Job Search Technology, Selection, and Hiring *

Sadhika Bagga
University of Rochester

Lukas F. Mann
Arizona State University

Version: March 27, 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 tools in artificial intelligence 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 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 average unemployment duration.

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 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). Finally, 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.

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. The paper contributes to three growing strands of literature: the literature on the effects of declining search frictions, the literature on the causes of long-term unemployment, and the literature on the effects of artificial intelligence on labor market matching.

First, the effects of declining search frictions on labor market outcomes have been a focus of both empirical work (Bhuller et al., 2023; Denderski and Sniekers, 2024) and theoretical work (Auster et al., 2025). Auster et al. (2025) highlight that, in a directed search framework with adverse selection, declining search frictions may inhibit the ability of firms to attract workers of their preferred type and can thus worsen matching outcomes. Our work bridges theory and empirics: Similar to Auster et al. (2025), we transparently show the conditions under which reducing search frictions may lead to adverse outcomes in matching markets, but do so in a random search setting that is designed to closely resemble a workhorse model of the labor market, the DMP model.[1] This allows us to link our mechanism to observable labor market trends, particularly the collapse in matching efficiency and the doubling in the average duration of unemployment since the 1990s.

Two closely related papers that study the labor market effects of easing search frictions are Martellini and Menzio (2020) and Birinci et al. (2025). Both papers study the impact of diminishing search frictions in settings where firms screen workers based on the id-

[1]We introduce applications and a screening process into a standard DMP model, keeping the model simple enough to be characterized graphically. Our model thus joins the ranks of papers that have recently integrated applications and interviews into models of search, such as Engbom (2021), Bradley (2025) and Cai et al. (2025).

iosyncratic quality of a match.[2] Reducing search frictions enables firms to select into better matches, which keeps unemployment exit rates stable in equilibrium even when the rate of encounters increases. Lower search frictions thus yield weakly higher matching rates and unambiguously positive welfare implications in these settings. We show that screening on permanent worker characteristics reverses this conclusion: diminishing search frictions can *decrease* matching efficiency, lengthen unemployment durations, and lower welfare. We find that, even though lower search frictions enhance the ability of an individual firm to select better workers, they reduce matching rates as firms tighten their admission criteria and lower their propensity to hire in response to a higher inflow of applications.[3] We provide empirical evidence for this channel.

Second, our paper relates to the literature on the causes of rising unemployment duration since the 1990s. A leading explanation in this literature is duration dependence, the idea that unemployment exit hazards decline as unemployment progresses (Blanchard and Diamond, 1994; Kroft et al., 2016; Jarosch and Pilossoph, 2019). Our theory offers a distinct perspective on this question: the rise in unemployment duration is driven by increasing selectivity of firms that widens the dispersion of job finding rates across workers. Using micro data from the NLSY, we provide evidence of this channel. Our findings are also consistent with the evidence from Ahn (2023) who documents that shifts in unobserved heterogeneity explain much of the increase in unemployment duration since 1980.

Third, we contribute to the literature evaluating the impact of artificial intelligence tools on matching markets. A newly emerging literature (Cowgill et al., 2024; Gans, 2024; Cui et al., 2025; Dhillon et al., 2025; Galdin and Silbert, 2025; Wiles and Horton, 2025) has focused on various aspects of how AI changes the signaling environment for job applicants and the employers who screen them. We contribute to this literature by conducting a general equilibrium analysis that accounts for two channels highlighted in this literature. First, AI makes it harder for employers to distinguish between applicants (Cowgill et al., 2024; Galdin and Silbert, 2025). Second, AI makes it easier to contact employers.

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

[2] Although the focus is different, Pries and Rogerson (2022) is another notable example of a screening model with idiosyncratic match productivity.

[3] A similar mechanism has previously been documented in empirical work by Modestino et al. (2020) who find that firms increased their job requirements in response to an inflow of applications in the aftermath of the Great Recession.

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 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 cross-sector evidence that suggests that aggregate trends 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

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. 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

