Job Search Technology, Selection, and Hiring *

Sadhika Bagga
University of Rochester

Lukas F. Mann
Arizona State University

Version: April 16, 2026

Abstract

Two key trends have characterized US labor markets since the 1990s: Unemployment durations have nearly doubled, while matching efficiency has halved. Yet, contacting jobs is easier than ever before. By introducing screening on worker characteristics into a standard labor market model, we reconcile these trends. Technology improvements that facilitate contacts, such as the rise of online job search, lead firms to be more selective and to keep vacancies open for longer. The resulting outcomes are quantitatively consistent with aggregate and micro-level data. We evaluate the recent impact of artificial intelligence on matching and find that it has accelerated trends set in motion by online job search.

*Bagga: University of Rochester, sadhika.bagga@rochester.edu, www.sadhikabagga.github.io;
Mann: Arizona State University, lfmann@asu.edu, www.lukasmann.com.
We thank seminar participants at Arizona State University, the Federal Reserve Bank of San Francisco, the Federal Reserve Bank of Saint Louis, the University of Rochester, and Duke University. A conference draft of this paper was circulated with the title "A Theory of Congestion and Rising Unemployment Durations."

1 Introduction

The technology through which workers search for jobs and firms search for workers is a central feature of labor markets. Over the past three decades, search technology has advanced substantially. Before the mid-1990s, job search relied on newspaper classifieds, mailed or faxed resumes, and in-person interviews. With the rise of the internet, classifieds migrated to online job boards and electronic applications replaced paper and fax, enabling job seekers to apply to positions instantly. By the 2010s, over 70% of unemployed job seekers in the US reported using the internet to search for jobs. More recently, the emergence of artificial intelligence tools have further reduced the cost and effort of applying, allowing job seekers to generate tailored application materials in seconds. Accompanying these technological shifts, the volume of applications submitted per job seeker has risen sharply.

The literature on frictional labor markets has long recognized that such improvements in search technology affect labor market outcomes. This literature has focused on the idea that diminishing search frictions either improve the quality of matches or increase the rate at which they are formed. In this paper, we instead focus on two distinct aspects of improvements in search technology and argue that they are important to understand major labor market developments that have unfolded since the 1990s. First, such improvements allow firms to become more selective in which types of worker they hire, generating more unequal labor market outcomes. Second, they incentivize firms to keep vacancies open for longer, lowering the rate at which contacts convert into hires.

Our motivation stems from developments in the US labor market over the last three decades which, by several measures, has become less efficient at converting contacts between workers and firms into productive matches. We document a few observations that illustrate this point. First, the average duration of unemployment spells has nearly doubled since the 1990s, driven primarily by the prevalence of long-term unemployment. Second, job finding rates have declined. Third, job openings take longer to fill with the job filling rate decreasing by roughly half. Fourth, a simple measure of matching efficiency, the ratio of hires to an aggregate of job seekers and vacancies, has also halved over this time period. These trends have persisted despite historically tight labor markets since the late 2010s.

Our hypothesis is that these labor market developments are a consequence of the decline in the costs of applying for jobs, brought about by advances in search technology. Firms, faced with a larger inflow of applications, devote more resources to screening and become more selective in which workers they hire. As a result, fewer contacts between job seekers and firms convert to hires. This depresses both matching efficiency, the rate at which workers find jobs, and the rate at which jobs are filled. The burden of higher hiring standards is distributed unevenly: some workers continue to find jobs at roughly the same rate as before, while others are increasingly screened out and remain persistently unemployed, driving up the average duration of unemployment spells.

We document two pieces of evidence that support this mechanism. First, firms now spend more resources on screening candidates than they did in the 1990s, evidenced by the increasing employment and vacancy shares of human resource (HR) personnel involved in screening candidates. Second, cross-sectional data reveals that sectors where online job search is more prevalent experienced more pronounced declines in matching efficiency, job finding rates, and job filling rates, larger increases in unemployment duration, and faster growth in HR employment.

To formalize the mechanism, we formulate a minimalist extension of the canonical Diamond-Mortensen-Pissarides (DMP) framework: a model of random matching between unemployed workers and vacancies. We modify this model in three ways. First, we introduce a notion of applications by assuming that workers can, at a cost, apply to vacancies they encounter. Second, we introduce a notion of screening: workers are permanently heterogeneous and firms can pay a cost to receive a signal about the worker's type before making a hiring decision. Third, we assume a sunk vacancy creation cost, which introduces a vacancy-level trade-off between filling the vacancy today or waiting for a better worker tomorrow. Despite this added complexity, the model remains simple enough to be characterized graphically and features intuitive comparative statics that we take to the data.

We show that the decline in application costs brought about by online job search generates outcomes that are consistent with both aggregate and micro-level data. In the aggregate, the model quantitatively matches the long-run decline in matching efficiency, job finding and filling rates, and the concurrent increase in the average duration of unemployment. To show that the model mechanism is consistent with micro-level data, we compare three model predictions against micro data. First, we document that unemployment duration has been increasing unevenly across different groups of workers in panel data from the National Longitudinal Survey of Youth 1979 and 1997. Second, the model calls for firms to devote more resources to screening applicants, consistent with the rising employment and vacancy shares of human resource personnel that we document in the data. Finally, the model predicts that the composition of hires shifts toward workers with higher pre-separation wages and lower unemployment durations, a pattern we confirm in the Current Population Survey (CPS).

We use our framework to study the equilibrium effects of more recent advances in search technology brought about by the deployment of artificial intelligence in matching markets. The rise of large language models (LLMs) has affected labor market matching through two distinct channels: (i) a further reduction in the cost of applying, as job seekers use AI tools to generate tailored application materials at scale; (ii) a deterioration in the informativeness of the signals available to firms, as AI-generated applications make it harder to distinguish between candidates (Cowgill et al., 2024; Galdin and Silbert, 2025). Quantifying the distinct effects of the shocks with experimental evidence from Cowgill et al. (2024), we find that the deterioration of signal quality offsets roughly half of the decline in application costs, and that the net effect of the technology is to accelerate the developments set in motion by online job search.

Related Literature. Our paper contributes to three growing strands of literature: the literature on the effects of diminishing search frictions on labor market outcomes, the literature on the causes of long-term unemployment, and an emerging literature on the effects of artificial intelligence on labor market matching.

The idea that costly search is a key determinant of labor market frictions dates back to Stigler (1961, 1962), and the question of how improvements in search technology affect these frictions has received renewed attention in the context of the internet revolution. Bhuller et al. (2023) provide quasi-experimental evidence on how the staggered rollout of broadband infrastructure in Norway affected labor market outcomes. On the theoretical side, Martellini and Menzio (2020) study the long-run consequences of declining search frictions in a model where firms screen matches based on idiosyncratic match quality. In their framework, lower frictions improve match selection but leave unemployment transition rates unchanged. Birinci et al. (2025) reach a similar conclusion: when applications increase but firms screen on match quality, the main effect is a decline in the separation rate while the job finding rate remains roughly unchanged. In both papers, declining search frictions thus lead to higher match quality, while matching rates stay roughly level. In contrast, we highlight several distinct implications that arise when firms screen on permanent worker characteristics instead of idiosyncratic match quality: a *decrease* in matching efficiency, lengthening unemployment durations, and concentrated negative welfare effects. We provide evidence that this mechanism explains recent developments in the US labor market over the last three decades. Auster et al. (2025) highlight a related mechanism in a directed search framework with adverse selection, showing that declining search frictions can worsen matching outcomes. We obtain a similar result in a random search setting that is designed to closely resemble the canonical DMP model, which allows us to link the mechanism to observable labor market trends.[1,2]

Our paper also contributes to the literature on the causes of rising unemployment duration, which has nearly doubled since the 1990s. Mukoyama and Şahin (2009) emphasize the role of rising wage dispersion in driving the trend, arguing that it raised reservation wages and lengthened unemployment spells through increased selectivity. The literature has also studied the role of duration dependence in shaping the duration of unemployment spells over the business cycle; that is, the idea that unemployment exit hazards decline as unemployment progresses (Kroft et al., 2016). More recent work has emphasized that duration dependence may reflect dynamic selection (Jarosch and Pilossoph, 2019; Mueller et al., 2021; Ahn, 2023). Our theory offers a distinct perspective on the rise in unemployment duration: increasing firm selectivity widens the *dispersion* of job finding rates across workers, which increases duration substantially even when job finding rates decline only moderately. Using micro data from the NLSY, we provide evidence that this channel is quantitatively important.

Finally, we contribute to a rapidly growing literature on the effects of artificial intelligence on matching markets. Cowgill et al. (2024) and Galdin and Silbert (2025) show that access to generative AI lowers the information content of job applications. Related work studies additional dimensions of how AI reshapes hiring (see, e.g., Gans, 2024; Cui et al., 2025; Dhillon et al., 2025; Wiles and Horton, 2025). We contribute to this literature by conducting a general equilibrium analysis that jointly accounts for two channels emphasized in this empirical work: AI makes it easier for workers to contact employers, and it makes it harder for employers to distinguish between applicants.

The remainder of the paper proceeds as follows. We begin by documenting aggregate labor market trends that have unfolded over the last three decades and compare them to sectoral measures of online job search exposure (Section 2). We then formulate a model that rationalizes these trends jointly (Section 3). In Section 4, we compare the predictions of the model with aggregate and micro data. Section 5 quantifies and analyzes an "AI

[1]Our model integrates applications and a screening process into a standard DMP framework, joining a number of papers that have recently incorporated these features into search models (see, e.g., Wolthoff, 2018; Engbom, 2021; Bradley, 2025; Cai et al., 2025).

[2]Pries and Rogerson (2022) is another notable example of a screening model with idiosyncratic match productivity although their focus is on explaining the secular decline in worker turnover and short-duration employment spells.

shock" to evaluate the effects of artificial intelligence on matching outcomes. Section 6 concludes.

2 Job search technology and the US labor market

In this section, we document patterns that have shaped US labor markets over the last five decades. We first establish long-run trends in four key labor market indicators: unemployment duration, job finding rates, job filling rates and matching efficiency (Section 2.1). We then show that these trends have unfolded alongside a rapid proliferation of internet-based job search and a sustained rise in the number of applications submitted per job seeker (Section 2.2). In Section 2.3, we show that employment and vacancy shares of occupations specifically tasked with screening job candidates have also risen since the late 1990s, suggesting that firms have increased the amount of resources they devote to screening. Finally, in Section 2.4, we show evidence that the aggregate trends we document and the increased screening activity are more pronounced in sectors where online job search is more prominent.

2.1 Evolution of the US labor market

2.1.1 Unemployment duration

A key development in US labor markets since the 1990s is that unemployment spells have become more persistent over time. The left panel of Figure 1 plots the average number of months spent in unemployment by job seekers between 1975 and 2025.[3] The series shows a sustained rise in the average duration of an unemployment spell after the early 2000s. Whereas during the expansions of the late 1970s an unemployed job seeker spent, on average, 2.5 months searching for a job; this had mildly increased to 3.2 months by the 1990s. By the late 2010s, this number was 5.3 months—a rise of 75% over the sample. Underlying the rise in mean unemployment duration was the increasing prevalence of long-term unemployment, defined as the share of job seekers who have been unemployed for six months or longer. We show in Appendix Figure A.1 that long-term unemployment grew from 8% in the late 1970s to 20% at the end of the sample period.

The rise in unemployment duration was not a result of slack labor markets, but persisted despite unprecedented increases in labor market tightness. The right panel of

[3] Unless otherwise indicated, we focus on the backward-looking duration of incomplete, ongoing spells among all unemployed workers.

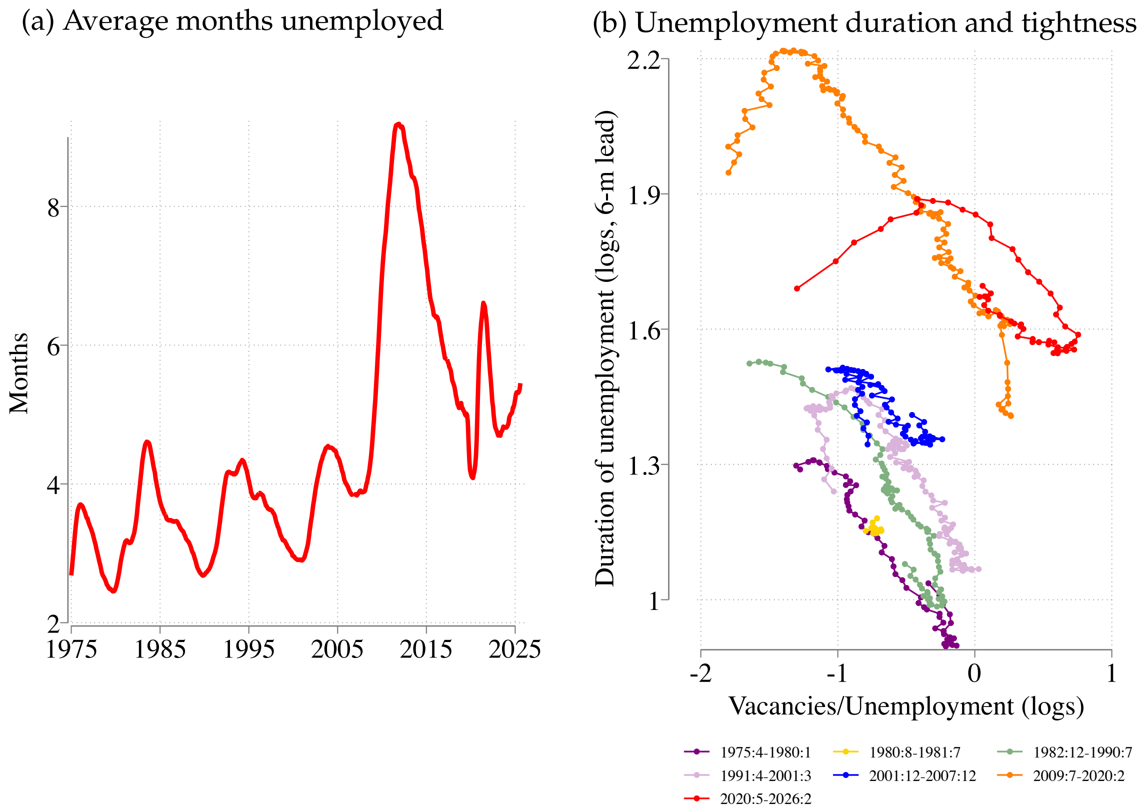


Figure 1: Unemployment duration

Notes: CPS, JOLTS and authors' calculations. Average weeks unemployed from the Current Population Survey. Weeks are converted to months and the figure reports 6-month leads of 12-month centered moving averages of monthly duration. The right panel uses vacancies from JOLTS and Petrosky-Nadeau and Zhang (2021), and unemployed from the CPS.

Figure 1 plots the average unemployment duration with a six month lead against the vacancy-to-unemployment ratio across the seven expansionary episodes of the last 50 years.[4] The figure exhibits an upward trend in unemployment duration conditional on the state of the cycle: in the earlier decades, for any given level of tightness in the recovery, the unemployment spells that followed had a shorter duration on average. By contrast, in the post-Great Recession and post-pandemic periods, unemployment duration left its prior path of recovery and shifted outwards, remaining elevated even when the vacancy-to-unemployment ratio was high.[5]

[4] An expansion is the time period between an NBER-dated trough and the following peak. Appendix Figure A.6 (a) plots an analogue of Figure 1 covering all expansionary and contractionary periods between 1975-2025.

[5] Appendix Figure A.5 (a) plots the same relationship using the unemployment rate as an indicator of

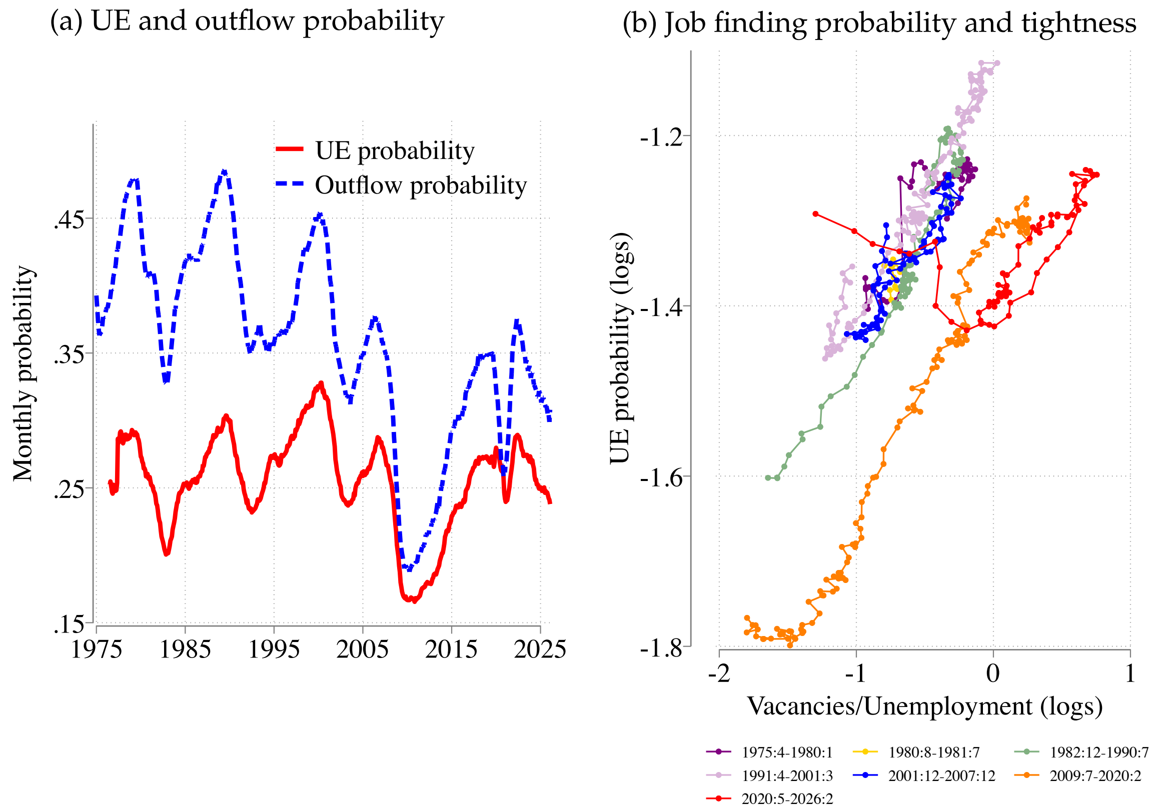


Figure 2: Job finding probability

Notes: CPS, JOLTS and authors' calculations. UE probability is defined as flows from unemployment to employment as a fraction of all unemployed. Outflow probability is defined per footnote 7. The right panel uses vacancies from JOLTS and Petrosky-Nadeau and Zhang (2021), and unemployed from CPS. The figure reports 12-month centered moving averages.

In Appendix A.2, we rule out two plausible explanations of this trend. First, we show that accounting for demographics does not resolve the puzzle: even when we condition on individual demographic characteristics and education, there is a clear upward trend in average unemployment duration.[6] Second, we show that long-run shifts in unemployment insurance are likewise unlikely to be responsible for the observed persistent trend.

the business cycle. The outward shift in unemployment duration remains apparent.

[6]This finding is consistent with Kroft et al. (2016), who rule out demographics-based compositional effects as a quantitatively important reason for the lengthening of unemployment spells after the Great Recession.

2.1.2 Job finding probability

A natural candidate explanation of the rise in unemployment duration is a decline in the rate at which unemployed workers find jobs. Figure 2 documents the evolution of the job finding probability and its relationship with labor market tightness. The left panel plots two measures of the monthly probability of transitioning out of unemployment following Shimer (2012): a flows-based Unemployment-to-Employment (UE) transition probability (solid red) and a stock-based outflow probability (dashed blue).[7] By construction, the stock-based outflow probability is broader, capturing all exits from unemployment, to either employment or non-participation, so it exceeds the UE probability. The figure shows that both series begin exhibiting a mild decline starting in the early 2000s. The outflow probability averaged 44% between 1998-1999 and decreased to 34% in the 2018-19 period. The UE probability displays a similar pattern, trending down from an average of nearly 32% in the pre-2000s period to 27% in the late-2010s.

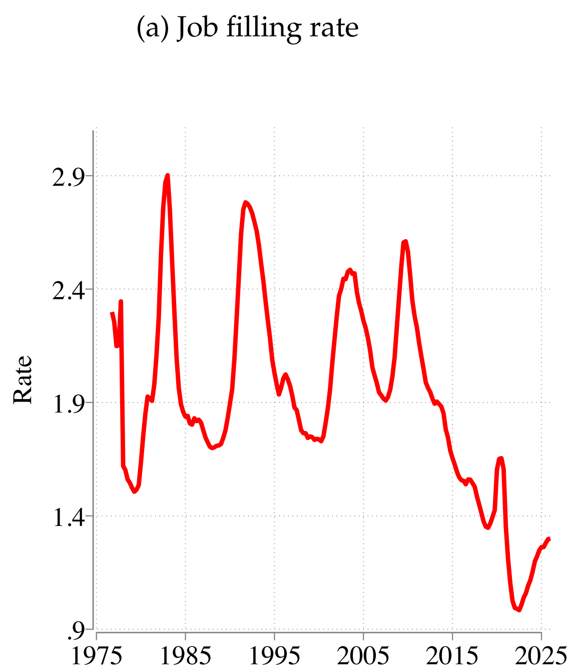
The right panel plots the UE probability against the vacancy-to-unemployment ratio for the recoveries of the last 50 years. The figure reveals a robust positive relationship between job finding probability and labor market tightness: as the vacancies per unemployed worker increases, jobs become easier to find. However, a key feature of this relationship has been that it has shifted downward over time, such that the most recent recoveries have traced paths that lie below the earlier ones. This downward shift suggests a deterioration in the efficiency with which matches are being formed in the labor market such that even when tightness is at the same levels as in the past, workers now find jobs at a lower rate.

The decrease in job finding probability contributed, in part, to the increase in unemployment duration. However, as we document in Appendix A.3, the rise in unemployment duration exceeded what the decline in the job finding probability alone can account for. This growing wedge between the two series signals a departure from the constant-hazard, homogeneous worker benchmark model, and therefore helps adjudicate between competing theories of rising unemployment duration. We return to this point in Section 4.2.

[7] The UE probability is the fraction of unemployed in month $t-1$ who transition to employment in month t, UE_t/U_{t-1}. The outflow probability F_t is derived from the law of motion of unemployment:

$$U_{t+1} = (1 - F_t)U_t + U_{t+1}^{<5w}$$

where U_t is the unemployment stock and $U_{t+1}^{<5w}$ approximates new entrants to unemployment using the stock of unemployed less than five weeks in $t+1$.



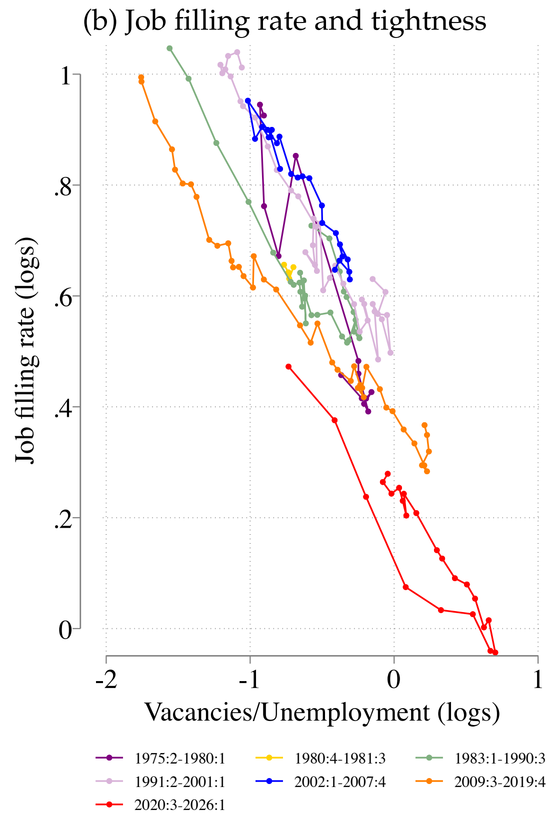


Figure 3: Job filling rate

Notes: CPS, JOLTS and authors' calculations. Job filling rate is defined as the ratio of hires to vacancies. Vacancies from JOLTS and Petrosky-Nadeau and Zhang (2021), unemployed from the CPS, and hires from JOLTS and CPS. The figure reports 12-month centered moving averages.

