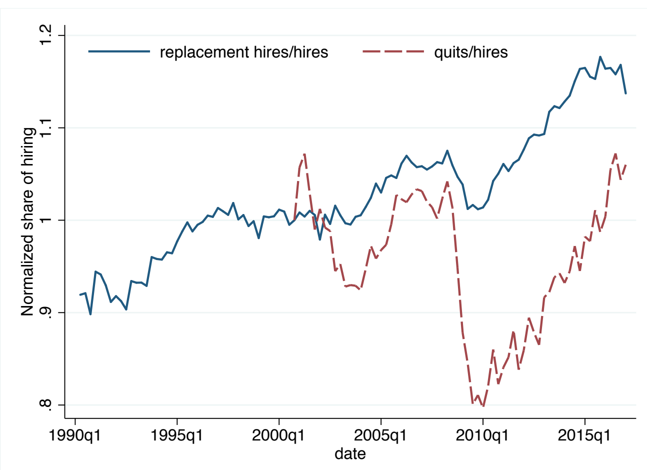
$$\text{Replacement Hiring Share} = \frac{\text{Replacement Hires}}{\text{Total Hires}} = \frac{\text{Hires - Job Gains}}{\text{Total Hires}}$$

We use the above definition to construct the replacement hiring share in Figure 1b. Overall we find that the replacement hiring share rose from about 0.33 in the 1990s to about 0.41 in 2017.[7] Separately, Appendix H shows that the rise in the share of replacement hiring is not limited to a single industry's experience but occurs broadly across all industries. Further, Appendix H also shows that changes in industry compositon only explain about one-seventh of the increase in the replacement hiring share. In contrast, changes in the replacement share within each industry explain the bulk of the increase in the replacement hiring share.

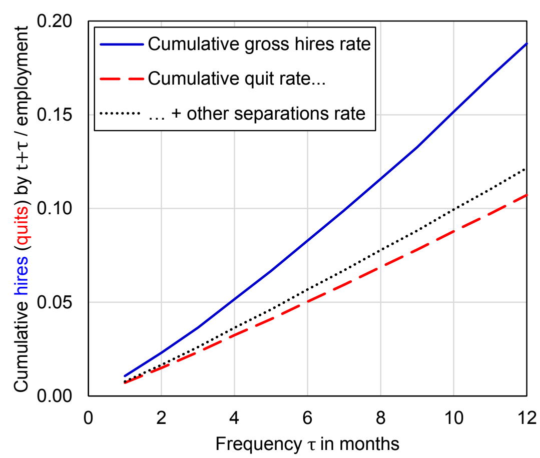
[7]Our finding that the replacement hiring share rose over time is robust even if we use a different measure of the replacement hiring share. Specifically, if we only consider replacement hires as a fraction of all hires that continue into next quarter (as opposed to all hires that include individuals who were hired in a quarter but who did not stay on until the next quarter), this alternative replacement hiring share rose from 0.55 in the 1990s to about 0.60 in 2017.

A replacement hire always coincides with a separation.[8] Recall there are two forms of separations that can give rise to a replacement hire. Firstly, replacement hires can occur whenever a firm seeks to re-fill a position that is vacated by its worker. If the primary reason for the occurrence of replacement hiring is to replace workers that have quit, one would expect that the ratio of quits to total hires would exactly mimic the trend in the replacement hiring share. As the QWI does not distinguish whether separations that occurred were due to quits or layoffs, we use information on the level of quits and hires from the Jobs Openings and Labor Turnover Survey (JOLTS) and compute the ratio of quits to total hires.[9]

Figure 2a plots how the replacement hiring share and the ratio of quits to hires has changed over time. We normalize both measures to be equal to 1 in 2001q1.[10] Reaffirming previous findings, the figure shows that quits are pro-cyclical and that the ratio of quits to hires observes no long-run upward trend. While the replacement hiring share has steadily grown since 2001q1, the ratio of quits to hires has only in recent quarters returned to the same level as it was in 2001q1. While the most recent Great Recession also observed a decline in the replacement hiring share, the fall in the former was muted in comparison to the stark drop in the ratio of quits-to-hires over the Great Recession period. Overall, the correlation between the two series is about 0.21, suggesting that quits alone cannot account for all of replacement hiring.



(a) Quits/hires vs. replacement hires share



Source: Elsby et al. (2016)

(b) Cumulative hires vs. cumulative quits

Figure 2: Ratio of Quits to Hires

The second way a replacement hire could occur is if a firm with a currently filled position meets a better match and chooses to replace his current worker with the new worker. Using the JOLTS micro-data, Elsby et al. (2016) focus on firms with zero net employment change and measure the cumulative

[8]To see this, consider the example of a firm that started the period with 1 worker. Suppose that worker left the firm and the firm hired a new worker to re-fill its vacated position. In this example, there was zero net employment change, 1 hire and 1 separation. In addition, the one hire that transpired was a replacement hire.

[9] Because the JOLTS data is a monthly series, we sum up quits and hires over a quarter and report the ratio of aggregated quits to hires at a quarterly frequency.

[10]We choose the date 2001q1 as this is the first full quarter of data available from JOLTS. We perform this normalization as the measures stem from two different data sources.

hires rate (solid blue line) and cumulative quits rate (dashed-red line) at such firms. While quits do affect the amount of replacement hiring, Figure 2b from Elsby et al. (2016) reveals that a sizable wedge exists between the cumulative hires rate and cumulative quits rates (plus other separations)[11], suggesting that a significant portion of replacement hiring also occurs alongside the event of a layoff. While much work has been done examining how replacement hiring stemming from quits can affect the volatility of unemployment rates and hiring over the business cycle, our findings indicate that the standard labor search model may be missing an important feature which is on-the-job search by firms. We outline in our model section how replacement hiring can occur in an environment where firms can do on-the-job search and when vacancies do not expire immediately.

3 Model

Workers and firms Time is continuous and runs forever. The economy comprises of a unit mass of infinitely-lived workers who are ex-ante identical. All workers are risk neutral and discount the future at a rate $\rho > 0$. Workers can either be employed or unemployed. Unemployed workers receive flow utility $b \geq 0$ per unit time. The other agents in the economy are firms each of which can employ at most one worker at any date. A firm-worker pair with match quality x produces x units of output at each date. The match quality $x \in [0, \overline{x}]$ is drawn from a time invariant distribution $\Pi(x)$ at the time the firm and worker meet and remains constant for the duration of the match.[12]

Vacancies and Firms Search is random. A firm that decides to enter the market must incur a fixed cost χ to post a vacancy. This implies that all firms enter the labor market initially as unfilled vacancies. Importantly, unlike the standard DMP setup, in our model, unfilled vacancies do not expire instantly. Instead, vacancies expire at a rate $\delta > 0$. This implies that a vacancy that goes unmatched today can still contact an applicant in the future as long as the vacancy has not expired. In addition, firms who were previously matched to a worker but whose worker separated from them at exogenous rate s, can still transition to become unfilled vacancies so long as their vacancy has not expired. Thus, firms can replace workers who separated from them without posting a new vacancy.

Further, firms who have unexpired vacancies, referred to as *recruiting matched firms*, can still continue to meet and accept new applicants *even if* they are currently matched with a worker. This allows recruiting matched firms to replace current workers without having to post a new vacancy or experience a separation. If the matched firm chooses to replace its current worker with the new job applicant, it releases its current worker into unemployment. Notice that this is akin to a model in which firms can search on-the-job. The left circle in Figure 3 describes the universe of vacancies in the labor market. The red area captures the unfilled vacancies while the blue area captures the recruiting matched firms who have unexpired vacancies. In the case where the worker leaves the firm and the firm is unable to find a replacement, recruiting matched firms become unfilled vacancies. If a firm with an unfilled vacancy

[11]The JOLTS data series defines other separations as separations stemming from retirements as well as discharges due to reasons of disability.

[12]The support of x is allowed to be unbounded above, i.e., $\overline{x}$ can be ∞. In fact, in our calibrated model, we assume that x is described by a log-normal distribution and hence, the support is unbounded above.

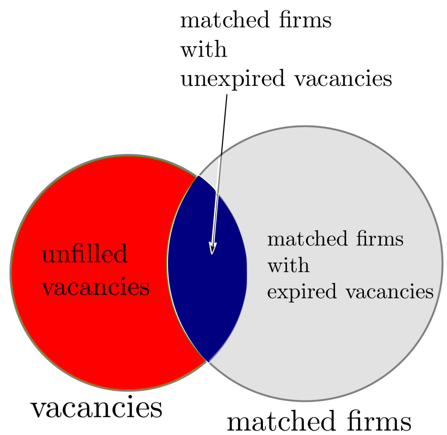


Figure 3: The universe of vacancies and firms in the labor market

hires a worker, it becomes a recruiting matched firm with an unexpired vacancy, i.e. it transitions from the red to blue area.

A matched firm with an expired vacancy, referred to as a *non-recruiting firm*, cannot meet any new applicants. As such, these firms are unable to do on-the-job search. The right circle in Figure 3 shows the universe of matched firms in the labor market. The gray area represents the set of non-recruiting matched firms while the blue area represents recruiting matched firms with unexpired vacancies. Notice that the blue area captures the intersection of both vacancies and matched firms. This intersection is empty when 1) vacancies expire instantaneously (as in the standard DMP model) and when 2) matched firms cannot do on-the-job search. If vacancies are infinitely-lived, then the blue area forms the entire set of matched firms. More generally, as the expiration rate of vacancies, δ, falls, the blue area becomes larger and the gray area shrinks ceteris paribus.

When a non-recruiting matched firm undergoes a separation from its worker, it shuts down and leaves the market. This is different from the experience of a recruiting matched firm, which as aforementioned, becomes an unfilled vacancy following a separation if it is unable to find a replacement. The expiration of a vacancy is not equivalent to the separation of a matched firm-worker pair. An expiration, which occurs at rate δ, merely destroys a firm's ability to meet new job applicants. Recruiting matched firms in the blue area transition to become non-recruiting matched firms in the gray area when a vacancy expires.

Labor Market Only unemployed workers can make contact with an unexpired vacancy. The rate at which meetings between applicants and vacancies take place is determined by a *meeting* technology which takes as its inputs total vacancies and the unemployed:

$$M = \xi v^{1-\alpha} u^{\alpha} \tag{1}$$

where ξ is a scaling parameter that affects meeting rates. v denotes total vacancies which is the sum of unfilled unexpired vacancies, v^u (the red area in Figure 3), and the mass of all recruiting matched firms, v^m (the blue area in Figure 3). u is the mass of unemployed workers. Unemployed workers meet vacancies at rate $p = \frac{M}{u} = \xi\theta^{1-\alpha}$ while vacancies meet unemployed workers at rate $q = \frac{M}{v} = \xi\theta^{-\alpha}$.

Importantly, meeting rates are not equivalent to hiring rates. In order for a meeting to result in a job, both the firm and the worker must agree to form a match. If the match is between an unfilled vacancy and an unemployed worker, then both parties agree to create a job as long as the match-specific productivity drawn is above a threshold $\widetilde{x}$, which is determined in equilibrium.[13] However, if the worker instead meets a recruiting matched firm, the new match quality x drawn must be at least as large as the match quality of the firm's incumbent worker. Thus, although the rate with which an unemployed worker meets a filled and unfilled vacancy is the same, the probability she will be hired is (weakly) lower for vacancies which are already filled.

Finally, workers can be both exogenously and endogenously separated from firms. The former occurs at rate s, while the latter occurs whenever a recruiting matched firm replaces their current worker with a better applicant. Having described the environment, we now proceed to describe the individual agents' problems.

3.1 Worker's Problem

Unemployed workers The asset value of a worker from unemployment, U, can be written as:

$$\rho U \quad = \quad b + p\left[\frac{v^u}{v}\int_{\widetilde{x}}^{\overline{x}}\left[W(y,1)-U\right]d\Pi(y) + \frac{v^m}{v}\int_{\widetilde{x}}^{\overline{x}}\int_{\varepsilon}^{\overline{x}}\left[W(y,1)-U\right]d\Pi(y)dF(\varepsilon)\right] \tag{2}$$

where $F(\varepsilon)$ denotes the fraction of recruiting matched firms whose worker possesses match quality ε or lower. $W(y,1)$ is the value of being employed at a recruiting firm, denoted by second argument of 1, and with match quality y.

The value of unemployment can be decomposed into two terms: b, the flow utility associated with home production and the second term in equation (2) which denotes the expected change in value that the worker enjoys in the event that he transitions to employment in the future. At a rate p, an unemployed worker meets a vacancy. With probability v^u/v, this vacancy is currently unfilled and the worker is accepted whenever he draws a match quality higher than $\widetilde{x}$. However, with probability v^m/v, the unemployed worker encounters a recruiting matched firm. In this case, the unemployed worker must draw a match quality x that is higher than the incumbent's value. The second term inside the parenthesis captures the unemployed worker's change in value when he is accepted by a recruiting matched firm with current match quality ε weighted by the probability of meeting such a firm.

Worker employed at a non-recruiting firm The asset value of an employed worker at a non-recruiting firm with match quality x can be written as:

$$\rho W(x,0) = w(x,0) - s\left[W(x,0)-U\right] \tag{3}$$

The asset value can be decomposed into two parts: (i) the wage $w(x,0)$ that the worker earns today and (ii) the change in value the worker receives whenever he is exogenously separated at a rate s and enters unemployment. The second argument of both the value-function $W(x,0)$ and the wage $w(x,0)$ indicates that the worker is situated in a non-recruiting firm.

[13] See Section 3.5 for details.

Worker employed at a recruiting firm The asset value of an employed worker at a recruiting firm with match quality x can be written as:

$$\rho W(x,1) = w(x,1) - \Big(s + q\left[1-\Pi(x)\right]\Big)\left[W(x,1) - U\right] + \delta\left[W(x,0) - W(x,1)\right] \tag{4}$$

where $w(x,1)$ denotes the wages paid to such a worker. As before, the second argument taking a value of 1 indicates that the worker is situated in a recruiting firm. There are three events that such a worker can experience in the future. Firstly, at rate δ, the vacancy of the firm expires, i.e the firm transitions from being a recruiting firm to a non-recruiting firm, and the worker receives a change in value $W(x,0) - W(x,1)$. Secondly, the worker is exogenously displaced into unemployment at rate s and receives a change in value of $U - W(x,1)$. Finally, the firm can meet a new applicant at a rate q and with probability $1 - \Pi(x)$ they draw a match quality above x, in which case the current worker is released into unemployment and receives the change in value $U - W(x,1)$.

For a given match quality x, the value of being employed at a recruiting matched firm, $W(x,1)$ can differ from that of being employed at a non-recruiting firm, $W(x,0)$. The worker at the former faces an additional risk of becoming unemployed as her firm has the ability to replace him. In contrast, a worker at the latter firm faces no risk of being replaced since her firm does not engage in on-the-job search.

3.2 Firm's Problem

Non-recruiting matched firms Consider a firm that is matched with productivity x but does not have a vacancy, i.e it is not actively recruiting. The value of such a firm can be written as:

$$\rho J(x,0) \quad = \quad x - w(x,0) + s\left[J(0) - J(x,0)\right] \tag{5}$$

where the '0' refers to the fact that the firm does not have a vacancy. The firm receives current profit $x - w(x,0)$ and observe a change of value whenever it separates from its worker. $J(0)$ refers to the value of creating a new vacancy.[14]

New firms/vacancies If a firm decides to post a vacancy, it incurs a cost χ and becomes an unfilled vacancy instantly. Under free-entry, firms post vacancies until the benefit of creating a vacancy is equal to its cost, i.e.

$$J(0) = 0 \qquad \Rightarrow \qquad J(1) = \chi \tag{6}$$

where $J(1)$ is the value of an unfilled unexpired vacancy.[15] Equivalently, the above also implies that all new vacancies created are subsumed into the stock of unfilled and unexpired vacancies.

[14]Notice that the problem of the non-recruiting matched firm is written as if the firm has no option of re-listing and creating an active vacancy. Appendix A.2 shows that in equilibrium, so long as re-listing costs are the same as vacancy posting costs, a matched firm with an expired vacancy has no incentive to re-list.

[15]Note that this is different than the standard DMP specification of flow costs of vacancy posting. Here χ is a one time fixed cost that the firm must pay in order to get the privilege of hiring.

Recruiting matched firms The value of such a firm with current match quality x can be written as:

$$\rho J(x,1)=x-w(x,1)+q\int_{x}^{\overline{x}}\left[J\left(y,1\right)-J\left(x,1\right)\right]d\Pi(y)+\delta\left[J(x,0)-J\left(x,1\right)\right]+s\left[J(1)-J(x,1)\right] \quad (7)$$

The firm receives current profits $x-w(x,1)$ and can undergo three possible events in the future. First, its vacancy expires at a rate δ and it receives a change of value $J(x,0)-J(x,1)$. Second, it may undergo an exogenous separation at rate s and become an unfilled vacancy with the associated change in value $J(1)-J(x,1)$. Finally, a firm may meet a new applicant at rate q and enjoys a change in value $J(y,1)-J(x,1)$ as long as the new match quality y exceeds current match quality x.