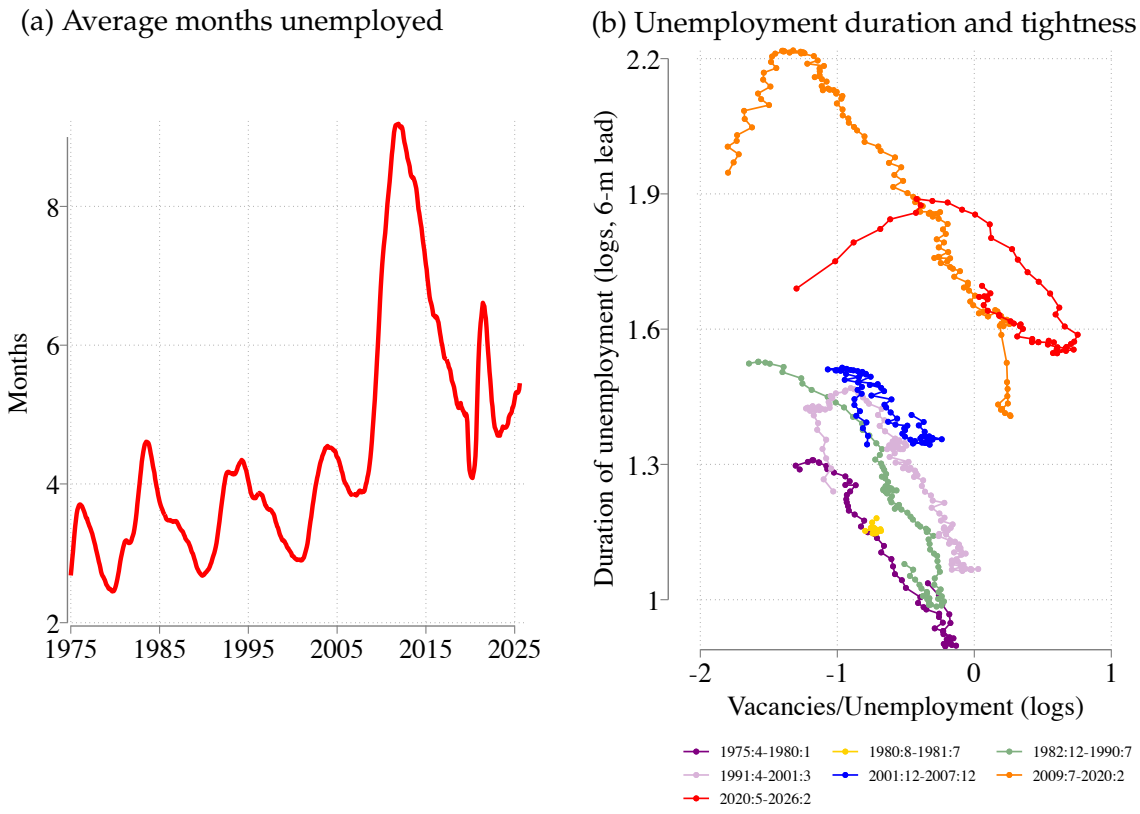


Figure 1: Unemployment duration

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.

In Appendix A.2, we rule out two plausible explanations of this trend. First, we show

[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.

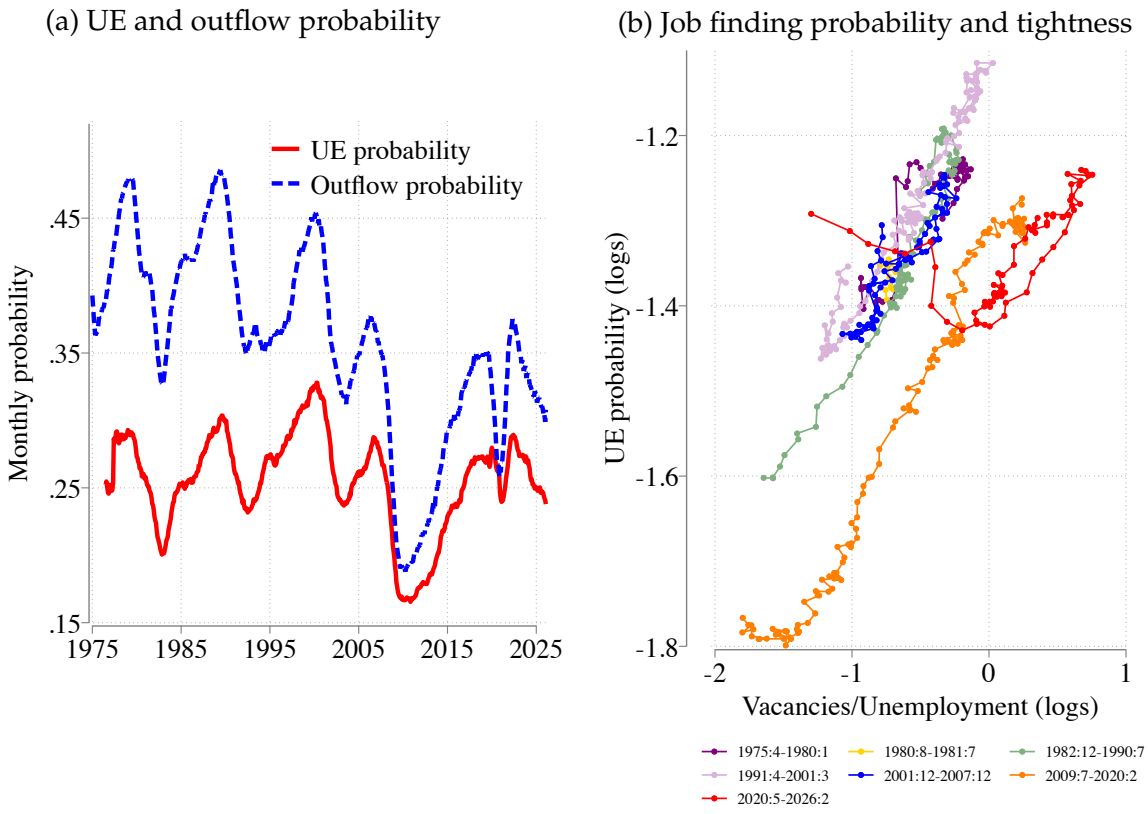


Figure 2: Job finding probability

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 6. 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.

that accounting for demographics does not resolve the puzzle: even conditional on individual demographic characteristics and education, the duration of unemployment has increased.[5] Second, we show that long-run shifts in unemployment insurance are likewise unlikely to be responsible for the observed persistent trend.

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 transi-

[5]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.

tioning 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).[6] 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.

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

[6]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 $t+1$.