2.1.3 Job filling rate

We next shift our focus to the firm side. The left panel of Figure 3 plots long-run changes in the job filling rate measured as a ratio of hires to vacancies. We measure hires as monthly flows from the CPS and vacancies as end-of-month stock from the Job Openings and Labor Turnover Survey (JOLTS).[8] Whereas the job filling rate displays strong fluctuations but a flat trend between 1975-2010, starting from the peak of the Great Recession, it follows a pronounced and sustained decline that does not recover until the end of

[8]We compute hires from the CPS defined as the sum of hires from unemployment, hires from non-participants, and hires from employment measured by Fujita, Moscarini, and Postel-Vinay (2024) after 1995 and Diamond and Şahin (2016) and Blanchard, Diamond, Hall, and Murphy (1990) before 1995. We adjust CPS hires following the methodology of Bagga et al. (2025) described in Appendix A.4. In Appendix Figure A.4 we plot job filling rate series using JOLTS and unadjusted CPS hires. The decreasing trend is observed in both series.

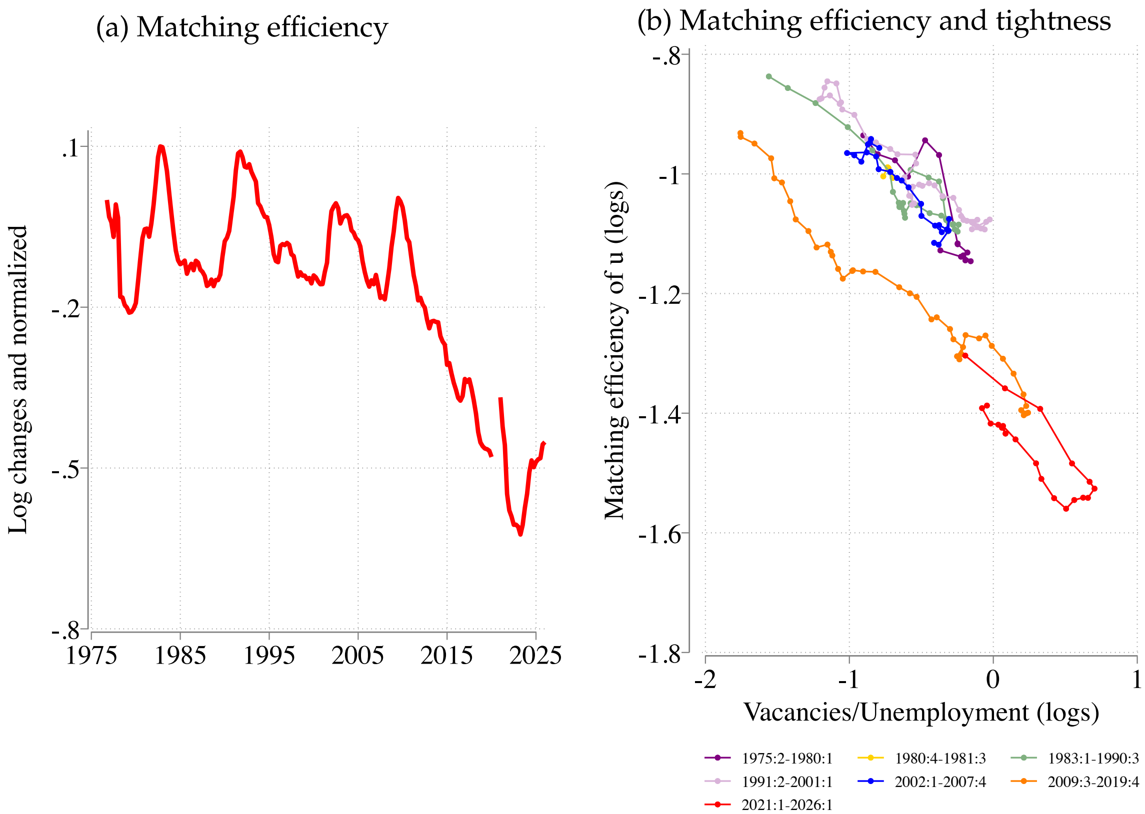


Figure 4: Matching efficiency

Notes: CPS, JOLTS and authors' calculations. Matching efficiency is defined per equation (1). Vacancies from JOLTS and Petrosky-Nadeau and Zhang (2021), and unemployed and hires from unemployment from the CPS. The figure reports 12-month centered moving averages.

the sample. Inverting the job filling rate captures time to fill job positions, revealing that in the 1970s-80s, on average, a vacancy was filled within two to three weeks.[9] However, by the end of the sample, vacancies required more than a month to generate a hire.

Plotting the job filling rate against labor market tightness in the right panel of Figure 3 shows a negative relationship between the two series: as the vacancy stock increases relative to unemployment, jobs become harder to fill. Whereas the recoveries between 1975-1990 trace similar paths, the post-2000 recoveries begin shifting downward, and for any level of tightness, the job filling rate is now lower than it was in earlier decades.

[9]This is consistent with Faberman and Menzio (2018) who use data from the Earnings and Opportunities Pilot Project and document that it took on average 16.2 days to fill a vacancy in 1982.

2.1.4 Matching efficiency

We next turn to "matching efficiency", a summary measure that captures how effectively the labor market brings together job seekers and prospective employers. Here, we define matching efficiency in its simplest form, assuming a Cobb-Douglas meeting technology that takes as inputs vacancies and unemployed job seekers, and gives as output UE hires:

$$\text{ME}_u = \frac{\text{UE hires}}{\sqrt{V \cdot U}} \tag{1}$$

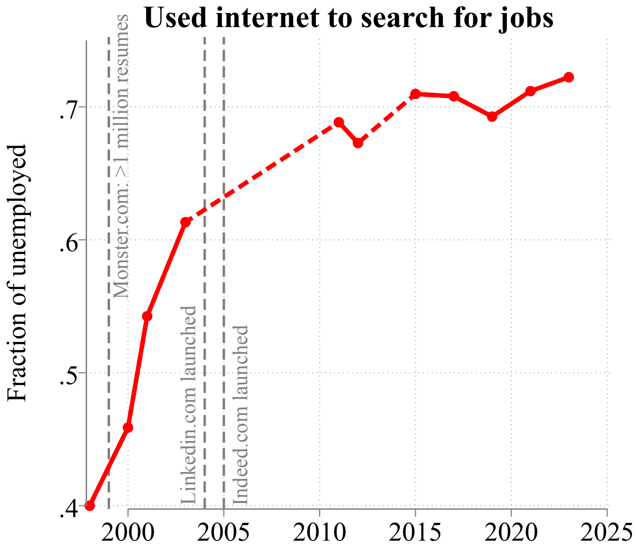
This definition is often motivated as the residual of a matching function: how many hires do we see in the data relative the number of hires we should expect for a given number of unemployed job seekers and vacancies?[10]

The left panel of Figure 4 reports matching efficiency, as defined by equation (1), which we report in logs and normalize to zero at the start of the sample. During the early decades of the sample, the series remains roughly stable around zero, showing no discernible trend. Beginning in the late 1990s, a gradual downward trend starts setting in, and after the Great Recession matching efficiency starts declining precipitously. The pandemic sets in motion a brief rebound but by the end of the sample matching efficiency settles at a level about 0.5 log points below its level in 1975. This implies that, for a given stock of job seekers and vacancies, the labor market now produces about half as many hires as it did five decades ago.

In the right panel of Figure 4, we again compare the dynamics of matching efficiency against the state of the cycle, as measured by the vacancy-to-unemployment ratio. A clear downward shift starts materializing after the recovery of the early 2000s, and the 2010s recovery is characterized by persistently lower levels of matching efficiency for comparable levels of market tightness.

Taking stock. We close this section by summarizing the facts we documented. Over the last five decades, unemployment spells have become more persistent, unemployed workers have taken longer to find jobs, firms have taken longer to fill vacancies, and the overall efficiency with which the labor market converts contacts into hires has declined. The bulk of these developments has taken shape over the last two to three decades, and they have persisted despite historically tight labor markets in the 2010s and early 2020s.

[10] In Appendix A.4, we report alternative measures of matching efficiency that also take into account employed job seekers and hires from employment. The aggregate trends remain similar across the different measures of matching efficiency.



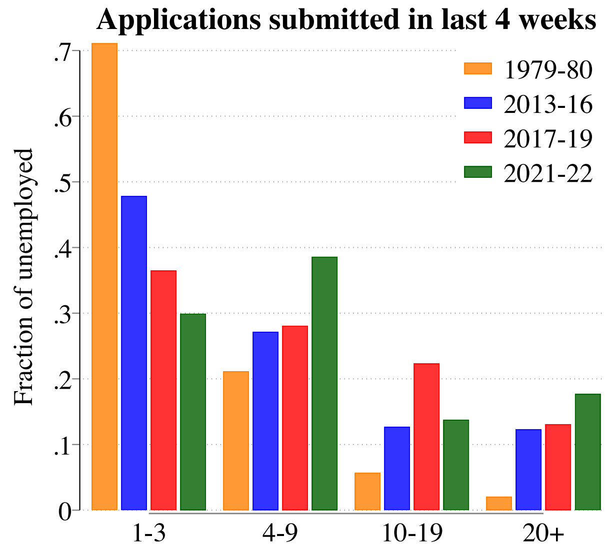


Figure 5: Internet use and applications submitted by the unemployed

Notes: CPS Computer and Internet Use Supplement, Survey of Consumer Expectations, Earnings and Opportunities Pilot Project and authors' calculations. The left panel plots the fraction of unemployed job seekers who reported using the internet to search for jobs. The right panel plots the distribution of applications submitted in the prior four weeks by unemployed job seekers across four time periods.

2.2 Online job search and applications

A parallel development to these aggregate labor market trends has been the proliferation of online job search and an accompanying shift toward online job postings and online applications since the mid-2000s. The left panel of Figure 5 illustrates this development by plotting the share of unemployed job seekers who reported using the internet for job search in the CPS Computer and Internet Use Supplement. This share was around 40% in 1998 and rose steeply over the following decade, coinciding with the emergence and rapid growth of major online job boards such as Monster.com, LinkedIn and Indeed. By the mid-2010s, internet-based job search had become ubiquitous, reportedly being used by over 70% of the unemployed and plateauing at that level through the end of the sample.

The shift in search technology translated into changes in application behavior. In the right panel of Figure 5, we plot the distribution of applications submitted by unemployed job seekers in the prior four weeks across four time periods spanning 1979-2022. For the 1979-80 period, we observe data from the Earnings and Opportunities Pilot Project (EOPP) which reveals that the distribution is concentrated at the lower end: about 70% of unemployed job seekers submitted between one and three applications and virtually none submitted more than ten applications during the prior month of their job spell. Starting in 2013, we observe a comparable measure from the Survey of Consumer Expectations (SCE). By 2013-16, we see that the mass on one to three applications had fallen below 50% while the upper tail had started thickening. This rightward shift of the distribution of applications continued over time, and in the most recent period almost 20% of job seekers reported submitting more than 20 applications over the preceding four weeks.[11] The increasing use of the internet for job search and subsequent rise in applications counts is consistent with the view that the technological shift to online job search lowered the cost of submitting job applications.

2.3 Screening

How did firms respond to the increasing volume of online job applications? The evidence from Figures 3 and 4 suggests that the excess flow of applications did not translate into vacancies being filled more quickly. Here, we present evidence that firms instead responded by intensifying their screening effort.

While it is not possible to track at an aggregate level how many resources firms spent on screening job candidates, we do observe employment and vacancy shares of the six-digit occupation "Human Resources (HR) Specialists" whose task is to "recruit, screen, interview, or place individuals within an organization".[12] We argue that the employment share of this occupation is well-suited as a proxy for screening activity because its SOC definition centers on the recruitment and evaluation of job candidates, and excludes ad-

[11]In the SCE, respondents report the number of applications submitted by answering the question "*How many potential employers, if any, did you apply to for employment within the LAST 4 WEEKS? Please include all applications made in person, online, or through other direct methods. Do not include inquiries that did not lead to a job application.*" The EOPP does not record a directly comparable measure. Instead, the EOPP Look-for-Work supplement asks respondents "*How many applications did you fill out as a result of [search method]?*" and records total applications submitted per search method across each looking-for-work spell. To make the EOPP measure comparable with the SCE, we first aggregate applications across methods within a spell, then prorate to a 28-day window by multiplying total applications in the spell by the share of the spell that falls within the 28 days prior to the interview date. We implicitly assume a constant application rate over the spell. We then sum across overlapping spells for each respondent to arrive at applications submitted per job seeker over the last four weeks preceding the interview. Both the EOPP and SCE distributions are computed among unemployed job seekers who submitted at least one application in the prior four weeks. Observations are weighted.

[12]In the 2010 and 2018 Standard Occupational Classification System (SOC), Human Resources Specialists (13-1071) is defined as those who "perform activities in the human resource area. Includes employment specialists who screen, recruit, interview, and place workers.". In the 2000 SOC, this occupation is referred to "Employment, Recruitment, and Placement Specialists" (13-1071) with definition "Recruit and place workers." In each of the SOC 2000, 2010 and 2018 this occupation excludes "Compensation, Benefits, and Job Analysis Specialists" and "Training and Development Specialists" who perform tasks related to other HR functions such as facilitating compensation and benefits, conducting employee training, etc. For more information please refer to 2000, 2010, and 2018 SOC definitions by the BLS (`https://www.bls.gov/soc/`).

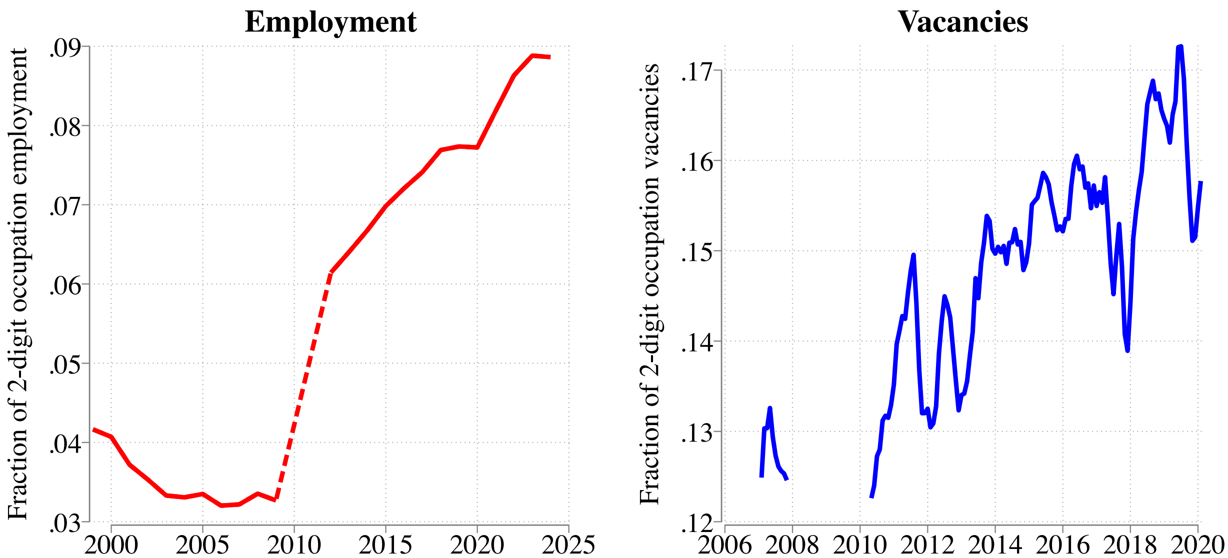


Figure 6: Employment and vacancies in "HR specialists" occupation as a fraction of all operations occupations

Notes: Occupational Employment and Wage Statistics, Lightcast and authors' calculations. HR specialists (SOC 13-1071) are defined as those whose task is to "recruit, screen, interview, or place individuals within an organization". The left panel plots employment of HR specialists as a fraction of all business and financial operations occupations (SOC 13). The right panel plots the share of online job postings advertising HR specialist positions as a fraction of all operations occupation postings.

jacent HR functions such as compensation and benefits administration, and employee training. The definition of this occupation has also remained stable over time.

In the left panel of Figure 6 we document the employment of HR specialists (SOC 13-1071) as a share of business and financial operations occupations (SOC 13) from the Occupational Employment and Wage Statistics (OEWS).[13] The series exhibits an upward trend rising from 4% in 1998 to nearly 9% in 2024 with a bulk of the increase taking place in the last decade, indicating that firms now employ more workers tasked with screening job candidates. The right panel of Figure 6 plots the share of online job postings advertising HR specialist positions in Lightcast data. This series also increases from around 12.5% in 2007-08 to nearly 17% by 2019 suggesting that the increase in HR employment at least in

[13]We focus on the share within business and financial operations occupations to account for the fact that secular changes in the firm size distribution over this time period may mechanically drive employment in all business operations occupations including HR specialists in the aggregate. Therefore, normalizing HR specialists within their two-digit occupational family nets out changing composition of two-digit occupations, and isolates reallocation of business operations toward HR specialists. We show in Appendix A.7 that the trends remain increasing when HR specialists are measured as a share of total employment and total vacancies.

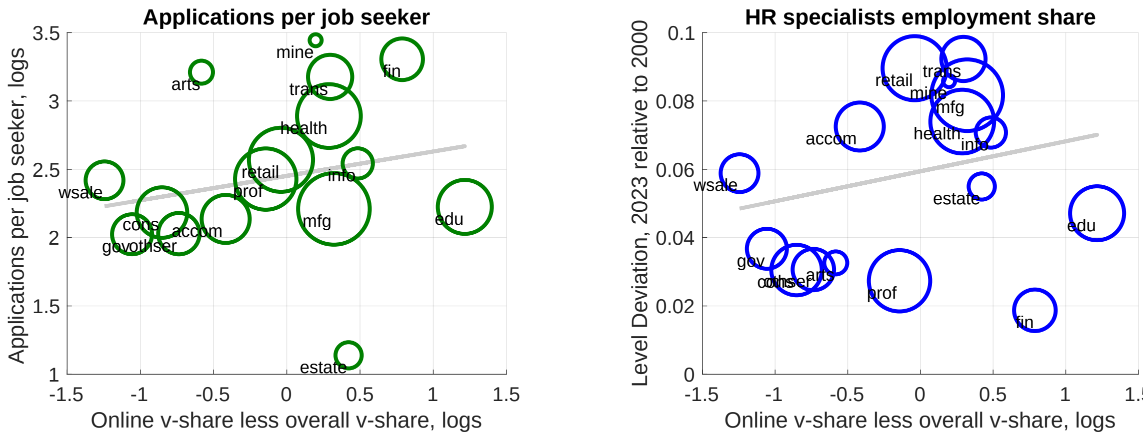


Figure 7: Applications and screening by industry

Notes: OEWS, SCE, Lightcast, JOLTS and authors' calculations. The x-axis plots the log difference between a sector's vacancy share in Lightcast online job postings and its overall vacancy share in JOLTS in 2019. The left panel plots applications per job seeker from the SCE Job Search Supplement. The right panel plots the level deviation in the HR specialist employment share within business operations occupations between 2000 and 2023 from the Occupational Employment and Wage Statistics. In both panels, circle areas are proportional to industry employment and the gray line is the employment-weighted linear fit.

part reflects an increasing firm demand for screening services. We argue that both series provide evidence that firms now commit more resources toward screening activities than they did in the 1990s.

2.4 Cross-sectoral evidence

The preceding sections documented aggregate trends in labor market outcomes alongside the rise of online job search and a firm-side screening response. We now examine whether these developments are linked by considering variation across sectors in their exposure to online job search.

We first construct a sector-level measure of online job search exposure in 2019 based on job postings data. For each two-digit NAICS sector, we measure the vacancy share of the sector in Lightcast online job postings data, and subtract its overall vacancy share in the JOLTS. A higher value of this measure therefore indicates that a sector's vacancies are disproportionately posted online relative to its total vacancy share. We find substantial variation in sectoral exposure to online vacancies. On one end of the spectrum, sectors such as Educational Services and Financial Activities post nearly twice as large a share of their vacancies online as their overall vacancy share would predict. At the other end, sectors such as Construction and Wholesale Trade are sharply underrepresented in online postings relative to their total vacancy share.

We begin by examining whether application behavior and screening activity vary with this measure of online job search exposure. The left panel of Figure 7 shows the level of applications per job seeker in 2013-21 from the SCE data for each two-digit sector, plotted against the sector's exposure to online job postings. The circle areas are proportional to the sectors employment and the gray line is the employment-weighted linear fit. Job seekers in sectors with a higher online vacancy share submit more applications.[14] The right panel plots the change in the HR specialist employment share of business operations occupations between 2000 and 2023 from the OEWS data against the same measure of exposure. We observe that sectors that were more exposed to online job search also experienced a larger increase in screening activity.

We next consider whether more exposed sectors were also the key drivers of the other long-run trends in labor market outcomes. In Figure 8 we plot log deviations between 2000 to 2023 in four key labor market outcomes across sectors against their exposure to online job search: average unemployment duration, the UE rate, the job filling rate, and matching efficiency. The figure shows that sectors with the largest online vacancy shares observed the greatest declines in job filling rates and job finding rates. These sectors were also the key drivers of the excess unemployment duration and experienced more severe declines in matching efficiency.

We argue that the patterns in Figures 7 and 8 are consistent with a mechanism in which online job search reduces the cost of applications, which induces firms to be more selective and to reduce their hiring rates, contributing to the labor market trends we observe. In sectors where vacancies are disproportionately posted online, the cost of applying is lower and job seekers submit more applications (Figure 7, left panel). The resulting increase in application volume does not, however, translate into more hires. Instead, firms respond to the additional inflow of applications by expanding their screening activity (Figure 7, right panel) and consequently, raise their hiring standards, becoming more selective in who they hire. As a result, a larger fraction of applications are rejected, fewer contacts between job seekers and vacancies convert into hires, and both the job finding and job filling rates decline (Figure 8, panels a and b). Both these forces lower matching efficiency (Figure 8, panel d). At the same time, because firms are now more selective, some workers who would have been more likely to find a job under slacker admission

[14]Given that our applications data is limited, we only plot the level of applications in this figure, but plot deviations on the y axis in other figures where we have corresponding data for an earlier time period.

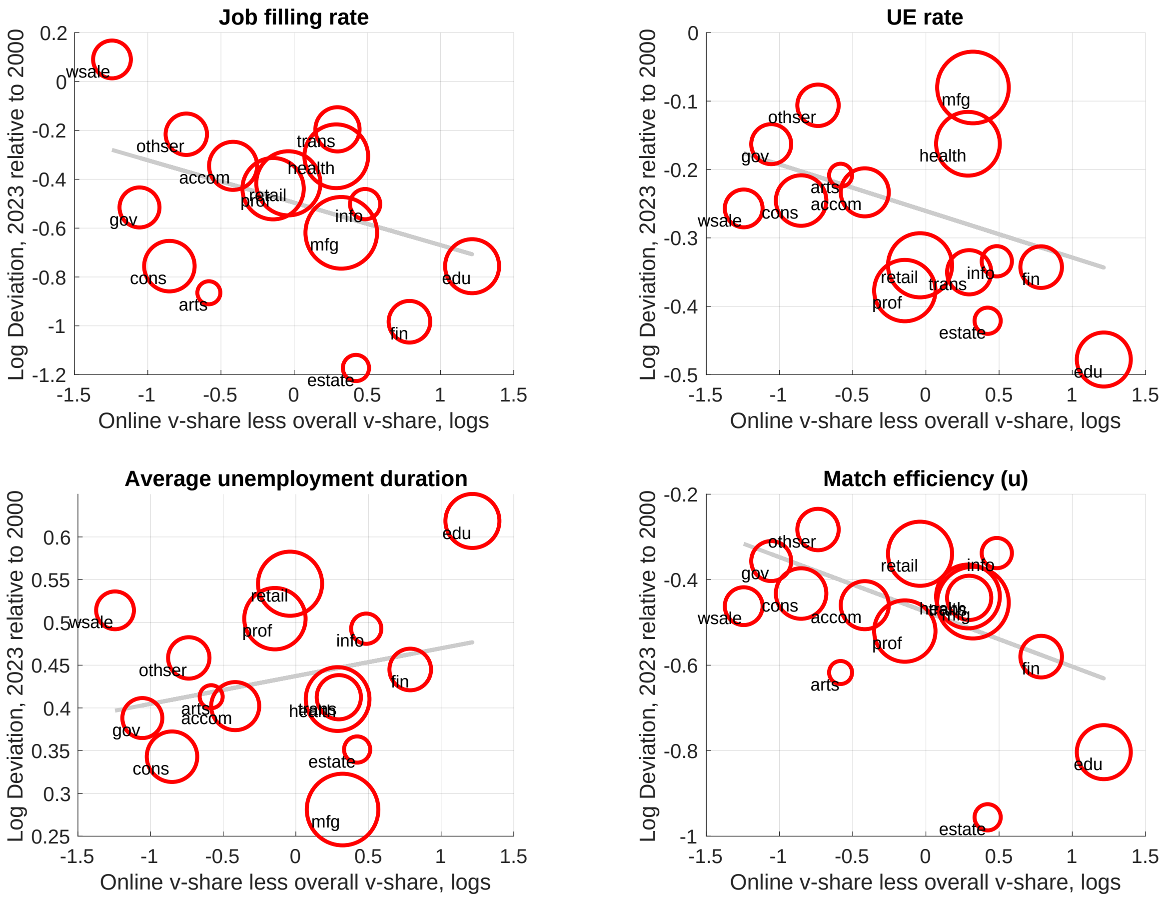


Figure 8: Labor market outcomes by industry

Notes: CPS, JOLTS, Lightcast and authors' calculations. Each panel plots the log deviation from 2000 to 2023 in a labor market outcome against the sector's online job search exposure, defined as the log difference between a sector's vacancy share in Lightcast online job postings and its overall vacancy share in JOLTS in 2019. Circle areas are proportional to industry employment and the gray line is the employment-weighted linear fit. The job filling rate is defined as hires divided by vacancies. The UE rate is defined as hires from unemployment divided by unemployed job seekers. Matching efficiency is computed per Equation 1.

criteria instead remain persistently unemployed, driving up the average duration of unemployment (Figure 8, panel c).