Firms with unfilled vacancies The value of an unfilled vacancy can be written as:

$$\rho J(1) \quad = \quad q\int_{\widetilde{x}}^{\overline{x}}\left[J\left(y,1\right)-J\left(1\right)\right]d\Pi(y)-\delta J(1) \qquad (8)$$

In the current period, an unfilled vacancy observes zero production and zero current profits. The vacancy expires at rate δ in which case it ceases to exist. At a rate q, the firm meets an unemployed applicant and forms a match as long as the match quality drawn by the pair is above reservation productivity $\widetilde{x}$. The firm's gain from matching with a worker of match quality x is given by $J(x,1)-J(1)$. The reservation match-productivity $\widetilde{x}$ is defined as the value of x which satisfies:

$$J(\widetilde{x},1)-J(1)=0 \qquad (9)$$

Recall from (6) that while the value of creating a vacancy, $J(0)=0$, the value of an unfilled vacancy $J(1)=\chi>0$. In other words, having an unexpired vacancy is associated with a positive option value for the firm. This positive option value affords the firm the ability to continue to search tomorrow even if it rejects or fails to meet a worker today. Further - and as we discuss next - the positive option value raises the matched recruiting firm's outside option when bargaining with the worker, allowing it to offer lower wages than a non-recruiting firm.

3.3 Surplus and Wage Formation

The joint payoff to a match is the total surplus shared by the firm and worker relative to continuing to search.[16] Appendix B.2 shows that the surplus for a non-recruiting matched firm-worker pair with match quality x is given by:

$$(\rho+s)S(x,0) \quad = \quad x-\rho U \qquad (10)$$

where ρU is as defined in equation (2). Since the non-recruiting firm has an outside option of zero and since the worker's value of continued search is the value of remaining unemployed, the joint surplus, appropriately discounted, is given by the total benefit of a match, here output, less the worker's outside option, the value of unemployment.

[16]We follow the standard definition of surplus: $S(x,i)=J(x,i)-J(i)+W(x,i)-U$ for $i\in\{0,1\}$

In contrast, the gain a recruiting matched firm derives from a match is his benefit to matching relative to the value of continuing to search for a better worker, where the latter is encapsulated by the value of an unfilled and unexpired vacancy, $J(1) > 0$. In this case, for a firm-worker pair with match quality x, the surplus for such a match is given by:[17]

$$\varrho(x)S(x,1) = x + \delta S(x,0) - \rho U - q\int_{\tilde{x}}^{x} [J(y,1) - J(1)]\, d\Pi(y) \tag{11}$$

where $\varrho(x) = \rho + \delta + s + q[1 - \Pi(x)]$. Focusing on the RHS of the first line of equation (11), the total benefits from a match include not only (the discounted value of lifetime) output produced but also take into account that at rate δ, the vacancy expires and the match achieves the surplus of a non-recruiting matched firm-worker pair. From these benefits, the surplus nets out the gain each party would receive if they continued to search. This is given by both the worker's outside option of unemployment, ρU, *and* the firm's relative gain from rejecting the worker of match quality x and continuing to search. When a firm-worker pair of match quality x is formed, the recruiting matched firm continues to meet applicants at rate q but is only willing to accept a new applicant who draws match quality greater than x. If the firm had instead chosen not to match, he would have had the additional ability to match with any applicant who drew a match quality between $\widetilde{x}$ and x. This is captured by the term $-q\int_{\tilde{x}}^{x} [J(y,1) - J(1)]\, d\Pi(y)$.

The differential option values of a recruiting matched firm versus a non-recruiting matched firm affect the total gains to matching. Specifically, one can show that $\forall x \geq \widetilde{x}$, we have:

$$(\rho + \delta + s)[S(x,1) - S(x,0)] = -q\int_{\tilde{x}}^{x} [J(y,1) - J(1)]\, d\Pi(y) - q[1 - \Pi(x)]S(x,1) \leq 0$$

The inequality is strict for $x > \widetilde{x}$. The surplus of a recruiting matched firm-worker pair is lower because of two reasons. First, as aforementioned, a recruiting matched firm with match quality x loses the potential surplus of matching with a new applicant who draws match quality between $\widetilde{x}$ and x, as captured by $-q\int_{\tilde{x}}^{x} [J(y,1) - J(1)]\, d\Pi(y)$. Secondly, the ability to do on-the-job search implies that the surplus of a match between a worker and a recruiting matched firm-worker pair is lower because of the potential value lost whenever the recruiting firm does a replacement hire, captured by $-q[1 - \Pi(x)]S(x,1)$.

Wage Determination Wages are determined at each date via Nash Bargaining:

$$w(x,i) = \underset{w(x,i)}{\arg\max} \Big[J(x,i) - J(i)\Big]^{1-\eta} \Big[W(x,i) - U\Big]^{\eta} \text{ for } i = 0,1 \tag{12}$$

where $\eta \in [0,1]$ denotes the bargaining power of a worker. Bargaining over wages takes place only after matches have been formed. This implies that whenever a recruiting matched firm chooses to hire a new applicant, he releases his current worker into unemployment prior to bargaining with the new applicant. We further assume that there are no recalls. As such, when both the firm and new applicant bargain over wages, the firm's outside option is simply the positive option value of an unfilled and unexpired

[17]See Appendix B.2 for derivation of the expression.

vacancy, $J(1)$, and not $J(y,1)$, the value of a matched firm who currently has match quality y.[18]

The difference in surpluses between recruiting matched firm-worker pairs and non-recruiting matched firm-worker pairs also affects the wages offered at these firms in equilibrium. The Proposition below summarizes the differential wage outcomes:

Proposition 1 (Wages)**.** *Consider a worker-firm pair with match quality x.*

1. *The wage the worker receives from a non-recruiting firm is given by:*

$$w(x,0) \quad = \quad \widetilde{x} + \eta(x - \widetilde{x}) \tag{13}$$

2. *The relative wage discount that workers receive from a recruiting firm relative is given by:*

$$w(x,1) - w(x,0) = -\eta\,(1-\eta)\,q \int_{\tilde{x}}^{x} S\,(y,1)\,d\Pi\,(y) \leq 0 \tag{14}$$

Proof. See Appendix B.3. □

Proposition 1 highlights the fact that there are in fact two dimensions of wage dispersion present in this economy. First, wages differ across jobs with differing match quality. Second, and unique to this model, wages differ across firms with different outside options. Intuitively, the positive outside option for a recruiting firm implies that it has a higher threat point, allowing it to pay lower wages than an identical non-recruiting firm. An implication of Proposition 1 is that an economy with a larger fraction of recruiting firms would have lower average wages, holding all else constant.

3.4 Labor Market Flows

Having described the relevant value functions, we proceed to describe labor market flows next. While Appendix C documents the labor market flows out of steady state, below we describe the corresponding steady state relationships.

Unemployed The steady state rate of unemployment u satisfies:

$$s(1-u) = qv^u\,[1 - \Pi\,(\tilde{x})] \tag{15}$$

The LHS denotes the inflows into unemployment driven by exogenous separations. The RHS denotes the outflows from unemployment. These occur when an unfilled, unexpired vacancy meets an unemployed worker at rate q and they draw a match quality above the reservation value. While unemployed job seekers also leave the pool of unemployed whenever they meet recruiting matched firms and draw a

[18] Notice that this assumption is without loss of generality since wages are determined via Nash Bargaining each period without commitment. Even if firms bargained before separating with their current worker, and effectively used $J(y,1)$ as their outside option, it would mean that at the instant the next match is formed, it would revert to having $J(1)$ as its current outside option and would have to pay workers wages commensurate with equation (12). Of course, if the the environment featured commitment by agents in the form of long-term contracts, then this would not be true and in that scenario, currently matched recruiting firms would offer a lower wage than unfilled vacancies for an applicant with the same match quality.

match quality that is above the incumbent's match quality, the firm in choosing to match with this new worker releases its incumbent worker into unemployment. This contributes to both gross flows in and out of unemployment which nets out to zero. As such, the steady state unemployment rate is unaffected by the replacement hiring that occurs when firms conduct on-the-job search.

Recruiting matched firms Since all recruiting matched firms continue to attract applicants, they contribute towards total vacancies and therefore, market tightness. The steady state stock of vacancies accounted for by recruiting matched firms, denoted v^m, can be expressed as:

$$(s+\delta)v^m = q[1-\Pi(\widetilde{x})]v^u \tag{16}$$

The LHS represents the outflows from the stock of matched vacancies due to exogenous separation or expiration of a vacancy while the RHS represents the inflows. These inflows occur whenever an unfilled, unexpired vacancy meets an applicant and forms a match. In addition to the total stock of matched vacancies, we can also characterize the steady state distribution of matched vacancies across the match-quality dimension. In equilibrium, the cdf of this distribution can be written as:

$$F(x) = \begin{cases} 0 & \text{for } x < \widetilde{x} \\ \left[\frac{[\Pi(x)-\Pi(\tilde{x})](s+\delta)}{q[1-\Pi(x)]+s+\delta}\right]\frac{1}{1-\Pi(\tilde{x})} & \text{else} \end{cases} \tag{17}$$

It is important to recognize that the distribution of matched vacancies by x is informative about the replacement hiring share. If the distribution of matched firms are skewed towards low values of x, then there is substantial room for such firms to find a better match in the future and conduct replacement hiring. In contrast, if most firm-worker pairs are of very high match quality, then it is very hard for a new applicant to replace an existing worker.

Non-recruiting matched firms The steady state mass of non-recruiting matched firms is implicitly given by:

$$sv^i = \delta v^m \tag{18}$$

The RHS of the above equation captures the inflows into the pool of non-recruiting matched firms. Notice that the only inflows into the mass of non-recruiting firms stem from the expiry of recruiting matched firms' vacancies. The LHS of the equation captures the outflows from the pool of non-recruiting matched firms which is driven solely by exogenous separations. Although the mass of non-recruiting matched firms is not the same as that of recruiting matched firms, the former still observes the same distribution as the latter as both δ and s are independent of match quality. Finally, it is important to notice that the universe of matched firms is given by $v^m + v^i$. This implies the following accounting identity: total employed equals the total number of matched firms:

$$1-u = v^m + v^i \tag{19}$$

Unfilled and unexpired vacancies Having described the steady state stock of matched vacancies and the unemployed, the final component in the labor market tightness that remains to be described is the stock of unfilled unexpired vacancies. This mass can be described by the equation below:

$$\Big(q\left[1-\Pi\left(\tilde{x}\right)\right]+\delta\Big)v^{u}=v^{new}+sv^{m} \tag{20}$$

where v^{new} represents the new vacancies that are created in the current period. The LHS of equation (20) represents the outflows from the stock of unfilled vacancies. This consists of those unfilled vacancies that expire at a rate δ and also includes those unfilled vacancies which successfully convert a meeting with an applicant into a hire. The RHS represents the inflows which come from two sources. First, all newly created vacancies add to the future stock of unfilled, unexpired vacancies. Second, all matched firms with unexpired vacancies which undergo an exogenous separation join the stock of unfilled, unexpired vacancies.

It is important to emphasize that v^{new} is not counted as part of the vacancies available for matching today.[19] In the continuous time limit, $\theta=(v^{u}+v^{m})/u\equiv v/u$. Thus, unemployed workers can only match with existing/old vacancies. New vacancies only add to the stock of unfilled vacancies in the future.

Using equations (15), (16) and (20), the ratio of unfilled vacancies to new vacancies, and the ratio of recruiting matched firms to new vacancies can be written as:

$$\frac{v^{u}}{v^{new}}=\frac{1}{\delta}\left[\frac{\delta+s}{q[1-\Pi(\widetilde{x})]+\delta+s}\right]\qquad\text{and}\qquad\frac{v^{m}}{v^{new}}=\frac{1}{\delta}\left[\frac{q[1-\Pi(\widetilde{x})]}{q[1-\Pi(\widetilde{x})]+\delta+s}\right] \tag{21}$$

which implies that ratio of matched vacancies to matched recruiting firms is given by:

$$\frac{v^{u}}{v^{m}}=\frac{\delta+s}{q\left[1-\Pi(\widetilde{x})\right]} \tag{22}$$

Given a level of q and $\widetilde{x}$, a lower v^{u}/v^{m} implies a much larger wedge between meeting and hiring rates as unemployed job-seekers who contact a recruiting matched firm must surpass the incumbent worker's match quality before they are hired.

Equation (21) allows us to relate the ratio of new vacancies to old vacancies to δ:

$$\frac{v^{new}}{v}=\delta \tag{23}$$

Ceteris paribus, the higher the expiration rate of vacancies δ, the larger the ratio of new vacancies to older vacancies. While the expiration rate δ is not directly observable given the data we use in this

[19] The total vacancies available for matching at time t are given by $v_t=(1-\delta\Delta)(v_{t-\Delta}^{u}+v_{t-\Delta}^{m})+v_t^{new}\Delta$ where Δ is the length of one period, $(1-\delta\Delta)(v_{t-\Delta}^{u}+v_{t-\Delta}^{m})$ is the stock of unexpired vacancies from the end of period $t-\Delta$. v_t^{new} is the number of new vacancies posted per unit time. Since each period is Δ units long, the total number of new vacancies posted in period t is $v_t^{new}\Delta$. Thus, market tightness can be written as:

$$\theta_t=\frac{(1-\delta\Delta)(v_{t-\Delta}^{u}+v_{t-\Delta}^{m})+v_t^{new}\Delta}{u_{t-\Delta}}$$

In the continuous time limit, $\Delta\to 0$, the term $v^{new}\Delta$ becomes vanishingly small implying that current vacancies available for matching at period t consist only of existing or *old* vacancies $v_t^{u}+v_t^{m}$.

paper, equation (23) from the model allows us to infer δ given data on the number of new vacancies posted and the stock of continuing vacancies. We use this relation when we test the empirical relation of our model in section 4 and also when we calibrate our model in section 4.2. Finally, we can derive the mass of new vacancies posted using the free entry condition (25) and the definition of the meeting function (1):

$$v^{new} = \frac{\delta \left(\xi/q\right)^{\frac{1}{\alpha}}}{1 + \frac{\delta+s}{s}\left(\xi/q\right)^{\frac{1}{\alpha}}\left[\frac{q[1-\Pi(\widetilde{x})]}{q[1-\Pi(\widetilde{x})+\delta+s}\right]} \tag{24}$$

3.5 Closing the Model

The entire model so far has been summarized by the surplus equations and the labor market flows. However, all these relationships depend critically on the reservation match quality, $\widetilde{x}$, and the job-filling rate q which in turn is a function of labor market tightness θ. Lemma 1 summarizes the key equations which pin down the equilibrium $(\tilde{x}, \theta)$:

Lemma 1. *In steady state, the equilibrium $\tilde{x}$ and θ are determined by the following equations:*

$$(\rho+\delta)\,\chi = q(1-\eta)\int_{\tilde{x}}^{\overline{x}} S\left(x,1\right) d\Pi\left(x\right) \tag{25}$$

$$\widetilde{x} = \rho U = b + \eta p\left[\left(\frac{v^u}{v}\right)\int_{\widetilde{x}}^{\overline{x}} S\left(y,1\right) d\Pi\left(y\right) + \left(\frac{v^m}{v}\right)\int_{\widetilde{x}}^{\overline{x}}\int_{x}^{\overline{x}} S\left(y,1\right) d\Pi\left(y\right) dF\left(x\right)\right] \tag{26}$$

where

$$S\left(x,0\right) = \frac{x-\tilde{x}}{\rho+s} \tag{27}$$

$$S\left(x,1\right) = \left[\frac{\delta+\rho+s}{\rho+s}\right]\int_{\tilde{x}}^{x} \frac{1}{\varrho(x)^{\eta}\varrho(y)^{1-\eta}} dy \tag{28}$$

Proof. See Appendix B.2. □

Equation (27) describes the surplus of a match between a worker and a non-recruiting firm and is the same as (10). Equation (28) describes the surplus of a match between a worker and a recruiting firm and is the solution to the equation (11).[20]

Equation (25) is the free-entry condition where we have used the fact that $J(1) = \chi$ and that the firm receives $1-\eta$ share of the match surplus in (8). Equation (25) describes the minimum level of match-productivity for which a firm with an unfilled vacancy is willing to form a match for a given θ - or how *selective* a firm is as a function of labor market tightness. Appendix D shows that (25) implies a negative relationship between $\widetilde{x}$ and θ - firms are more selective when the rate of contacting applicants is high. In a slack labor market (lower θ, high q), holding out for a better worker is relatively costless for the firm, and hence the firm raises the minimum level of match quality $\widetilde{x}$ for which it is willing to accept a worker. Conversely, in a tight labor market (low q), holding out for a better applicant is

[20]See Appendix B.2 for details.

more costly as the firm is unlikely to meet another applicant soon. As such, tight labor markets are associated with lower firm selectivity.