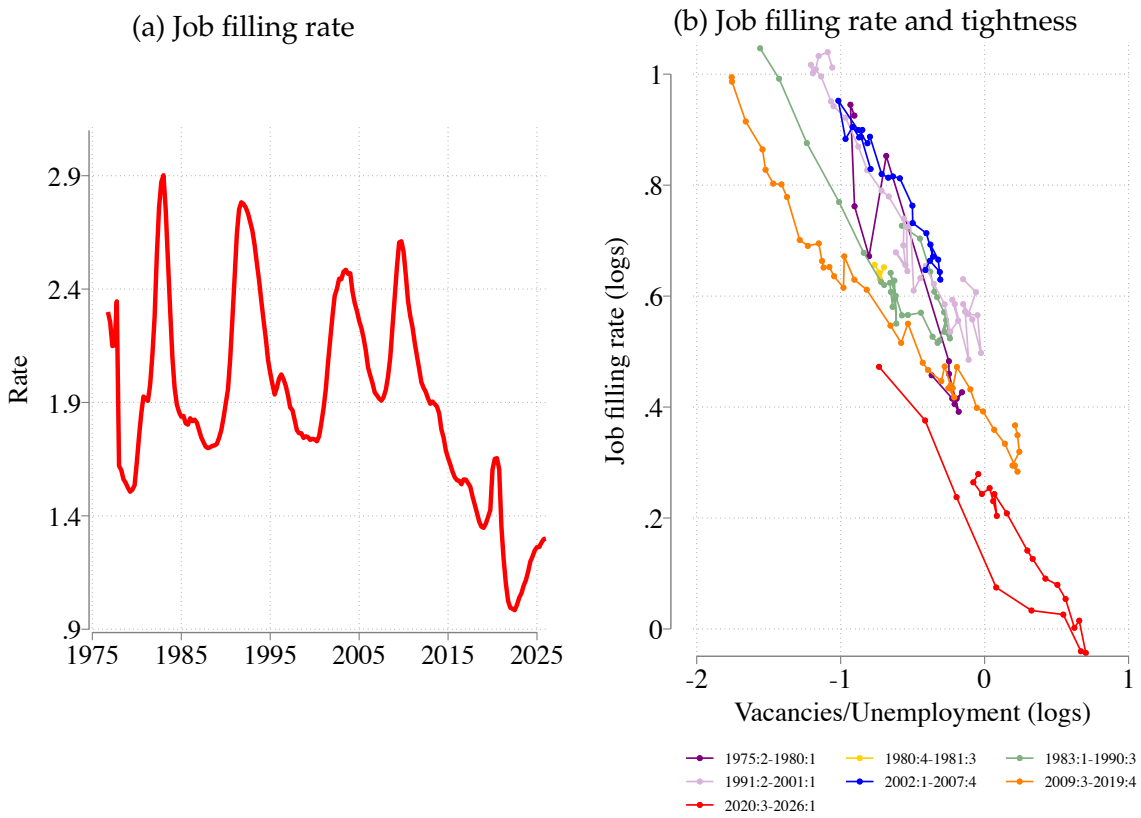


Figure 3: Job filling rate

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.

Job Openings and Labor Turnover Survey (JOLTS).[7] 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 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.[8] However, by the end of the sample, vacancies required more than a month to generate a

[7]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 where hires are measured from JOLTS and unadjusted CPS hires. The decreasing trend is observed in both series.

[8]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.

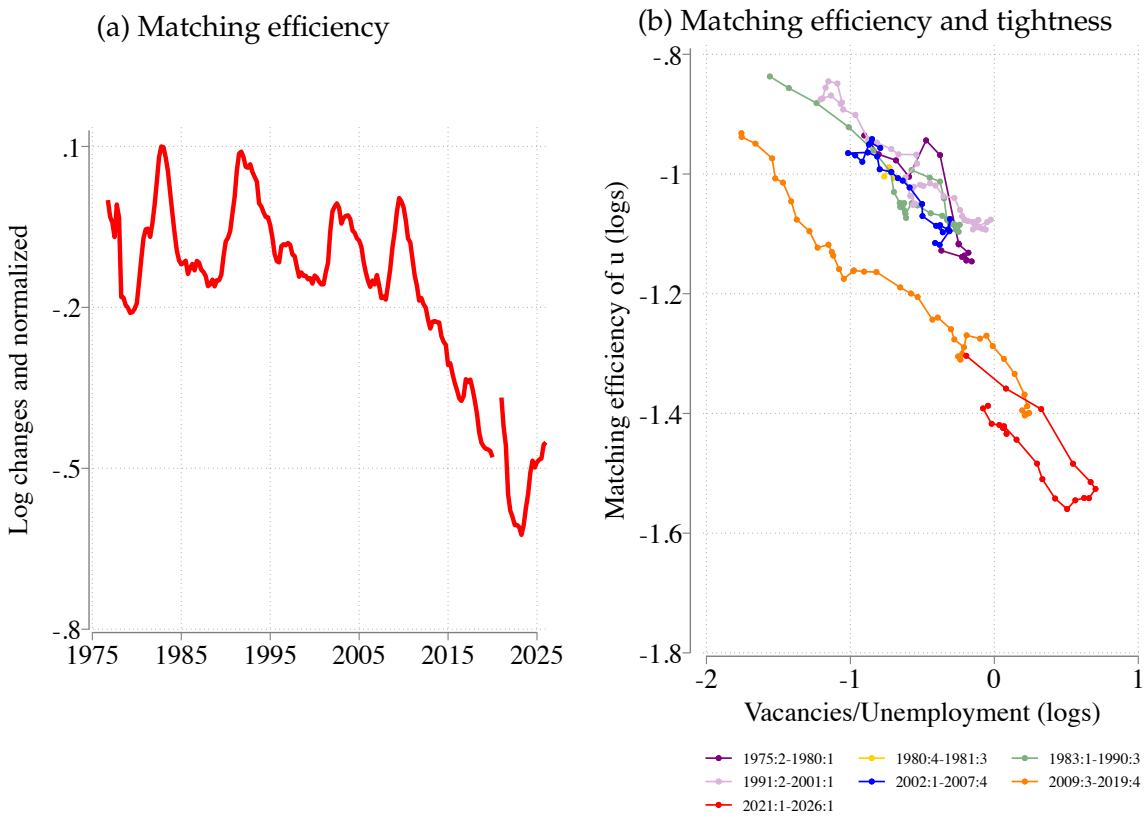


Figure 4: Matching efficiency

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.

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.

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?[9]

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.

2.2 Online job search and applications

A parallel development to the aggregate labor market trends has been the proliferation of online job search and an accompanying shift toward online job postings and online

[9] 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.