In the next section, we formalize these ideas through a model that captures the relationship between declining application costs, endogenous screening, and aggregate labor market outcomes.

3 Model

The model is an extension of the canonical Diamond-Mortensen-Pissarides (DMP) model with three modifications. First, we introduce a notion of applications: upon meeting a vacancy, workers draw a random cost and decide whether to apply. Second, we introduce a notion of screening. There are two permanent, unobserved worker types who differ only in their productivity. Firms can pay a random cost to receive a signal about an applicant's type before making a hiring decision. Third, vacancy costs are sunk, which introduces a trade-off between filling a job immediately and waiting for a worker of higher expected productivity. We begin by outlining the environment (Section 3.1), derive optimal decision rules (Section 3.2) and flows (Section 3.3), and define an equilibrium (Section 3.4). We then use the model to explore the effects of online job search (Section 3.5) and take it to the data (Section 3.6).

3.1 Environment

Demographics. Time is continuous. Throughout the paper, we focus on stationary equilibria and thus suppress time subscripts. The economy consists of a unit measure of infinitely-lived workers who can be either employed or unemployed. Workers permanently belong to one of two types $i \in \{L, H\}$, with type-L workers comprising an exogenous share α_L^{pop} of the population. Employed workers produce output y_i with $y_H \geq y_L$. Unemployed workers receive a flow payoff $b_i = (1-\rho)y_i$, where ρ denotes the replacement rate, and search for a match. When employed, workers of type i receive a wage w_i and do not search. On the other side of the market are one-worker jobs that can be filled and producing or vacant and searching for a worker.[15] Firms and workers discount future at rate r.

Job creation. Jobs are homogeneous. There is an infinite supply of potential entrants who can create new vacant jobs ("vacancies") at will. To create a new vacancy, an entering firm must pay a fixed cost, k.[16]

[15] We will also refer to jobs as firms. In our economy, as in most search models, ownership across jobs is irrelevant for outcomes and each job operates independently.

[16] This assumption is a departure from the more standard flow cost assumption. With a flow instead of sunk vacancy cost, the value of a vacancy is driven to zero in equilibrium, and it is never optimal for a firm to forego a match with positive surplus. Thus, a sunk cost is a simple way to generate a positive vacancy value and model the firm's trade-off between matching with a worker immediately and waiting to match with a worker of better expected quality in the future, which will be key to the mechanism we emphasize.

Meeting technology. Vacancies and heterogeneous workers meet at random in a frictional labor market to form matches. Let u denote the measure of unemployed workers and v the measure of vacancies. The total number of meetings at any instant of time is given by an aggregate meeting technology that takes a Cobb-Douglas functional form, $m = Zv^{\omega}u^{1-\omega}$ where Z is the aggregate meeting efficiency and ω is the elasticity of meetings with respect to vacancies. The aggregate meeting rates for vacancies and workers are

$$q = m/v, \quad \lambda = m/u = Z^{\frac{1}{1-\omega}} q^{\frac{\omega}{\omega-1}} \tag{2}$$

Applications. Upon encountering a vacancy, an unemployed worker draws an application cost c from a distribution with cdf $F(c) = \gamma_a c^{\eta_a}$. If the worker pays cost c, the firm receives their application; otherwise, the encounter is destroyed. We refer to γ_a as the "ease of applying": a higher value corresponds to a lower cost of the average application. We denote the (endogenous) probability that a type-i worker submits an application, conditional on an encounter, as p_i^{app}.

Screening. Upon receiving an application, the firm does not observe the worker's type.[17] The firm has access to a screening technology that reveals a signal (L or H) about the worker's type. Operating this technology entails a one-time cost κ, drawn from a distribution with cdf $G(\kappa) = \gamma_s \kappa^{\eta_s}$. We refer to γ_s as the "ease of screening": a higher value corresponds to a lower cost of the average signal. We use p^{sig} to denote the (endogenous) probability that the firm decides to purchase a signal.[18] The signal reveals the worker's true type with probability $s \geq 0.5$, and we refer to s as the "signal quality". The signal is thus imperfect but weakly more likely than not to reveal the workers true type.

[17]For simplicity, we assume that the firm cannot observe anything about the worker's individual state unless it decides to obtain a costly signal about the worker. In particular, we depart from some recent work, for example by Jarosch and Pilossoph (2019) and Bradley and Mann (2024), in assuming that the firm cannot observe a worker's employment or unemployment history. These papers emphasize that *duration dependence* arises when a worker's unemployment history is observable to the firm; that is, the likelihood that a worker exits unemployment changes with their duration of unemployment. We abstract from this assumption for tractability and transparency reasons, and argue in Appendix B.7 that true duration dependence does not drive the trend in unemployment duration that we seek to explain.

[18]In Appendix B.1, we establish that this screening technology is equivalent to a formulation where the firm controls the probability of receiving a signal p^{sig} and pays a deterministic screening cost that is convex in p^{sig}.

Hiring and production. After the application and screening stages conclude, the firm decides whether to hire the applicant. It may do so regardless of whether it has obtained a signal. If the firm does not hire, the worker returns to unemployment and the vacancy remains open for future applicants. Upon hiring, the worker's type is revealed and the match produces flow output y_i. The firm retains a flow profit of $y_i - w_i$, where wages w_i are determined via Nash bargaining with worker bargaining weight β. Both filled and unfilled jobs are destroyed at rate δ. Once a firm matches with a worker, it relinquishes the option to search for free: if it wishes to search again, it has to pay again the sunk cost k.[19]

Surplus function and wage setting. The surplus function for a match of type i is defined as

$$S_i = W_i - U_i + J_i, \tag{3}$$

where U_i and W_i, respectively, denote the value of unemployment and employment to a type-i worker. J_i is the value to the firm of forming a match with type-i worker.[20]

We assume that wages are set by Nash bargaining – that is, they solve

$$w_i = \text{arg max}_{w_i} J_i(w_i)^{1-\beta}(W_i(w_i) - U_i)^{\beta}.$$

Thus, the surplus accruing to the firm and worker in a match is, respectively,

$$W_i - U_i = \beta S_i, \qquad J_i = (1-\beta)S_i. \tag{4}$$

3.2 Value functions and optimal decisions

We now describe the value functions of workers and firms and derive optimal application and screening decisions.

[19] This assumption is a simple way to ensure that all matches are maintained even after a worker's type is revealed, which would also be true under any sufficiently large cost of re-posting. An alternative setup would be one in which firms learn about a worker's type over time and separate if their beliefs become sufficiently pessimistic, as in Baley et al. (2022). We argue that the distinction between both approaches is not central to the mechanism we emphasize.

[20] Note that the vacancy value does not enter the surplus expression because we assume that filled jobs cannot costlessly return to search. The consequence of this assumption is that it is never optimal for the firm to fire a type-L worker and return to search.

Value of an unemployed worker. The value to an unemployed worker of type i, U_i, satisfies

$$rU_i = b_i + \lambda \mathbb{E}\left[\max\left\{ 0, \underbrace{\left[p^{\text{sig}} \left(\mathbb{I}_{i=L}(1-s) + \mathbb{I}_{i=H} s \right) + (1-p^{\text{sig}})(1-R) \right]}_{p_i^{\text{acc}} \equiv \text{Acceptance probability}} (W_i - U_i) - c \right\} \right] \tag{5}$$

An unemployed worker receives a flow value of b_i and meets vacancies at rate λ. Upon meeting, the worker draws an application cost c and decides whether to apply. Equation (5) captures this decision through the max operator: the worker applies whenever the expected gain from an application, which we denote as $\Delta_i^w \equiv p_i^{\text{acc}} \cdot (W_i - U_i)$, exceeds the cost c. The endogenous probability of applying is therefore given by $p_i^{\text{app}} = F(\Delta_i^w)$. The acceptance probability p_i^{acc} is the sum of two terms. If the firm obtains a signal about the worker, which occurs with probability p^{sig}, then type-L workers are accepted with probability $1-s$ and type-H workers with probability s.[21] If the firm does not obtain a signal, which occurs with probability $1-p^{\text{sig}}$, the firm's hiring decision does not depend on the worker's true type. We define $R \in [0,1]$ as the probability that the firm rejects applicants in this scenario.

In what follows, a useful object is the worker's job finding rate $f_i = \lambda F(\Delta_i^w) p_i^{\text{acc}}$. It is the product of the meeting rate, the application probability and the acceptance probability.

Value of an employed worker. The value of an employed worker of type i, is characterized by the following Bellman equation:

$$rW_i = w_i + \delta(U_i - W_i). \tag{6}$$

An employed worker receives a flow wage w_i and faces exogenous separation at rate δ, upon which the worker returns to unemployment.

Value of a filled job. A firm matched to a worker of type i receives a value J_i where

$$(r+\delta)J_i = y_i - w_i \tag{7}$$

[21] This follows from an argument we outline below: the optimal acceptance strategy of a firm that has purchased a signal is to admit the worker if and only if the signal is H.

The flow value of a filled job is the output y_i produced by the worker net of the wage w_i. The match dissolves at rate δ, at which point the firm exits with zero value.

Value of an unfilled job. The value of an unfilled job opening, V, satisfies the following Bellman equation:

$$(r+\delta)V = \underbrace{q(a_L^u p_L^{\text{app}} + (1-a_L^u)p_H^{\text{app}})}_{\text{encounter rate}} \cdot \mathbb{E}\left[\max\{V_{\text{signal}} - V - \kappa, V_{\text{nosignal}} - V\}\right]. \qquad (8)$$

An unfilled job meets applicants at rate q. The unemployed pool is composed of type-L workers (with share a_L^u) and type-H workers who apply with probabilities p_L^{app} and p_H^{app}, respectively. Conditional on receiving an application, the firm draws a screening cost κ and decides whether to purchase a signal. The max operator captures this decision: the firm screens whenever the net gain from obtaining the signal, $V_{\text{signal}} - V_{\text{nosignal}}$, exceeds the cost κ. We define this threshold as $\Delta^f = V_{\text{signal}} - V_{\text{nosignal}}$, so the probability of purchasing a signal is $p^{\text{sig}} = G(\Delta^f)$.

The firm's hiring decision depends on whether it has purchased a signal. Consider first the case in which it has. A firm that purchases a signal optimally accepts the worker if and only if the signal is H. This follows from two arguments: first, the value of a signal lies in its ability to change the hiring decision, so a firm that ignores the signal outcome would have no reason to purchase it; second, since, $J_H \geq J_L$ and $s \geq 0.5$, accepting on L signals but not on H signals is a weakly dominated strategy. The expected value to a firm that has purchased the signal is therefore

$$V_{\text{signal}} = \underbrace{(1-a_L^{\text{app}})sJ_H + a_L^{\text{app}}(1-s)J_L}_{\text{average value of accepting}} + \underbrace{\Big(1-(1-a_L^{\text{app}})s - a_L^{\text{app}}(1-s)\Big)V}_{\text{average value of rejecting}}, \qquad (9)$$

where a_L^{app} is the share of type-L workers among all applicants that we characterize in Section 3.3. The first term captures the case in which the worker is truly type H and the signal is correct: this occurs with probability $(1-a_L^{\text{app}})s$ and yields value J_H. The second term captures the case in which the worker is truly type L but the signal is incorrect: this occurs with probability $a_L^{\text{app}}(1-s)$ and yields value J_L. In the remaining cases, the signal is L and the firm rejects the applicant, retaining the vacancy value V.

We now consider the case in which the firm decides not to purchase a signal. We express the value of a firm as:

$$V_{\text{nosignal}} = \max_R (1-R) \cdot [(1-a_L^{\text{app}})J_H + a_L^{\text{app}} J_L] + R \cdot V = max\{V, (1-a_L^{\text{app}})J_H + a_L^{\text{app}} J_L\}. \tag{10}$$

Without a signal, the firm cannot distinguish between worker types and faces a binary choice. It can accept the applicant, receiving the expected match value, or reject the applicant and remain vacant. By accepting applicants without a signal, the firm fills the vacancy more quickly but with a worker of lower expected productivity. By rejecting such applicants, the firm preserves the option to fill the job with an applicant of higher expected productivity later on. The rejection probability without a signal, R, captures how the firm resolves this trade-off. When the expected value of an unscreened hire, $(1-a_L^{\text{app}})J_H + a_L^{\text{app}} J_L$, exceeds the vacancy value V, the firm accepts all unscreened applicants ($R = 0$). When it falls below, the firm rejects them ($R = 1$). For R to lie in the interior, the firm must be indifferent between accepting and rejecting workers without a signal.

Free entry. To close the firm side of the model, we turn to entry. To create a new vacancy, a potential entrant must pay a sunk entry cost k. Free entry implies that entrants create vacancies (and thereby change q) until the value of a vacancy is driven to the entry cost, so that $V = k$ in equilibrium.

3.3 Flows

Having characterized the optimal decisions of workers and firms, we now derive the steady-state conditions that pin down the aggregate composition of the unemployed and the applicant pool. For type-i workers the outflow from unemployment must equal inflow:

$$u_L f_L = \delta(\alpha_L^{\text{pop}} - u_L), \quad u_H f_H = \delta((1-\alpha_L^{\text{pop}}) - u_H) \tag{11}$$

where u_i denotes the steady-state mass of unemployed type-i workers.

The share of type-L workers among the unemployed follows directly from the steady-state unemployment levels:

$$a_L^u \equiv \frac{u_L}{u_L + u_H} = \frac{\alpha_L^{\text{pop}}(f_H + \delta)}{\alpha_L^{\text{pop}} f_H + (1 - \alpha_L^{\text{pop}}) f_L + \delta} \tag{12}$$

Similarly, the share of type-L workers among applicants depends on the composition of the unemployed pool and the type-specific application probabilities:

$$a_L^{\text{app}} = \frac{a_L^u p_L^{\text{app}}}{a_L^u p_L^{\text{app}} + (1 - a_L^u) p_H^{\text{app}}} \tag{13}$$

3.4 Equilibrium

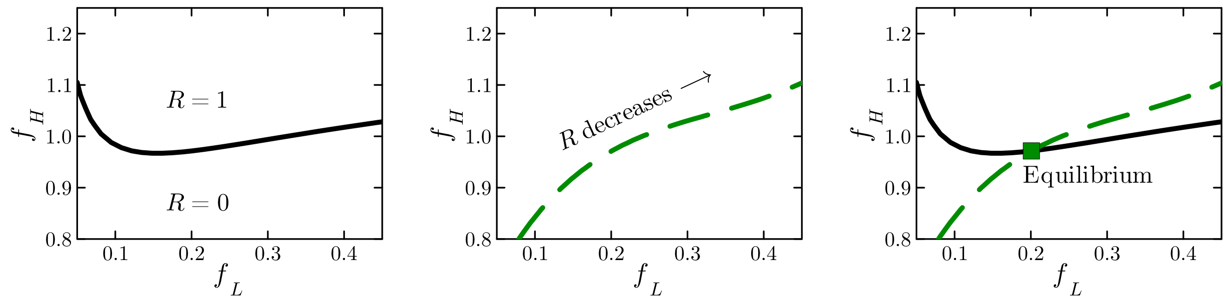
Definition (equilibrium).
An *equilibrium* is a tuple $(\lambda, q, V, V_{\text{signal}}, V_{\text{nosignal}}, \{W_i, U_i, J_i, S_i\}_{i \in \{L,H\}}, \{w_i\}_{i \in \{L,H\}}, R, a_L^u, a_L^{\text{app}})$, such that the above Bellman equations ((3), (4), (5), (6), (7), (8), (9), (10)) and flow equations ((11), (12), (13)) are satisfied, and λ and q are consistent with the meeting function (equation (2)).

As discussed above, the firm's rejection decision without a signal, R, governs the key trade-off faced by the firm: filling a job quickly but with workers of lower expected productivity, or filling it slowly but with workers of higher expected productivity. Since changes in this trade-off will be central to the mechanism we study, we focus on equilibria in which firms choose $R \in (0, 1)$. That is, the firm follows a mixed strategy when deciding whether to reject applicants without a signal.

Definition (mixed strategy equilibrium).
A *mixed strategy equilibrium* is an equilibrium that features $R \in (0, 1)$.

As we show in Appendix B.2, a mixed strategy equilibrium can be characterized as the solution to a system of two non-linear equations in two variables, f_L and f_H:

$$k = \overbrace{a_L^{\text{app}} J_L + (1 - a_L^{\text{app}}) J_H}^{\text{value of admitting without signal}} \tag{IC}$$



(a) Indifference curve (b) Screening curve (c) Equilibrium

Figure 9: Graphical model intuition

Notes: Left panel: indifference curve, defined by all points (f_L, f_H) that solve equation (IC). $R = 0$ and $R = 1$ indicate regions of the state space in which the firms optimally accept and reject no-signal encounters, respectively. Middle panel: screening curve, defined by all points (f_L, f_H) that solve equation (SC). As indicated, points in the top-right direction correspond to a lower value of R. Parameters are taken from Table 1. Right panel: (IC) and (SC), plotted together. Equilibrium job finding rates lie at the intersection of both curves.

$$f_H = \left(\frac{p_H^{\text{app}}}{p_L^{\text{app}}}\right) \cdot f_L + \underbrace{\lambda p_H^{\text{app}} p^{\text{sig}}(2s-1)}_{\text{screening advantage of } H \text{ types}} \tag{SC}$$

where $J_i = J_i(f_i), \lambda = \lambda(\vec{f}), p_i^{\text{app}} = p_i^{\text{app}}(\vec{f}), p^{\text{sig}} = p^{\text{sig}}(\vec{f})$ and $a_L^{\text{app}} = a_L^{\text{app}}(\vec{f})$ are closed-form functions of $\vec{f} = (f_L, f_H)$ defined in Appendix B.2. We refer to equation (IC) as the "indifference curve" and to equation (SC) as the "screening curve". Figure 9 plots both curves for our baseline calibration, discussed in Section 3.6.

The indifference curve. The indifference curve (IC) states that, for a mixed strategy equilibrium, the firm must be indifferent between accepting an applicant without a signal (which yields a weighted average of job values) and rejecting them (which yields value $V = k$). The left panel of Figure 9 traces out all combinations (f_L, f_H) for which this condition is satisfied. Two margins operate along the curve. First, the job finding rates affect the composition of the applicant pool: as f_L rises relative to f_H, the share of type-L applicants falls. Second, the job finding rates affect each type's bargaining position and therefore the rents that the firm can extract from a match: as f_i rises for any type, J_i falls.

For combinations of (f_L, f_H) above the curve, the applicant pool tilts toward type-L workers, which reduces the expected value of hiring a worker without a signal. Therefore, the firm strictly prefers to reject workers without a signal ($R = 1$). Below the curve, the applicant pool is more favorable and the firm strictly prefers to accept workers without a signal ($R = 0$). The roughly horizontal shape of the curve reflects the interaction between the two margins: along the curve to the right, the likelihood of encountering type-H workers increases (f_L rises), but the bargaining position of type-L workers also improves. These two effects largely offset, keeping the firm indifferent.

The screening curve. The screening curve (SC) captures the relationship between the job finding rates of type-L and type-H workers. Two margins introduce a wedge between f_L and f_H: First, the two types apply at different rates, captured by the ratio of application probabilities. Second, type-H workers enjoy a screening advantage: since $s \geq 0.5$, a type-H worker is more likely to produce an H signal to the firm. The middle panel of Figure 9 traces out the screening curve; that is, all combinations of (f_L, f_H) consistent with equation (SC). Each point on the screening curve corresponds to a particular value of R. It is upward sloping: a lower value of R (greater willingness to admit workers without a signal) raises job finding rates of both worker types. However, reducing R disproportionately raises the job finding rate of type-L workers, because their chances of being hired depend more strongly on the likelihood that a firm admits workers without a signal.

Equilibrium. The equilibrium of the economy lies at the intersection of the two curves, as shown in the right panel of Figure 9. The position of the intersection determines equilibrium job finding rates (f_L, f_H), from which all other equilibrium variables can be derived in closed form (see Appendix B.2). The equilibrium value of R is pinned down by the position at which the indifference curve intersects the screening curve.

3.5 Declining application costs through the lens of the model

We now study how the economy responds to a decline in applications costs, modeled as an increase in γ_a. Figure 10 illustrates the resulting shift in equilibrium, using parameter values described in Section 3.6 below.

A decline in application costs leaves the indifference curve unchanged: the firm's decision between accepting and rejecting an applicant without a signal depends on the composition of the applicant pool and the value of unfilled jobs, neither of which is directly affected by the cost of applying. The screening curve, however, shifts up: as applying becomes cheaper, the intercept term in equation (SC) increases.

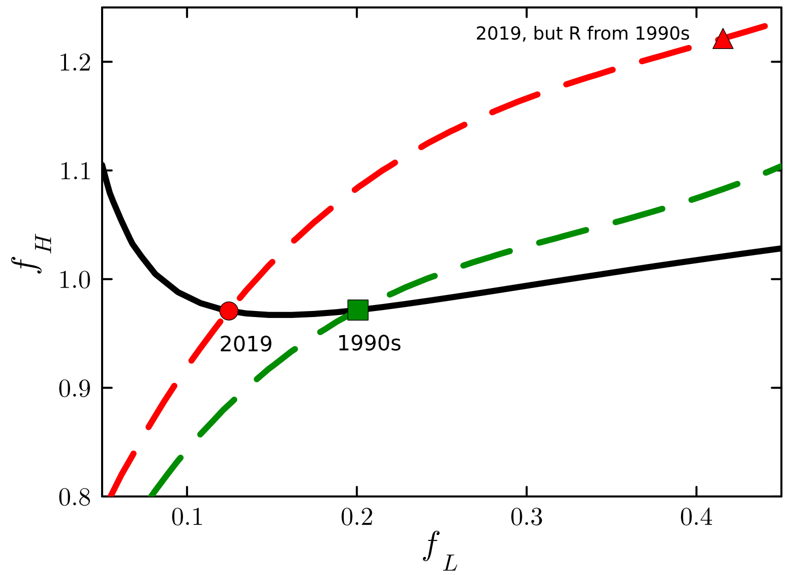


Figure 10: Effect of an application cost shock

Notes: Black line: Indifference curve (IC) for the 1990s and 2019 steady state. Green dashed line: Screening curve (SC) in 1990s steady state. Red dashed line: Screening curve (SC) in 2019 steady state (identical parameters except $\gamma_a' > \gamma_a$). Parameters are taken from Table 1. Dots labeled "1990s" and "2019" indicate respective equilibria.

To build intuition, suppose first that R, the rejection probability of workers without a signal, remains at its initial value. The increase in applications raises job finding rates for both worker types, moving the economy from the initial equilibrium to the red triangle in the upper right corner of Figure 10. To a first order, cheaper applications benefit both worker types by nearly the same amount, so this point lies above the initial screening curve.

The red triangle, however, is not an equilibrium. In this state of the economy, firms strictly prefer to reject applicants without a signal in order to match with workers of higher expected quality later on. As firms become more selective, R rises and the economy moves down along the new screening curve toward the the new equilibrium. At this point, job finding rates for both worker types are strictly lower than in the initial steady state. The decline is especially severe for type-L workers, whose chances of being hired more strongly depend on being admitted without a signal. The fall in application costs thus produces a perhaps counterintuitive result: cheaper applications do not lead to more hires, but to fewer, as the increase in application volume is more than offset by the endogenous rise in firm selectivity.