Equation (26) defines the reservation match quality for which a match is formed and is derived by evaluating (11) at $x = \widetilde{x}$. Given a value of unemployment ρU, (26) describes the lowest realization of match quality for which an unemployed worker will be willing to transition into employment. Clearly, the higher the value of unemployment, the more *selective* a worker is. Since a tighter labor market (higher θ) implies a higher value of unemployment ρU, (26) implies a positive relationship between $\widetilde{x}$ and θ.[21] We refer to this curve as the *workers' indifference condition*.

Finally, the unique equilibrium level of selectivity $\widetilde{x}$ and labor market tightness θ is given by the intersection of the two aforementioned curves describing workers' selectivity vs firms' selectivity respectively. Next, we describe how changes in key parameters affect equilibrium outcomes.

3.6 Forces at play

We uncover three channels through which increased replacement hiring can cause a divergence between productivity and wages. Briefly, the three channels can be broadly thought of as (i) a change in *firms' market power*, (ii) *increased job insecurity* faced by workers and (iii) *measured matching efficiency*. Within the model, we measure *market power* as the value of an unfilled vacancy $J(1)$. Next, we define *job insecurity* as the fraction of total separations that stem from a firm replacing its current worker with a better applicant. Formally, this can be written as:

$$\textit{job insecurity} = \frac{v^m q \int_{\widetilde{x}}^{\overline{x}} [1 - \Pi(\varepsilon)]\, dF(\varepsilon)}{s(1-u) + v^m q \int_{\widetilde{x}}^{\overline{x}} [1 - \Pi(\varepsilon)]\, dF(\varepsilon)} \tag{29}$$

The denominator of the expression above represents total separations in the economy while the numerator captures only endogenous separations which arise when a firm does on-the-job search. Finally, *measured matching efficiency* is defined as the wedge between the meeting and hiring rates faced by workers and can be written as:[22]

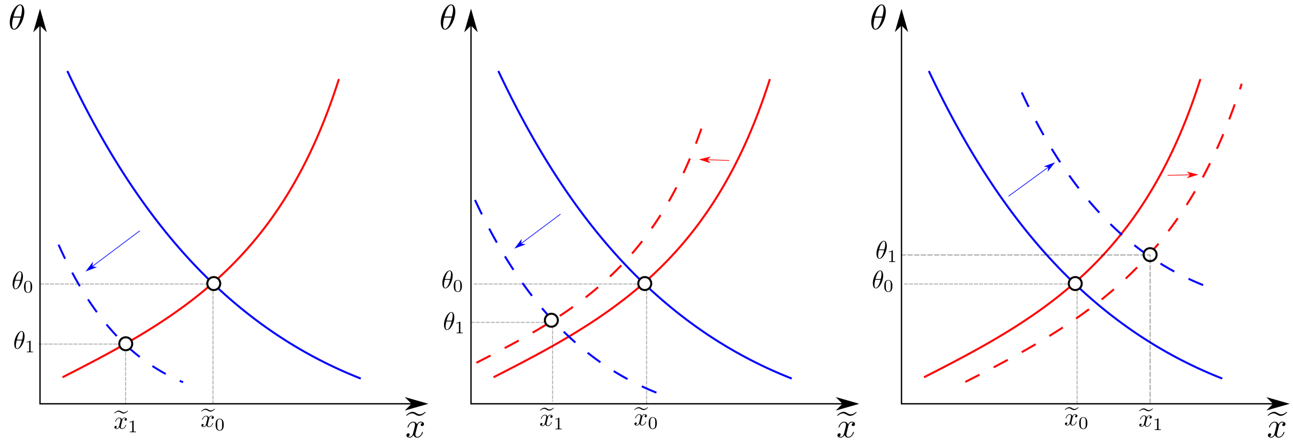
$$\textit{measured matching efficiency} = \frac{v^u}{v} [1 - \Pi(\widetilde{x})] + \frac{v^m}{v} \int_{\widetilde{x}}^{\overline{x}} [1 - \Pi(\varepsilon)]\, dF(\varepsilon) \tag{30}$$

where the first term denotes the probability that an applicant meets an unfilled vacancy multiplied by the probability of being accepted by an unfilled vacancy while the second term is the probability that an applicant meets a currently matched recruiting firm multiplied by the average probability of being accepted by such a firm. The market power of firms derives from the longevity of vacancies while the other two channels stem from the ability of firms to conduct on-the-job search.

In our model, three parameters are key for affecting the replacement hiring share: χ, the value of an unfilled vacancy, δ, the expiration rate of vacancies, and s, the exogenous separation rate. We explain how changes in each of these three parameters which impact the replacement hiring share, operate

[21] See Appendix D for a proof.

[22] Note that the rate at which applicants meet vacancies is p while the rate at which applicants get hired is just the product of p and measured matching efficiency. In other words, measured matching efficiency is the same as the unconditional acceptance rate.



(a) Increase in χ (b) Fall in δ (c) Fall in s

Figure 4: Comparative statics

through the channels we highlighted above to affect the productivity-wage gap.

Increase in χ The solid blue curve in Figure 4a corresponds to the free-entry curve with cost χ_0 while the red curve corresponds to the workers' indifference condition. Equilibrium $(\widetilde{x}_0, \theta_0)$ are given by the intersection of these two solid curves in Figure 4a. Now consider an increase in the vacancy posting cost from χ_0 to χ_1. A higher χ, which also corresponds to a higher value of an unfilled vacancy, lowers a firm's gain to matching, reducing the number of vacancies posted. This corresponds to an inward shift in the free entry curve (solid blue curve to the dashed blue curve). A change in χ does not directly affect the workers' indifference condition, leaving the red curve unchanged. However, since there are fewer vacancies to match with, the rate at which workers meet firms falls, lowering the value of unemployment. As such, this results in workers reducing how selective they are in terms of reservation match quality, causing $\widetilde{x}$ to fall (represented by a south-west movement along the red curve). Thus, in the new equilibrium, firms post fewer vacancies and workers are less selective, i.e. θ and $\widetilde{x}$ are both lower.

Since there is greater slack in labor markets in the new equilibrium, recruiting firms now contact unemployed job-seekers at a higher rate (higher q). The higher q allows: (i) currently recruiting matched firms more opportunities to replace their current worker with a better applicant and (ii) recently separated recruiting firms to more easily refill their vacated positions. At the same time, lower $\widetilde{x}$ implies that the recruiting matched firms also have an increased likelihood of meeting a new applicant with match quality superior to that of their current worker. Both these forces combined result in a higher replacement hiring share.

These forces which led to an increase in the replacement hiring share in turn depress wages while raising average productivity in the economy. When matched recruiting firms have greater opportunities to meet new applicants, they have an increased incidence of replacing current workers with better applicants of higher match quality, causing average labor productivity in the economy to increase. At the same time, higher χ raises the value of an unfilled vacancy and hence the recruiting firm's outside option, allowing recruiting firms to extract a larger share of the surplus and offer lower wages to workers.

The higher outside option of firms is akin to an increase in firms' *market power* and is the first channel through which an increase in χ results in a greater incidence of replacement hiring while simultaneously widening the productivity-wage gap.

Higher q and lower $\widetilde{x}$ also imply greater *job insecurity* faced by workers. With some math, observe that Equation (29) can also be expressed at:

$$\text{job insecurity} = \frac{q \int_{\widetilde{x}}^{\overline{x}} \left[1 - \Pi\left(\epsilon\right)\right] dF\left(\epsilon\right)}{s + \delta + q \int_{\widetilde{x}}^{\overline{x}} \left[1 - \Pi\left(\epsilon\right)\right] dF\left(\epsilon\right)} \tag{31}$$

Since the term in the numerator appears also in the denominator, for a given distribution of workers, a higher q and lower $\widetilde{x}$ causes the numerator to increase by more than the increase in the denominator. Intuitively, job insecurity increases when q rises and $\widetilde{x}$ falls as recruiting matched firms have greater opportunities to meet new applicants. In addition, because workers are now less selective over their reservation match quality, existing matched workers are now more likely to be displaced into unemployment whenever the recruiting matched firm meets a new applicant.

Finally, higher q and lower $\widetilde{x}$ also imply that most vacancies that applicants meet are filled vacancies. Equation (22) shows that a lower $\widetilde{x}$ and higher q tilts the composition of vacancies towards recruiting matched firms, resulting in a larger wedge between meeting rates and hiring rates, i.e. lower *measured matching efficiency*. Measured matching efficiency falls as unemployed job-seekers must now draw match quality higher than that of the firm's current employee before they are hired. Both the increase in job insecurity and fall in measured matching efficiency further drive the value of unemployment downwards, and consequently wages become more depressed while average productivity rises with the increase in replacement hiring.

Decrease in δ Next, consider a decrease in the expiration rate of vacancies, δ, keeping the flow cost of vacancies constant.[23] A lower δ raises the value of an unfilled vacancy as vacancies last for longer. As above, this higher value of an unfilled vacancy lowers the firm's gain to matching, causing the free entry curve to shift inward from the solid blue curve to the dashed blue curve in Figure 4b. Longer-lived vacancies also lower the value of unemployment since workers now transition at a lower rate from a recruiting matched firm-worker pair to a non-recruiting firm-worker pair where surplus in the latter is higher. This lower unemployment value shifts the workers' indifference condition inwards from the red solid curve to the dashed red curve in Figure 4b. In other words, workers become less selective over $\widetilde{x}$ for a given level of market tightness. Overall, the shift in both curves implies that the new equilibrium features an unambiguous fall in $\widetilde{x}$. The impact on θ is less clear and depends on the relative shifts in the firm's free entry curve and the worker's indifference condition. In particular, if the value of unemployment falls a lot, workers are willing to accept very low wages. This in turn makes it increasingly profitable for firms to post vacancies, increasing θ. In contrast, if the fall in the value of unemployment is relatively muted, then workers are unwilling to suffer a large wage cut. As such, firms post fewer vacancies and θ is lower in the new equilibrium. For brevity, we focus on the second case in Figure 4b as this is the relevant case in our counterfactual exercises as shown in section 4.2.2.

The decrease in δ encourages replacement hiring since longer-lived vacancies (i) afford matched firms

[23] We refer to $(\rho + \delta)\chi$ as the flow cost of a vacancy in contrast to the fixed cost χ.

greater opportunity to meet more applicants and thus, find better workers and (ii) a greater opportunity for firms to replace workers who quit (exogenous separation into unemployment). Because the fall in δ also triggers a decline in equilibrium $(\widetilde{x}, \theta)$, these declines in θ and $\widetilde{x}$ further reinforce the rise in replacement hiring, for reasons as discussed in the case where χ increased.

From equation (31), a lower δ alongside a higher q and lower $\widetilde{x}$ causes the numerator in (31) to increase more relative to the denominator. As such, endogenous separations account for a larger share of total separation, raising the level of *job insecurity* faced by workers. Intuitively, the fall in δ increases the job insecurity faced by workers since recruiting matched firms not only observe longer intervals of time to conduct on-the-job search, but workers are also more likely to be replaced when existing match quality is already lower (lower $\widetilde{x}$) and when firms have a higher rate of contacting applicants (higher q).

At the same time, a lower δ elevates the firms' outside option by raising the value of an unfilled vacancy while the decline in θ lowers the value of unemployment by reducing the unemployed job-seeker's rate of contacting vacancies. This higher *market power* of firms alongside workers' lower unemployment values allows firms to push down wages. Finally, (22) highlights that a lower $\widetilde{x}$ and δ together with a higher q shift the composition of vacancies towards recruiting matched firms, causing *measured matching efficiency* to fall. The combined effect of lower job security, falling matching efficiency and a higher value of an unfilled vacancy causes wages to fall with the decline in δ. In contrast, average labor productivity in the economy rises because recruiting matched firms have a longer window to search for better applicants to replace their current workers.

Decline in s A decline in s causes both the firm's gain to matching and the unemployment value of a worker to increase as matches now last for longer. The higher unemployment value causes workers to be more selective over the minimum match quality job they are willing to accept. This shifts the workers indifference condition outwards (solid red curve to dashed red curve in Figure 4c). At the same time, since the benefit of creating a job has increased for any $\widetilde{x}$, firms post more vacancies, shifting the free-entry curve outwards. The shift in the two curves unambiguously raises the reservation match quality $\widetilde{x}$ in the new equilibrium but the effect on θ is ambiguous. If the value of unemployment rises relatively more than the firms' value from creating a job, the higher wages demanded by workers cause firms to post fewer vacancies in the new equilibrium. Alternatively, if the value of unemployment rises relatively less, then firms take advantage of the lower wages that they get to enjoy for longer durations of a match and post more vacancies. In what follows, we focus on the case where the worker's unemployment value rises by less than the firm's value of creating a job because this is the relevant case in our counterfactual exercises using the calibrated model.

When matches last longer, firms need to replace fewer workers due to exogenous separations. As such the replacement hiring share falls. At the same time, a higher θ and $\widetilde{x}$ tends to raise wages and productivity. The higher wages are partly due to the higher unemployment value and partly due to the fact that the average match quality of jobs is higher. Further, from (18), a fall in s corresponds to a fall in s/δ which tilts the composition of matched firms relatively more towards non-recruiting firms, further raising average wages in the economy since non-recruiting firms have lower (zero) outside options. This effectively reduces the average market power of firms.

Overall, a lower s tends to reduce the replacement hiring share, raise average wages and labor productivity prdoductivity. A higher s does the opposite. Alongside the fact that job destruction rates have been declining over time in the data, the analysis above makes clear that the key parameters which can account for a higher replacement hiring share and a widening gap between wages and labor productivity are going to primarily be χ and δ and not s. We confirm this in our counterfactual exercises using the calibrated model in Section 4.2.2.

4 Empirical Investigation

4.1 Testing the Predictions of the Model

Our model provides some testable predictions which we now verify against the data. For this purpose, we use the information in the QWI and the HWOL dataset to test some of the predictions of the model. The HWOL provides monthly information on the total volume of job advertisements posted online for the periods spanning May 2005 to the present at both the industry, state and the national level. In addition, the HWOL data contains information on the volume of job postings that are new. A job posting is counted as a 'new' job advertisement only in the month that it first appears. Because different states participated in the QWI and started supplying data at different points in time, we focus on cross-sectional variation rather than variation across time when comparing data from the HWOL to data from the QWI.[24]

4.1.1 Expiration Rates and Replacement Share

Equation (23) provides an exact mapping from the ratio of new-to-old vacancies to the vacancy expiration rate, δ. Thus, while we do not have direct measures of the expiration rate, we can compute the model implied expiration rate by using information on the volume of old and new vacancies. Using the HWOL data, we can construct an estimate of the average quarterly expiration rate, δ.[25]

The replacement hiring share is positively correlated with v^m/v, the fraction of vacancies that are made up of recruiting matched firms. Using equation (21), this ratio can be expressed

$$\frac{v^m}{v} = \frac{q[1 - \Pi(\widetilde{x})]}{q[1 - \Pi(\widetilde{x})] + \delta + s}$$

While q and $\widetilde{x}$ are endogenous objects and will change whenever δ changes, the term $q\,[1 - \Pi(\widetilde{x})]$ appears in both the numerator and denominator of v^m/v. Thus, no matter the sign and size of the change in $q\,[1 - \Pi(\widetilde{x})]$, a decline in δ always makes v^m/v larger.[26] Consequently, a lower δ by raising v^m/v, is associated with a higher replacement hiring share.

Figures 5a and 5b show how the average replacement hiring share is falling with δ across industries and state respectively. Using information on the replacement hiring share across industries over time,

[24] States like Massachusetts joined the QWI only in 2010 while the earliest data in the QWI dates back to 1990.

[25] As the HWOL data is monthly, we first sum up the total volume of job ads and the total volumes of new jobs ads within a quarter. The difference between the total volume of job ads and the volume of new job ads form the stock of old vacancies, $v^m + v^u$. We then compute the δ from these quarterly measures.

[26] To see this, observe that if in equilibrium, a decline in δ causes $q[1 - \Pi(\widetilde{x})]$ to fall, then the denominator falls faster than the numerator, causing v^m/v to rise. If instead the decline in δ causes $q[1 - \Pi(\widetilde{x})]$ to rise, then the denominator increases less than the numerator, causing v^m/v to again rise.

column 1 of Table 1 shows a 1 percentage point increase in δ lowers the replacement hiring share by about 0.36 percentage points. A similar negative relationship exists between the replacement hiring share across states and δ although the coefficient on δ in the state-level regression is not statistically significant.[27] Overall, the data suggests that the replacement hiring share decreases when vacancies expire more quickly and firms have a shorter window to re-match with better applicants.