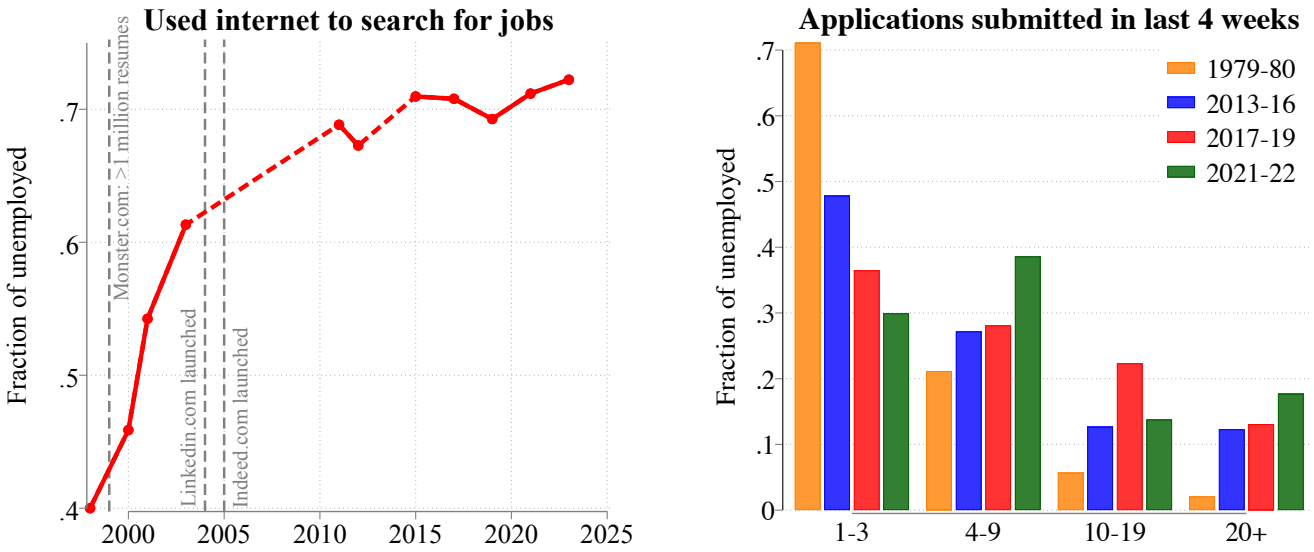


Figure 5: Internet use and applications submitted by the unemployed

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.

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 started 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.

A natural question is whether 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.[10] 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".[11] 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 adjacent HR functions such as compensation and benefits administration, and employee training. The definition of this occupation has also remained stable over time.

[10] 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.

[11] 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/).

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).[12] 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 part reflects an increasing demand for screening labor by firms. 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

[12] 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.

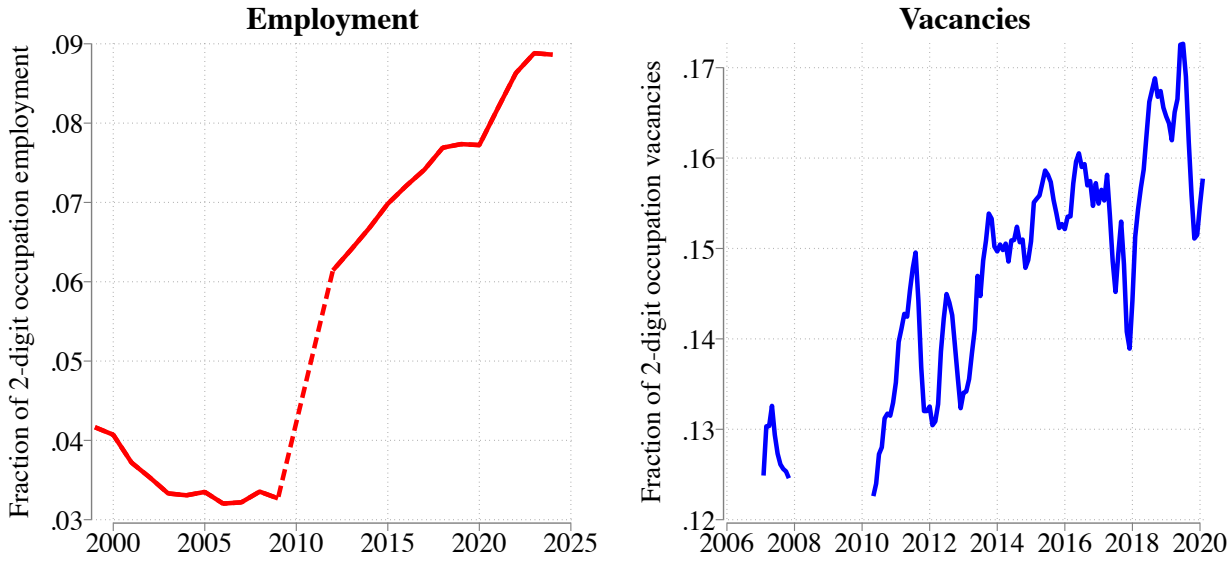


Figure 6: Employment and vacancies in "HR specialists" occupation as a fraction of all operations occupations

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.

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.[13] 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

[13]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.

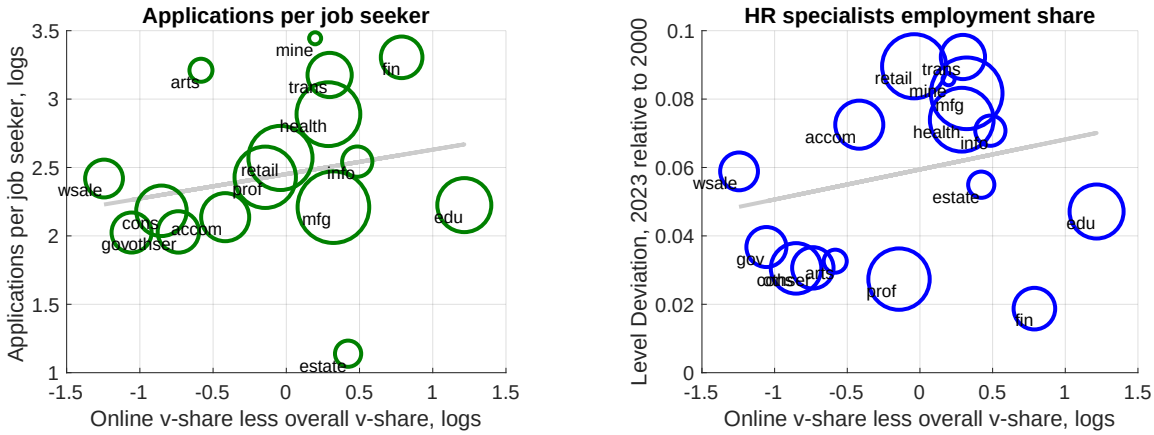


Figure 7: Applications and screening by industry

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.