Parameter	Value	Target	Target value	Time period
Exogenously set / normalizations				
Signal quality s	1.0	Normalization	—	—
Encounter productivity Z	$\approx 1.4 \cdot 10^3$	Normalization	—	—
Low–type productivity y_L	1.0	Normalization	—	—
Worker bargaining weight β	0.5	Convention (DMP)	—	—
Matching elasticity ω	0.3	Petrongolo and Pissarides (2001)	—	—
Discount rate r (monthly)	0.05/12	Convention	—	—
Destruction rate δ (monthly)	0.035	JOLTS	—	—
Internally calibrated				
H–type productivity y_H	9.4	Wage dispersion (CV)	0.47	1990
Replacement factor ρ	0.599	Replacement rate	0.4	1990
Share of L–type workers α	0.189	Unemployment duration (months)	3.0	1990
Entry cost k	3.9046	Job-finding rate	0.588	1990
Ease of screening γ_s	0.331	Job-filling rate	1.02	1990
Application cost dispersion η_a	0.402	Excess dispersion of app's	1.42	Present
Screening cost dispersion η_s	4.38	Differential job finding elast.	0.76	Present
Ease of applying (1990s) γ_a	0.0120	Application rate (per month)	6.8	1990
Ease of applying (2019) γ_a'	0.0337	Application rate	13.0	Present

Table 1: Parameter values and targets

Notes: Calibration table. "Parameter" indicates the name and symbol of the parameter. "Value" indicates the parameter value. "Target" indicates the targeted variable or the justification for exogenously set or normalized parameters. "Target value" indicates the value of targeted variables. "Time period" indicates the targeted time period. Aggregate targets are averaged over the periods Dec. 1998 - Dec. 2000 and Dec. 2017 - Dec. 2019, respectively.

3.6 Calibration

We now discipline the model quantitatively. Table 1 summarizes all parameters values and their targets. We organize parameters into two groups: those set externally by normalization or based on existing literature, and those calibrated internally to match moments from the data.

We calibrate the model to two steady states, corresponding to the late 1990s (Dec 1998 - Dec 2000) and late 2010s (Dec 2017 - Dec 2019), respectively. Both periods correspond to similar points in the business cycle with low unemployment rates (4.1% and 3.8%, respectively), alleviating concerns that differences in labor market outcomes between the two periods are due to cyclical factors. All parameters are held constant across steady states and target a single data moment in one of the two periods, mostly the 1990s. The only exception is the ease of applying, γ_a, which is allowed to change between steady states and targets the application rate in both periods.[22] This identification strategy allows us to isolate the effects of online job search through the lens of the model.

We begin with normalizations and externally set parameters. The signal quality s is normalized to one. The choice of s affects only the mapping between points on the screening curve and the implied value of the rejection probability R, not the shape of the curve itself. Since $R \in (0, 1)$ is necessary for a mixed strategy equilibrium, and s imposes restrictions only on the *length* of the screening curve, we set $s = 1$ to keep the parameter space for which mixed strategy equilibria exist maximally large, and note that any value of s sufficiently close to one would deliver observationally equivalent outcomes. Next, we normalize the encounter productivity parameter Z to a value that guarantees a worker encounter rate of 1000, which ensures that $F(\Delta_i^w) < 1$ but has no impact on the model beyond this.[23] Our final normalization is to set the productivity of the low type y_L to one. We then impose four parameters using conventional values from the literature: the worker bargaining weight $\beta = 0.5$, the elasticity of the encounter function $\omega = 0.3$ (Petrongolo and Pissarides, 2001), the discount rate r equal to 5% annually, and the job destruction rate $\delta = 0.035$, consistent with a long-run average of the separation rate according to data from the Bureau of Labor Statistics.

The second parameter block is disciplined by matching data moments. We choose the productivity of H-type workers, y_H, to target a coefficient of variation of wages equal to 0.4, which we take from Hornstein et al. (2011). The replacement factor ρ targets a replacement rate of 40%.[24] The population share of L-type workers α_L^{pop} matches the the average duration of unemployment, the entry cost parameter k targets the average job finding rate, and the the screening cost parameter γ_s targets the average job filling rate, all in the 1990s steady state.[25,26]

The three remaining parameters are the application and screening elasticities, η_a and η_s, and the ease of applying, γ_a. We discuss each in turn. The application cost dispersion

[22] No direct measure of the application rate exists for our initial calibration period. We therefore rely on EOPP data and compute the application rate in 1979–80 following Birinci, See, and Wee (2025).

[23] Z and γ_a are not separately identified in this model as long as $F(\Delta_i^w) < 1$.

[24] We define the replacement rate as the average production of unemployed workers relative to the average wage.

[25] As discussed in Section 4.2, the implied wedge between the job finding rate and the average duration of unemployment is informative about the degree to which job finding rates are heterogeneous in the model, which is directly moderated by α_L^{pop}.

[26] We match the unemployment outflow series from Shimer (2012).

parameter η_a targets the cross-sectional overdispersion in applications, defined as

$$\mathcal{D} = \frac{\sqrt{\text{Var}(N) - \mathbb{E}[N]}}{\mathbb{E}[N]}$$

where N is the number of applications sent over a month. As we show in Appendix B.3, $\mathcal{D}$ is a sufficient statistic that directly identifies the coefficient of variation in application rates. This makes $\mathcal{D}$ a relevant target for η_a which governs the dispersion of application rates across L and H types. We compute $\mathcal{D}$ from the 2018 and 2019 waves of the Survey of Consumer Expectations.[27] Next, the screening cost dispersion η_s governs the slope of the screening curve and thus the relative sensitivity of type-specific job finding rates to productivity shocks. We choose η_s to match the difference in the elasticities of job finding rates with respect to a marginal productivity shock between H and L type workers. We target a value of 0.76 consistent with evidence in Mann (2026). Finally, we choose the ease of applying in the initial and final steady state, γ_a and γ_a', to match an average monthly application rate of 6.8 and 13, respectively, which we take from Birinci et al. (2025).

4 The effects of online job search in the model and data

We now evaluate the macroeconomic and microeconomic consequences of online job search, conceptualized as a shock that decreases the cost of applications, and compare the predictions of the model with data. We begin by studying the effects of the shock on aggregate moments (Section 4.1). We then show that the micro-level predictions of the model are consistent with the data: first, on the worker side, the gap between type-specific job finding rates widens, contributing to a rise in unemployment duration (Section 4.2); second, we validate two model predictions on the firm side: firms commit more resources to screening applicants and the composition of hires shifts toward more desirable workers with shorter unemployment durations (Section 4.3). Finally, we explore welfare consequences of the shock (Section 4.4).

4.1 Macro implications

Figure 11 compares the changes of aggregate labor market variables between the two steady states in the model with the corresponding changes observed in the data. We

[27] Since we have data from the SCE only for the later period, we target the model counterpart of $\mathcal{D}$ in the late 2010s steady state of the model.

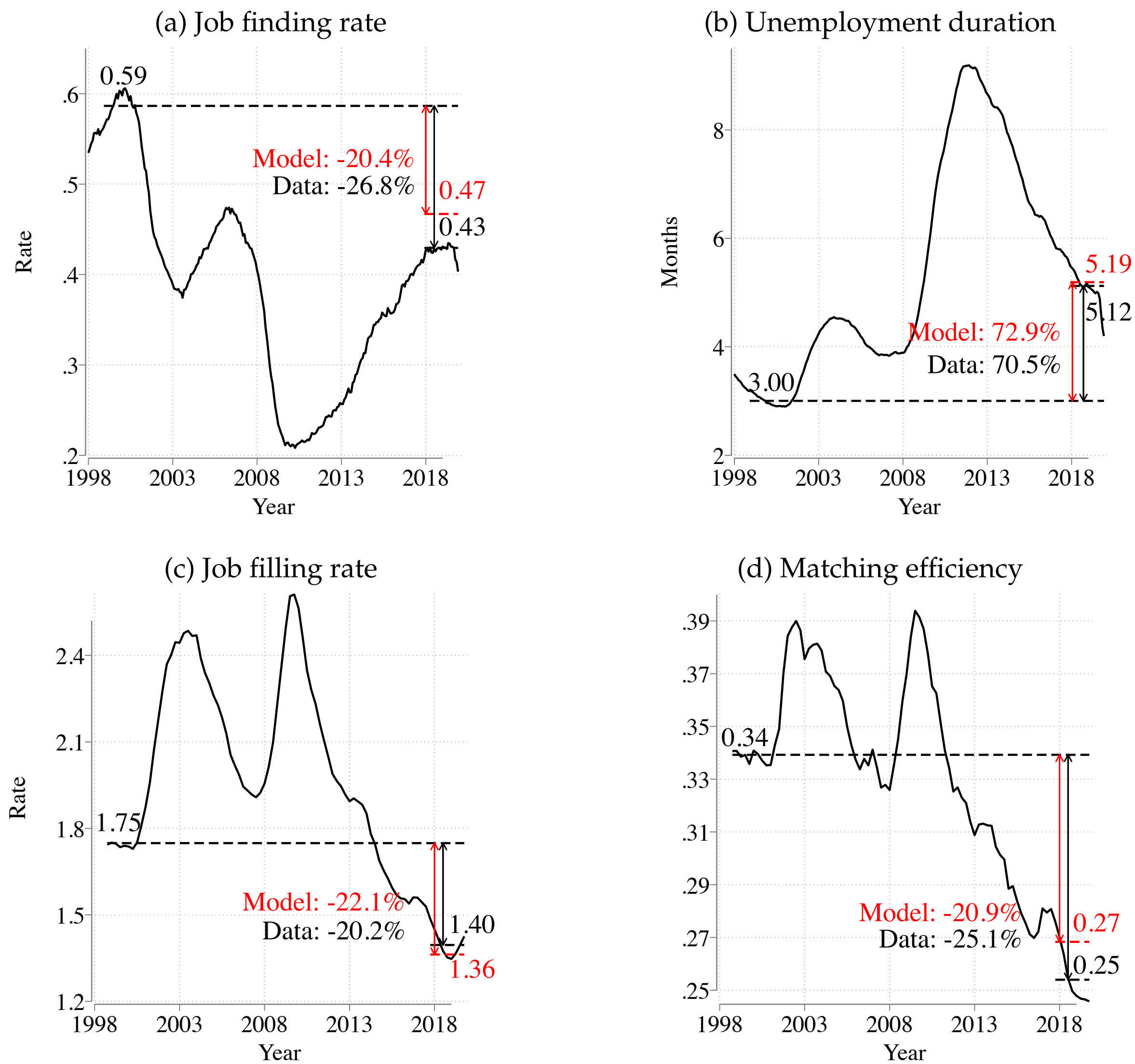


Figure 11: Comparison of aggregates across the two steady states

Notes: CPS, JOLTS and authors' calculations. Each panel plots a data series (solid black) alongside its averages in Dec. 1998 - Dec. 2000 and Dec. 2017 - Dec 2019 (dashed black lines for data, red for model). Percentage changes between the 1990s and 2010s averages in the data and model are shown.

discuss each moment in turn.

4.1.1 Job finding rate

The job finding rate declines by 20% in the model compared to 27% the data. In the model, two distinct forces contribute to this decline. First, as illustrated in Figure 10, the job finding rate of L-type workers falls substantially, as firms become more selective by raising the rejection probability of workers without a signal. Second, the composition of the unemployment pool shifts toward L-type workers, who have lower job finding rates.

4.1.2 Unemployment duration

The average duration of unemployment rises by 73% in the model, closely matching the 71% increase in the data. Notably, in the model and in the data, the rise in unemployment duration exceeds the decline in the job finding rate. This is a pattern that would be absent in a model with homogeneous workers. Intuitively, heterogeneity matters for the duration of unemployment because workers with a low job finding rate play a larger role for the average duration of unemployment than for the average job finding rate.[28] We discuss this intuition in Section 4.2 and formalize it in Appendix B.5, where we show that the growing dispersion in type-specific job finding rates plays a central role in generating the wedge.

4.1.3 Job filling rate

The job filling rate declines by 22% in the model and by 20% in the data. To understand this result, note that the rate at which vacancies are filled depends on two margins: the rate at which firms encounter applicants, and the probability that an encounter converts into a hire. This is captured by the following formula which we derive in Appendix B.4:

$$\text{job filling rate} = \overbrace{k(r+\delta)(1+\eta_s)(G(\Delta^f)\Delta^f)^{-1}}^{\text{encounter rate}} \cdot \left(\underbrace{\left(1-G(\Delta^f)\right)(1-R)}_{\text{admissions without signal}} + \underbrace{G(\Delta^f)}_{=p^{\text{sig}}} \underbrace{[a_L^{\text{app}}(1-s)+(1-a_L^{\text{app}})s]}_{\text{conversion prob. of signals}} \right). \quad (14)$$

After the shock, the encounter rate rises.[29] The intuition follows from free entry: as the applicant pool deteriorates, signals become less valuable to the firm (Δ^f falls). To keep

[28]For example, a homogeneous population of workers with an average job finding rate of 0.5 will observe an average unemployment duration of 2 months. However, if the same population is composed of two types with equal shares among the unemployed and type-specific job finding rates of $1-\varepsilon$ and ε respectively, then the aggregate job finding rate will still equal 0.5, but the average unemployment duration tends to infinity as $\varepsilon \to 0$, since the latter type reports infinite duration in the limiting steady state.

[29]The encounter rate is defined as $q(a_L^u p_L^{\text{app}} + (1-a_L^u)p_H^{\text{app}})$, like in equation (8). Here, we write it as an expression that substitutes in the free entry condition. See Appendix B.4 for derivations.

the value of a vacancy equal to the sunk entry cost, the encounter rate must therefore rise. The application shock thus exerts upward pressure on the job filling rate through this channel. However, two opposing forces dominate. First, firms reject a larger share of workers for whom they do not receive a signal. Second, the composition of the applicant pool shifts toward L-type workers, which reduces the probability of a favorable signal and resulting hire. Together, these forces produce a net decline in the job filling rate.

4.1.4 Matching efficiency

Matching efficiency declines by 21% in the model, compared to 25% in the data. Matching efficiency measures the number of hires the economy produces for a given stock of job seekers and vacancies. In our model, the application cost shock lowers the fraction of contacts that result in a hire, producing fewer hires per unit of search activity and thereby reducing matching efficiency.

4.2 Worker-level effects

We now explore and test the implications of the shock for individual worker types. First, we show the model predictions on type-specific applications rates, job finding rates, and unemployment durations. Then, we use micro-data from the National Longitudinal Survey of Youth (NLSY) to document that unemployment duration heterogeneity has increased over time, as predicted by the model.

4.2.1 Type-specific effects in the model

Table 2 reports model-predicted worker-level variables across the two steady states. Both worker types apply more frequently when application costs fall and the application rate roughly doubles, which is a targeted moment and therefore consistent with the data.

This increase in the application rate, however, does not translate into higher job finding rates, which are a product of the application rate and the acceptance probability. Even though application probabilities rise by roughly the same proportion, acceptance probabilities fall, and do so unequally. Because firms reject more workers without a signal, L-type workers especially struggle to find work: even though they send about twice as many applications, they experience a success rate per application that is four times lower, and their job finding rate falls by 38%. For H-type workers, the rise in the application rate roughly offsets the increase in rejections.

Variable	Level (1998-00)	Level (2017-19)	Change
Application rate (L)	4.41	8.73	+97.9%
Application rate (H)	9.23	19.3	+109.3%
Application rate	6.82	13.0	+90.8%
Job finding rate (L)	0.20	0.13	-37.9%
Job finding rate (H)	0.97	0.97	-0.1%
Job finding rate	0.59	0.47	-20.4%
Unemployment duration (L)	4.98	8.02	+61.0%
Unemployment duration (H)	1.03	1.03	+0.1%
Unemployment duration	3.00	5.19	+72.9%

Table 2: Model-predicted worker-side variables across steady states

Notes: Each row corresponds to a variable. "Level" columns correspond to model values of the corresponding variable in the indicated steady state. The "Change" column denotes the percentage change between both steady states.

The average unemployment duration of each worker type equals the inverse of the type-specific job finding rate. For L-type workers, it rises by 61%, from approximately 5 months to 8 months. For H-type workers, it remains at about 1 month.

In the aggregate, average unemployment duration rises by 73%, closely matching the 71% increase observed in the data. We now show that the divergence in job finding rates is important to understand this large increase. To do this, we derive the following expression for average unemployment duration in the model. We relegate derivations to Appendix B.5.

$$\text{unemployment duration} = (\text{job finding rate})^{-1} \cdot \left(1 + a_L^u(1 - a_L^u)\frac{(f_H - f_L)^2}{f_H f_L}\right). \quad (15)$$

The average duration is determined by two factors. First, the inverse aggregate job finding rate. Second, the heterogeneity in job finding rates which pushes the average duration of unemployment above the inverse job finding rate. The logic behind this result follows directly from Jensen's inequality: as heterogeneity increases, the average of the inverse job finding rates ($\mathbb{E}_i\left[1/f_i\right]$), becomes larger than the inverse of the average job finding rate ($1/\mathbb{E}_i\left[f_i\right]$). The inverse job finding rate increases by nearly 25% from the shock, while increasing heterogeneity contributes the remainder of the increase in unemployment duration.[30]

[30]In Appendix B.6, we generalize the set of assumptions under which unemployment duration can be decomposed as in Equation (15). Specifically, we show that even when job finding rates vary over the

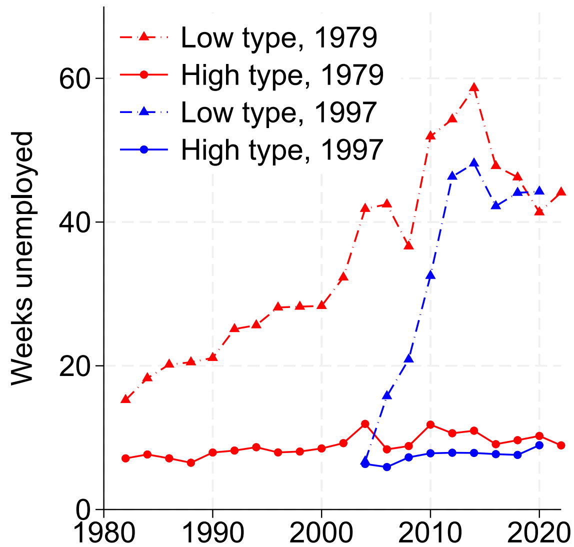


Figure 12: Type-specific heterogeneity in the NLSY

Notes: Group-specific average completed unemployment spell duration in the NLSY 1979 (red) and NLSY 1997 (blue). A spell is assigned to the year in which it is completed. “Low type” indicates spells belonging to workers with an above median lifetime average duration of completed unemployment spells. “High type” indicates spells belonging to workers with below median lifetime average duration of completed unemployment spells.

4.2.2 Type-specific unemployment duration in the data

The model predicts that the gap in unemployment duration across worker types has widened over time. We test this prediction using micro-data from the NLSY. The NLSY 1979 and 1997 cohorts follow about 6000 and 8000 individuals, respectively, from the time they first enter the labor market until the present day. The surveys record employment status at a weekly frequency, allowing us to measure the duration of unemployment spells at a granular level. We restrict the sample to individuals aged 25 and older and exclude the minority and military over-samples.

To proxy for worker types, we classify individuals into two groups based on their lifetime average duration of completed unemployment spells—above median workers are designated as “low type” and below-median workers as “high type”. Through the lens

unemployment spell, and the number of worker types is unlimited, such a decomposition can be derived. The generalized decomposition reveals that changes in duration dependence can also contribute for the increasing wedge between unemployment duration and the job finding rate. However, in Appendix B.7, we analyze micro-data from the National Longitudinal Survey of Youth (NLSY) and find no evidence of trends in duration dependence.

of the model, lifetime average spell duration is a natural proxy for a worker's underlying type, as it reflects persistent differences in job finding rates that may not be observable to the firm at the time of encounter.[31] Figure 12 plots for each NLSY cohort the average completed spell duration of each group over time. Two patterns stand out. First, the duration gap between the two groups widens over time in both cohorts: low-type workers experience a sharp increase in spell duration, while high-type workers experience only modest changes. Second, the gap opens almost immediately for the 1997 cohort, whereas it builds gradually for the 1979 cohort. This is consistent with the model which implies that the dispersion in unemployment duration across types has grown.

4.3 Firm-level effects

We now turn to the firm side of the market. We first describe the model's predictions for the resources firms devote to screening, and then discuss how firms' screening behavior affects the composition of hires. Finally, we test the model's predictions that hires have become more positively selected over time directly in the data.

4.3.1 Firm-level effects in the model

Screening expenditures. We begin by examining the model's implications for screening expenditures reported in the upper panel of Table 3. A decline in application costs affects the flow of screening resources through two opposing effects: firms receive more applications per unit of time, but purchase signals for a smaller fraction of them. Under our parametric assumptions, these two forces exactly offset, so that the flow cost of screening is constant per unit of time.[32] Total screening expenditures per vacancy therefore depend only on how long the vacancy remains open. Vacancies durations lengthen from the shock: vacancies receive nearly twice as many applications after the shock, but firms lower their acceptance rates more than proportionally, leading to a 28% increase in vacancy duration. Screening resources per hire therefore also rise by 28%. This implication is qualitatively consistent with the evidence presented in Section 2.3: the employment and vacancy shares of HR specialists, that is, workers whose task is to recruit, screen, and place candidates, have increased since the late 1990s (Figure 6).

[31] Whether heterogeneity is assessed in terms of job finding rates (f_i) or expected unemployment duration (f_i^{-1}) is irrelevant: both equation (15) and the more general decomposition in equation (30) are invariant to replacing f_i with f_i^{-1}.

[32] In Appendix B.9 we derive the screening expenditures per unit time in the model.

Variable	Level (1998-00)	Level (2017-19)	Change
Screening expenditures			
Screening expenditure per hire	0.38	0.49	+28.4%
Vacancy duration	0.57	0.73	+28.4%
Monthly applications per vacancy	20.3	37.9	+86.7%
Acceptance probability per vacancy	0.08	0.03	-61.8%
Composition of hires			
P(Applicant is L-type)	0.32	0.40	+23.7%
P(Signal | Hire)	0.47	0.60	+27.9%
P(Hire is L-type)	0.17	0.16	-7.02%
Duration rank of average hire	0.44	0.42	-4.07%

Table 3: Model-predicted firm-side variables across the two steady states

Notes: Each row corresponds to a variable. "Level" columns correspond to model values of the corresponding variable in the indicated steady state. The "Change" column denotes the percentage change between both steady states.

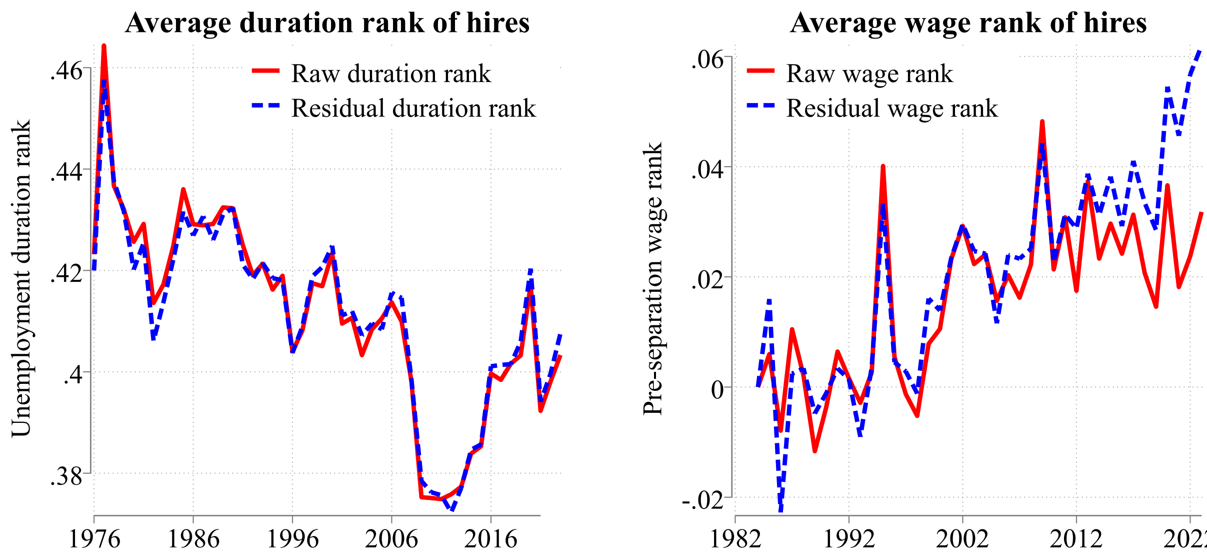
Composition of hires. The lower panel of Table 3 reports the model's predictions for the composition of hires across the two steady states. The share of L-type workers among applicants rises by 24% as their job finding rate decreases. Firms respond by rejecting a larger fraction of applicants for whom they have not received a favorable signal: the probability that a hire is preceded by a signal rises from 47% to 60%. Therefore, even though the applicant pool is more negatively selected after the shock, hires are more positively selected: the share of L-type workers among hires falls by 7%.