(a) Variation Across Industries (b) Variation Across State

Figure 5: Replacement hiring share declining in δ

Table 1: Replacement Hiring Share And Earnings in the Data

Variable	**Industry**	**State**	**Industry**	**State**
δ	-0.359***	-0.014	-	-
	(0.122)	(0.017)	-	-
s/δ	-	-	-0.779***	-1.031**
	-	-	(0.200)	(0.424)
N	918	2421	918	2421
R^2	0.139	0.071	0.253	0.057

Notes: In columns (2) and (3), the dependent variable is the replacement hiring share as measured in the QWI while the independent variable is the ratio of new-to-old vacancies as measured from the HWOL. In columns (4) and (5), the dependent variable is the natural log of deflated earnings while the independent variable is ratio of separation rates to expiration rates. In columns (2) and (3), the unit of observation is an industry, while in columns (4) and (5), the unit of observation is a state. All regressions include unemployment rates and an indicator for the great recession as controls. Standard errors are clustered by industry (state) and are reported in parentheses. $***$ denotes significance at 1% while $**$ denotes significance at 5%.

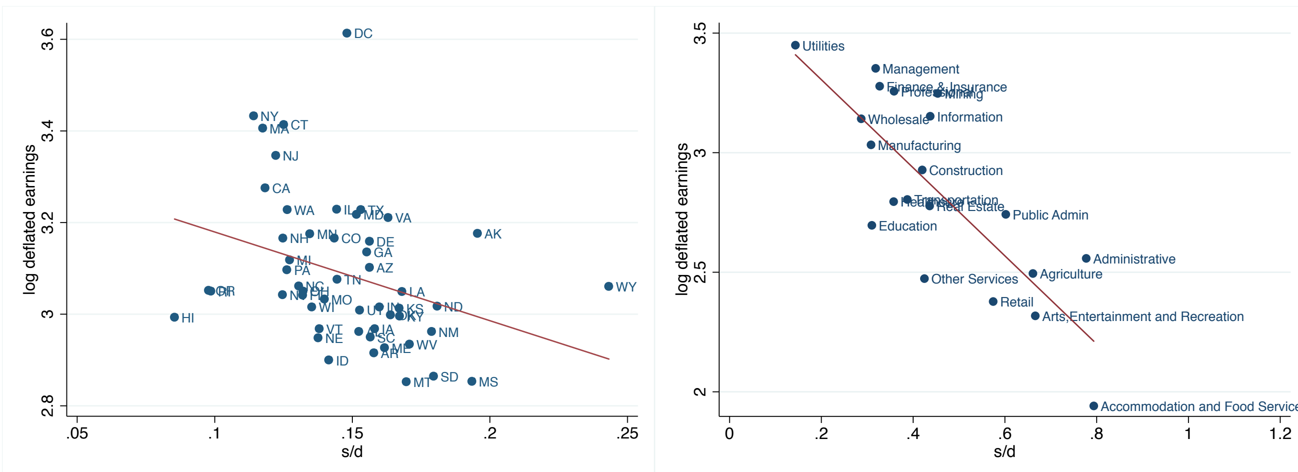
4.1.2 Earnings and Composition of Matched Firms

Our model also predicts that average earnings are declining in the ratio of recruiting matched firms to non-recruiting firms, v^m/v^i. Intuitively, the former set of firms observe a positive option value of a

[27] Specifically, we regress the replacement hiring share against δ, the national unemployment rate and a dummy for the Great Recession. All standard errors are clustered at the industry or state level. In our regressions, we use all available time-series information on replacement hiring shares at the industry (state) level. Since different states joined the QWI at different points in time, this gives us an unbalanced panel for the coverage period of 1990q1-2016q4.

vacancy, allowing them to offer wage discounts to their workers. Holding all else constant, the average wage rate and earnings in the economy are lower whenever the composition of firms tilts towards recruiting matched firms. In our model, equation (18) relates the ratio of recruiting to non-recruiting matched firms to the ratio of exogenous separation rates to vacancy expiration rates, $v^m/v^i = s/\delta$.

The QWI provides information on the beginning-of-quarter separation rates as well as information on the average earnings of all employees. We use the beginning-of-quarter separation rates as our proxy for s. Combining this measure with our estimate of δ from the HWOL data, our model suggests that industries (states) with a higher s/δ should have lower wages on average. The QWI lacks information on the number of hours worked so we use earnings to proxy for wages. We deflate the measure of average monthly earnings of all employees by the Consumer Price Index (CPI) and take the natural log of deflated earnings.



(a) Variation Across State

(b) Variation Across Industries

Figure 6: Log earnings are declining in s/δ

Figure 6 shows that (log) earnings are in fact declining in our estimated measure of v^m/v^i for both industries and states. Regressing (log) deflated earnings on s/δ at the industry level reveals that a 1 percentage point increase in v^m/v^i lowers earnings of all employees by about 0.78 log points. Similarly, at the state level, a 1 percentage point increase in v^m/v^i lowers earnings of all employees by about 1 log point. Overall, the data reaffirms our model's prediction that firms' market power - as measured by the fraction of firms with a positive outside option - is an important factor explaining why wages in some states and industries are so low.

4.2 Numerical Exercise

4.2.1 Calibration

We set a period in our model to be one month but time-aggregate all our model generated moments to a quarterly frequency.[28] The discount rate ρ is accordingly set to 0.004 to reflect an annual interest rate of about five percent. We set the elasticity of the meeting function with respect to unemployment

[28]We do this since the replacement hiring share is measured at a quarterly frequency from the QWI.

α to be 0.5 as is standard in the literature.

The empirical literature has typically found that wages are log-normally distributed. Thus, we assume that the distribution of match quality, $\Pi(x)$ is given by the log of $N(-\sigma_x^2/2, \sigma_x^2)$. Overall, this leaves us with six key parameters to calibrate: the expiration rate of vacancies δ, the exogenous separation rate s, the unemployment benefit b, the scaling parameter, ξ, the dispersion in match quality σ_x, and η the bargaining weight of the worker. The last parameter of the model, the vacancy posting cost, χ, is calculated as the residual to equation (25).

The set of parameters to be calibrated: $s, b, \delta, \sigma_x, \xi$ and η affect key labor market outcomes in our model. We jointly calibrate these parameters by targeting the following moments: the exit rate of employed individuals, the job-finding rate of the unemployed, the unemployment rate, the unemployment insurance replacement rate and the labor share of output. Since there is a distinction between gross and net flows in our model, targeting the exit rate from employment, the job-finding rate and the unemployment rate does not automatically make any one of the moments a linear combination of the others. The exit rate from employment is a function of total gross separations which is composed of both exogenous separations as well as endogenous separations stemming from firm on-the-job search. The mass of unemployed, however, is only affected by exogenous separations (net inflows into unemployment).

We also target the replacement hiring share in our model. As per the definition in the QWI, we calculate replacement hires as the amount of hires *in excess* of net employment change and divide this by total hires to get the share. The time aggregation in our model allows us to capture within a quarter, 1) the replacement hiring conducted by recruiting firms whose workers exogenously separated into unemployment in the previous month and 2) the replacement hiring conducted to replace current workers with higher match productivity workers in the current month. Our computed measure of replacement hiring thus captures both margins of replacement hiring due to "quits" (replacing workers who have left) and replacement hiring for the purposes of finding a better worker.

We conduct two separate exercises. We first calibrate the model to data moments for the period 1990m1 to 2005m5.[29] We then re-calibrate the model to data moments for the period post-2005m5. We choose the date of 2005m5 as our cut-off point as that is the date with which the HWOL data starts, and for which we have data on our model implied vacancy expiration rate, δ.[30] For both time periods, we follow Shimer (2005) and target a 40 percent unemployment insurance replacement rate to pin down the value of home production b. For the period of 1990m1-2005m5, we target a labor share of 0.67 to pin down η. We fix η in the post 2005m5 period to be the same as our calibrated value of η for the pre-2005 period so that any change in the labor share stems only from changes in the factors that affected the replacement hiring share and not from a change in the worker's bargaining weight over the two time periods.

Using data from the Current Population Survey (CPS) on employment, unemployment and short term unemployment, we find, for the period 1990m1-2005m5, that the average monthly exit probability of an employed individual is about 0.032 while the average monthly job finding probability of an

[29] Our data on replacement hiring comes from the QWI where the earliest year when information was recorded is in 1990

[30] As aforementioned, Barnichon (2010) documents that the share of print advertising observes a sharp drop-off during the 2000s and a shift towards online advertising during that period. We assume that any changes in the expiration rate of vacancies that arise from the different technology of posting vacancies can be broadly captured by calibrating the model to the period before and after online vacancies became predominant.

unemployed individual is given by 0.44.[31] In continuous time, this would imply that workers leave employment with at a quarterly rate of $-3*\log(1-0.032)$ and find jobs at a rate of $-3*\log(1-0.44)$. The average unemployment rate during this period is about 5.6%. Using data from the QWI, we calculate that the average share of replacement hires for the period 1990Q1 to 2005Q2 is about 0.34.[32]

For the period post-2005, we find an average unemployment rate of 6.1%, an average exit probability of 0.023, an average job-finding probability of 0.29 and a replacement hiring share of about 0.37. Because the ratio of new-to-old vacancies is related to the vacancy expiration rate in our model, we use information from the HWOL to fix δ in the post-2005 period. In the HWOL data, the average ratio of new-to-old vacancies is about 0.92.

Table 2: Model Parameters

Fixed Parameters			
Parameter	Description	Value	
ρ	discount rate	0.004	
z	agg. productivity	1	
α	matching function elasticity	0.5	
Calibrated Parameters (1990-2005)			
Parameter	Target	Value	Model Moment
η	labor share of 0.67	0.100	0.692
δ	replacement hiring share of 0.34	2.370	0.341
b	40% UI replacement rate	0.329	0.454
ξ	job finding rate of 1.79	2.055	1.798
s	exit rate of 0.098	0.032	0.109
σ_x	unemployment rate of 0.056	0.088	0.056
Calibrated Parameters (post-2005)			
Parameter	Target	Value	Model Moment
δ	Set to ratio of $\frac{v^{new}}{v^{old}}$	0.92	0.92
b	40% UI replacement rate	0.290	0.433
ξ	job finding rate of 1.03	1.0611	1.064
s	exit rate of 0.070	0.017	0.065
σ_x	unemployment rate of 0.061	0.087	0.059
	replacement hiring share of 0.37	-	0.372

Table 2 summarizes our calibrated parameters. It is worth noting that while not an explicit target of our calibration procedure, our calibrated value of the exogenous separation rate in the post-2005 period, which is our proxy for quits into unemployment, comes close to the quit rate recorded in JOLTS over that same period. In particular, we calibrate $s = 0.017$ in the post-2005 period while in the JOLTS data,

[31] We calculate the unemployment outflow and inflows rates by following the method proposed in Shimer (2012).

[32] Our calibrated parameters would not change much if we expand the sample to be between 1951m1 to 2005m5 as the average unemployment rate during that period is about 5.5%, the exit probability is about 0.034 and the job finding probability is about 0.45. The targeted labor market moments for the period 1990m1-2005m5 is not substantially different from the period 1951m1 to 2005m5.

the quits rate is about 0.018. Thus, our model replicates the frequency with which firms experience quits from its workers.

Notably, our calibrated measure of δ for the period 1990-2005m5 is significantly higher than the average value of δ observed in the data for the post-2005 period. This accords with the general downward trend in the ratio of new-to-old vacancies observed in the HWOL data over the last decade.[33] Overall, our results suggest that prior to 2005, 90% of vacancies used to expire within a month. In contrast, our model suggests that 60% of vacancies expire within a month post-2005, implying that vacancies last longer with the advent of online vacancies.

In addition, the value of an unfilled active vacancy for the period 1990-2005m5 is given by $\chi = 5.29$ and for the period post-2005 by $\chi = 11.67$. While these fixed costs are not directly comparable to the flow cost of posting a vacancy, our calibrated values of χ generate vacancy rates - defined as the ratio of vacancies to all jobs - that are comparable with the data. In particular, the model predicts vacancy rates of 2.3 and 2.8 percent in the pre- and post-2005 periods respectively, similar to the vacancy rates observed in JOLTS of 2.7 and 3.0 percent over the same time periods. Thus, while χ was solved for as a residual to equation (25), our calibrated values of χ give rise to vacancy rates similar to those observed in the data. Further, the increase in χ over time is equivalent to the value of a firm increasing over the two periods. As per Hall (2017), the value of a new firm confers information on the financial valuation of the firm. Using data on the stock market index[34] and taking the average for the periods 1990m1-2005m5 and for post-2005 separately, we note that the S&P 500 stock index rose by about 94%, which is similar to the 120% increase in our model calibrated χ across the two time periods.

Finally, our calibration exercise suggests that reservation match qualities, $\widetilde{x}$, declined between the two time periods (see Table 3).

4.2.2 Quantitative Predictions

A key question is whether our model, which is calibrated to match the replacement hiring share across the two time periods, can also replicate the widening productivity-wage gap over the two time periods. Table 3 shows the results for the non-targeted moments of our model.

It follows directly from our discussion in Section 3.6 that the lower δ and higher χ observed in the second time period create a more favorable environment for on-the-job search by recruiting matched firms and supports a higher replacement hiring share. Since the lower δ and higher χ observed in the second time period gives rise to a lower $\widetilde{x}$ and higher q, the composition of vacancies shifts more towards recruiting matched firms (from (22)). Alongside the observed decrease in s between the two time periods, this implies that most of the increase in the replacement hiring share stems from firms who are replacing current workers with better applicants rather than refilling vacated positions.[35] This increased replacement hiring share which stems mainly from firms doing on-the-job search facilitates a widening gap between productivity and the average wage, as discussed in Section 3.6. Accordingly, we observe that labor productivity across the two time periods increases by approximately 1% while

[33] Figure 10b in the appendix demonstrates that δ observes a downward trend in the HWOL

[34] Following Hall (2017), we take the data for the S&P 500 stock portfolio from Robert Shiller's website, http://www.econ.yale.edu/~shiller/data.htm

[35] In fact, in our numerical exercises, the fraction of replacement hires conducted to replace workers lost due to exogenous separations approximately halved between the time periods.

average wages decline by about 7% during the same period.[36] Measuring the productivity-wage gap as the ratio of labor productivity to mean wage earnings, our model predicts that the productivity-wage gap widened by 8% with the 3 percentage point increase in the share of replacement hiring. Importantly, this predicted divergence in the productivity-wage gap due to the rise in the replacement hiring share is associated with a decline in the labor share. In our simulations, the labor share declines by about 5 percentage points across the two time periods. This fall in the share of income to labor is reminiscent of the US experience. In particular, Elsby et al. (2013) find that the labor share of income has declined by about 6.4 percentage points since the 1980s.

Recall from the discussion in Section 3.6 that there are three channels through which increased replacement hiring can give rise to a widening productivity-wage gap. First, market power of recruiting firms, as measured by the value of an unfilled vacancy, more than doubles between the two time periods. Next, job insecurity, as measured by the proportion of separations that are due to firms conducting on-the-job search, rises by about 3 percentage points between the two time periods. Finally, measured matching efficiency declines about 6.6% between the two time periods which further lowers the value of unemployment and hence average wages. All these three forces work towards pushing wages down while raising productivity.

Table 3: Non-targeted Model-Generated Moments

	Pre-2005	Post-2005	Post-2005 (no OTJ)
$\widetilde{x}$	0.697	0.639	0.876
q	3.329	1.580	1.974
Job Insecurity	0.332	0.365	0
Measured Matching Efficiency	0.627	0.586	0.931
Mean Wage	0.722	0.670	0.655
Labor Productivity, Y/N	1.04	1.05	1.01
Labor Share	0.692	0.639	0.647

To understand what portion of the rise in the productivity-wage gap stems from the increased outside option of firms (greater market power) and which portion stems from the more favorable conditions for firms to do on-the-job search, we impose the parameters from the post-2005 period but shut down on-the-job search by firms.[37] This exercise allows us to isolate how much an increased outside option of the firm alone affects the productivity-wage gap in the economy. A priori, it is unclear if shutting off on-the-job search necessarily increases the average wage in the economy. On one hand, the absence of on-the-job search implies lower job insecurity[38] for workers and that all vacancies available to unemployed job-seekers are unfilled vacancies. The former raises the value of unemployment through longer spells of employment and the latter raises the value of unemployment through a smaller wedge between the rate at which a worker meets a vacancy and the rate at which he is accepted (higher measured matching

[36] We measure labor productivity in terms of total market output divided by total number of employed individuals. Note that this does not count output produced from home production by the unemployed and is consistent with the way GDP and labor productivity is measured in the National Income accounts.

[37] Appendix E shows the equations for the model without firm on-the-job search.