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

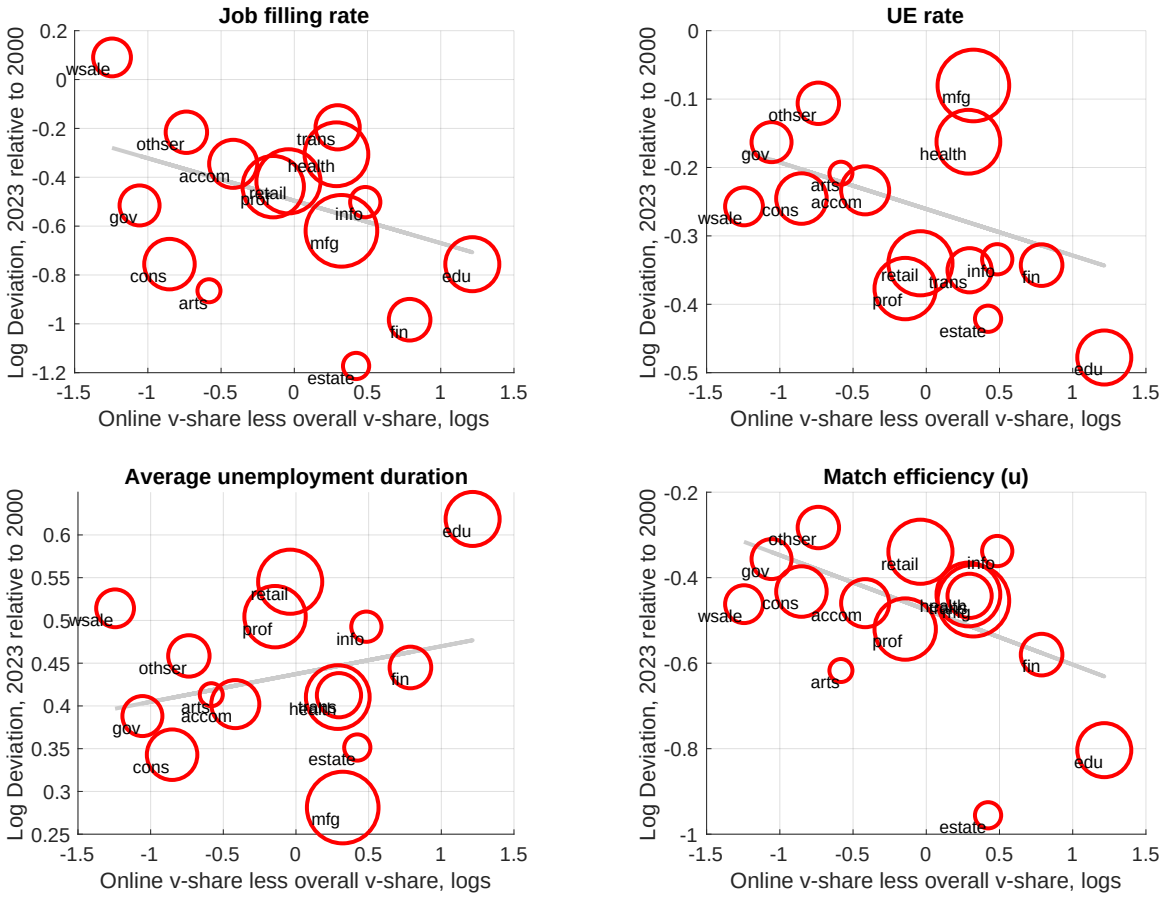


Figure 8: Labor market outcomes by industry

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.

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, workers are ex ante heterogeneous. For simplicity, we will focus on two permanent worker types who differ only in their productivity. Second, we introduce a notion of applications and screening that captures the idea that encounters between workers and firms do not automatically produce hires. Upon an encounter, workers draw a random cost and decide whether to apply. If the worker applies, the firm does not observe the worker's type but can pay a random cost to receive a signal about their 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 $b_i = (1-\rho)y_i$ 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. We will also refer to jobs as firms.[14] 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.[15]

[14] In our economy, as in most search models, one could alternatively define firms as collections of multiple jobs. However, such ownership across jobs is irrelevant for the model's outcomes. A notion of a job and a firm are therefore not cleanly distinguishable in our setting and we refer to both terms interchangeably.

[15] We model job creation through free entry with a sunk cost. 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

Meeting technology. Vacancies and heterogeneous workers meet at random in a frictional labor market to form matches. We use u to denote the measure of unemployed workers and v to denote 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. Whenever unemployed workers encounter a vacancy, they draw an application cost c from a distribution with cdf $F(c) = \gamma_a c^{\eta_a}$. We refer to γ_a as the "ease of applying" since a higher value corresponds to a lower cost of the average application. If a worker decides to pay c, the firm receives their application; otherwise, the encounter is destroyed. We denote the (endogenous) probability that a type-i worker submits an application, conditional on an encounter, as p_i^{app}.