The compositional shift among hires toward H-type workers implies that hired workers increasingly come out of comparatively short unemployment spells. We measure this in the model using the percentile rank that each hired worker occupies in the duration distribution of the unemployed at the time of hire (see Appendix B.8 for a derivation). In the model, the average duration rank of hires falls from 0.44 to 0.42 across steady states.[33]

4.3.2 Firm-level effects in the data

Turning to the data, we now test the model's prediction that the composition of hires has become positively selected over time using micro-data from the Current Population Survey (CPS). The CPS is a rotating panel that interviews individuals for four consecutive

[33]Heterogeneity in job finding rates among the unemployed plays a central role in average duration rank: more dispersion in job finding rates lowers the rank because H-type workers, who exit unemployment quickly and therefore occupy lower duration ranks, account for a growing share of hires.



(a) Duration rank of hires in the CPS

(b) Wage rank of hires in the CPS

Figure 13: Rank measures of hires in the CPS

Notes: CPS and authors' calculations. The left panel plots the average unemployment duration rank of workers hired from unemployment. The raw (solid red) and residual (dashed blue) lines are the within-year percentile rank of elapsed unemployment duration among all unemployed. Residual ranks remove effect of occupation, age and education. The right panel plots the average pre-separation wage rank of workers hired from non-employment. The raw (solid red) and residual (dashed blue) lines are the within-year percentile rank of log hourly wages observed in the fourth interview of CPS respondents.

months, leaves them out for eight months, and then interviews them again for another four months. This structure allows us to track labor market transitions across interviews and to observe workers' wages and unemployment duration at different points in the panel. Our empirical approach builds on Mueller (2017), who uses the CPS panel structure to study compositional changes in the pool of unemployed over the business cycle.

Duration rank of hires. The model predicts that the average duration rank of hires declines as the composition of hires shifts toward H-type workers. To test this prediction, we construct an empirical measure of duration rank in the CPS as follows. For each unemployed worker, we observe their elapsed unemployment duration. We then compute each worker's within-year percentile rank relative to all other unemployed workers. The average duration rank of hires is the mean of this rank among workers who transition from unemployment in a given month to employment in the following month. To ensure that the measure captures duration composition rather than shifts in the occupation or demographics of the unemployed, we also construct a residual rank that removes the effects of occupation, education, and age.

The left panel of Figure 13 plots both measures of duration rank from 1976 to 2021. The average duration rank of hires declines steadily, from approximately 0.42 in the 1990s to 0.40 in the early 2020s, matching the model decline from 0.44 to 0.42 closely. The decline is present in both the raw and residualized series, indicating that it is not driven by changes in the observable characteristics of the unemployed.

Pre-separation wage rank of hires. We complement the duration-based analysis with an independent measure of compositional change, following the approach of Mueller (2017), who uses pre-displacement wages of the unemployed as a summary indicator of worker quality. Rather than measuring where a newly hired worker ranks in the duration distribution, this exercise asks where they rank in the wage distribution prior to becoming unemployed.

We use wages observed in the Outgoing Rotation Group (ORG, interview month 4) as a measure of pre-separation earnings and track each worker's labor force status in interview months 5 through 8. A worker is classified as hired from non-employment if they are employed in interview month 8 but were unemployed or out of the labor force in at least one of interview months 5 through 7.[34] Within each year, we compute the percentile rank of each worker's log hourly wage among all ORG-eligible workers and residualize with respect to occupation and education.

The right panel of Figure 13 plots the average pre-separation wage rank of hires from 1983 to 2023. Both series rise steadily from the 1990s, suggesting that hires have increasingly shifted toward workers with higher pre-separation wages. This pattern provides another piece of evidence that the composition of hires has shifted toward more productive workers, as implied by the model.[35]

Variable	Level (1998-00)	Level (2017-19)	Change
Worker welfare (average)	1803.9	1801.8	-0.1%
Worker welfare (L)	207.5	196.8	-5.2%
Worker welfare (H)	2176.8	2176.8	0%
Wage (average)	7.82	7.92	+1.2%
Wage (L)	0.93	0.91	-2.5%
Wage (H)	9.24	9.24	0%

Table 4: Welfare and wages across the two steady states

Notes: Each row corresponds to a variable. "Level" columns correspond to model values of the corresponding variable in the indicated steady state. The "Change" column denotes the percentage change between both steady states.

4.4 Welfare

We conclude the analysis of the application cost shock by examining its welfare implications. We define worker welfare as the population-weighted average of type-specific expected lifetime utility:

$$\mathcal{W} = \alpha^L_{\text{pop}}\mathcal{W}(L) + (1-\alpha^L_{\text{pop}})\mathcal{W}(H), \qquad \mathcal{W}(i) = \frac{\delta}{\delta + f_i}U_i + \frac{f_i}{\delta + f_i}W_i, \tag{16}$$

where $\mathcal{W}(i)$ is the average welfare of type-i workers, weighting the values of unemployment and employment by the steady-state population shares in each state.

Table 4 reports the welfare changes between the 1990s and 2010s steady states. Average welfare declines by 0.1%. This modest aggregate figure, however, masks substantial heterogeneity across types: welfare is essentially unchanged for H-type workers but falls by over 5% for L-type workers.[36] The loss for L-type workers stems from two reinforcing channels. First, their job finding rates plummet, reducing their labor market access. Second, their lower job finding rates also erode their bargaining position, lowering their

[34]Hourly wages are measured directly for workers paid by the hour and computed as weekly earnings divided by usual hours for salaried workers. Wages are trimmed at the 1st and 99th percentiles and deflated by the CPI. We combine unemployed and non-participants in our sample to increase its size.

[35]A caveat of the wage rank exercise is that conditioning on employment in interview month 4 introduces selection that is difficult to control for. An ideal version of this experiment would require measuring pre-displacement wages for every worker, not only those observed as employed approximately a year ago. The duration rank measure avoids this issue, as elapsed duration is observed for every unemployed worker regardless of employment history.

[36]The welfare level of L-type workers is an order of magnitude smaller in our calibration compared to that of H-type workers, which reduces their overall impact on aggregate welfare.

wages by 2.5%. Overall, wages increase by 1.2% as the composition of employed workers shifts toward H-types.

Diminishing search frictions thus generate a Pareto deterioration in our model: lower application costs endogenously trigger more intensive screening, which congests the labor market enough to make every participant at least weakly worse off. Workers who are at the receiving end of increased selectivity by firms experience substantial losses. For other workers, the consequences are more benign.[37]

We conclude by noting that these results can be viewed as a lower bound on the net welfare effects of declining search frictions. Our model features screening on worker characteristics rather than on match quality. If screening also improved the quality of worker-firm matches, as in Martellini and Menzio (2020) where declining frictions contribute to productivity growth, an additional welfare-improving force would counteract the congestion effects documented here.

5 The impact of artificial intelligence on matching markets

In the previous sections, we introduced, quantified, and validated a model to assess the positive and normative consequences of changes in job search technology. In this final section, we now apply this model to understand the effects of a very recent technological innovation: the rise of large language models (LLMs) and its use in job search. We begin by studying the impacts of changes in job search technology beyond the ease of applying (Section 5.1) before mapping the LLM shock to changes in those parameters (Section 5.2) and presenting results (Section 5.3)

5.1 Equivalence of technology shocks

To apply our model in the context of AI, we begin by extending our analysis to consider shocks to parameters that may characterize the job search or screening technology beyond γ_a. Specifically, in addition to the ease of applying (γ_a), we consider shocks to the ease of screening (γ_s) and the quality of the signal (s).

It turns out that all three types of shocks are equivalent in our model. The underlying reason is that, while the indifference curve is independent of (γ_a, γ_s, s) the screening curve

[37] Auster et al. (2025) obtain a related result in a directed search model with adverse selection: lowering search frictions inhibit firms' ability to attract their preferred types, which can lower welfare. Our analysis provides a quantitative counterpart: the losses are modest in the aggregate but substantial for the workers who bear the brunt of increased firm selectivity.

can be written as a function of a sufficient statistic $X(\gamma_a, \gamma_s, s)$, as Proposition 1 formalizes:

Proposition 1 (Sufficient statistic for screening)**.** *The screening curve* (SC) *can be written as*

$$f_H = M_1(\vec{f}) \cdot f_L + X(\gamma_a, \gamma_s, s)^{\frac{1}{1+\eta_a-\omega}} \cdot M_2(\vec{f}).$$

where

$$X(\gamma_a, \gamma_s, s) = \gamma_a \; \gamma_s^{1+\eta_a} \; (2s-1)^{(1+\eta_s)(1+\eta_a)}.$$

and $M_1(\vec{f})$, $M_2(\vec{f})$ do not depend on (γ_a, γ_s, s).

Proof. See Appendix B.10. □

This proposition implies that, for any shock to γ_a, there are corresponding shocks to γ_s and s that introduce the same movement in all important aggregate variables. We formalize this insight in Proposition 2:

Proposition 2 (Equivalence of shocks)**.** *There exists $\varepsilon > 0$ such that, for any pair (γ_a, γ_a') with associated mixed strategy equilibria (E, E') that feature $(f_L', f_H') \in B_\varepsilon((f_L, f_H))$, the value*

$$s' = \frac{1}{2}\left[1 + (2s-1)\left(\frac{\gamma_a'}{\gamma_a}\right)^{\frac{1}{(1+\eta_s)(1+\eta_a)}}\right]$$

and the value

$$\gamma_s' = \gamma_s \left(\frac{\gamma_a'}{\gamma_a}\right)^{\frac{1}{1+\eta_a}}$$

are associated with mixed strategy equilibria E_s' and E_{γ_s}'. The three equilibria $(E', E_s', E_{\gamma_s}')$ feature the same type-specific job finding rates, distribution of unemployment duration, type-specific unemployment stocks, vacancy stocks, job filling rates, and match efficiency. The equilibria E_s' and E_{γ_s}' feature identical type-specific application rates.

Proof. See Appendix B.10. □

Proposition 2 states that shocks to γ_a, γ_s and s induce the same changes in aggregate variables (though they may have different implications for application rates). Intuitively, this result arises because positive changes to all three variables shift the screening curve upwards, which makes firms optimally increase their selectivity and therefore induces an endogenous increase in R and an overall equilibrium shift to the left along the indifference curve.

Assessing the impact of LLM technology on matching markets thus requires assessing the relative strength of these competing channels. We now turn to this quantification.

5.2 Quantifying the AI shock

To quantify the impact of AI, we make three assumptions pertaining to each of the three parameters γ_s, γ_a, and s.

First, we assume that AI leaves γ_s unchanged.[38] We do this primarily because reliable estimates on how much the availability of AI has reduced the cost of screening remain scarce. However, we note that by Proposition 2, any decrease in the cost of screening would push in the same direction as a larger application cost shock.

Next, we quantify the shock to γ_a. For this, we rely on reporting from Greenhouse, a large job search platform, that the number of applications per job posting has approximately doubled since the release of ChatGPT in November 2022.[39] Motivated by this report, we assume the change in γ_a necessary to double the number of applications per position: $\gamma_a^{\text{AI}}/\gamma_a^{\text{noAI}} = 2.56$.

Lastly, we quantify the change in s. An emerging literature (Galdin and Silbert, 2025; Cowgill et al., 2024) argues that the availability of LLMs for job search has made it more difficult for employers to distinguish suitable candidates from less suitable ones. The model permits us to evaluate the general equilibrium effects of such a mechanism. In quantifying s, we rely on experimental evidence by Cowgill et al. (2024) and relegate details to Appendix B.11. We find that, in the parlance of our model, the signal accuracy $2s-1$ declines by 6.5%.

5.3 Effects of LLMs on matching markets

To evaluate the combined results of the two shocks we impose, we begin by evaluating their effects graphically. Figure 14 is a a version of Figure 10, updated with the effects unfolding from easier applications (γ_a^{AI}) and the combined effects of both shocks ($\gamma_a^{\text{AI}}, s^{\text{AI}}$). Because AI makes applying to jobs cheaper, it reinforces the effects of online job search.

[38] In ongoing work, we seek to relax this assumption.

[39] See Business Insider, "How Tech Broke the Job Market", accessed March 19, 2026.

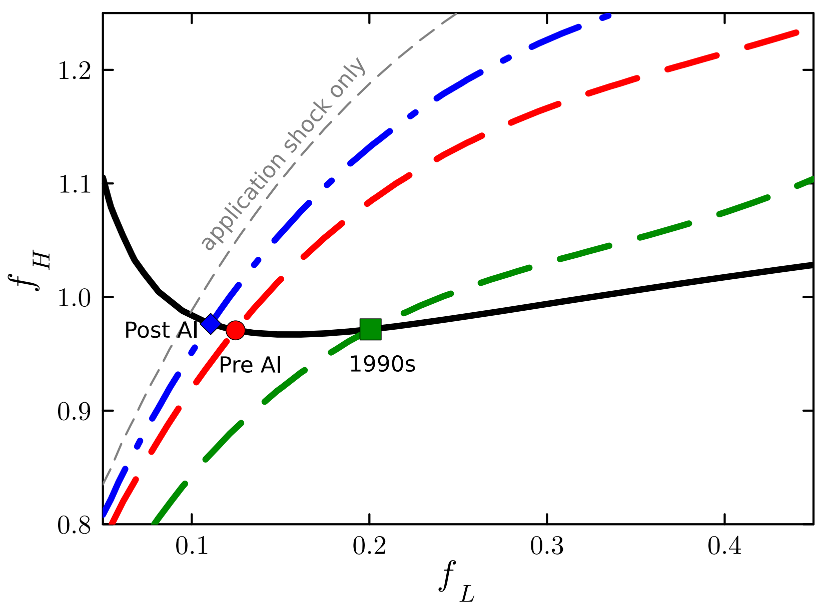


Figure 14: Shifts in the screening curve resulting from AI

Notes: The black, green, and red lines are the same as in Figure 10. We label the 2019 steady state "Pre AI". The gray dashed line displays the screening curve (SC) with parameters as in Table 1 except a higher value of γ_a equal to $\gamma_a^{\text{AI}} > \gamma_a^{\text{noAI}} = \gamma_a'$. The blue dash-dotted line displays the screening curve (SC) with parameters as in Table 1 except a higher value of γ_a equal to $\gamma_a^{\text{AI}} > \gamma_a^{\text{noAI}} = \gamma_a'$ and a lower value of s equal to $s^{\text{AI}} < s^{\text{noAI}} = s$. We quantify γ_a^{AI} and s^{AI} as described in the main text.

Variable	Effect of $\gamma_a^{\mathbf{AI}}$	Effect of $s^{\mathbf{AI}}$	Model change (2019–2025)	Data change (2019–2025)
Job finding rate	-10.6%	+5.6%	-5.6%	-12.3%
Unemployment duration	+31.9%	-12.7%	+15.2%	+5.1%
Job filling rate	-1.9%	+0.1%	-1.8%	-8.3%
Matching efficiency	-8.1%	+3.9%	-4.5%	-3.6%
Worker welfare (avg)	-0.0%	-0.0%	-0.0%	—
Worker welfare (L)	-2.8%	+1.4%	-1.4%	—
Worker welfare (H)	+0.1%	-0.0%	+0.0%	—

Table 5: Labor market and welfare impacts of the LLM technology

Notes: Each row corresponds to a model variable. The "Effect of γ_a^{AI}" column indicates the percentage change of the variable between the 2019 steady state and a steady state in which $\gamma_a = \gamma_a^{\text{AI}}$ but all other variables are as in the 2019 steady state. The "Effect of s^{AI}" column indicates the residual percentage change of the variable between the 2019 steady state and a steady state in which $\gamma_a = \gamma_a^{\text{AI}}$ and $s = s^{\text{AI}}$. Changes in the "Data" column refer to changes in averages between (December 2017 - December 2019) and (October 2024 - December 2025).

The screening curve shifts up and the equilibrium shifts further to the left, with negative effects on the job finding rates of L-type workers but small (mildly positive) effects on the job finding rates of H-type workers. The negative shock to the quality of the signal offsets some, but not all of the effects of cheaper applications. The final equilibrium lies to the left of the initial one.

In Table 5, we explore the effects of both shocks on aggregate variables and welfare. The two types of shock work in opposite directions. We find that the shock matches, at least qualitatively, the direction of trends aggregate labor market variables, comparing an average of December 2024 to December 2025 with the pre-pandemic steady state (December 2017 to December 2019). We find that changes in job search technology fueled by AI explain about half of the decline in the job finding rate, overexplain the rise in unemployment duration, explain a quarter of the decline in the job filling rate, and the full decline in match efficiency. Welfare effects are approximately neutral overall but the welfare of L type workers decreases by 1.4%.

6 Conclusion

Improvements in job search technology can account for long-term labor market trends observed since the 1990s, in particular the lengthening of unemployment spells and the fall in matching efficiency. Our framework helps understand the role of internet job search in amplifying search frictions and lets us interpret recent changes brought about by AI. By continuing to make applications easier and screening less costly, AI has the potential to further intensify congestion in hiring.

The broader lesson of our paper is that lower search frictions brought about by recent improvements in search technology may, paradoxically, worsen congestion frictions. In markets where firms select their preferred type of worker among a large pool of applicants, technologies that facilitate contacts can generate more selectivity, fewer hires, and more unequal job finding prospects. Identifying policies and market designs that preserve the benefits of easier search without amplifying these congestion forces is an important direction for future work.

References

Ahn, H. J. (2023). The role of observed and unobserved heterogeneity in the duration of unemployment. *Journal of Applied Econometrics 38*(1), 3–23.

Auster, S., P. Gottardi, and R. Wolthoff (2025). Simultaneous search and adverse selection. *Review of Economic Studies*, rdaf014.

Bagga, S., L. F. Mann, A. Şahin, and G. L. Violante (2025). Job amenity shocks and labor reallocation. Technical report, National Bureau of Economic Research.

Baley, I., A. Figueiredo, and R. Ulbricht (2022). Mismatch cycles. *Journal of Political Economy 130*(11), 2943–2984.

Bhuller, M., D. Ferraro, A. R. Kostøl, and T. C. Vigtel (2023). The internet, search frictions and aggregate unemployment. Technical report, National Bureau of Economic Research.

Birinci, S., K. See, and S. L. Wee (2025). Job applications and labour market flows. *Review of Economic Studies 92*(3), 1438–1496.

Blanchard, O. J., P. Diamond, R. E. Hall, and K. Murphy (1990). The cyclical behavior of the gross flows of us workers. *Brookings papers on economic activity 1990*(2), 85–155.

Bradley, J. (2025). Worker-firm screening and the business cycle. *Review of Economic Dynamics 57*, 101272.

Bradley, J. and L. Mann (2024). Learning about labor markets. *Journal of Monetary Economics 148*, 103612.

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

Cowgill, B., P. Hernandez-Lagos, and N. L. Wright (2024). Does ai cheapen talk? theory and evidence from global entrepreneurship and hiring. *Theory and Evidence From Global Entrepreneurship and Hiring (July 26, 2024). Columbia Business School Research Paper* (4702114).

Cui, J., G. Dias, and J. Ye (2025). Signaling in the age of ai: Evidence from cover letters. *arXiv preprint arXiv:2509.25054*.

Dhillon, P., Y. Gao, T. Lu, and Y. Zhang (2025). From signal to noise: How widespread llm usage transforms evaluator behavior in credit screening. *Available at SSRN 5616650*.

Diamond, P. A. and A. Şahin (2016). Disaggregating the matching function. Technical report, National Bureau of Economic Research.

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

Faberman, R. J. and G. Menzio (2018). Evidence on the relationship between recruiting and the starting wage. *Labour Economics 50*, 67–79.

Feller, W. (1991). *An introduction to probability theory and its applications, Volume 2*, Volume 2. John Wiley & Sons.

Fujita, S., G. Moscarini, and F. Postel-Vinay (2024). Measuring employer-to-employer reallocation. *American Economic Journal: Macroeconomics 16*(3), 1–51.

Galdin, A. and J. Silbert (2025). Making talk cheap: Generative ai and labor market signaling. *arXiv preprint arXiv:2511.08785*.

Gans, J. S. (2024). How will generative ai impact communication? *Economics Letters 242*, 111872.

Hornstein, A., P. Krusell, and G. L. Violante (2011). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review 101*(7), 2873–2898.

Jarosch, G. and L. Pilossoph (2019). Statistical discrimination and duration dependence in the job finding rate. *The Review of Economic Studies 86*(4), 1631–1665.

Kroft, K., F. Lange, M. J. Notowidigdo, and L. F. Katz (2016). Long-term unemployment and the great recession: the role of composition, duration dependence, and nonparticipation. *Journal of Labor Economics 34*(S1), S7–S54.

Landais, C., P. Michaillat, and E. Saez (2018). A macroeconomic approach to optimal unemployment insurance: Theory. *American Economic Journal: Economic Policy 10*(2), 152–181.

Mann, L. (2026). Labor market selection and the dynamics of a recovery.

Martellini, P. and G. Menzio (2020). Declining search frictions, unemployment, and growth. *Journal of Political Economy 128*(12), 4387–4437.

Mueller, A. I. (2017). Separations, sorting, and cyclical unemployment. *American Economic Review 107*(7), 2081–2107.

Mueller, A. I., J. Spinnewijn, and G. Topa (2021). Job seekers perceptions and employment prospects: Heterogeneity, duration dependence, and bias. *American Economic Review 111*(1), 324–363.

Mukoyama, T. and A. Şahin (2009). Why did the average duration of unemployment become so much longer? *Journal of Monetary Economics 56*(2), 200–209.

Petrongolo, B. and C. A. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic literature 39*(2), 390–431.

Petrosky-Nadeau, N. and L. Zhang (2021). Unemployment crises. *Journal of Monetary Economics 117*, 335–353.

Pries, M. J. and R. Rogerson (2022). Declining worker turnover: The role of short-duration employment spells. *American Economic Journal: Macroeconomics 14*(1), 260–300.

Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics 15*(2), 127–148.

Stigler, G. J. (1961). The economics of information. *Journal of political economy 69*(3), 213–225.

Stigler, G. J. (1962). Information in the labor market. *Journal of political economy 70*(5, Part 2), 94–105.

Wiles, E. and J. J. Horton (2025). Generative ai and labor market matching efficiency. *Available at SSRN 5187344*.

Wolthoff, R. (2018). Applications and interviews: Firms recruiting decisions in a frictional labour market. *The Review of Economic Studies 85*(2), 1314–1351.

A Empirical appendix

A.1 Long-term vs. short-term unemployed

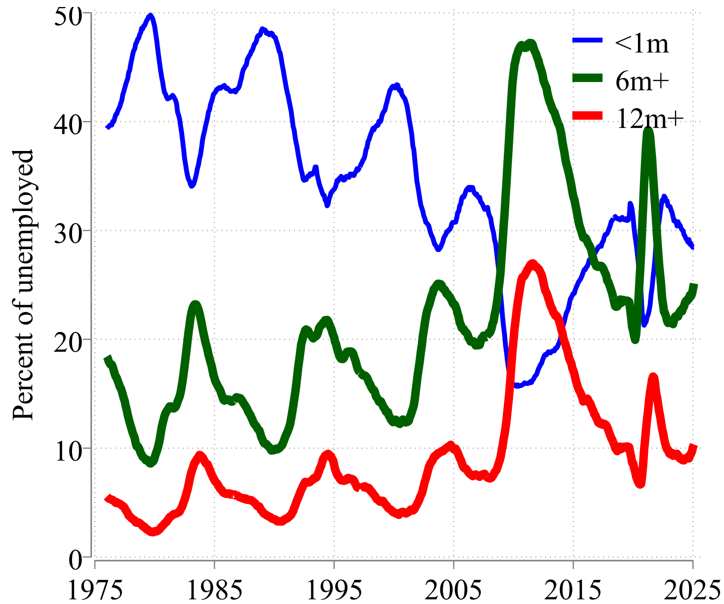


Figure A.1: Long-term vs. short-term unemployed

Notes: CPS and authors' calculations. Duration groups are defined by elapsed weeks of unemployment reported in the CPS, converted to months. 12 month centered moving averages are reported.