[38] By definition, job insecurity, which is the fraction of separations due to firms replacing current workers with better applicants, is equal to zero when on-the-job search is shut down. In other words, only exogenous separations occur when matched firms are prevented from re-matching with new applicants.

efficiency). Thus, these two effects provide a force towards higher wages in the absence of on-the-job search. On the other hand, the higher value of unemployment drives up the reservation match productivity workers are willing to accept. This higher selectivity by workers implies that, on average, more jobs are rejected, lowering the expected net benefit of creating a job for the firm and consequently the number of vacancies posted. This in turn acts towards reducing the value of unemployment for workers. Given these countervailing forces, the impact on wages in the absence of on-the-job search is ambiguous. Our numerical exercises suggest that the reduction in surplus dominates the fall in job insecurity and rise in matching efficiency, mean wages in the post-2005 period is lower at 0.66 when on-the-job search is shut off.[39]

Although the removal of on-the-job search further lowers the mean wage in our exercise, the productivity-wage gap is actually 3% lower than what the model would predict with on-the-job search. This is because recruiting firms can no longer receive productivity gains from re-matching with higher match productivity applicants. As such, our model shows that labor productivity falls by about 4 percent in the post-2005 period when on-the-job search is removed. This fall in labor productivity, and thus output, is far larger than the fall in wage income, causing the productivity-wage gap to be smaller in the economy without firm on-the-job search. Consequently, the labor share is also about one percentage point higher in the post-2005 period when on-the-job search is removed.

4.2.3 Identifying the key forces at play

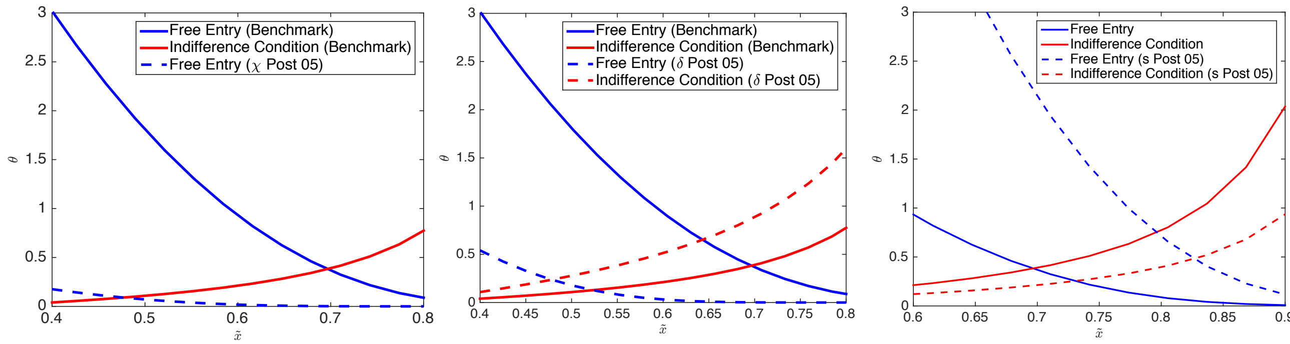
As foreshadowed in Section 3.6, we now analyze the factors that were most important towards contributing to the rise in the replacement hiring share and the divergence between productivity and wages across the two time periods. Our calibration exercise in section 4.2 revealed that χ increased while δ and s fell across the two time periods. Setting as our benchmark economy the model that we parameterized to match the aggregate moments in the period pre-2005, we ask how much the replacement hiring share and the productivity-wage gap would change if only one of the key parameters, (χ, δ, s), were updated to its post-2005 value. All results are summarized in Table 4. Figure 7 depicts how the worker indifference and free entry curves move in response to changes in χ, δ and s separately.[40]

A change in χ only Figure 7a shows that the higher χ in the second time period shifts the free-entry curve inward and causes equilibrium reservation match productivity, $\widetilde{x}$, and labor market tightness, θ, to fall. The second and third columns of Table 4 compares outcomes of interest in this counterfactual to the benchmark.

The inward shift in the free-entry condition causes the mass of new vacancies to decline, reducing the amount of competition and raising the rate q at which vacancies meet unemployed job applicants. Low job-meeting rates for workers in turn cause unemployed job-seekers to be less selective over reservation match quality, causing $\widetilde{x}$ to fall by about 30%. The resulting decline in vacancy posting (higher q) and lower rejection rate by workers (lower $\widetilde{x}$) afford recruiting firms a higher rate of re-matching with new applicants. Holding all other parameters at their values in the first time period, the rise in χ alone

[39] Overall, we observe that the average surplus, $\int_{\widetilde{x}}^{\overline{x}} S(y,1)d\Pi(y)$, is lower without on-the-job search in post-2005 at 6.07 compared to an average surplus of 7.58 when firms can conduct on-the-job search.

[40] Notice that Figure 7 is simply the quantitative counterpart to Figure 4.



(a) Rise in χ (b) Fall in δ (c) Fall in s

Figure 7: New vs. old equilibria

would have caused the replacement hiring share to increase by about 13 percentage points.

This higher replacement hiring share also coincides with higher job insecurity faced by workers. Since unemployed job-seekers are willing to accept lower quality jobs (lower $\widetilde{x}$), this allows recruiting matched firms to easily replace such workers whenever they encounter a new applicant. Together with the higher meeting rate of recruiting firms, job insecurity faced by workers rises by close to 36%. Moreover, as discussed in section 3.6, a higher χ implies that a greater share of vacancies an applicant meets are filled vacancies. As such, the shift in the composition of vacancies towards matched recruiting firms causes measured matching efficiency to fall by about 25 percent and unemployment rate to rise to about 17 percent.

All three forces - a higher value of an unfilled vacancy, higher job insecurity and lower matching efficiency - act towards lowering a worker's wage relative to her labor productivity in a particular match. In our simulations, average wages fall by about 29%. At the same time, the higher χ allows recruiting matched firms to face less congestion in the labor market and to re-match with better applicants at a higher rate, causing labor productivity to rise by 3 percent. Overall, the productivity-wage gap would have risen by 44% relative to its value in the pre-2005 period. This widening productivity-wage gap in turn causes the labor share to fall by close to 20 percentage points. Clearly, if only χ had changed, i.e the only thing that changed between the two time periods was the outside option of firms, then we would have observed a much larger increase in the productivity-wage gap.

A change in δ only In this counterfactual we change the value of δ to its lower value in the second time period while keeping other parameters at their values in the baseline.[41] The third column of Table 4 compares various outcomes of interest in this counterfactual to the benchmark. Recall from section 4.2.2 that while a decrease in δ unambiguously causes $\widetilde{x}$ to decline, the effect on θ is ambiguous. Given the parameterization of our model, Figure 7b features a decline in θ because the value of unemployment declined relatively less than the firm's gain to matching. Under this configuration, the fall in θ and $\widetilde{x}$ further reinforces the impact that a decline in δ has on the replacement hring share. Not only do recruiting matched firms have a longer interval to conduct on-the-job search, but reservation match quality is lower and the rate at which recruiting firms contact applicants is higher. As such, updating

[41]Throughout this counterfactual we hold the flow cost of vacancies constant.

Table 4: Counterfactuals

Variable	Benchmark	$\Delta\chi$	$\Delta\delta$	Δs	Post 2005
$\widetilde{x}$	0.697	0.477	0.478	0.826	0.639
q	3.329	6.844	4.224	2.968	1.580
replacement share	0.341	0.469	0.547	0.316	0.372
u	0.056	0.165	0.147	0.026	0.059
mean wage	0.722	0.513	0.511	0.844	0.670
labor productivity	1.04	1.07	1.09	1.04	1.05
labor share	0.692	0.480	0.471	0.810	0.639
matching efficiency	0.627	0.473	0.382	0.642	0.586
job insecurity	0.332	0.451	0.518	0.311	0.365

Notes: The first column represents our results from our benchmark economy calibrated to match aggregate moment in the US economy for the period 1990m1-2005m5. Columns represented with a "ΔY" represents the model's outcomes when only parameter Y is updated to its post-2005 value. Finally the last column presents the model's outcomes when all parameters are updated to their post-2005 values.

δ to its post-2005 value causes the replacement hiring share to rise by about 20 percentage points.

As outlined in Section 3.6, the declines in θ and $\widetilde{x}$ cause both measured matching efficiency to fall and job insecurity faced by workers to rise. In this case, the decline in δ causes measured matching efficiency to fall by close to 40% and job insecurity faced by workers to rise by about 56%. Both these forces act towards depressing wages while the longer interval for firms to conduct on-the-job search encourages further productivity gains. As such, when δ falls to its post 2005 value, the productivity-wage gap widens as productivity increases by about 5% while wages decline by about 29%. This greater gap in productivity and wages causes labor share to decline by 22 percentage points. Thus, without any countervailing forces, the decline in δ to its post-2005 value would have elevated the replacement hiring share while simultaneously driving a wedge in the productivity-wage gap far above what is observed in the US economy.

A change in s only Finally, we focus on what would have been the implied labor market outcomes if only s were updated to its post-2005 value. As discussed in Section 3.6, a decline in s unambiguously raises $\widetilde{x}$. Figure 7c shows that the increase in the firm's gain to matching quantitatively overshadows the increase in the worker's unemployment value when s is updated to its post-2005 value. As such θ rises and q falls.

When workers and firms are more selective over reservation match quality (higher $\widetilde{x}$), and when the rate at which recruiting matched firms contact new applicants decreases (lower q), the replacement hiring share falls. This lower replacement hiring share is coincident with lower levels of job insecurity faced by workers and higher measured matching efficiency. The lower level of job security and higher measured matching efficiency in turn boosts the average wage by about 17%. In contrast, labor productivity remains relatively unchanged. Intuitively, although recruiting matched firms have less opportunity to replace current workers with better workers, higher levels of reservation match productivity imply that the average worker hired out of unemployment possesses higher productivity. As such, the higher

$\widetilde{x}$ counteracts the loss in productivity gains from reduced replacement hiring. Overall, relative to the its pre-2005 value, the productivity-wage gap would have shrunk by about 15%. This narrower productivity-wage gap implies a higher labor share of 0.81, about 11 percentage points above the labor share observed in the benchmark economy. Overall, our results suggest that the decline in the exogenous separation rate would have served to reverse the widening of the productivity-wage gap and increase the labor share.

5 Is Replacement Hiring Efficient?

A natural question that arises is whether the outcomes in such an economy are socially efficient. We explore these questions through the lens of a constrained social planner's problem. Since all agents are risk-neutral, the planner seeks to maximize lifetime discounted value of output net of vacancy creation costs by choosing the number of new vacancies and the reservation productivity $\widetilde{x}$. The objective function can be written as:

$$\max \int_0^{\infty} e^{-\rho t}\left\{-\chi v_t^{new} + bu_t + v_t^m \int_{\underline{x}}^{\overline{x}} \varepsilon dF_t\left(\varepsilon\right) + v_t^i \int_{\underline{x}}^{\overline{x}} \varepsilon dF_t\left(\varepsilon\right)\right\} dt \tag{32}$$

The first term in the parenthesis is the cost associated with posting new vacancies while the second term denotes the value of home-production from the unemployed. Because the universe of matched firms consists of both recruiting and non-recruiting firms, market output comes from two sources. The third term represents the output produced by recruiting matched firms while the last term represents that produced by non-recruiting firms. $F_t(\varepsilon)$ denotes the fraction of firms at date t with match quality less than or equal to ε while f_t denotes the corresponding density.[42]

In maximizing the objective function above, the planner must respect the constraints imposed by matching technology on labor market flows:

$$\begin{aligned}
v^m \dot{f}(x) + \dot{v}^m f(x) &= \left[q\pi\left(x\right)\left(v^u + v^m \int_{\widetilde{x}}^{x} f\left(\epsilon\right) d\epsilon\right) - \left(s + \delta + q\left[1 - \Pi\left(x\right)\right]\right) v^m f\left(x\right)\right] \forall x \geq \widetilde{x} && (33)\\
v^i \dot{f}\left(x\right) + \dot{v}^i f(x) &= \delta v^m f\left(x\right) - s v^i f\left(x\right) && (34)\\
\dot{v}^u &= v^{new} + s v^m - \left(q\left[1 - \Pi\left(\widetilde{x}\right)\right] + \delta\right) v^u && (35)\\
u &= 1 - v^m - v^u && (36)
\end{aligned}$$

where the dots represent time derivatives. These equations are simply the dynamic analogs of equations (15) - (20) and are presented in detail in Appendix C.[43] (33) denotes how the mass of matched recruiting firms with match quality x evolves over time given the planner's choice of $\widetilde{x}$ and q. Similarly, (34) describes how the mass of non-recruiting firms with match quality x evolves over time. (35) describes the evolution of unfilled vacancies which the planner can affect by choosing v^{new}. (36) simply defines the level of unemployment as the mass of individuals in the economy minus those who are employed. These constraints are the same as those which describe labor market flows in the decentralized economy

[42] Recall that the distribution of match quality across matched recruiting firms and non-recruiting firms is the same at any date. See the discussion around equation (18) for details.

[43] These equations are also available in Appendix F which characterizes the solution to the planner's problem.

but outcomes may differ due to differences in the planner's choice of reservation match quality $\widetilde{x}$ and new vacancies v^{new}.

The solution to the planner's problem is detailed in Appendix F. In a similar fashion to the decentralized model, the efficient outcomes can also be summarized by the surplus equations, the reservation threshold $\widetilde{x}$ and labor market-tightness θ. Lemma 2 below summarizes these conditions.

Lemma 2. *In steady state, the planner's optimal decisions can be summarized by the following:*

$$(\rho+\delta)\,\chi+(\tilde{x}-b)\left(\frac{q}{\xi}\right)^{\frac{1}{\alpha}}=q\int_{\tilde{x}}^{\overline{x}}\mathbb{S}\left(x,1\right)d\Pi\left(x\right) \tag{37}$$

$$\widetilde{x}=b+\alpha p\left[\left(\frac{v^{u}}{v}\right)\int_{\widetilde{x}}^{\overline{x}}\mathbb{S}\left(y,1\right)d\Pi\left(y\right)+\left(\frac{v^{m}}{v}\right)\int_{\widetilde{x}}^{\overline{x}}\int_{x}^{\overline{x}}\left[\mathbb{S}\left(y,1\right)-\mathbb{S}\left(x,1\right)\right]d\Pi\left(y\right)dF\left(x\right)dx\right] \tag{38}$$

where

$$\mathbb{S}\left(x,0\right)=\frac{x-\tilde{x}}{\rho+s} \tag{39}$$

$$\mathbb{S}\left(x,1\right)=\frac{\rho+s+\delta}{\rho+s}\int_{\widetilde{x}}^{x}\frac{dy}{\varrho(y)} \tag{40}$$

$\mathbb{S}(x,0)$ *refers to the social value of a non-recruiting matched firm-worker pair with match quality* x *and is the analog of* $S(x,0)$ *described in* (27)*. Similarly,* $\mathbb{S}(x,1)$ *denotes the social value of a recruiting matched firm-worker pair with match quality* x *and is the analog of* $S(x,1)$ *described in* (28)*.*

Proof. See Appendix F. □

Equations (37), (38), (39) and (40) are the counterparts to (25), (26), (27) and (28) respectively in the decentralized economy. (37) is the analog of the free-entry condition in the decentralized economy (25) and shows that the planner, in choosing $\widetilde{x}$ and v^{new}, balances the gain from having another matched firm-worker pair, $q\int_{\widetilde{x}}^{\overline{x}}\mathbb{S}\left(x,1\right)d\Pi\left(x\right)$ against the loss of having one less unemployed applicant in the job-seeking pool for firms to match with, $(\widetilde{x}-b)\left(q/\xi\right)^{1/\alpha}$. This stands in contrast to the decentralized economy where the firm does not internalize how meeting rates are affected when an additional vacancy is created. These are the usual *congestion* and *thick market* externalities present in many random search models.

In addition to these externalities, (38) - which is the analog to the worker's indifference condition in the decentralized economy (26) - reveals that the planner takes into account a third externality which we label the *rat race* externality. Specifically, the RHS of (38) corresponds to the social value of unemployment in the efficient allocation. The second term in the square brackets on the RHS shows that the planner only values the net gain associated with recruiting firms rematching with better applicants. When a currently matched recruiting firm replaces its current worker with match quality x with a better applicant of match quality $y>x$, the net social gain to matching is given by $\mathbb{S}(y,1)-\mathbb{S}(x,1)$. since this replacement hire also destroys value by removing the current worker from the workforce (generating a loss of $\mathbb{S}(x,1)$). In contrast, the unemployed agent and recruiting matched firm in the decentralized economy only take into account the gross gain of $S(y,1)$ and do not internalize the accompanying loss from displacing the current worker.

Since the planner only values the net gain from a replacement hire, the social value of a recruiting matched firm-worker pair is weakly less than its valuation in the decentralized economy. For any $\widetilde{x}$ and θ, we have:

$$S(x,1) = \frac{\rho+\delta+s}{\rho+s}\int_{\widetilde{x}}^{x} \frac{dy}{\varrho(x)^{\eta}\varrho(y)^{1-\eta}} \geq \frac{\rho+\delta+s}{\rho+s}\int_{\widetilde{x}}^{x} \frac{dy}{\varrho(y)} = \mathbb{S}(x,1)$$

since $\varrho(x)^{\eta}\varrho(y)^{1-\eta} \leq \varrho(y)$ for $x \geq y$. In contrast, the social and private valuation of a non-recruiting matched firm-worker pair are the same $S(x,0) = \mathbb{S}(x,0)$. Thus, the planner places a higher value on non-recruiting matched firm-worker pairs than recruiting firm-worker pairs and would like to allocate more workers to non-recruiting firms, thereby reducing the incidence with which a worker is replaced.