Screening. When the firm receives an application, it does not observe the worker's type.[16] However, the firm may choose to purchase a signal which reveals some information about the worker's type to the firm. Purchasing a signal entails a one-time cost κ, which is drawn from a distribution with cdf $G(\kappa) = \gamma_s \kappa^{\eta_s}$. We refer to γ_s as the "ease of screening" since 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.[17] The signal reveals the worker's true type with probability $s \geq 0.5$. This means that

would be zero in equilibrium, which means that it would never be optimal to forego a match with positive surplus. Thus, a sunk cost is a simple way to model the 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.

[16] For simplicity, we assume that the firm cannot observe anything about the worker's individual state unless it decides to screen. 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 leaves unemployment changes with their duration of unemployment. We depart from this assumption mostly for tractability and transparency reasons, but argue in Appendix B.7 that true duration dependence does not seem to drive the trend in unemployment duration that we seek to explain.

[17] 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}.

the screening signal may be imperfect but is weakly more likely than not to reveal the worker's true type. We refer to s as the "signal quality".

Production. After the application and screening stage concludes, the firm may decide to form a match with the worker. If it does, 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 and we use β to denote the bargaining weight of workers. 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.[18]

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.[19]

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

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

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.

[18] 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. Our modeling choice has the advantage that it keeps the model simple and the mechanism transparent. 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.

[19] 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 a worker. The value to an unemployed worker of type i, U_i, satisfies

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

$$= b_i + \underbrace{\lambda F(\Delta_i^w) p_i^{\text{acc}}}_{f_i \equiv \text{job finding rate}}(W_i - U_i) - \lambda \mathbb{E}[\mathbb{I}\{c \leq \Delta_i^w\}c]. \tag{5}$$

In this Bellman equation, we introduce two new pieces of notation: First, R denotes the probability that a worker is rejected in cases where the firm does not receive a signal. Second, we define $\Delta_i^w = p_i^{\text{acc}}(W_i - U_i)$ as the net gain in value for an unemployed worker of type i from submitting an application.

An unemployed worker receives a flow value of b_i and an expected continuation value which depends on the option value of applying and potentially being accepted into a job. The acceptance probability is the sum of two terms: Either the worker's potential employer obtains a signal, in which case a worker of type L is accepted with probability $1-s$ and a worker of type H is accepted with probability s.[20] Or they do not, in which case the firm rejects the worker with probability R, regardless of type.

The optimal strategy of the unemployed worker is to apply whenever the cost c is below the expected return of an application, Δ_i^w. Thus, the probability of sending an application is $p_i^{\text{app}} = F(\Delta_i^w)$. We denote the job finding rate of an unemployed worker of type i as f_i. It is the product of the meeting rate of unemployed workers (λ), the probability of submitting an application (p_i^{app}), and the acceptance probability of a type-i worker conditional on applying (p_i^{acc}).

For the value of a matched worker of type i, the Bellman equation is

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

which captures the flow wage and the possibility that the match dissolves and the worker ends up unemployed.

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

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

[20] We argue below that 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.

which captures the flow output to the firm net of the wage paid to the worker.

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

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

where V_{signal} and V_{nosignal}, defined below, are the expected value of the firm after it decides to purchase a signal and its value without the signal, respectively. We define $\Delta^f = V_{\text{signal}} - V_{\text{nosignal}}$ as the firm's expected return to obtaining a signal about the worker. The firm optimally chooses to purchase a signal only if its cost κ is less than its net benefit, Δ^f. Therefore, the probability of purchasing a signal is $p^{\text{sig}} = G(\Delta^f)$.

We now consider the firm's optimal hiring decision in two cases: when it has or has not decided to purchase a signal. We start with the former case.

If the firm optimally purchases a signal about the worker, then it optimally accepts the worker if and only if the signal is H. This follows from two insights: First, the reason why the signal has any value to the firm is that it has an effect on the firm's optimal hiring decision: it would be irrational for the firm to buy a signal only to accept or reject the worker regardless of its outcome. Second, since $J_H \geq J_L$ and $s \geq 0.5$, the value of a policy that only accepts on L-signals is clearly dominated by a policy that never extracts a signal and accepts with the same probability.

Given these considerations, we can express the expected value of a firm that has purchased a signal as

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

where a_L^{app} is the share of L-type workers among all applicants that we characterize in Section 3.3. The firm hires the applicant in two cases: first, if they are truly type H and the signal is H (probability $(1-a_L^{\text{app}})s$, value J_H); second if they are truly type L and the signal is H (probability $a_L^{\text{app}}(1-s)$, value J_L). The remaining share of matches is rejected, yielding the value of a vacancy, V.

We now turn to the case in which the firm decides not to purchase a signal. We express

the value of a firm in this situation as

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

The firm faces a choice between two options: it can accept the applicant or reject the applicant and remain vacant. R is the probability that the firm chooses the latter option. The firm faces a trade-off between filling a job more quickly and filling a job with an applicant of higher expected quality. It can choose $R = 0$, accepting all unscreened applicants, which fills vacancies more quickly but with applicants of lower expected value to the firm. Or it can set $R = 1$, accepting only those applicants for which it has retrieved an H signal: this strategy lowers the job filling rate but increases the value that the average accepted applicant provides to the firm. Finally, it can move between these two extremes by setting $R \in (0,1)$. For the latter to be optimal, the firm must be indifferent between rejecting and accepting workers without an associated signal.

Free entry. Entrants decide whether or not to create a job by comparing the sunk entry cost with the expected value of entering and being vacant. Free entry thus implies that q adjusts to ensure $k = V$ in equilibrium.

3.3 Flows

Stationary flows imply that the following conditions must hold in equilibrium:

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

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

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

where a_L^u denotes the share of L-types among the unemployed, a_L^{app} denotes the share of L-types among applicants, and u_i denotes the mass of unemployed type i workers in steady state.