Figure A.1 plots the fraction of unemployed with elapsed duration less than one month, more than six months, and more than twelve months in unemployment. The short duration share trends downward over the sample, falling from nearly 45% in the late 1970s to about 30% by the end of the sample. The long-duration shares offset this decrease by moving in the opposite direction. This Figure illustrates that the rise in average unemployment duration documented in Figure 1 is driven by a thickening of the right tail of the duration distribution. We focus on average duration throughout the paper since it admits a clean analytical decomposition (see equation (30)), but note that the rise in "long-term unemployment" is another expression of the same trend.

A.2 Demographics-adjusted duration and the role of UI

In this appendix we test two alternative explanations that could potentially account for the rise in mean unemployment duration. First, over the last 50 years, the composition

(a) Demographics-adjusted duration (b) UI replacement rate

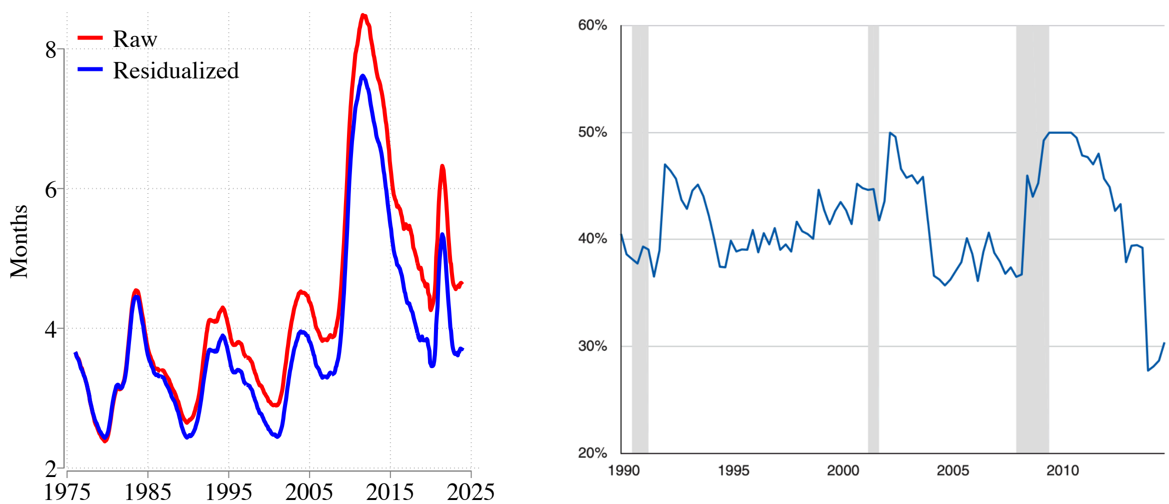


Figure A.2: Demographics-adjusted duration and UI replacement rate

Notes: Left panel: CPS and authors' calculations. The raw series (red) plots average months unemployed. The residualized series (blue) plots average months unemployed after removing the component explained by age, age squared, gender, race, and education. 12 month centered moving averages are reported. The residualized series is normalized to match the level of raw unemployment duration at the start of the sample. Right panel: Effective UI replacement rate from Landais, Michaillat, and Saez (2018), Figure 2.

of labor force has shifted toward older and more educated workers. If these groups experience higher than average unemployment duration, then these compositional changes could drive the aggregate trend in duration. To control for this effect, we regress individual elapsed unemployment duration on education and demographics such as age, gender and race and plot the residualized unemployment duration in the left panel of Figure A.2. We normalize the residualized series to match the level of raw unemployment duration at the start of the sample. Both the series track each other closely and the upward trend persists even after removing the contribution of demographics.

Another potential explanation is that unemployment insurance (UI) has become more generous over time, which could have prolonged job search. Landais, Michaillat, and Saez (2018) construct the effective UI replacement rate which captures the level and duration of benefits, and we report their Figure 2 in the right panel of Figure A.2. The figure shows that the effective UI replacement rate fluctuates countercyclically around a mean of 42% but shows no secular trend, ruling out increasing UI generosity as a potential driver for the long-run increase in unemployment duration.

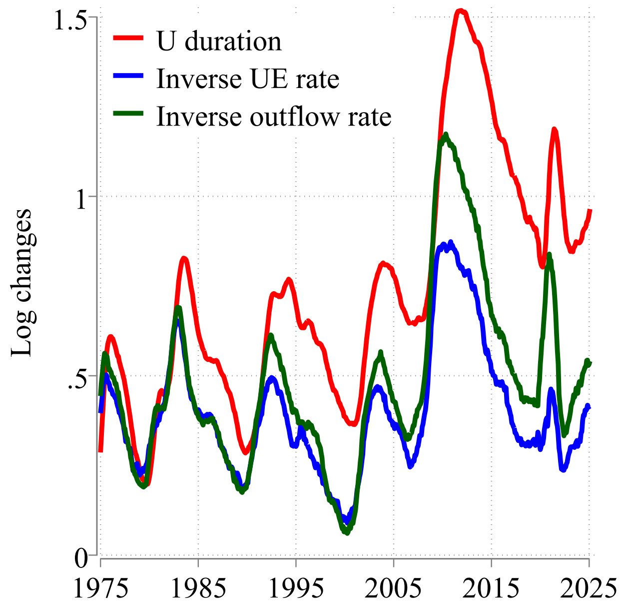


Figure A.3: Unemployment duration and inverse job finding rate

Notes: CPS and authors' calculations. Unemployment duration is average months unemployed. The inverse UE and outflow rates are reciprocals of the UE transition and outflow probabilities converted to hazard rates. The figure reports 12-month centered moving averages and all series are expressed as log deviations.

A.3 Job finding rate and unemployment duration

Figure A.3 reframes the aggregate series from Figures 1 and 2 plotting unemployment duration alongside the inverse hazard rates, with all three series expressed as log deviations. The three series move almost in lockstep from the beginning of the sample through the mid-1980s. Following the recovery from the "twin recessions" of the 1980s, a gap emerges between the unemployment duration and inverse job finding rate. The gap persists through the 2000s and starts widening during the 2010s expansion, continuing through the end of the sample.

The rising gap indicates that average unemployment duration has risen by more than can be explained by the fall in job finding rates alone. This signals a departure from the constant-hazard, homogeneous-worker benchmark. As we discuss in Section 4.2, two potential mechanisms can generate a wedge. First, heterogeneity in hazard rates. Second, a change in duration dependence. We demonstrate in Appendix B.7 that changes in duration dependence play only a limited role in explaining the rising wedge.

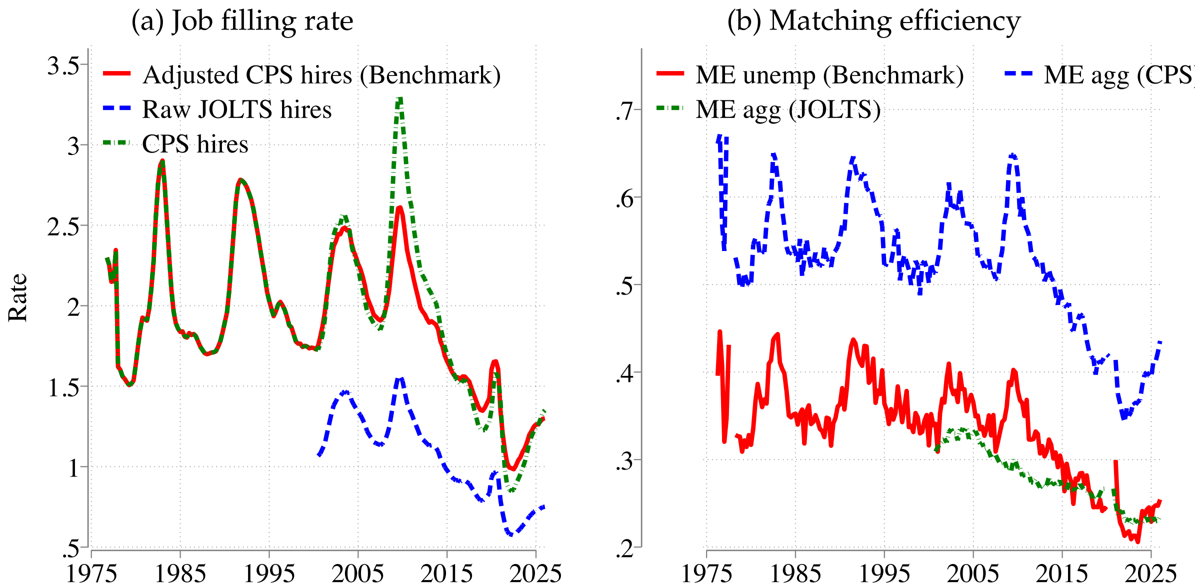


Figure A.4: Alternative measures of job filling rate and matching efficiency

Notes: CPS, JOLTS and authors' calculations. Left panel: The adjusted CPS hires series (benchmark, red) computes job filling rate by rescaling CPS hires to preserve CPS levels while matching the JOLTS cyclical variation following the methodology of Bagga, Mann, Şahin, and Violante (2025). The raw JOLTS series (blue dashed) uses JOLTS hires divided by JOLTS vacancies. The CPS series (green dash-dot) uses unadjusted CPS hires divided by vacancies. Right panel: ME unemp (benchmark) is defined per Equation 1 using only unemployed job seekers and hires from unemployment in the CPS. ME agg (CPS) and ME agg (JOLTS) additionally include the search effort of employed workers with relative search intensity $s = 0.4$ and all hires including those from employment and non-participation.

A.4 Alternative measures of matching efficiency and job filling rate

In the main text, we compute hires from the CPS as a sum of hires from unemployment, hires from non-participation, and hires from employment. The first two are constructed from CPS gross flows as the product of transition probabilities and relevant population stocks. Hires from employment are constructed by multiplying the employment stock with the Employer-to-Employer transition probability constructed from Fujita, Moscarini, and Postel-Vinay (2024) post 1995, and Diamond and Şahin (2016) and Blanchard, Diamond, Hall, and Murphy (1990) prior to 1995. After constructing total hires in the CPS, we adjust the hires following the methodology of Bagga, Mann, Şahin, and Violante (2025). Over the period when the JOLTS and CPS overlap from 2001 onward, we rescale CPS hires to preserve the level of hires while matching JOLTS fluctuations in hires. We combine the adjusted hires series with the unadjusted CPS hires series before 2001 when JOLTS was not available. This yields a series that has a longer frequency, is consistent with levels of the CPS data, and shares higher-frequency movements with the JOLTS data.

Figure A.4 shows that our qualitative results do not hinge on this measurement choice. The left panel plots three measures of job filling rate: our benchmark measure using adjusted CPS hires divided by vacancies (red solid), unadjusted CPS hires divided by vacancies (green dash-dot), and JOLTS hires divided by JOLTS vacancies (blue dash). The unadjusted and adjusted CPS-based series track each other closely throughout the sample, while the JOLTS-based series lies below both, but exhibits the same cyclical properties and the same post-2000 downward trend.

The right panel of Figure A.4 reports three measures of matching efficiency. Our benchmark measure (red solid) uses unemployed job seekers (and vacancies), and hires from unemployment as, respectively, inputs and output in the matching function. The other two series additionally account for employed job seekers with a search intensity of $s = 0.4$, and hires from employment and non-participation as inputs and outputs in the matching function. The CPS series (blue dash) uses unadjusted hires from CPS, while the JOLTS series (green dash-dot) uses JOLTS hires. Again, all three series share the same qualitative properties and demostrate a roughly flat trend through the 1990s followed by a sustained decline beginning in the early 2000s. The choice between CPS-based and JOLTS-based hires, and between unemployed-only and aggregate matching efficiency, thus affects levels but not the secular trend that is the focus of our analysis.

A.5 Unemployment rate as alternative measure of the business cycle

In the main text, the right panels of Figures 2-4 plot aggregate variables against vacancy-by-unemployment as a measure of the state of the cycle. In Figure A.5, we instead use the unemployment rate instead of labor market tightness on the horizontal axis. Similar patterns emerge. We observe a structural upward shift in unemployment duration over time, accompanied by a structural downward shift in the job finding probability, job filling rate, and matching efficiency. Therefore, the shifts documented in the main text are robust to the choice of cyclical indicator.

A.6 Covariation of aggregate variables with labor market tightness

In the main text, the right panels of Figures 2-4 restrict attention to expansionary episodes. In Figure A.6 we reproduce the same plots but include both expansionary and recessionary episodes. The downward shifts in the job finding probability, job filling rate, and

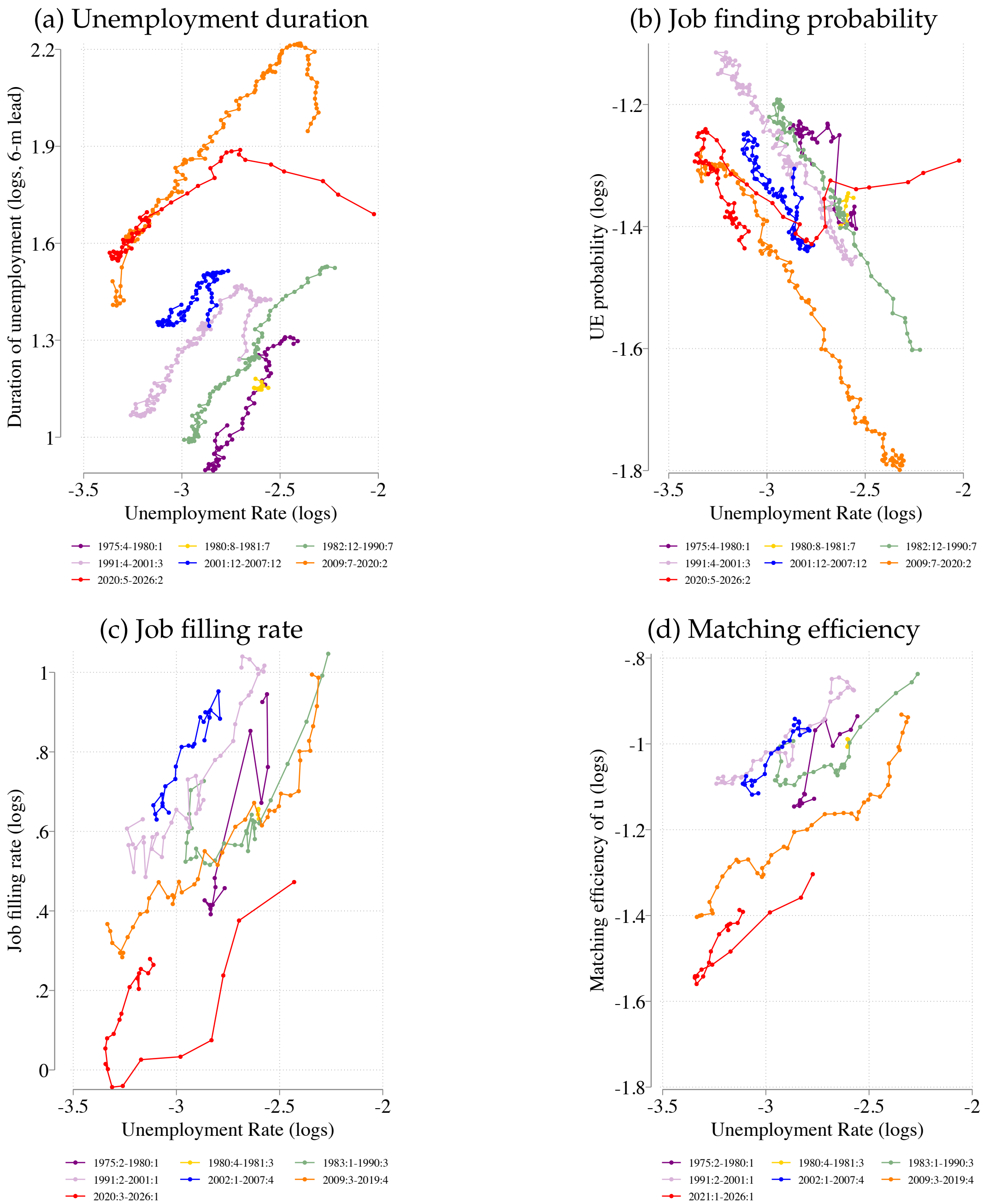


Figure A.5: Covariation of aggregate labor market variables with unemployment rate

Notes: CPS, JOLTS and authors' calculations. Each panel plots a labor market variable (in logs) against the log unemployment rate, with each recovery traced separately.

matching efficiency, and the upward shift in unemployment duration are apparent even when we consider all business cycles between 1975-2026.

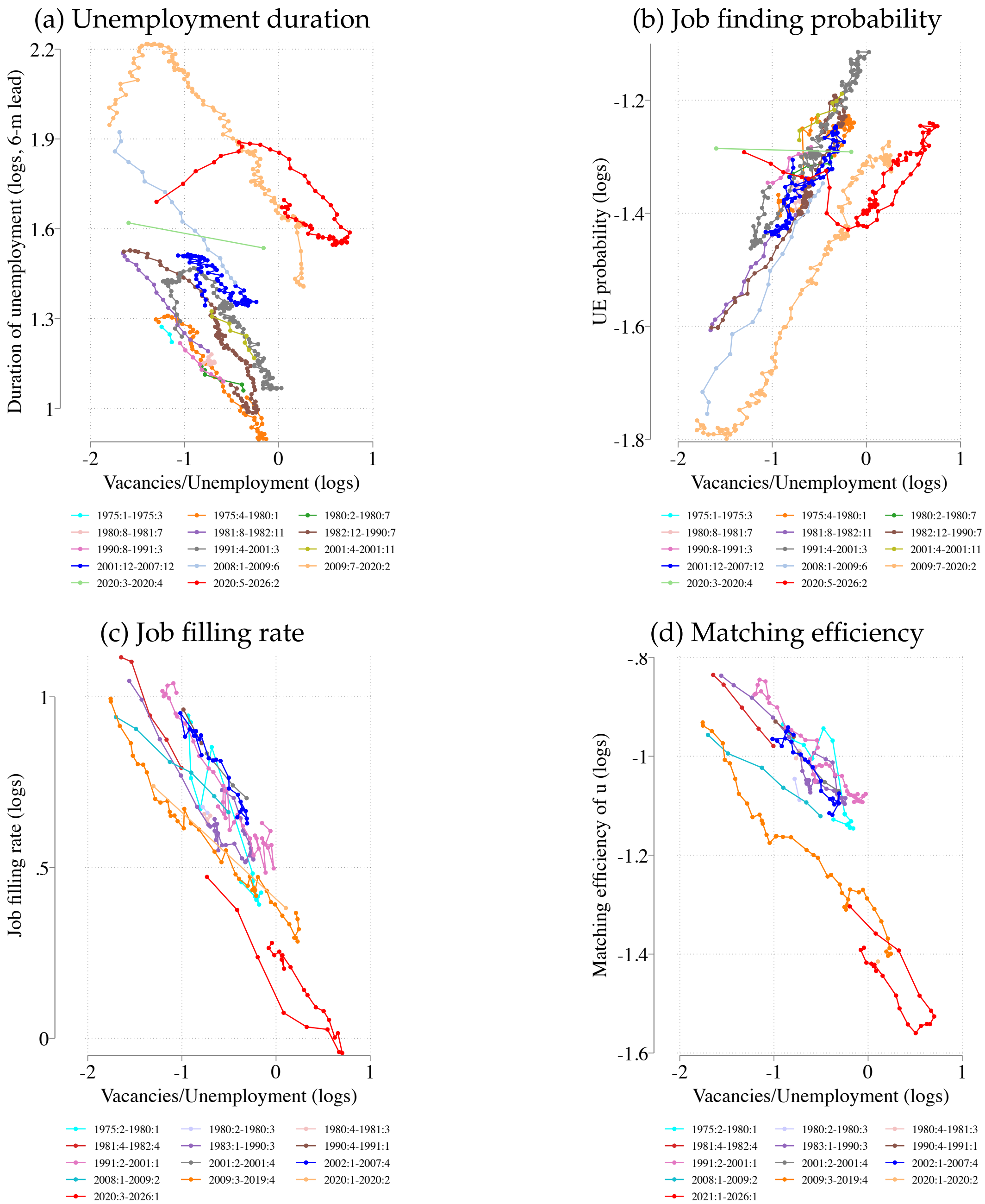


Figure A.6: Covariation of aggregate variables with labor market tightness

Notes: CPS, JOLTS and authors' calculations. Each panel plots a labor market variable (in logs) against the log vacancy-to-unemployment ratio, with each expansionary and recessionary episode traced separately.

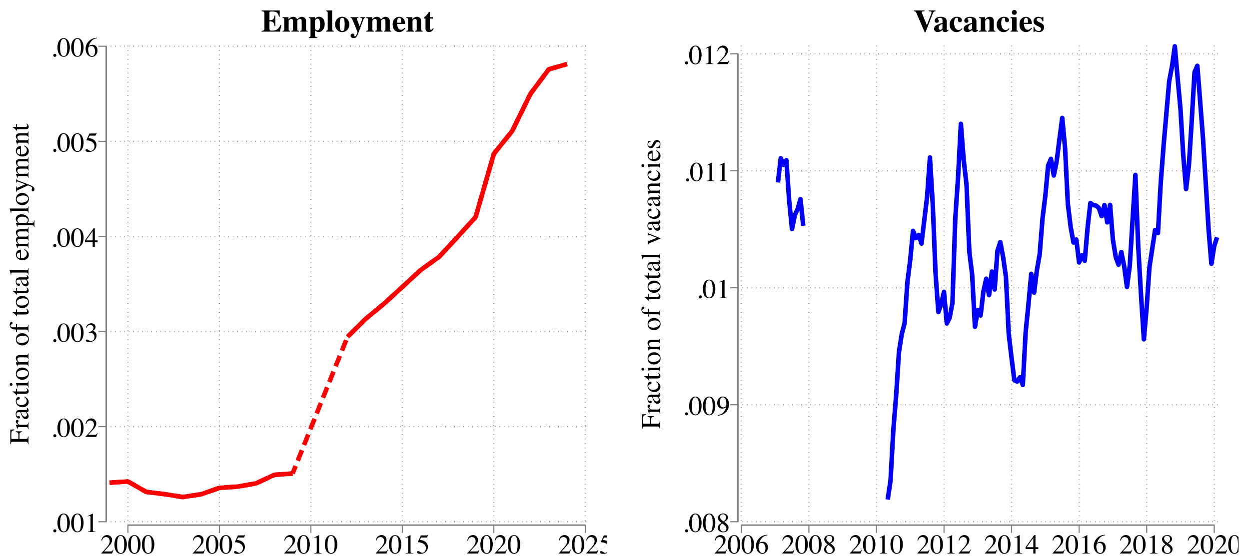


Figure A.7: HR specialists as a fraction of all occupations

Notes: Occupational Employment and Wage Statistics, Lightcast and authors' calculations. The left panel plots employment of HR specialists (SOC 13-1071) as a fraction of total employment. The right panel plots the share of online job postings advertising HR specialist positions as a fraction of all online job postings.

A.7 HR specialists relative to all occupations

In the main text, we showed the employment and vacancy share of HR specialists as a fraction of business operation occupations. The rationale behind this was that over the last 30 years, employment has become concentrated within larger firms, which may mechanically drive the demand for operational staff. An increase in the overall employment share of HR specialists, therefore, may simply reflect changes in the organizational structure of firms over this time period. HR specialist employment as a fraction of business operations occupations is not subject to this critique.

In Figure A.7, we show that the share of HR specialists as a fraction of total employment exhibits the same qualitative pattern, with an even more substantial increase observed since the 1990s. The left panel shows that HR specialist employment roughly tripled as a share of total employment between 2000 and 2024. The right panel shows a similar upward trend in the online vacancy share of HR specialists, though the trend is somewhat noisy due to an outlier in 2007-08. This outlier reflects the fact that many business operations positions, including HR specialists, were among the first occupational groups advertised extensively on online job boards, which could have temporarily inflating their share in online postings before other occupations caught up.

B Theory appendix

B.1 Alternative formulation of the screening technology

In the main text, we assume that a firm draws a cost of screening a worker and may then decide to screen or not conditional on the realization of the cost draw. Here, we demonstrate that this model is equivalent to one where the firm decides to optimally set a probability of screening subject to a convex cost.