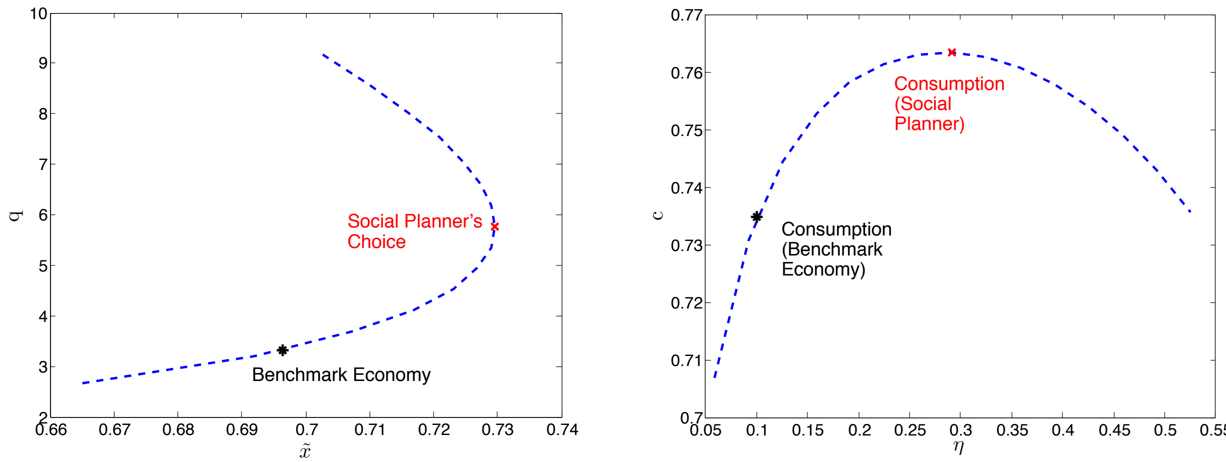
This desire to minimize the losses from replacement hiring drives the planner to reduce the mass of recruiting matched firms, v^m. Equation (21) (replicated below for convenience) shows that the planner can accomplish this by reducing v^{new} and increasing $\widetilde{x}$.

$$v^m = \frac{1}{\delta}\frac{q\left(1-\Pi[\widetilde{x}]\right)}{q\left(1-\Pi[\widetilde{x}]\right)+s+\delta}v^{new}$$

The equation demonstrates that holding all else constant, higher selectivity $\widetilde{x}$ lowers v^m as this causes more applicants to be rejected on average. Similarly, holding all else constant, a lower v^{new} causes v^m to fall, limiting the amount of replacement hiring that occurs for the purposes of replacing current workers with better applicants. Unlike lowering v^{new}, a higher $\widetilde{x}$ has the additional property that only applicants with relatively high match quality are hired in the first place. Consequently, this lowers the incidence of replacement hiring as likelihood that a current worker is replaced by a better applicant is reduced when $\widetilde{x}$ is high.

Overall, because the planner takes into account the rat race externality, she tends to raise $\widetilde{x}$ above the levels prevailing in the decentralized economy to limit the losses that arise when unemployed job-seekers displace current workers. Figure 8a depicts the planner's choice of $(\widetilde{x}, q)$ (depicted with a red "x") relative to the locus of decentralized outcomes of the same quantities for $\eta \in [0,1]$. Unlike our benchmark model where η is parameterized to match the labor share in the US economy (depicted with a black "$*$"), the planner chooses a higher $\widetilde{x}$. While there exists an $\eta = \eta_{SP}$ that can decentralize the planner's solution, in general a weakly lower $\widetilde{x}$ prevails in the decentralized economy. Again, this is because private agents in the decentralized economy do not take into account the rat race externality and hence do not internalize the losses that arise when job-seekers displace workers in lower productivity matches.

Depending on η, the decentralized economy can choose to create more or less vacancies than the planner. When η is low ($\eta \to 0$), firms receive a large share of the surplus, prompting them to create too many vacancies. Since the private recruiting firm does not internalize the congestion it imposes on other recruiting firms when it creates a vacancy, this causes q to be lower than what the planner would choose. Conversely, when η is very high, firms receive a small share of the surplus, and create too few jobs, causing q to be much higher relative to what the planner would choose. Thus, depending on η, the choice of q in the decentralized economy can be higher or lower than what the planner would choose but the decentralized economy's choice of $\widetilde{x}$ tends to overall be lower than that chosen by the planner.



(a) Choice of $(\widetilde{x}, q)$ for $\eta \in [0,1]$

(b) Losses from inefficient choice of $(\widetilde{x}, q)$

Figure 8: Efficient vs. decentralized choices

The differential choice of $\widetilde{x}$ and q implies that the decentralized economy, in general, need not replicate the socially optimal allocations for any arbitrary η. This in turn implies that there are welfare losses from inefficient choices of $\widetilde{x}$ and q. Figure 8b plots the level of steady state consumption in the decentralized economy for different values of $\eta \in [0,1]$.[44] The black $*$ which denotes consumption in our benchmark decentralized economy lies strictly below the red x which denotes the consumption under the planner's solution. For the case of our benchmark economy, Table 5 shows that the welfare loss is between 3 to 4% depending on the time period.

Table 5: Efficiency losses

	Pre-2005			Post-2005		
Variable	Benchmark	Planner	% Loss	Benchmark	Planner	% Loss
η	0.100	0.292	-	0.100	0.275	-
$\widetilde{x}$	0.697	0.730	-	0.639	0.671	-
q	3.329	5.761	-	1.580	2.624	-
consumption	0.735	0.763	0.039	0.717	0.740	0.032
mean wage	0.722	0.800	0.107	0.670	0.750	0.118
y/n	1.04	1.06	0.017	1.05	1.07	0.017
labor share	0.692	0.753	0.088	0.639	0.702	0.100

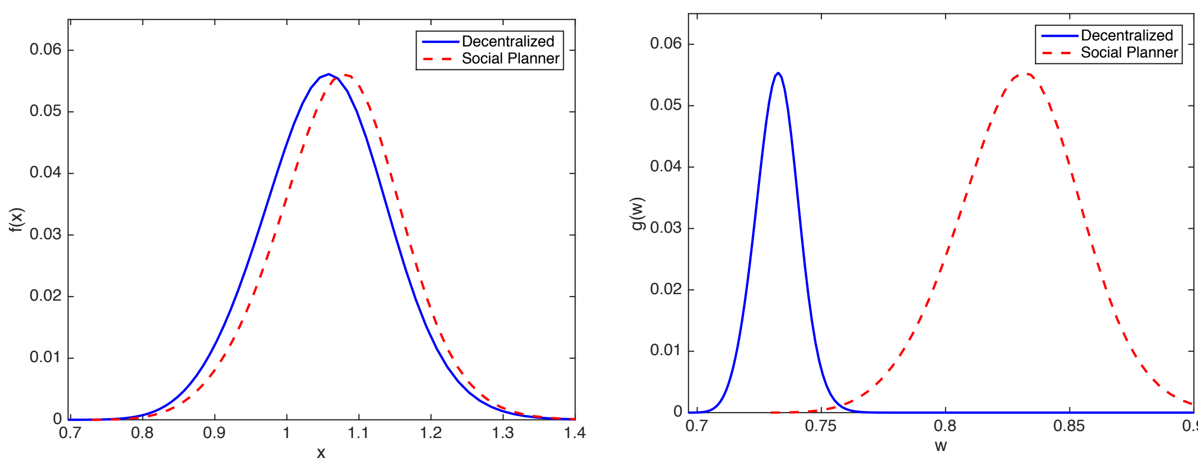
Notes: Columns (2)-(3) characterize the outcomes under the benchmark decentralized economy and under the bargaining weight $\eta = \eta_{SP}$ that implements the planner's solution for the pre-2005 period while column (4) calculates the losses in outcomes between these two economies. Columns (5)-(7) replicate the same exercise but for the post 2005 period.

Table 5 also shows how much larger the productivity-wage gap is in our benchmark economy against

[44] We truncate the graph here at $\eta < 0.55$ for exposition purposes. In general, the consumption associated with different levels of η exhibits a hump-shaped behavior.

a decentralized economy which replicates the planner's choices. To compute wages consistent with the planner's solution, we compute the bargaining weight, η_{SP}, that implements the planner's outcomes. The first row of Table 5 shows that η_{SP} is larger than the η in our benchmark economy. We then compute the wages and productivity for the decentralized economy with bargaining weight η_{SP}. While productivity is only about 2% higher under the planner's solution, wages are about 11% lower in our benchmark economy. These lower wages stem from the lack of selectivity in match quality.

Figure 9a highlights that relative to our benchmark economy (blue solid line) the distribution of firms over match quality is further shifted to the right for the economy that replicates the planner's solution (red dashed line), demonstrating that the planner, on average, chooses to create more productive matches than the decentralized benchmark economy. The higher selectivity along with the higher bargaining power of workers η_{SP} which decentralizes the planner's solution imply much larger differences in wages. Figure 9b shows that the wage distribution under η_{SP} first order stochastically dominates the wage distribution in the benchmark economy. Overall, the productivity-wage gap in the efficient economy is smaller by close to 10%, giving rise to a labor share that larger by about 9 to 10% than that observed in our benchmark economy.



(a) Distribution of accepted match quality

(b) Distribution of accepted wages

Figure 9: Efficient vs. benchmark distributions of match quality and wages

6 Discussion

Worker on-the-job search While our paper has focused on firm on-the-job search and abstracted from worker on-the-job search, it should be noted that 1) not all replacement hiring is accounted for by quits and 2) employment-to-employment (EE) transitions have been declining as a fraction of total hires, whereas the replacement hiring share has been increasing. As such, we have focused on firm on-the-job search rather than worker on-the-job search.[45] Further, even if one were to argue that increased replacement hiring shares were due to increased worker on-the-job search, standard models featuring the latter (see Burdett and Mortensen (1998), Menzio and Shi (2011) for example) typically predict

[45]Using CPS data, Figure 10c in Appendix G.1 demonstrates that the EE share of hires has declined since the early 1990s.

that wages and productivity move in the same direction since workers switch jobs in these models for the purposes of attaining higher productivity matches that offer higher wages.

Vacancy duration and the relevant measure of market tightness The standard DMP model assumes that vacancies expire instantaneously, implying that only the current flow of vacancies are relevant for computing market tightness. In contrast, we argue that our assumption of long-lived vacancies better accords with how the data on job openings is collected. In particular, the Bureau of Labor Statistics (BLS) which conducts the monthly JOLTS states that the information it collects on job openings are "a stock,or point-in-time, measurement for the last business day of each month".[46] Further, our measure of vacancies is not inconsistent with the JOLTS definition of a job-opening. Specifically, JOLTS requires a job opening to satisfy three criteria: 1) a position exists, 2) work can start in 30 days, and 3) the firm is actively recruiting where active recruiting implies that the firm has undertaken "steps to fill a position". Recruiting matched firms in our model satisfy these three conditions. Our model calls to attention that the proportion of unfilled vacancies in addition to labor market tightness are important for rationalizing job-finding rates.

Additional hiring costs Our paper also assumes a fixed cost of posting vacancies and abstracts from other types of hiring costs. Inclusion of a hiring cost paid at the time of matching with a worker would not qualitatively change the predictions of our model. Its inclusion would only serve to raise hiring thresholds. So long as the replacement hiring share increases, this will still drive a larger productivity-wage gap. Moreover, the one time fixed vacancy posting cost χ, rather than the standard flow cost of a vacancy, can be interpreted as the cost of maintaining a human resources department for the duration of the vacancy posted and therefore subsumes the additional costs that could have been incurred at the time of matching.

7 Conclusion

We document that the replacement hiring share in the US has risen over time alongside a widening productivity wage-gap. We find that an increases in the value of an unfilled vacancy and decreases in the vacancy expiration rate are key factors raising the replacement hiring share and consequently, the widening gap between labor productivity and wages. Declines in the quit rate actually temper the increase in the replacement hiring share. We show that with a larger replacement hiring share, the higher market power of firms, larger job insecurity faced by workers, and lower measured matching efficiency in the labor market allow firms to reap productivity gains from replacing workers while still keeping wages low. Compared to the efficient benchmark, private agents in the decentralized economy do not internalize the rat race externality and tend to create too many low productivity jobs. This in turn raises the incidence of replacement hiring and widens the gap between productivity and wages.

[46]See the JOLTS chapter of the "BLS Handbook of Methods". https://www.bls.gov/opub/hom/pdf/homch18.pdf

References

Albrecht, James, Bruno Decreuse, and Susan Vroman, "Directed Search with Phantom Vacancies," AMSE Working Papers 1707, Aix-Marseille School of Economics, Marseille, France March 2017.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," NBER Working Papers 23396, National Bureau of Economic Research, Inc May 2017.

Azar, Jose, Ioana Marinescu, and Marshall I. Steinbaum, "Labor Market Concentration," NBER Working Papers 24147, National Bureau of Economic Research, Inc December 2017.

Barnichon, Regis, "Building a composite Help-Wanted Index," *Economics Letters*, December 2010, *109* (3), 175–178.

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

Cheron, Arnaud and Bruno Decreuse, "Matching with Phantoms," *Review of Economic Studies*, 2017, *84* (3), 1041–1070.

Davis, Steven J. and Brenda Samaniego de la Parra, "Application Flows," Technical Report 2017.

Elsby, Michael, Bart Hobijn, and Aysegul Sahin, "The Decline of the U.S. Labor Share," *Brookings Papers on Economic Activity*, 2013, *44* (2 (Fall)), 1–63.

Elsby, Michael W. L., Ryan Michaels, and David Ratner, "Vacancy Chains," Technical Report 2016.

Fujita, Shigeru and Garey Ramey, "Job matching and propagation," *Journal of Economic Dynamics and Control*, November 2007, *31* (11), 3671–3698.

Haefke, Christian and Michael Reiter, "Long Live the Vacancy," Technical Report 2017.

Hall, Robert E., "High Discounts and High Unemployment," *American Economic Review*, February 2017, *107* (2), 305–30.

Karabarbounis, Loukas and Brent Neiman, "The Global Decline of the Labor Share*," *The Quarterly Journal of Economics*, 2013, *129* (1), 61.

Menzio, Guido and Espen R. Moen, "Worker replacement," *Journal of Monetary Economics*, September 2010, *57* (6), 623–636.

_ **and Shouyong Shi**, "Efficient Search on the Job and the Business Cycle," *Journal of Political Economy*, 2011, *119* (3), 468–510.

Shimer, Robert, "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, March 2005, *95* (1), 25–49.

_ , "Reassessing the Ins and Outs of Unemployment," *Review of Economic Dynamics*, April 2012, *15* (2), 127–148.

Appendix

A Re-listing

A.1 Re-listing short-lived vacancies

Consider the standard random search model where vacancies expire instantly. Consider the value of a matched firm with worker of match quality x, $J(x)$. The firm makes current profit $x - w(x)$, is exogenously separated from his worker at rate s, and receive an opportunity to re-list his vacancy and look for a better worker at rate ψ. If the firm chooses to relist, the firm pays a flow vacancy posting κ and can meet with a new applicant in the current period. Observe that the value of $J(x)$ is given by:

$$\rho J(x) = x - w(x) - sJ(x) \tag{41}$$

$$+\psi \max_{r\in\{0,1\}} \left\{ r \left[-\kappa + q \int_{x}^{\overline{x}} [J(y) - J(x)]\, d\Pi(y) \right] + (1-r) \times 0 \right\} \tag{42}$$

If the firm chooses to relist, $r = 1$, he gets the net expected benefit from re-listing, $-\kappa + q \int_x^{\overline{x}} [J(y) - J(x)]\, d\Pi(y)$. Observe that for new unattached firms, free entry implies that the flow cost of vacancy must equal to the expected benefit of a vacancy.

$$q \int_{\underline{x}}^{\overline{x}} J(y)\, d\Pi(y) = \kappa$$

which we can rewrite as:

$$-\kappa + q \int_{x}^{\overline{x}} [J(y) - J(x)]\, d\Pi(y) = -q \int_{\underline{x}}^{x} J(y)\, d\Pi(y) - q[1 - \Pi(x)]\, J(x) < 0$$

Notice that the LHS of the above equation is exactly the matched firm's gain in value from re-listing in equation (42). Since this term is always negative, the matched firm optimally never chooses to re-list. In an environment with short-lived vacancies, matched firms optimally choose to never post vacancies to replace current workers with better applicants.