3.4 Equilibrium

Definition (equilibrium).
An *equilibrium* is a tuple (λ, q, V, V_{signal}, V_{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^{app}), such that the above Bellman equations ((3), (4), (5), (6), (7), (8), (9), (10)) and flow equations ((11), (12), (13)) hold and λ and q are consistent with the meeting function (equation (2)).

A key trade-off faced by the firm is the choice between either filling a job quickly but with workers of lower expected productivity, or filling it more slowly but with workers of higher expected productivity. Since changes in this trade-off will be key to the mechanism we study, we focus on parametrizations of the model that feature equilibria in which firms choose $R \in (0,1)$. We call these equilibria mixed strategy equilibria, because the firm follows a mixed strategy when deciding whether to reject workers without a signal.

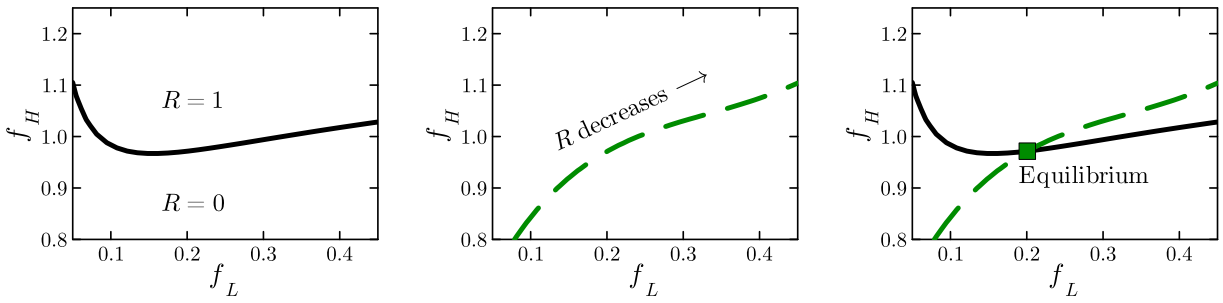
Definition (mixed strategy equilibrium).
A *mixed strategy equilibrium* is an equilibrium that features $R \in (0,1)$.

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

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

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

where $J_i(f_i), \lambda(\vec{f}), p_i^{\text{app}}(\vec{f}), p^{\text{sig}}(\vec{f})$ and $a_L^{\text{app}}(\vec{f})$ 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 shows the two curves for our baseline calibration, discussed in Section 3.6.



(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.

The indifference curve. The indifference curve imposes an indifference condition on the firm: In a mixed strategy equilibrium, firms have to be indifferent between accepting a no-signal worker (and obtaining the RHS value of equation (IC)) and rejecting them (yielding value $V = k$). The curve thus traces out combinations (f_L, f_H) for which the firm obtains a value of k by accepting no-signal workers. Along different points of the state space, two margins affect this value: first, f_L and f_H affect the type-mix of applicants $(a_L^{\text{app}}(\vec{f}))$; second, f_L and f_H affect each type's bargaining position and therefore change the rents that a firm can extract from a match $(J_i(f_i))$. At all points along the indifference curve, a value of R strictly between 0 and 1 is consistent with firm optimality. In contrast, the firm strictly prefers rejecting unscreened workers ($R = 1$) for states of the economy where (f_L, f_H) lies above the curve and strictly prefers accepting unscreened workers ($R = 0$) for states below the curve.

The indifference curve is downward-sloping on the left hand side of the diagram. There, f_H is much larger than f_L and thus most of the applicants in equilibrium are type-L workers with a poor bargaining position. Along the curve to the right, the likelihood of encountering a type-H worker increases (f_L rises, f_H falls) but the bargaining position of type-L workers improves, keeping the firm indifferent. Further to the right still, any increases in f_L now lead to a substantially increased likelihood of randomly encountering an H-type worker in equilibrium and thus have to be compensated by a mildly larger f_H to keep the firm indifferent.

The screening curve. The screening curve imposes the condition that, at any value of R, f_H must equal f_L, adjusted for the relative likelihood of applying, plus a screening advantage that H-type workers enjoy relative to L-type workers. The screening curve traces out how changes in R affect the job finding rates of L and H type workers, taking into account optimal application rates, encounter rates, and screening strategies. It is upward sloping: a larger willingness of the firm to admit workers without a signal (a lower value of R) raises the job finding rates of both worker types in equilibrium. A lower value of R disproportionately raises the job finding prospects of L-type workers relative to H-type workers, since their chances of finding a job critically depend on their likelihood of being admitted without a signal.

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

3.5 The consequences of an application cost shock

We now study how this economy responds to a fall in applications costs. Figure 10 illustrates the resulting shift of the equilibrium.

We first consider the partial equilibrium effect if the firm's strategy were to remain unchanged: Holding R fixed, the effect on the job finding rates of both worker types is positive. More encounters lead to more applications which to first order raises the job finding rates of both worker types. Imposing an unchanged value of R, the economy would move from the initial equilibrium to the red triangle in the upper right corner. This point lies to the top of the screening curve, because cheaper applications benefit both types similarly, while moving along the screening curve would disproportionately raise the job finding rates of L-type workers.