We begin by setting up the alternative firm problem. Let Δ^f denote the return to screening, as it does in the main text. In our alternative formulation, we assume that the cost of screening with probability p^{sig} is

$$\phi(p^{\text{sig}}) = \gamma_s^{-1} \frac{\eta_s}{1+\eta_s} (p^{\text{sig}})^{\frac{1+\eta_s}{\eta_s}}$$

and thus, defining $\tilde{q} = q(a_L^u p_L^{\text{app}} + (1 - a_L^u) p_H^{\text{app}})$ as the rate at which firms encounter applicants, the value function of an unmatched vacancy is

$$(r+\delta)V = \max_{p^{\text{sig}}} \tilde{q}(p^{\text{sig}} \Delta^f - \phi(p^{\text{sig}}))$$

and the first order condition of this problem is

$$\begin{aligned} \Delta^f &= \phi'(p^{\text{sig}}) = \gamma_s^{-1} (p^{\text{sig}})^{\frac{1}{\eta_s}} \\ \implies p^{\text{sig}} &= \gamma_s (\Delta^f)^{\eta_s} = G(\Delta^f) \end{aligned}$$

which is the same relationship between the return to screening and the probability of screening that is featured in the baseline model. Thus, the two models are equivalent.

B.2 Derivation of mixed-strategy equilibrium conditions

In this appendix, we define the functions referenced by the screening curve (SC) and indifference curve (IC) from the main text, as well as a number of auxiliary functions that are useful for those definitions. All functions we define are functions of $\vec{f} = (f_L, f_H)$ that take parameters as given, and are defined recursively. By this we mean that, in what follows below, every equation encodes instructions on how to compute a new object using only $\vec{f}$, parameters, and previously defined objects. Thus, all objects defined in equations (17)-(25) are closed-form functions of $\vec{f}$. One implication of this is that, fixing some parameters, knowing the solution $\vec{f}$ of the system defined by equations (IC) and (SC) is enough to derive all equilibrium objects defined below in closed form. The functions are defined as follows:

$$J_i(f_i) = \left(\frac{(1-\beta)\rho y_i}{r+\delta+\beta(1+\eta_a)^{-1} f_i} \right) \tag{17}$$

$$a_L^u(\vec{f}) = \frac{\alpha_L^{\text{pop}}(f_H+\delta)}{\alpha_L^{\text{pop}} f_H + (1-\alpha_L^{\text{pop}}) f_L + \delta} \tag{18}$$

$$a_L^{\text{app}}(\vec{f}) = a_L^u(\vec{f}) \left(a_L^u(\vec{f}) + (1-a_L^u(\vec{f})) \left(\frac{f_H}{f_L} \right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{r+\delta+\beta(1+\eta_a)^{-1} f_L}{r+\delta+\beta(1+\eta_a)^{-1} f_H} \right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{y_H}{y_L} \right)^{\frac{\eta_a}{1+\eta_a}} \right)^{-1} \tag{19}$$

$$\begin{aligned} \Delta^f(\vec{f}) &= a_L^{\text{app}}(\vec{f})(1-a_L^{\text{app}}(\vec{f}))(1-\beta)(2s-1) \cdot \\ &\left(\frac{1}{r+\delta+\beta\,(1+\eta_a)^{-1} f_H} (\rho y_H) - \frac{1}{r+\delta+\beta(1+\eta_a)^{-1} f_L} (\rho y_L) \right) \end{aligned} \tag{20}$$

$$\begin{aligned} \lambda(\vec{f}) = \left(\frac{Z}{((r+\delta)k)^{\omega}} \right)^{\frac{(1+\eta_a)}{1+\eta_a-\omega}} &\left[\gamma_a^{\frac{1}{1+\eta_a}} (1+\eta_a)^{\frac{\eta_a}{1+\eta_a}} \left(a_L^u(\vec{f}) \left(\frac{\beta(1+\eta_a)^{-1} f_L}{r+\delta+\beta(1+\eta_a)^{-1} f_L} (\rho y_L) \right)^{\frac{\eta_a}{1+\eta_a}} \right.\right. \\ &\left.\left. + (1-a_L^u(\vec{f})) \left(\frac{\beta(1+\eta_a)^{-1} f_H}{r+\delta+\beta(1+\eta_a)^{-1} f_H} (\rho y_H) \right)^{\frac{\eta_a}{1+\eta_a}} \right) (1+\eta_s)^{-1} \gamma_s \Delta^f(\vec{f})^{(1+\eta_s)} \right]^{\frac{\omega(1+\eta_a)}{1+\eta_a-\omega}} \end{aligned} \tag{21}$$

$$\Delta_L^w(\vec{f}) = \left(\gamma_a^{-1}(1+\eta_a)\lambda(\vec{f})^{-1} \left(\frac{\beta(1+\eta_a)^{-1} f_L}{r+\delta+\beta(1+\eta_a)^{-1} f_L} \right) (\rho y_L) \right)^{\frac{1}{1+\eta_a}} \tag{22}$$

$$\Delta_H^w(\vec{f}) = \left(\gamma_a^{-1}(1+\eta_a)\lambda(\vec{f})^{-1} \left(\frac{\beta(1+\eta_a)^{-1} f_H}{r+\delta+\beta(1+\eta_a)^{-1} f_H} \right) (\rho y_H) \right)^{\frac{1}{1+\eta_a}} \tag{23}$$

$$p_i^{\text{app}}(\vec{f}) = F(\Delta_i^w(\vec{f})) \tag{24}$$

$$p^{\text{sig}}(\vec{f}) = G(\Delta^f(\vec{f})) \tag{25}$$

In what follows, we derive these equations alongside the screening and indifference curves from the main text.

Worker side. With $F(c) = \gamma_a c^{\eta_a}$,

$$\int_0^{\Delta} c\, dF(c) = \gamma_a \eta_a \int_0^{\Delta} c^{\eta_a}\, dc = \frac{\eta_a}{1+\eta_a} F(\Delta)\Delta,$$

and hence

$$rU_i = b + \lambda \int_0^{\Delta_i^w} (\Delta_i^w - c)\, dF(c) = b + \lambda \left(\Delta_i^w F(\Delta_i^w) - \int_0^{\Delta_i^w} c\, dF(c) \right)$$
$$= b + \lambda F(\Delta_i^w)(1+\eta_a)^{-1} \Delta_i^w.$$

Plugging this into the surplus Bellman equation, we get

$$(r+\delta) S_i = y_i - rU_i = y_i - b - \frac{\lambda \gamma_a}{1+\eta_a} (\Delta_i^w)^{1+\eta_a}.$$

We know that by Nash bargaining and the definition of Δ_i^w, $S_i = \Delta_i^w / (\beta p_i^{\text{acc}})$. Substituting this into the previous equation, we get

$$(r+\delta) \frac{\Delta_i^w}{\beta p_i^{\text{acc}}} = y_i - b - \frac{\lambda \gamma_a}{1+\eta_a} (\Delta_i^w)^{1+\eta_a}.$$

Since $f_i = \lambda F(\Delta_i^w) p_i^{\text{acc}} = \lambda \gamma_a (\Delta_i^w)^{\eta_a} p_i^{\text{acc}}$, we can eliminate p_i^{acc} and write

$$(r+\delta) \Delta_i^w = \frac{\beta f_i}{\lambda \gamma_a (\Delta_i^w)^{\eta_a}} \rho y_i - \frac{\beta f_i}{1+\eta_a} \Delta_i^w.$$

Solving for Δ_i^w yields

$$\Delta_i^w = \left[\gamma_a^{-1} (1+\eta_a) \lambda^{-1} \frac{\beta (1+\eta_a)^{-1} f_i}{r+\delta+\beta(1+\eta_a)^{-1} f_i} \rho y_i \right]^{\frac{1}{1+\eta_a}}.$$

which establishes equations (22), (23) and (24). We can immediately use this result to derive

$$\frac{F(\Delta_H^w)}{F(\Delta_L^w)} = \left(\frac{f_H}{f_L} \right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{r+\delta+\beta(1+\eta_a)^{-1} f_L}{r+\delta+\beta(1+\eta_a)^{-1} f_H} \right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{y_H}{y_L} \right)^{\frac{\eta_a}{1+\eta_a}}. \tag{26}$$

which will be useful below.

Firm side. From $J_i = (1-\beta) S_i$ and the surplus equation one obtains the compact form

$$J_i = \frac{(1-\beta)\rho y_i}{r+\delta+\beta(1+\eta_a)^{-1} f_i}.$$

which establishes equation (17). We now derive Δ^f. The value when screening an applicant is

$$V_{\text{signal}} = (1 - a_L^{\text{app}})sJ_H + a_L^{\text{app}}(1-s)J_L + \Big(1 - (1 - a_L^{\text{app}})s - a_L^{\text{app}}(1-s)\Big)V.$$

For a mixed strategy equilibrium, the firm is indifferent between rejecting and accepting an unscreened worker, which implies

$$V_{\text{nosignal}} = (1 - a_L^{\text{app}})J_H + a_L^{\text{app}}J_L = V = k.$$

Compute the return to screening from its definition:

$$\begin{aligned}\Delta^f &= V_{\text{signal}} - V_{\text{nosignal}} \\ &= (1 - a_L^{\text{app}})s(J_H - V) + a_L^{\text{app}}(1-s)(J_L - V).\end{aligned}$$

Write $J_H - V$ and $J_L - V$ as

$$\begin{aligned}J_H - V &= J_H - \Big((1 - a_L^{\text{app}})J_H + a_L^{\text{app}}J_L\Big) = a_L^{\text{app}}(J_H - J_L), \\ J_L - V &= J_L - \Big((1 - a_L^{\text{app}})J_H + a_L^{\text{app}}J_L\Big) = -(1 - a_L^{\text{app}})(J_H - J_L).\end{aligned}$$

Substitute into Δ^f:

$$\begin{aligned}\Delta^f &= (1 - a_L^{\text{app}})s \cdot a_L^{\text{app}}(J_H - J_L) + a_L^{\text{app}}(1-s) \cdot \Big(-(1 - a_L^{\text{app}})(J_H - J_L)\Big) \\ &= a_L^{\text{app}}(1 - a_L^{\text{app}})(J_H - J_L)\Big(s - (1-s)\Big) \\ &= a_L^{\text{app}}(1 - a_L^{\text{app}})(2s-1)(J_H - J_L).\end{aligned}$$

Using the Bellman equations for J_L and J_H yields

$$\Delta^f = a_L^{\text{app}}(1 - a_L^{\text{app}})(1-\beta)(2s-1)\left[\frac{\rho y_H}{r + \delta + \beta(1+\eta_a)^{-1}f_H} - \frac{\rho y_L}{r + \delta + \beta(1+\eta_a)^{-1}f_L}\right].$$

which establishes equations (20) and (25).

The aggregate encounter rate of the worker. Free entry implies

$$(r+\delta)k = q\Big(a_L^{\text{app}}F(\Delta_L^w) + (1-a_L^{\text{app}})F(\Delta_H^w)\Big)\frac{\gamma_s}{1+\eta_s}(\Delta^f)^{1+\eta_s}.$$

Solving for q:

$$q = \frac{(r+\delta)k}{\Big(a_L^{\text{app}}F(\Delta_L^w) + (1-a_L^{\text{app}})F(\Delta_H^w)\Big)(1+\eta_s)^{-1}\gamma_s(\Delta^f)^{1+\eta_s}}.$$

Now use $\lambda = Z^{\frac{1}{1-\omega}}q^{\frac{\omega}{\omega-1}}$ and plug in Δ_i^w from above:

$$\lambda^{1+\frac{\omega\eta_a}{(1+\eta_a)(1-\omega)}} = \left(\frac{Z}{((r+\delta)k)^{\omega}}\right)^{\frac{1}{1-\omega}}\left[\gamma_a^{\frac{1}{1+\eta_a}}(1+\eta_a)^{\frac{\eta_a}{1+\eta_a}}\Big(a_L^u A_L + (1-a_L^u)A_H\Big)(1+\eta_s)^{-1}\gamma_s(\Delta^f)^{1+\eta_s}\right]^{\frac{\omega}{1-\omega}}.$$

where

$$A_i = \left(\frac{\beta(1+\eta_a)^{-1}f_i}{r+\delta+\beta(1+\eta_a)^{-1}f_i}\rho y_i\right)^{\frac{\eta_a}{1+\eta_a}}.$$

Solving this equation for λ gives

$$\lambda = \left(\frac{Z}{((r+\delta)k)^{\omega}}\right)^{\frac{1+\eta_a}{1+\eta_a-\omega}}\cdot \left[\gamma_a^{\frac{1}{1+\eta_a}}(1+\eta_a)^{\frac{\eta_a}{1+\eta_a}}\Big(a_L^u A_L + (1-a_L^u)A_H\Big)(1+\eta_s)^{-1}\gamma_s(\Delta^f)^{1+\eta_s}\right]^{\frac{\omega(1+\eta_a)}{1+\eta_a-\omega}}.$$

which establishes equation (21).

Flows and stocks. We now derive the flow identities. Let u_i be the steady-state mass unemployed of type i and $u \equiv u_L + u_H$. Stationarity implies

$$u_i f_i = \delta(\alpha_i^{\text{pop}} - u_i) \quad\Rightarrow\quad u_i = \frac{\delta\alpha_i^{\text{pop}}}{f_i+\delta}.$$

Hence the share of L among the unemployed is

$$a_L^u = \frac{u_L}{u_L+u_H} = \frac{\alpha_L^{\text{pop}}(f_H+\delta)}{\alpha_L^{\text{pop}}f_H + (1-\alpha_L^{\text{pop}})f_L+\delta},$$

which establishes equation (18). The share of L-types among applicants follows Bayes' Law:

$$a_L^{\text{app}} = \frac{a_L^u F(\Delta_L^w)}{a_L^u F(\Delta_L^w) + (1-a_L^u)F(\Delta_H^w)} = \frac{a_L^u}{a_L^u + (1-a_L^u)\frac{F(\Delta_H^w)}{F(\Delta_L^w)}},$$

which, together with equation (26), establishes equation (19).

Screening curve. Start from the definition of type-specific job finding rates

$$f_i = \lambda F(\Delta_i^w)\, p_i^{\text{acc}} \qquad (i \in \{L, H\}).$$

The acceptance probabilities satisfy

$$\begin{aligned} p_L^{\text{acc}} &= (1 - G(\Delta^f))(1-R) + G(\Delta^f)(1-s) \\ p_H^{\text{acc}} &= (1 - G(\Delta^f))(1-R) + G(\Delta^f)s \\ \implies p_H^{\text{acc}} - p_L^{\text{acc}} &= G(\Delta^f)\Big(s - (1-s)\Big) = G(\Delta^f)(2s-1). \end{aligned}$$

Now express f_H in terms of f_L:

$$\begin{aligned} f_H &= \lambda p_H^{\text{app}}\, p_H^{\text{acc}} \\ &= \lambda p_H^{\text{app}} \left(p_L^{\text{acc}} + \left(p_H^{\text{acc}} - p_L^{\text{acc}} \right) \right) \\ &= \lambda p_H^{\text{app}}\, p_L^{\text{acc}} + \lambda p_H^{\text{app}} p^{\text{sig}} (2s-1). \end{aligned}$$

Replace $p_L^{\text{acc}} = f_L / (\lambda F(\Delta_L^w))$ (from the definition of f_L) to obtain

$$f_H = \frac{p_H^{\text{app}}}{p_L^{\text{app}}} f_L \; + \; \lambda p_H^{\text{app}}\, p^{\text{sig}}\, (2s-1).$$

This is the screening curve (SC) used in the main text.

Indifference curve (IC). By equation (10), in a mixed strategy equilibrium, the value of accepting an unscreened applicant equals the vacancy value V, which by free entry equals k in equilibrium:

$$V = k = (1 - a_L^{\text{app}})J_H + a_L^{\text{app}} J_L.$$

This is the indifference curve (IC) used in the main text.

B.3 Overdispersion as a sufficient statistic

Let μ be some non-negative random variable and $N \mid \mu$ be a count variable generated by a Poisson mixture with mean μ:

$$N \mid \mu \sim \text{Poisson}(\mu)$$

By the law of total variance,

$$\text{Var}(N) = \mathbb{E}[\text{Var}(N \mid \mu)] + \text{Var}(\mathbb{E}[N \mid \mu]) = \mathbb{E}[\mu] + \text{Var}(\mu),$$

since $\text{Var}(N \mid \mu) = \mathbb{E}[N \mid \mu] = \mu$ for a Poisson. Hence,

$$\mathbb{E}[N] = \mathbb{E}[\mu], \qquad \text{Var}(N) - \mathbb{E}[N] = \text{Var}(\mu).$$

Define overdispersion as

$$\mathcal{D} \equiv \frac{\sqrt{\text{Var}(N) - \mathbb{E}[N]}}{\mathbb{E}[N]}.$$

Then $\mathcal{D}$ identifies the coefficient of variation of the mixing distribution:

$$\mathcal{D} = \frac{\sqrt{\text{Var}(\mu)}}{\mathbb{E}[\mu]} = \text{CV}(\mu).$$

In our baseline model, the number of applications sent by a worker over a month is Poisson with a scale $\mu_i = \lambda p_i^{\text{app}}$ that varies by type. Thus, our environment is a special case of the Poisson-mixture setup above and the observed overdispersion statistic $\mathcal{D}$ equals the coefficient of variation of equilibrium application intensities in the model.

We compute our target using self-reported data on applications sent from the SCE.

B.4 Deriving the job filling rate

For the equilibria we focus on, the job filling rate is pinned down by the indifference condition of the firm

$$k = V = (r + \delta)^{-1} \tilde{q} G(\Delta^f)(\Delta^f - \mathbb{E}[\kappa \mid \kappa \leq \Delta^f])$$

where $\Delta^f = V_{\text{signal}} - V_{\text{nosignal}} = a_L^{\text{app}}(1 - a_L^{\text{app}})(2s - 1)\,(J_H - J_L)$ and $\tilde{q} = q(a_L^u p_L^{\text{app}} + (1 - a_L^u) p_H^{\text{app}})$ is the rate at which firms encounter applicants. Given the functional form

$G(x) = \gamma_s x^{\eta_s}$,

$$G(\Delta^f)(\Delta^f - \mathbb{E}[\kappa \mid \kappa \leq \Delta^f]) = (1+\eta_s)^{-1} G(\Delta^f)\Delta^f,$$

and thus,

$$k(r+\delta) = (1+\eta_s)^{-1}\tilde{q}G(\Delta^f)\Delta^f. \tag{27}$$

Equation (27) illustrates how the encounter rate $\tilde{q}$ is determined in equilibrium: Since only encounters during which a signal is obtained produce value in excess of k to the firm, the encounter rate of firms is solely determined by the excess value of such events, Δ^f. This has parallels to the free-entry logic of a standard DMP model: a higher value of encounters implies lower encounter rates for firms. The job filling rate depends on the encounter rate, multiplied by the probability that an encounter converts into a hire:

$$\text{job filling rate} = \tilde{q}\left(\left(1 - G(\Delta^f)\right)(1-R) + G(\Delta^f)[a_L^{\text{app}}(1-s) + (1-a_L^{\text{app}})s]\right)$$
$$= \underbrace{k(r+\delta)(1+\eta_s)(G(\Delta^f)\Delta^f)^{-1}}_{\text{encounter rate}}\left(\underbrace{\left(1-G(\Delta^f)\right)(1-R)}_{\text{admissions without signal}} + \underbrace{G(\Delta^f)}_{=p^{\text{sig}}}\underbrace{[a_L^{\text{app}}(1-s) + (1-a_L^{\text{app}})s]}_{\text{conversion prob. of signals}}\right). \tag{28}$$

Equation (28) shows how the job filling rate is pinned down in equilibrium. It is the product of the encounter rate, determined by Δ^f, and the probability of admission with and without a signal. The relative strength of these channels depends on the quantitative properties of the model.

B.5 Model formula for unemployment duration

In the model, the average duration of unemployment is

$$\text{unemployment duration} = a_L^u \frac{1}{f_L} + (1-a_L^u)\frac{1}{f_H}$$

and the aggregate job-finding rate is

$$\text{job finding rate} = \left(a_L^u f_L + (1-a_L^u) f_H\right).$$

Therefore,

$$\begin{aligned}\frac{\text{unemployment duration}}{(\text{job finding rate})^{-1}} &= \left(a_L^u \frac{1}{f_L} + (1-a_L^u)\frac{1}{f_H}\right) \cdot \left(a_L^u f_L + (1-a_L^u) f_H\right) \\ &= \left((a_L^u)^2 + (1-a_L^u)^2 + a_L^u(1-a_L^u)\left(\frac{f_H}{f_L} + \frac{f_L}{f_H}\right)\right) \\ &= \left(1 + a_L^u(1-a_L^u)\left(\frac{f_H}{f_L} + \frac{f_L}{f_H} - 2\right)\right) \\ &= \left(1 + a_L^u(1-a_L^u)\frac{(f_H - f_L)^2}{f_H f_L}\right). \end{aligned} \tag{29}$$

B.6 A generalized unemployment duration decomposition

In this appendix section, we derive a general decomposition under assumptions that are consistent with the model but at the same time robust to more general environments. Specifically, our decomposition holds in environments with an arbitrary number of worker types and an arbitrary degree of duration dependence, provided that certain regularity conditions are met.

We make the following assumptions:

(A1) The unemployed population consists of a mass u of workers that are indexed by their type $i \in I$ and distributed according to a stationary cdf $H(\cdot) : i \mapsto [0,1]$, which has finite moments.

(A2) The length c of eac h completed unemployment spell for a worker of type i is independently distributed according to $\mathcal{F}_i$ with finite expectation $\mathbb{E}\left[c \mid i\right] = \theta_i$.

(A3) $\mathcal{F}_i$ has a finite and homogeneous coefficient of variation: $\frac{\sqrt{\text{Var}(c|i)}}{\mathbb{E}[c|i]} = \text{CV}_i = \text{CV}$.

These assumptions warrant some discussion. (A1) imposes stationarity: The distribution of types in unemployment, H, is unchanging and flows are consistent with the stock. While this is unlikely to be exactly true in real world data, it is a good approximation whenever flow rates in the data are sufficiently high and sufficiently stable, which tends to be true for the flow rates in and out of unemployment.[40] (A2) imposes the assumption

[40] Without assuming stationarity, it is typically impossible to make any claims about the relationship between flows and stocks, as we do here.

that unemployment spell lengths are independently drawn from a family of distributions $\mathcal{F}_i$. This is a rather weak assumption that lends meaning to the notion of a worker type i in this context. It does not impose much structure beyond that because we do not restrict the set of worker types I.

(A3) is a stronger assumption. It imposes that the coefficient of variation of $\mathcal{F}_i$, defined as the standard deviation of type-conditional unemployment spell lengths divided by the type-conditional mean, is independent of the worker type i. How restrictive is assumption (A3)? To answer this question by means of example, we first note that in a model without duration dependence and thus constant hazards, such as our baseline model, unemployment durations are exponentially distributed. This means that in such an environment the coefficient of variation is 1, and (A3) is satisfied. Next, consider an environment with duration dependence, specifically one where the hazard rate of leaving unemployment is decreasing over time according to a homogeneous power law ($h(t) = \lim_{\Delta \downarrow 0} \frac{\Pr(t \leq T < t+\Delta | T \geq t)}{\Delta} = c_i t^{k-1}$, $k > 0$). In this environment, $\mathcal{F}_i$ is Weibull with shape parameter k and a coefficient of variation of

$$\text{CV}_i = \text{CV} = \frac{\sqrt{\Gamma\left(1+\frac{2}{k}\right) - \Gamma\left(1+\frac{1}{k}\right)^2}}{\Gamma\left(1+\frac{1}{k}\right)}$$

where Γ is the incomplete Gamma function. Thus, (A3) is satisfied in this type of environment, too. Now, consider an example for which (A3) is violated: we consider the same data generating process as before, but now make the amount of duration dependence a function of the worker type. That, is if we assume that the hazard rate satisfies $h(t) = a_i t^{k_i - 1}$, then (A3) is violated whenever there are $i, j \in I$ such that $k_i \neq k_j$. In a sense, (A3) thus imposes orthogonality between type heterogeneity and duration dependence. This structure makes it possible to disentangle type heterogeneity in terms of the mean expected unemployment duration on the one hand and deviations from the constant hazard model on the other.