A.2 Re-listing long-lived vacancies

Suppose the non-recruiting matched firm receives an option to re-list at a rate $\psi > 0$. If the firm chooses to avail this opportunity, the firm has to repay the cost χ and transitions back into a recruiting matched firm. The value of such a firm is given by:

$$\rho J(x,0) \quad = \quad x - w(x,0) + s\left[J(0) - J(x,0)\right] + \psi \max_{r\in\{0,1\}} \left[r\left\{J(x,1)-\chi\right\} + (1-r)\,J(x,0) - J(x,0)\right]$$

For any match quality x, we have:

$$J(x,1) - \chi = (1-\eta)S(x,1) \leq (1-\eta)S(x,0) = J(x,0)$$

where the inequality is strict for $x > \widetilde{x}$. The LHS is the gain from relisting net of the cost while the LHS of the inequality is the value of not relisting. The above condition shows that the value of a firm from relisting is weakly lower than not listing. As such, no non-recruiting firm would choose to relist in equilibrium.

B Surplus and Wage Determination

B.1 Nash Bargaining

Wages are determined by Nash Bargaining and wages can be expressed as a solution to the problem:

$$\max_{w(x,i)} \Big[J(x,i) - J(i)\Big]^{1-\eta} \Big[W(x,i) - U\Big]^{\eta} \text{ for } i = 0,1 \tag{43}$$

The Nash-Bargaining solution features:

$$(1-\eta)[W(x,i) - U\Big] = \eta\Big[J(x,i) - J(i)\Big] \tag{44}$$

Then using the definition of surplus $S(x,i) = J(x,i) - J(i) + W(x,i) - U$ for $i \in \{0,1\}$, under the Nash Bargaining solution, the worker gets a fraction η of the surplus while the firm receives the rest:

$$W(x,i) - U = \eta S(x,i) \qquad \text{and } J(x,i) - J(i) = (1-\eta)S(x,i) \tag{45}$$

B.2 Surplus

Non-recruiting firm-worker pair Free entry $J(0) = 0$ implies that (5) can be written as:

$$(\rho + s)J(x,0) = x - w(x,0) \tag{46}$$

Subtracting ρU from both sides of (3), we can re-write that equation as:

$$(\rho + s)\left[W(x,0) - U\right] = w(x,0) - \rho U \tag{47}$$

Then using the definition of surplus $S(x,0) = J(x,0) - J(0) + W(x,0) - U$, the surplus of a non-recruiting matched firm-worker pair can yields the expression in (10) in the main text:

$$(\rho + s)S(x,0) = x - \rho U \tag{48}$$

Recruiting firm-worker pair Subtract (8) from (7) and rearrange to get:

$$\varrho(x)\left[J(x,1)-J(1)\right] \;=\; x-w(x,1)-q\int_{\widetilde{x}}^{x}\left[J\left(y,1\right)-J(1)\right]d\Pi(y)+\delta\left[J(x,0)-J(0)\right] \tag{49}$$

where $\varrho(x)=\rho+\delta+s+q\left[1-\Pi(x)\right]$. Next, subtracting ρU from both sides of (4) yields:

$$\varrho(x)\left[W(x,1)-U\right]=w(x,1)+\delta\left[W(x,0)-U\right]-\rho U \tag{50}$$

Adding the two expressions above yields an expression identical to (11) in the main text:

$$\varrho(x)S(x,1)=x-\rho U-q\int_{\widetilde{x}}^{x}\left[J\left(y,1\right)-J(1)\right]d\Pi(y)+\delta S(x,0) \tag{51}$$

where $S(x,1)=J(x,1)-J(1)+W(x,1)-U$. Next using the Nash Bargaining solution (45), and the expression for $S(x,0)$ from (48) the equation above can be re-written as:

$$\varrho(x)S(x,1)=x-\rho U-q(1-\eta)\int_{\widetilde{x}}^{x}S\left(y,1\right)d\Pi(y)+\frac{\delta}{\rho+s}(x-\rho U) \tag{52}$$

Evaluating (52) at $x=\widetilde{x}$ yields the property that $\widetilde{x}=\rho U$ which further implies that

$$(\rho+s)S(x,0)=x-\widetilde{x} \tag{53}$$

which is the same as equation (27) in the text. Notice that just like a match with an active vacancy, the surplus of a match with an inactive vacancy also goes to zero at $\widetilde{x}$. Next, take the derivative of (52) with respect to x:

$$-\frac{\eta q\Pi'(x)}{\varrho(x)}S\left(x,1\right)+S'(x,1) \;=\; \left[\frac{\delta+\rho+s}{\rho+s}\right]\frac{1}{\varrho(x)} \tag{54}$$

which is an ordinary differential equation in x. Using the boundary condition $S(\widetilde{x},1)=0$, we can derive the expression for $S(x,0)$ in (28) in the main text:

$$S\left(x,1\right) \;=\; \left[\frac{\delta+\rho+s}{\rho+s}\right]\int_{\widetilde{x}}^{x}\frac{1}{\varrho(x)^{\eta}\varrho(y)^{1-\eta}}dy \tag{55}$$

B.3 Wages

Wages paid by a firm with an inactive vacancy and match productivity x To derive expressions for wages, we start by using the fact that $J(x,0)=(1-\eta)S(x,0)$ in (53) and rearranging:

$$(\rho+s)J(x,0)=(\rho+s)(1-\eta)S(x,0)=(1-\eta)(x-\widetilde{x}) \tag{56}$$

Using (46), we can rewrite the above equation as:

$$(\rho+s)J(x,0)=(1-\eta)(x-\widetilde{x})=x-w(x,0) \tag{57}$$

Rearranging yields: $w(x,0) = \widetilde{x} + \eta(x - \widetilde{x})$ for $x \geq \widetilde{x}$ which is identical to (13) in the main text.

Wages paid by a firm with an active vacancy and match productivity x In a similar fashion as above, using $J(x,1) - \chi = (1-\eta)S(x,1)$ in (52) and rearranging:

$$\varrho(x)[J(x,1) - \chi] = (1-\eta)(x - \widetilde{x}) - q(1-\eta)\int_{\widetilde{x}}^{x} [J(y,1) - \chi] d\Pi(y) + \delta J(x,0)$$

Using (49) with the equation above and rearranging, we can get the expression for wages in (14), Proposition 1 in the main text:

$$w(x,1) = \widetilde{x} + \eta(x - \widetilde{x}) - \eta(1-\eta)q\int_{\widetilde{x}}^{x} [J(y,1) - \chi] d\Pi(y) = w(x,0) - \eta(1-\eta)q\int_{\widetilde{x}}^{x} S(y,1)\, d\Pi(y) \quad (58)$$

C Labor Market Flows

Unemployed Out of steady state, the law of motion for the unemployed is given by:

$$\dot{u} = -qv^u [1 - \Pi(\widetilde{x})] + s(1-u) \quad (59)$$

Similar to (15), the first term on the RHS of the equation above represents the outflows from unemployment while the second term represents inflows. We use the fact that $p_t(v_t^u/v_t)u_t = q_t\theta_t(v_t^u/v_t)u_t = q_t v_t^u$ and replace the worker's rate of meeting unfilled vacancies with the unfilled vacancy's rate of meeting a job-seeker.

Unfilled Vacancies The law of motion for unfilled vacancies is given by:

$$\dot{v}^u \quad = \quad [-q[1 - \Pi(\widetilde{x})] - \delta] v^u + v^{new} + sv^m \quad (60)$$

As per equation (20), the first term on the RHS represents the outflows from unfilled vacancies while the second and third terms on the RHS of the above equation represents the inflows from new vacancies created and from recruiting matched firms being exogenously separated from their workers.

Recruiting Matched Firms The mass of recruiting matched firms with match quality less than or equal to x evolves according to:

$$\dot{F}(x)\, v^m + F(x)\, \dot{v}^m = -\left(q[1 - \Pi(x)] + s + \delta\right) F(x)\, v^m + q[\Pi(x) - \Pi(\tilde{x})]\, v^u \quad (61)$$

The first term on the RHS of the equation represents the outflows. Outflows occur whenever 1) recruiting matched firms meet a new applicant whose match quality y is higher than x, 2) recruiting matched firms become non-recruiting firms at vacancy expiration rate δ, and 3) recruiting matched firms become unfilled vacancies at exogenous separation rate s. Inflows into this group stem from the mass of unfilled vacancies that meet an unemployed applicant and draw match quality between x and $\widetilde{x}$. Alternatively, this relationship can be described in terms of the evolution of the measure of the recruiting matched firms with match quality x:

$$v^m \dot{f}(x) + \dot{v}^m f(x) = \left[q\pi(x) \left(v^u + v^m \int_{\widetilde{x}}^{x} f(\epsilon)\, d\epsilon \right) - \Big(s + \delta + q\left[1 - \Pi(x)\right] \Big) v^m f(x) \right] \forall x \geq \widetilde{x}$$

where f denotes the density. Evaluating (61) at $\overline{x}$, we get the law of motion for the total mass of recruiting matched firms:

$$\dot{v}^m = -(s+\delta) v^m + q\left[1 - \Pi(\tilde{x})\right] v^u \tag{62}$$

Non-recruiting matched firms The mass of non-recruiting matched firms evolves according to:

$$\dot{v}^i = -s v^i + \delta v^m \tag{63}$$

Outflows from this group stem from exogenous separations, while inflows into this group stem from the expiration of vacancies at recruiting matched firms.

D Determination of $\tilde{x}$ and θ

Free entry Using the expression for $S(x,1)$ in (25), one gets:

$$(\rho+\delta)\chi = (1-\eta)\left(\frac{\delta+\rho+s}{\rho+s}\right) \int_{\tilde{x}}^{\overline{x}} \int_{\tilde{x}}^{x} \frac{q}{\varrho(x)^{\eta} \varrho(y)^{1-\eta}} \pi(x)\, dy dx \tag{64}$$

Using the implicit function theorem:

$$\frac{d\theta}{d\tilde{x}} = \frac{\int_{\tilde{x}}^{\overline{x}} \frac{q\pi(x)}{\varrho(x)^{\eta}\varrho(\tilde{x})^{1-\eta}} dx}{\frac{dq}{d\theta} \int_{\tilde{x}}^{\overline{x}} \int_{\tilde{x}}^{x} \left[1 - \left(\frac{\eta[1-\Pi(x)]}{\varrho(x)} + \frac{(1-\eta)[1-\Pi(y)]}{\varrho(y)} \right) \right] \frac{\pi(x)}{\varrho(x)^{\eta}\varrho(y)^{1-\eta}} dy dx} \tag{65}$$

where $dq/d\theta < 0$. Thus, (25) describes a negative relationship between $\tilde{x}$ and θ. Since we have one strictly decreasing and one strictly increasing curve, they can only intersect once, implying that there is a unique equilibrium ($\tilde{x}$ and θ).

Indifference condition Using the definition of ρU, (26) can be written as:

$$\widetilde{x} = b + \eta \left[\underbrace{\left(1 - \frac{v^m}{v}\right) \int_{\widetilde{x}}^{\overline{x}} pS(y,1)\pi(y)\, dy}_{\texttt{term1}} + \underbrace{\left(\frac{v^m}{v}\right) \int_{\widetilde{x}}^{\overline{x}} \int_{x}^{\overline{x}} pS(y,1)\, d\Pi(y)\, dF(x)}_{\texttt{term2}} \right] \tag{66}$$

Start by considering the term labelled `term1` on the RHS. This term can be rewritten as:

$$\texttt{term1} = \left(\frac{\delta+\rho+s}{\rho+s}\right) \left(\frac{s+\delta}{s+\delta+q\left[1-\Pi(\tilde{x})\right]}\right) \int_{\tilde{x}}^{\overline{x}} \int_{\tilde{x}}^{y} \frac{p}{\varrho(y)^{\eta} \varrho(\varepsilon)^{1-\eta}} \pi(y)\, d\varepsilon dy$$

The integrand $\frac{p}{\varrho(y)^{\eta}\varrho(\varepsilon)^{1-\eta}}$ is decreasing in q and thus `term1` is also decreasing in q (increasing in θ):

$$\frac{\partial}{\partial q}\left[\frac{p}{\varrho\left(y\right)^{\eta}\varrho\left(\varepsilon\right)^{1-\eta}}\right]=\frac{1}{\varrho\left(y\right)^{\eta}\varrho\left(\varepsilon\right)^{1-\eta}}\left[\frac{dp}{dq}-\eta\frac{p[1-\Pi(x)]}{\varrho(x)}-(1-\eta)\frac{p[1-\Pi(y)]}{\varrho(y)}\right]<0$$

Next, `term1` is decreasing in $\tilde{x}$ since

$$\frac{\partial}{\partial\tilde{x}}\left[\int_{\tilde{x}}^{\overline{x}}\int_{\tilde{x}}^{y}\frac{p}{\varrho\left(y\right)^{\eta}\varrho\left(\varepsilon\right)^{1-\eta}}\pi\left(y\right)d\varepsilon dy\right]=-\int_{\tilde{x}}^{\overline{x}}\frac{p}{\varrho\left(y\right)^{\eta}\varrho\left(\tilde{x}\right)^{1-\eta}}\pi\left(y\right)<0$$

Now, consider `term2` on the RHS of (66). This can be rewritten as:

$$\texttt{term2}=\frac{\left(\delta+\rho+s\right)\left(s+\delta\right)}{\rho+s}\int_{\tilde{x}}^{\overline{x}}\int_{x}^{\overline{x}}\int_{\tilde{x}}^{y}\underbrace{\frac{qp\Big(\rho+\delta+s+q\left[1-\Pi\left(\varepsilon\right)\right]\Big)^{\eta-1}}{\Big(\rho+\delta+s+q\left[1-\Pi\left(y\right)\right]\Big)^{\eta}}\frac{d\varepsilon}{\left(q\left[1-\Pi\left(x\right)\right]+s+\delta\right)^{2}}}_{\Xi(q,\tilde{x})}d\Pi(y)d\Pi(x)$$

Again the integrand of this expression is decreasing in q (increasing in θ):

$$\begin{aligned}\frac{q}{\Xi}\frac{\partial\Xi}{\partial q} &= -\frac{\alpha}{(1-\alpha)}-(1-\eta)\frac{q\left[1-\Pi\left(\varepsilon\right)\right]}{\rho+\delta+s+q\left[1-\Pi\left(\varepsilon\right)\right]}-\eta\left[\frac{q\left[1-\Pi\left(y\right)\right]}{\rho+\delta+s+q\left[1-\Pi\left(y\right)\right]}\right]\\ &\quad -2\left[\frac{q\left[1-\Pi\left(x\right)\right]}{q\left[1-\Pi\left(x\right)\right]+s+\delta+\rho}\right]\left[\frac{q\left[1-\Pi\left(x\right)\right]+s+\delta+\rho}{q\left[1-\Pi\left(x\right)\right]+s+\delta}\right]<0\end{aligned}$$

and similarly, it is straightforward to show that `term2` is non-increasing in $\tilde{x}$. Finally, the RHS of (66) is increasing in $\tilde{x}$ and flat in q (θ). Thus, (66) describes a increasing relation between θ and $\tilde{x}$.

E Without On-the-Job Search

We re-characterize the main equations that change when there is no on-the-job search. Because the values of non-recruiting firm-worker pairs or workers hired at firms that have expired vacancies stays the same, we only highlight how the equations change for recruiting firm-worker pairs.

Unemployed The value of unemployment becomes:

$$\rho U\;=\;b+p\int_{\tilde{x}}^{\overline{x}}\left[W\left(y,1\right)-U\right]d\Pi\left(y\right)$$

Worker at Recruiting Matched Firm The value of working at recruiting matched firm is:

$$\rho W\left(x,1\right)=w\left(x,1\right)-s\left[W\left(x,1\right)-U\right]-\delta\left[W\left(x,1\right)-W\left(x,0\right)\right]$$

Unfilled, Unexpired Vacancy The value of unfilled vacancy becomes:

$$\rho J\left(1\right)=-\delta\left[J\left(1\right)-J\left(0\right)\right]+q\int_{\tilde{x}}^{\overline{x}}\left[J\left(y,1\right)-J\left(1\right)\right]d\Pi\left(y\right)$$

Recruiting Matched Firm The value of a recruiting matched firm becomes:

$$\rho J(x,1) \quad = \quad x - w(x,1) - \delta [J(x,1) - J(x,0)] - s[J(x,1) - J(1)]$$

Surplus of Recruiting Matched Firm-Worker Pair Since $S(x,1) = J(x,1) - J(1) + W(x,1) - U$, we have:

$$(\rho + s + \delta) S(x,1) \quad = \quad x + \delta S(x,0) - \rho U - q \int_{\widetilde{x}}^{\overline{x}} [J(y,1) - J(1)]\, d\Pi(y)$$

Laws of motion: recruiting matched firms Only the law of motion describing the distribution of recruiting matched firms changes when on-the-job search is shut off. As such, we highlight how the distribution of firms changes.

$$\dot{F}(x) v^m + F(x) \dot{v^m} = -(s+\delta) F(x) v^m + q v^u [\Pi(x) - \Pi(\widetilde{x})]$$

In steady state, this implies:

$$F(x) = \frac{[\Pi(x) - \Pi(\widetilde{x})]}{[1 - \Pi(\widetilde{x})]}$$

F Efficient Benchmark

Rather than solving the planner's problem as choosing $\widetilde{x}$ and v^{new}, it is easier to solve an isomorphic problem in which instead of $\widetilde{x}$, the planner chooses an *acceptance function* $a(x)$ for each level of match productivity x. This acceptance function $a(x) \in [0,1]$ for all $x \in [\underline{x}, \overline{x}]$. By definition, $\widetilde{x}$ is the value of match productivity above which $a(x) = 1$. It is also convenient to define $\mu^m(x) = v^m dF(x)$ as the measure of matched recruiting firms with match quality x. Correspondingly, define $\mu^i(x) = v^i dF(x)$ as the measure of non-recruiting firms with match quality x. This reformulation means that the constraints given by equations (33) and (35) have to be altered to:

$$\dot{\mu}_t^m(x) = \left[a_t(x) q_t \pi(x) \left(v_t^u + \int_{\underline{x}}^{x} \mu_t^m(\epsilon)\, d\epsilon \right) - \left(s + \delta + q_t \int_{x}^{\overline{x}} a_t(\epsilon) \pi(\epsilon)\, d\epsilon \right) \mu_t^m(x) \right] \tag{67}$$

and

$$\dot{v}_t^u = v_t^{new} + s \int_{\underline{x}}^{\overline{x}} \mu_t^m(\epsilon)\, d\epsilon - \left(q_t \int_{\underline{x}}^{\overline{x}} a(\epsilon) \pi(\epsilon)\, d\epsilon + \delta \right) v_t^u \tag{68}$$

respectively. Other constraints are unchanged. The planner's problem can be written as a Hamiltonian:

$$
\begin{aligned}
\mathcal{H}_t &= -\chi v^{new} + \int_{\underline{x}}^{\overline{x}} (x-b)\,\mu^m(x)\,dx + \int_{\underline{x}}^{\overline{x}} (x-b)\,\mu^i(x)\,dx + b \\
&\quad + \int_{\underline{x}}^{\overline{x}} \Lambda(x,1)\left[a(x)\,q\pi(x)\left(v^u + \int_{\underline{x}}^{x}\mu^m(\epsilon)\,d\epsilon\right) - \left(s+\delta+q\int_{x}^{\overline{x}} a(\epsilon)\,\pi(\epsilon)\,d\epsilon\right)\mu^m(x)\right]dx \\
&\quad + \int_{\underline{x}}^{\overline{x}} \Lambda(x,0)\left[\delta\mu^m(x) - s\mu^i(x)\right]dx \\
&\quad + \varphi\left[v^{new} + s\int_{\underline{x}}^{\overline{x}}\mu^m(\epsilon)\,d\epsilon - \left(q\int_{\underline{x}}^{\overline{x}} a(x)\,\pi(x)\,dx + \delta\right)v^u\right] \\
&\quad - \psi\left[\ln q - \ln\xi + \alpha\ln\left(v^u + \int_{\underline{x}}^{\overline{x}}\mu^m(\epsilon)\,d\epsilon\right) - \alpha\ln\left(1 - \int_{\underline{x}}^{\overline{x}}\mu^m(\epsilon)\,d\epsilon - \int_{\underline{x}}^{\overline{x}}\mu^i(\epsilon)\,d\epsilon\right)\right]
\end{aligned}
$$

The choice variables are v^{new}, q and $a(x)$. State variables are: mass of (i) matched recruiting and (ii) non-recruiting firm-worker pairs with match quality x denoted by $\mu^m(x)$ and $\mu^i(x)$ resp., and (iii) stock of unfilled vacancies. $\Lambda(x,1)$ is the co-state associated with $\mu^m(x)$ and denotes the gross social value of a matched recruiting firm-worker pair. Similarly, $\Lambda(x,0)$ is the co-state associated with $\mu^i(x)$ and denotes the gross social value of a matched non-recruiting firm-worker pair with match quality x. φ_t is the co-state associated with v^u and denotes the benefit of an additional unfilled vacancy. ψ is the multiplier on the meeting function. The planner's choices in steady-state are described by:

$$
\chi = \varphi \tag{69}
$$

$$
\frac{\psi}{q} = \int_{\underline{x}}^{\overline{x}} \Lambda(x,1)\left[a(x)\,\pi(x)\left(v^u + \int_{\underline{x}}^{x}\mu^m(\epsilon)\,d\epsilon\right) - \int_{x}^{\overline{x}} a(\epsilon)\,d\Pi(\epsilon)\,\mu^m(x)\right]dx - \varphi\int_{\underline{x}}^{\overline{x}} a(x)\,d\Pi(x)\,v^u \tag{70}
$$

$$
a(x) = \begin{cases} 0 & \text{if } q\pi(x)\left[\Lambda(x,1)\,v^u + \int_{\underline{x}}^{x}\left[\Lambda(x,1)-\Lambda(\epsilon,1)\right]\mu^m(\epsilon)\,d\epsilon - \varphi v^u\right] < 0 \\ 1 & \text{if } q\pi(x)\left[\Lambda(x,1)\,v^u + \int_{\underline{x}}^{x}\left[\Lambda(x,1)-\Lambda(\epsilon,1)\right]\mu^m(\epsilon)\,d\epsilon - \varphi v^u\right] \geq 0 \end{cases} \tag{71}
$$

$$
\begin{aligned}
\rho\Lambda(x,1) &= x - b + q\int_{x}^{\overline{x}}\Lambda(\epsilon,1)\,a(\epsilon)\,d\Pi(\epsilon) - \Lambda(x,1)\left(s+\delta+q\int_{x}^{\overline{x}} a(\epsilon)\,d\Pi(\epsilon)\right) + \delta\Lambda(x,0) + s\varphi \\
&\quad - \alpha\left[\frac{\psi}{v} + \frac{\psi}{u}\right]
\end{aligned} \tag{72}
$$

$$
\rho\Lambda(x,0) = x - b - s\Lambda(x,0) - \alpha\frac{\psi}{u} \tag{73}
$$

$$
\rho\varphi = \int_{\underline{x}}^{\overline{x}}\Lambda(x,1)\,a(x)\,q\pi(x)\,dx - \varphi\left[q\int_{\underline{x}}^{\overline{x}} a(x)\,d\Pi(x) + \delta\right] - \alpha\frac{\psi}{v} \tag{74}
$$

where (69)-(71) are the FOC's w.r.t. v^{new}, q and $a(x)$ respectively. (72)-(74) describe the evolution of co-state variables $\Lambda(x,1), \Lambda(x,0)$ and φ respectively. Using (71), $\widetilde{x}$ is implicitly defined as:

$$
\Lambda(\widetilde{x},1) = \chi = \varphi \tag{75}
$$

In other words, the planner doesn't want a match which is worth less than a unfilled vacancy to exist. Imposing the optimal $a(x)$, (70), (77) and (74) can be rewritten as:

$$\frac{\psi}{u} = q\theta\left[\left(\frac{v^u}{v}\right)\int_{\tilde{x}}^{\overline{x}} \mathbb{S}(x,1)d\Pi\left(x\right) + \frac{v^m}{v}\int_{\tilde{x}}^{\overline{x}}\int_{x}^{\overline{x}} \left[\mathbb{S}\left(\epsilon,1\right) - \mathbb{S}\left(x,1\right)\right] d\Pi\left(\epsilon\right)\frac{\mu^m\left(x\right)}{v^m}dx\right] \tag{76}$$

$$\varrho(x)\Lambda\left(x,1\right) = x - b - \frac{\alpha\psi}{u} + \delta\Lambda\left(x,0\right) + s\chi + q\int_{x}^{\overline{x}} \Lambda\left(\epsilon,1\right)d\Pi\left(\epsilon\right) - \frac{\alpha\psi}{v} \tag{77}$$

$$\left(\rho + \delta + q\left[1 - \Pi\left(\tilde{x}\right)\right]\right)\chi = q\int_{\tilde{x}}^{\overline{x}} \Lambda\left(x,1\right)d\Pi\left(x\right) - \alpha\frac{\psi}{v} \tag{78}$$

where $\mathbb{S}(x,1) = \Lambda(x,1) - \chi$ is the social value of a matched vacancy net of the value of an unfilled vacancy. Using (78) and (73) in (77), we can get:

$$\varrho(x)\mathbb{S}\left(x,1\right) \quad = \quad \left(\frac{\rho+s+\delta}{\rho+s}\right)\left[x - b - \frac{\alpha\psi}{u}\right] - q\int_{\tilde{x}}^{x} \mathbb{S}\left(y,1\right)d\Pi\left(y\right) \tag{79}$$

Differentiating this equation with respect to x yields:

$$\varrho(x)\mathbb{S}'\left(x,1\right) \quad = \quad \left(\frac{\rho+s+\delta}{\rho+s}\right) \tag{80}$$

Additionally, $\mathbb{S}(\widetilde{x},1) = \Lambda(\widetilde{x},1) - \chi = 0$. This boundary value problem has a unique solution:

$$\mathbb{S}\left(x,1\right) = \left(\frac{\rho+s+\delta}{\rho+s}\right)\int_{\tilde{x}}^{x} \frac{1}{\rho+s+\delta+q\left[1-\Pi\left(y\right)\right]}dy \tag{81}$$

which is identical to (40), Lemma 2 in the main text. In addition, evaluating (79) at $x = \widetilde{x}$, it must be the case that $\widetilde{x} = b + \frac{\alpha\psi}{u} = 0$. Using this fact, we can rewrite equation (73) as:

$$\mathbb{S}(x,0) = \left[\Lambda(x,0) - 0\right] = \frac{x - \widetilde{x}}{\rho+s} \tag{82}$$

which is identical to equation (39), Lemma 2 in the main text. Next, using the definition of $\mathbb{S}(x)$ and rearranging equation (78), we can derive the analog to the free entry condition:

$$(\rho+\delta)\chi \quad = \quad q\int_{\tilde{x}}^{\overline{x}} \mathbb{S}\left(x,1\right)d\Pi\left(x\right) - \frac{\alpha\psi}{v} \tag{83}$$

Since we know that $\widetilde{x} = b + \frac{\alpha\psi}{u}$, we can rewrite the expression above as:

$$(\rho+\delta)\chi \quad = \quad q\int_{\tilde{x}}^{\overline{x}} \mathbb{S}\left(x,1\right)d\Pi\left(x\right) - \frac{\widetilde{x} - b}{\theta} = q\int_{\tilde{x}}^{\overline{x}} \mathbb{S}\left(x,1\right)d\Pi\left(x\right) - \left(\widetilde{x} - b\right)\left(\frac{q}{\xi}\right)^{\frac{1}{\alpha}} \tag{84}$$

which is identical to equation (37), Lemma 2 in the main text (we have used the meeting function in going from the first equality to the second). Using the definition of $\mathbb{S}(x,1)$ in (76) and rearranging:

$$\tilde{x} = b + \alpha q\theta\left[\left(\frac{v^u}{v}\right)\int_{\tilde{x}}^{\overline{x}}\mathbb{S}\left(y,1\right)d\Pi\left(y\right) + \frac{v^m}{v}\int_{\tilde{x}}^{\overline{x}}\int_{x}^{\overline{x}}\left[\mathbb{S}\left(y,1\right)-\mathbb{S}\left(x,1\right)\right]d\Pi\left(y\right)\frac{\mu^m\left(x\right)}{v^m}dx\right] \tag{85}$$

which is identical to equation (38), Lemma 2 in the main text.

G Worker on-the-job Search

G.1 Employment-to-employment hires as a fraction of total hires

Using data from the CPS for the coverage period 1994-2017, we calculate the fraction of total hires that stem from employment-to-employment hires. The CPS re-design in 1994 includes a question which asks individuals if they are employed at the same job. We compute the level of employment-to-employment hires as the number of individuals who were employed in period $t-1$ and are still employed in period t but at a different job. We construct total hires as all the individuals who found a new job in period t, these include both those who were in non-employment in period $t-1$ but who are employed in period t, as well as all those who did employment-to-employment transitions. We sum-up all monthly data to a quarterly frequency. We calculate our quarterly measure of the employment-to-employment (EE) share of total hires as:

$$\text{EE share} = \frac{\text{EE hires}}{\text{Total hires}}$$

Figure 10c demonstrates that the EE share has been in decline since the 1990s and has only recently seen an uptick. This stands in stark contrast to the upward trend in the replacement hiring share recorded in Figure 1b, suggesting that employment-to-employment transitions are not behind the increase in the replacement hiring share. It should be noted that the EE share is not equivalent to the quits to hires ratio presented in Figure 2a as EE transitions are only a subset of all quits (excluded in EE transitions are quits into unemployment).

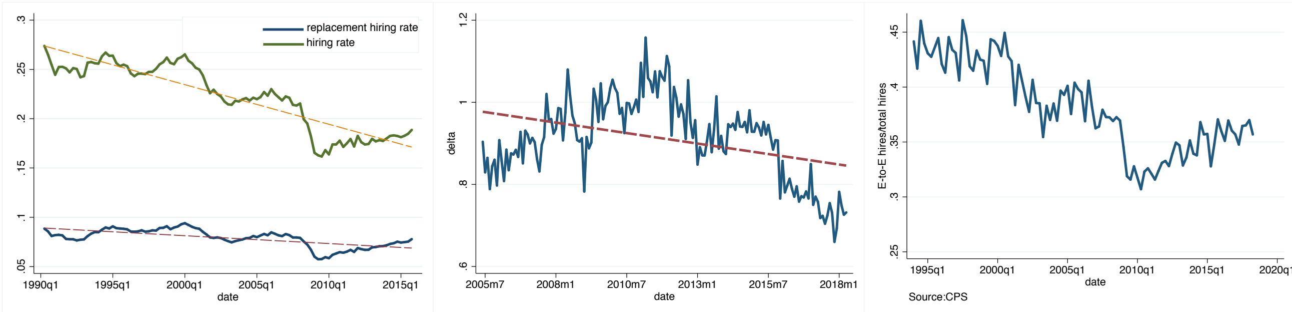
H Additional Data Work

H.1 Examining Replacement Hires

Figure 10a plots the replacement hiring rate against the hiring rate over time. The hiring rate is defined as the ratio of total hires to total employed while the replacement hiring rate is given by the ratio of replacement hires to total employed. Figure 10a highlights that higher replacement hiring share is a result of hires declining faster than replacement hires.

H.2 Replacement hiring within industry

To see if compositional changes in sectors is the primary factor behind the rise in the replacement hiring share, we conduct a shift-share analysis. We compute the increase due to changes in industry composition by holding fixed the share of replacement hires in each industry to its average level for the



(a) Replacement vs. total hiring rate (b) δ over time (c) EE hires/total hires

Figure 10: Additional graphs

period covering 1990-2016, and allowing only each industry's contribution to private GDP to change over time. We refer to this change in the replacement share as the "between" component. Separately, we allow the share of replacement hiring to change within each industry across the time period 1990-2016 and hold fixed the average weight of each industry across that time period. We label this second term the "within" component. Formally, the equation below denotes how we compute each component:

$$\Delta\text{Replacement Share} = \Delta \sum_{it} rr_{it} \frac{VA_{it}}{GDP_{it}^{private}} \approx \underbrace{\sum_{it} \hat{rr} \Delta \left(\frac{VA_{it}}{GDP_{it}^{private}} \right)}_{\text{between}} + \underbrace{\sum_{it} \Delta rr_{it} \left(\frac{\hat{VA}}{GDP^{private}} \right)}_{\text{within}}$$

where Δ refers to the change, rr_{it} denotes the replacement share in industry i in period t, and VA_{it} denotes the value added of industry i in period t. Terms with a "hat" denote averages. The top panel of Table 6 demonstrates that most of the change stems from an increase in the replacement share of hiring within industries rather due to changes in industry composition. The bottom panel of Table 6 shows that almost all industries experienced a rise in the share of replacement hiring, suggesting that the rise in the share of replacement hiring is ubiquitous and not limited to only one industry's experience.

Table 6: Decomposing the change

	Between	Within	Total
Δreplacement share	0.013	0.065	0.077
By Industry	1990	2016	% Change
National	0.313	0.435	39.23
Agriculture	0.084	0.204	141.3
Mining	0.191	0.372	94.55
Utilities	0.273	0.463	69.83
Construction	0.223	0.321	44.08
Manufacturing	0.332	0.468	41.12
Wholesale	0.292	0.433	48.44
Retail	0.383	0.525	37.03
Transportation	0.311	0.437	40.66
Information	0.403	0.371	-8.05
Finance and Insurance	0.402	0.480	19.39
Real Estate	0.317	0.423	33.52
Prof., Scientific, and Technical Svcs	0.328	0.405	23.59
Management	0.321	0.572	78.14
Administrative	0.318	0.393	23.64
Education	0.265	0.375	41.86
Healthcare	0.393	0.510	29.78
Arts, Entertainment, Recreation	0.191	0.378	97.63
Accommodation and Food	0.361	0.471	30.36
Other Services	0.319	0.357	11.89