Before we are ready to state the theorem that establishes our decomposition, we define

$$\bar{f} = \mathbb{E}_H\left[\theta_i^{-1}\right]$$

which in the stationary distribution corresponds to the average rate at which workers find jobs. For this reason, we also refer to $\bar{f}$ as the UE rate. We can now state our decomposition theorem:

Proposition 3 (Decomposition of unemployment duration). *Under assumptions (A1)-(A3), the cross-sectional stationary mean elapsed unemployment duration $\mathbb{E}[d]$ in the population satisfies*

$$\mathbb{E}[d] = \mathbb{E}_H[\mathbb{E}\left[d \mid i\right]] = \underbrace{\frac{1}{\bar{f}}}_{\text{(i) inv. job finding rate}} \times \underbrace{\frac{1}{2}(1 + CV^2)}_{\substack{\text{(ii) type-conditional spell dispersion} \\ \text{(duration dependence)}}} \times \underbrace{\left(1 + Cov_H(\theta, -\theta^{-1})\right)}_{\text{(iii) worker type heterogeneity}} \tag{30}$$

Proof of Proposition 3. The proof relies on the famous inspection paradox identity (Feller, 1991):

Lemma 1 (Inspection paradox identity). *Let N be a renewal process with random interval lengths c. Let $\mathbb{E}[c] = \theta_i$ and $CV = \frac{\sqrt{Var(c)}}{\mathbb{E}[c]}$. Let d denote the elapsed (backward) duration when sampling from the process at a random time. Then*

$$\mathbb{E}[d] = \frac{\mathbb{E}[c^2]}{2\,\mathbb{E}[c]} = \frac{1 + \mathrm{CV}^2}{2}\,\theta_i.$$

If H is stationary, the unemployment duration of each worker in unemployment follows a renewal process. By Lemma 1, $\mathbb{E}[d \mid i] = \frac{1+\mathrm{CV}^2}{2}\theta_i$. Thus,

$$\mathbb{E}[d] = \tfrac{1+\mathrm{CV}^2}{2}\mathbb{E}_H[\theta_i].$$

Expanding $\mathbb{E}_H[\theta_i]$, we get

$$\begin{aligned}\mathbb{E}_H[\theta_i] = \mathbb{E}_H[\theta_i^{-1}]^{-1}\mathbb{E}_H[\theta_i^{-1}]\mathbb{E}_H[\theta_i] &= \frac{1}{\mathbb{E}_H[\theta_i^{-1}]}\left(\mathbb{E}[\theta_i\theta_i^{-1}] - \mathrm{Cov}_H\left(\theta_i, \theta_i^{-1}\right)\right) \\ &= \frac{1}{\bar{f}}\left(1 + \mathrm{Cov}_H\left(\theta_i, -\theta_i^{-1}\right)\right).\end{aligned}$$

Hence,

$$\mathbb{E}[d] = \frac{1}{\bar{f}} \cdot \frac{1 + \mathrm{CV}^2}{2} \cdot \left(1 + \mathrm{Cov}_H(\theta_i, -\theta_i^{-1})\right),$$

which is the claimed decomposition. □

Proposition 3 shows that the average unemployment duration in the population can be decomposed into three constituent parts.

Term (i) captures the inverse UE rate, a measure of aggregate conditions. In a representative agent model with no duration dependence, this is the only operative term: the average unemployment duration in this simple environment equals the inverse rate of finding a job.

Term (ii) captures that, in the presence of duration dependence, the dispersion of spell lengths conditional on type starts mattering for unemployment duration. Specifically, the relationship between the average duration of complete spells and the average duration of incomplete spells changes depending on the predictability of the complete spell length: When spells are fully deterministic, for example, $CV = 0$ and $\mathbb{E}[d] = \frac{1}{2}\mathbb{E}[c]$. On the other hand, when spell lengths are exponentially distributed, $CV = 1$ and $\mathbb{E}[d] = \mathbb{E}[c]$.

The final term (iii) captures a measure of worker type heterogeneity. The term is always weakly greater than one and strictly greater than one in the presence of heterogeneity. A more heterogeneous composition of workers ceteris paribus leads to longer average unemployment duration. This is formalized in the following corollary of Proposition 3:

Corollary 1 (Heterogeneity increases duration). *Let $\theta \sim H_1$, $\theta' \sim H_2$, $\theta^{-1} \sim H_1^{inv}$, $(\theta')^{-1} \sim H_2^{inv}$, and let H_2^{inv} be a mean preserving spread of H_1^{inv}.*
Then $\bar{f}_{H_1} = \bar{f}_{H_2}$ but $\mathbb{E}_{H_1}[\mathbb{E}[d \mid i]] < \mathbb{E}_{H_2}[\mathbb{E}[d \mid i]]$.

Proof of Corollary 1. Let $\lambda = \theta^{-1}$ and $\lambda' = (\theta')^{-1}$. The assumption that H_2^{inv} is a mean-preserving spread of H_1^{inv} means that $\mathbb{E}_{H_2}[\lambda'] = \mathbb{E}_{H_1}[\lambda] = \bar{f}$ and that there exists a random variable ε such that

$$\lambda' = \lambda + \varepsilon, \qquad \mathbb{E}[\varepsilon \mid \lambda] = 0.$$

Consider the convex function $\varphi(x) = 1/x$ on $(0, \infty)$. Jensen's inequality yields

$$\mathbb{E}\left[\frac{1}{\lambda'} \;\middle|\; \lambda\right] = \mathbb{E}\left[\frac{1}{\lambda + \varepsilon} \;\middle|\; \lambda\right] \geq \frac{1}{\lambda}.$$

Taking expectations and using the law of iterated expectations yields

$$\mathbb{E}[\theta'] = \mathbb{E}\left[\frac{1}{\lambda'}\right] \geq \mathbb{E}\left[\frac{1}{\lambda}\right] = \mathbb{E}[\theta].$$

with strict inequality whenever the spread is non-degenerate. Proposition 3 then implies

$$\mathbb{E}[d'] = \frac{1}{\bar{f}} \cdot \frac{1 + \text{CV}^2}{2} \cdot \mathbb{E}_{H_2}[\theta']\mathbb{E}_{H_2}[(\theta')^{-1}] \geq \frac{1}{\bar{f}} \cdot \frac{1 + \text{CV}^2}{2} \cdot \mathbb{E}_{H_1}[\theta]\mathbb{E}_{H_1}[\theta^{-1}] = \mathbb{E}[d].$$

□

This decomposition shows that heterogeneity matters for the average duration of unemployment.

It is worth noting that the literature has proposed and discussed a number of models of duration dependence in which unemployment duration is governed by the first two channels (terms (i) and (ii)) (Kroft et al., 2016, see e.g.). In fact, Proposition 3 states that *any* model without worker heterogeneity that adheres to assumptions (A1)-(A3) must necessarily move the average duration of unemployment via the first two channels. In contrast, our model predicts that permanent worker heterogeneity plays a major role in generating changes in unemployment duration over time. As we saw in Section 4, this has major implications for the incidence and sign of intertemporal welfare changes across different worker groups.

B.7 Estimating trends in duration dependence

Using the same classification from the main text, we construct a consistent estimator of the coefficient of variation (CV) of unemployment duration conditional on worker type as follows: First, we bin workers into 5 groups according to their lifetime average duration. Then, within each bin and decade, we compute the sample standard deviation of unemployment spells and divide it by the sample mean. Finally, we average the resulting object across bins. This yields a consistent estimator for the coefficient of variation of type-conditional spell lengths.

Figure B.1 plots our estimate of CV over time. Two important features of the series are worth noting: First, it is, if anything, mildly decreasing over time. Thus, a rise in duration dependence cannot account for the observed rise in unemployment duration. Second, it is a little bit less than one in most periods, though we acknowledge that finite sample bias will generally lower our estimates of this object. Given that the estimates are stable and close to one, we argue that the constant hazard model of Section 3 does not suffer from major misspecification issues arising from duration dependence.

B.8 Model formula for unemployment duration rank of hires

For notational brevity, in this section, let $a \in [0, 1]$ be the share of type-L workers in the stock of unemployed workers, and let $b \in [0, 1]$ be the share of type-L workers among

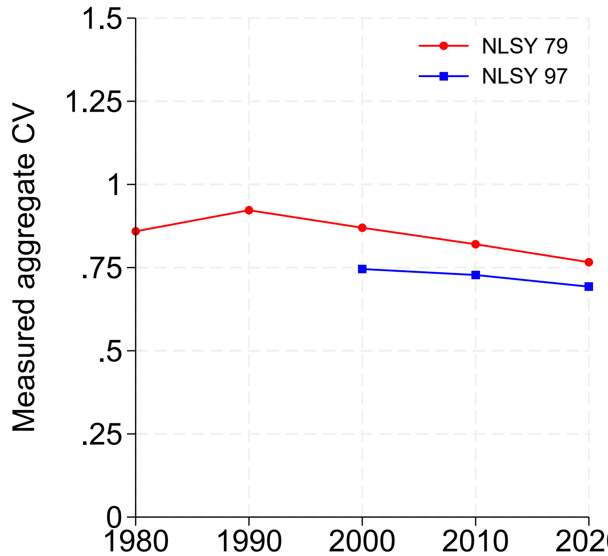


Figure B.1: Estimated CV in the NLSY

Notes: Estimate of the coefficient of variation of unemployment spells (CV), estimated from the NLSY. We construct the estimator as described in Appendix B.7.

hires. Then the unemployment-duration distribution in the unemployed stock is

$$F_U(x) = 1 - ae^{-f_L x} - (1-a)e^{-f_H x},$$

and the unemployment-duration density among hires is

$$f_{\text{hire}}(x) = b f_L e^{-f_L x} + (1-b) f_H e^{-f_H x}.$$

The average percentile rank (CDF-rank) of a typical hire's unemployment duration within the unemployed distribution is

$$\mathcal{R} := \mathbb{E}\left[F_U(X_{\text{hire}})\right] = \int_0^\infty F_U(x)\, f_{\text{hire}}(x)\, dx.$$

Expanding and integrating term-by-term,

$$\begin{aligned}
\mathcal{R} &= \int_0^\infty \left(b f_L e^{-f_L x} + (1-b) f_H e^{-f_H x} \right) dx \\
&\quad - a \int_0^\infty e^{-f_L x} \left(b f_L e^{-f_L x} + (1-b) f_H e^{-f_H x} \right) dx \\
&\quad - (1-a) \int_0^\infty e^{-f_H x} \left(b f_L e^{-f_L x} + (1-b) f_H e^{-f_H x} \right) dx \\
&= 1 - a \left(b f_L \frac{1}{2 f_L} + (1-b) f_H \frac{1}{f_L + f_H} \right) - (1-a) \left(b f_L \frac{1}{f_L + f_H} + (1-b) f_H \frac{1}{2 f_H} \right) \\
&= 1 - \frac{ab}{2} - \frac{(1-a)(1-b)}{2} - \frac{a(1-b) f_H + (1-a) b f_L}{f_L + f_H}.
\end{aligned}$$

Now impose that

$$b = \frac{a f_L}{a f_L + (1-a) f_H}, \qquad 1 - b = \frac{(1-a) f_H}{a f_L + (1-a) f_H}.$$

Then

$$\frac{ab}{2} = \frac{a^2 f_L}{2\left(a f_L + (1-a) f_H\right)}, \qquad \frac{(1-a)(1-b)}{2} = \frac{(1-a)^2 f_H}{2\left(a f_L + (1-a) f_H\right)},$$

and

$$a(1-b) f_H + (1-a) b f_L = \frac{a(1-a)\left(f_H^2 + f_L^2\right)}{a f_L + (1-a) f_H}.$$

Plugging this into the formula for $\mathcal{R}$ above, we get

$$\mathcal{R} = 1 - \frac{a^2 f_L + (1-a)^2 f_H}{2\left(a f_L + (1-a) f_H\right)} - \frac{a(1-a)\left(f_H^2 + f_L^2\right)}{(f_L + f_H)\left(a f_L + (1-a) f_H\right)}.$$

We now simplify this expression by using the identity

$$a^2 f_L + (1-a)^2 f_H = \left(a f_L + (1-a) f_H\right) - a(1-a)(f_L + f_H),$$

so

$$1 - \frac{a^2 f_L + (1-a)^2 f_H}{2\left(a f_L + (1-a) f_H\right)} = \frac{1}{2} + \frac{a(1-a)(f_L + f_H)}{2\left(a f_L + (1-a) f_H\right)}.$$

Therefore,

$$\mathcal{R} = \frac{1}{2} + \frac{a(1-a)}{af_L + (1-a)f_H}\left[\frac{f_L + f_H}{2} - \frac{f_H^2 + f_L^2}{f_L + f_H}\right].$$

Finally, we simplify the term in brackets:

$$\frac{f_L + f_H}{2} - \frac{f_H^2 + f_L^2}{f_L + f_H} = \frac{(f_L + f_H)^2 - 2(f_L^2 + f_H^2)}{2(f_L + f_H)} = -\frac{(f_H - f_L)^2}{2(f_L + f_H)},$$

which yields the formula

$$\mathcal{R} = \frac{1}{2}\left(1 - \frac{a(1-a)(f_H - f_L)^2}{(f_L + f_H)\Big(af_L + (1-a)f_H\Big)}\right).$$

A few observations on $\mathcal{R}$ are in order. First, without heterogeneity, $\mathcal{R} = \frac{1}{2}$. Second, $\mathcal{R} \leq \frac{1}{2}$ for any (f_L, f_H), which indicates that the average duration rank of hires is always less than the average duration rank in the pool of unemployed (which by definition equals $\frac{1}{2}$). Lastly, when job finding rates are more dispersed, the average duration rank of hires falls.

B.9 Deriving the rate of screening expenditures

In this appendix, we show that the rate of screening resource consumption depends only on k, r, δ, and η_s. We first note that the rate of screening is defined by

$$\text{screening expenditure rate} = \tilde{q} \cdot \mathbb{E}\left[\kappa \cdot \mathbb{I}\{\kappa \leq \Delta^f\}\right]$$

where $\tilde{q} = q(a_L^u p_L^{\text{app}} + (1 - a_L^u)p_H^{\text{app}})$ is the rate at which vacancies encounter applicants. Substituting in the free entry condition from equation (27) in Appendix B.4 yields

$$\text{screening expenditure rate} = k(r + \delta)\eta_s.$$

Therefore, the rate at which firms spend resources on screening does not depend on changes in equilibrium that are driven by parameters other than k, r, δ, and η_s.

B.10 Proofs of propositions in the main text

Proof of Proposition 1. We start by defining

$$\widehat{\Delta}^f(\vec{f}) := a_L^{\text{app}}(\vec{f})\Big(1 - a_L^{\text{app}}(\vec{f})\Big)(1-\beta)\left[\frac{\rho y_H}{r+\delta+\beta(1+\eta_a)^{-1}f_H} - \frac{\rho y_L}{r+\delta+\beta(1+\eta_a)^{-1}f_L}\right],$$

which yields

$$\begin{aligned} \Delta^f(\vec{f}) &= (2s-1)\widehat{\Delta}^f(\vec{f}) \\ \implies G(\Delta^f(\vec{f}))(2s-1) &= \gamma_s(2s-1)^{1+\eta_s}\Big(\widehat{\Delta}^f(\vec{f})\Big)^{\eta_s}. \end{aligned} \tag{31}$$

Next, define

$$\widehat{\lambda}(\vec{f}) := \left(\frac{Z}{((r+\delta)k)^{\omega}}\right)^{\frac{1+\eta_a}{1+\eta_a-\omega}}\left[(1+\eta_a)^{\frac{\eta_a}{1+\eta_a}}\Gamma(\vec{f})(1+\eta_s)^{-1}\Big(\widehat{\Delta}^f(\vec{f})\Big)^{1+\eta_s}\right]^{\frac{\omega(1+\eta_a)}{1+\eta_a-\omega}},$$

where

$$\begin{aligned} \Gamma(\vec{f}) := {}& a_L^u(\vec{f})\left[\frac{\beta(1+\eta_a)^{-1}f_L}{r+\delta+\beta(1+\eta_a)^{-1}f_L}(\rho y_L)\right]^{\frac{\eta_a}{1+\eta_a}} \\ &+\Big(1-a_L^u(\vec{f})\Big)\left[\frac{\beta(1+\eta_a)^{-1}f_H}{r+\delta+\beta(1+\eta_a)^{-1}f_H}(\rho y_H)\right]^{\frac{\eta_a}{1+\eta_a}}, \end{aligned}$$

and write

$$\begin{aligned} \lambda(\vec{f}) &= \left(\frac{Z}{((r+\delta)k)^{\omega}}\right)^{\frac{1+\eta_a}{1+\eta_a-\omega}}\left[\gamma_a^{\frac{1}{1+\eta_a}}(1+\eta_a)^{\frac{\eta_a}{1+\eta_a}}\Gamma(\vec{f})(1+\eta_s)^{-1}\gamma_s\Big((2s-1)\widehat{\Delta}^f(\vec{f})\Big)^{1+\eta_s}\right]^{\frac{\omega(1+\eta_a)}{1+\eta_a-\omega}} \\ &= \widehat{\lambda}(\vec{f})\ \gamma_a^{\frac{\omega}{1+\eta_a-\omega}}\ \gamma_s^{\frac{\omega(1+\eta_a)}{1+\eta_a-\omega}}\ (2s-1)^{\frac{\omega(1+\eta_a)(1+\eta_s)}{1+\eta_a-\omega}}. \end{aligned} \tag{32}$$

Combining the second term of the screening curve, we get

$$\lambda(\vec{f})F(\Delta_w^H(\vec{f}))G(\Delta^f(\vec{f}))(2s-1)$$
$$= \lambda(\vec{f}) \cdot \left[\gamma_a^{\frac{1}{1+\eta_a}} (1+\eta_a)^{\frac{\eta_a}{1+\eta_a}} \lambda(\vec{f})^{-\frac{\eta_a}{1+\eta_a}} \left(\frac{\beta(1+\eta_a)^{-1} f_H}{r+\delta+\beta(1+\eta_a)^{-1} f_H} (\rho y_H) \right)^{\frac{\eta_a}{1+\eta_a}} \right]$$
$$\cdot \left[\gamma_s (2s-1)^{1+\eta_s} (\widehat{\Delta}^f(\vec{f}))^{\eta_s} \right]$$
$$= \gamma_a^{\frac{1}{1+\eta_a}} \gamma_s (2s-1)^{1+\eta_s} \lambda(\vec{f})^{\frac{1}{1+\eta_a}} \cdot (1+\eta_a)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{\beta(1+\eta_a)^{-1} f_H}{r+\delta+\beta(1+\eta_a)^{-1} f_H} (\rho y_H) \right)^{\frac{\eta_a}{1+\eta_a}} (\widehat{\Delta}^f(\vec{f}))^{\eta_s}.$$

Use equation (32) to substitute in for $\lambda(\vec{f})$:

$$\lambda(\vec{f})^{\frac{1}{1+\eta_a}} = \widehat{\lambda}(\vec{f})^{\frac{1}{1+\eta_a}} \ \gamma_a^{\frac{\omega}{(1+\eta_a)(1+\eta_a-\omega)}} \ \gamma_s^{\frac{\omega}{1+\eta_a-\omega}} \ (2s-1)^{\frac{\omega(1+\eta_s)}{1+\eta_a-\omega}}.$$

This finally yields

$$\lambda(\vec{f})F(\Delta_w^H(\vec{f}))G(\Delta^f(\vec{f}))(2s-1) = X(\gamma_a, \gamma_s, s)^{\frac{1}{1+\eta_a-\omega}} \cdot M_2(\vec{f}),$$

where

$$X(\gamma_a, \gamma_s, s) = \gamma_a \ \gamma_s^{1+\eta_a} \ (2s-1)^{(1+\eta_s)(1+\eta_a)}.$$

and

$$M_2(\vec{f}) := (1+\eta_a)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{\beta(1+\eta_a)^{-1} f_H}{r+\delta+\beta(1+\eta_a)^{-1} f_H} (\rho y_H) \right)^{\frac{\eta_a}{1+\eta_a}} (\widehat{\Delta}^f(\vec{f}))^{\eta_s} \widehat{\lambda}(\vec{f})^{\frac{1}{1+\eta_a}},$$

which does not depend on (γ_a, γ_s, s).

Finally, $M_1(\vec{f}) = \frac{F(\Delta_H^w)}{F(\Delta_L^w)}$, which does not depend on (γ_a, γ_s, s) by equation (26). This establishes the claim. □

Proof of Proposition 2. Starting from the result of Proposition 1, if $\vec{f}'$ constitutes an equilibrium, then the indifference curve holds and

$$f'_H = M_1(\vec{f}')\, f'_L + X(\gamma'_a, \gamma_s, s)^{\frac{1}{1+\eta_a-\omega}} \cdot M_2(\vec{f}').$$

Note that neither $M_1(\vec{f})$ nor the indifference curve depend on (γ_a, γ_s, s). Thus, if

$$\gamma_s' = \gamma_s \left(\frac{\gamma_a'}{\gamma_a}\right)^{\frac{1}{1+\eta_a}}, \tag{33}$$

$$s' = \frac{1}{2}\left[1 + (2s-1)\left(\frac{\gamma_a'}{\gamma_a}\right)^{\frac{1}{(1+\eta_s)(1+\eta_a)}}\right], \tag{34}$$

then the indifference curve holds and

$$\begin{aligned} f_H' =& M_1(\vec{f}')\, f_L' + X(\gamma_a, \gamma_s', s)^{\frac{1}{1+\eta_a-\omega}} \cdot M_2(\vec{f}') \\ =& M_1(\vec{f}')\, f_L' + X(\gamma_a, \gamma_s, s')^{\frac{1}{1+\eta_a-\omega}} \cdot M_2(\vec{f}') \end{aligned}$$

and thus the screening curve holds.

We still need to verify that there is ε such that $R \in (0,1)$ for any $(f_L', f_H') \in B_\varepsilon(\vec{f})$. This claim follows directly from the fact that (f_L, f_H) are part of a mixed strategy equilibrium and the fact that the expression

$$R(x) = 1 - \frac{\frac{f_L}{\lambda(x)\, F(\Delta_w^L(x))} - G(\Delta^f(x))\,(1-s)}{1 - G(\Delta^f(x))}.$$

is continuous in $x = (\vec{f}, \gamma_a, \gamma_s, s)$. Therefore, the screening curve and indifference condition hold at $\vec{f}'$ for E', E_s', and E_{γ_s}' and $R \in (0,1)$ holds true in each case. This establishes the claim that the equilibria exist and feature the same type-specific job finding rates. From this claim it follows directly that the distribution of unemployment duration is identical across all economies.

Next, we establish that all economies also feature the same job filling rates, and the same unemployment and vacancy stocks. As established above, $\vec{f}' = \vec{f}_s' = \vec{f}_{\gamma_s}'$, and hence the type-specific unemployment stocks (and total unemployment) are identical across the three equilibria. Moreover, combining (32) with (33) and (34) implies that $\lambda(\vec{f}')$ is identical across E', E_s', and E_{γ_s}'. From $\lambda = Z \cdot (v/u)^\omega$, it follows that tightness v/u is identical, and therefore the vacancy stock v is identical as well. Total hires are $H = \sum_{i \in \{L,H\}} u_i f_i$, which is identical since both u_i and f_i are identical. Thus the job filling rate H/v is identical across the three equilibria. The same applies to match efficiency, since H, v and u are all identical.

We now show that E'_s and E'_{γ_s} feature the same type-specific application rates ($\lambda F(\Delta_w^L)$, $\lambda F(\Delta_w^H)$). Given the insight that λ is constant across E', E'_{γ_s} and E'_s, this claim follows directly from the fact that the expression for $F(\Delta_w^H)$ depends on s and γ_s only through λ.

□

B.11 Mapping Cowgill et al. (2024) into the model

Cowgill et al. (2024) run a screening experiment in which "senders" produce short written pitches and "receivers" evaluate them. Some senders have access to an LLM when writing the pitch, while receivers do not. Receivers then report a posterior belief $p \in [0,1]$ that the sender is an "expert" (where true expertise is a binary variable). The authors report that $p(1-p)$ increases from 0.17 (without LLMs) to 0.18 (with LLMs), indicating that signals contained in the pitch are less informative when the LLM is available.

To map this to our model, we interpret the firm's signaling technology as analogous to the "pitch" technology in CHW's experiment. Given a symmetric prior, the firm's posterior probability of the worker being H upon receiving an H signal is s, which implies that p and s can be treated as analogous. The reported change of $p(1-p)$ implies that

$$\frac{2p'-1}{2p-1} = \frac{2\sqrt{0.25 - p'(1-p')}}{2\sqrt{0.25 - p(1-p)}}$$

and therefore the corresponding proportional change in $2s - 1$ is

$$\frac{2s^{\mathrm{AI}}-1}{2s^{\mathrm{noAI}}-1} = \frac{2\sqrt{0.25-0.18}}{2\sqrt{0.25-0.17}} = \sqrt{\frac{0.07}{0.08}} \approx 0.935,$$

which corresponds to a 6.5% decline in $2s - 1$.