However, this point is not an equilibrium, because the old value of R is no longer optimal from the perspective the firm. In this state of the economy, the firm is strictly better off waiting rather than admitting a worker without a signal. A "flight to quality" ensues: as firms reject more workers without a signal, R rises and equilibrium is restored. At the new equilibrium point, the job finding rates of both worker types are now strictly lower than before. For L-type workers, whose job finding prospects critically depend on

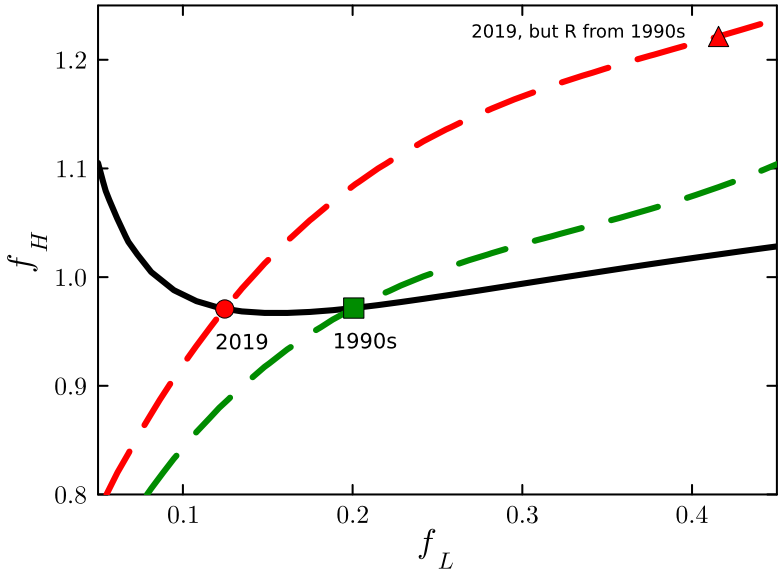


Figure 10: Effect of an application cost shock

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.

their chances of being admitted without a signal, the job finding rate becomes much lower than the job finding rate in the initial equilibrium.

3.6 Calibration

We now discipline the model quantitatively by targeting salient moments from the data. Table 1 summarizes parameters and their targets.

We calibrate the model to two periods, corresponding to the 1990s and a pre-AI modern steady state (2019), respectively. Our aggregate targets are averaged over the periods Dec. 1998 - Dec. 2000 and Dec. 2017 - Dec. 2019, respectively.

We organize parameters into two groups: First, parameters that can be normalized or set externally; second, parameters calibrated to match moments that belong to either the initial or final steady state. All parameters are assumed constant across steady states and target only one 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 the 1990s and the 2019 steady state. This allows us to isolate the effects of online job search through the lens of the model.

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.

We begin by discussing the parameters that can be normalized or are set by convention. First, the signal quality s can be normalized to one. The fact that this is in fact a normalization may seem surprising at first. However, the choice of s only changes the mapping between different points on the screening curve and the implied value of R. Since $R \in (0,1)$ is necessary for a mixed strategy equilibrium, the choice of s thus imposes restrictions only on the *length* of the screening curve, not its shape. In order to keep the parameter space for which we observe mixed strategy equilibria maximally large, we thus impose $s = 1$ and note that any value of s sufficiently close to one would deliver observationally equivalent model outcomes. Next, we normalize the encounter productivity parameter Z to a value that guarantees a worker encounter rate of 1000. This choice, too, guarantees only that, in our calibration, $F(\Delta_i^w) < 1$ but has no impact on the model beyond this.[21] The final normalization is to set the productivity of the low type y_L to one. We then impose 4 parameters using conventional values used in the search and match-

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

ing literature: We set the worker bargaining weight β to 0.5, the elasticity parameter of the encounter function ω to 0.3 (Petrongolo and Pissarides, 2001), the discount rate to 5% annually, and the job destruction rate to 0.035, consistent with a long-run average of the separation rate according to the Bureau of Labor Statistics.

We quantify a second group of parameters by targeting 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%.[22] We use the share of L-type workers, α, to match the average duration of unemployment.[23] The entry cost parameter k targets the average job finding rate in the 1990s steady state.[24] The screening cost parameter γ_s targets the average job filling rate in the 1990s steady state.

The last two remaining parameters to pin down are the application and screening elasticities, η_a and η_s. We discuss both in order.

First, the application cost dispersion parameter η_a targets the cross-sectional overdispersion in applications, which is 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. The logic behind using this calibration target is that $\mathcal{D}$ is a sufficient statistic that directly identifies the coefficient of variation in application rates, as we discuss in Appendix B.3. η_a governs the dispersion of application rates of L versus H types, so it is directly relevant for the model counterpart of $\mathcal{D}$. In the data, we compute $\mathcal{D}$ from the SCE using the 2018 and 2019 surveys.[25]

Next, we choose a target for the screening cost dispersion parameter η_s. We find that, in practice, η_s determines the slope of the screening curve and thus the relative sensitivity of type-specific job finding rates to productivity shocks. Given this observation, we choose η_s to match the difference in elasticity of the job finding rates between H and L types in response to a marginal productivity shock. We target a value of 0.76, consistent with evidence in Mann (2026).

[22] We define the replacement rate as the average production of unemployed workers relative to the average wage. We find that our quantitative results are not sensitive to the targeted value.

[23] 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.

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

[25] Since we have data from the SCE only for the later period, we target the model counterpart of $\mathcal{D}$ in the 2019 steady state of the model.

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 13.0, respectively, which we take from Birinci et al. (2025).

4 The effects of online job search in 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 hold true in 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: the composition of hires shifts towards more desirable workers with shorter unemployment durations and firms commit more resources to screening applicants (Section 4.3). Finally, we explore welfare consequences of the shock (Section 4.4).

4.1 Aggregate variables

Figure 11 shows the changes of aggregate labor market variables induced by the application shock in the model and compares it with the corresponding changes in the data. We discuss each moment in turn.

First, the job finding rate decreases by 20% in the model versus 27% the data. Two effects contribute to this result: On the one hand, as illustrated in Figure 10, job finding rates decline substantially for L-type workers, which make up part of the pool of unemployed. On the other, the share of L-worker types in unemployment also rises, which leads to a compositional effect since L-type workers have lower job finding rates.

Second, the duration of unemployment rises by 73% in the model versus 71% the data. Both in the model and the data, this increase is larger than the decline in the job finding rate. We revisit this finding in Section 4.2 where we show that heterogeneity plays a large role in generating a wedge that increases the average duration of unemployment relative to the inverse job finding rate.

Third, the job filling rate declines by 22% in the model and by 20% in the data. As we

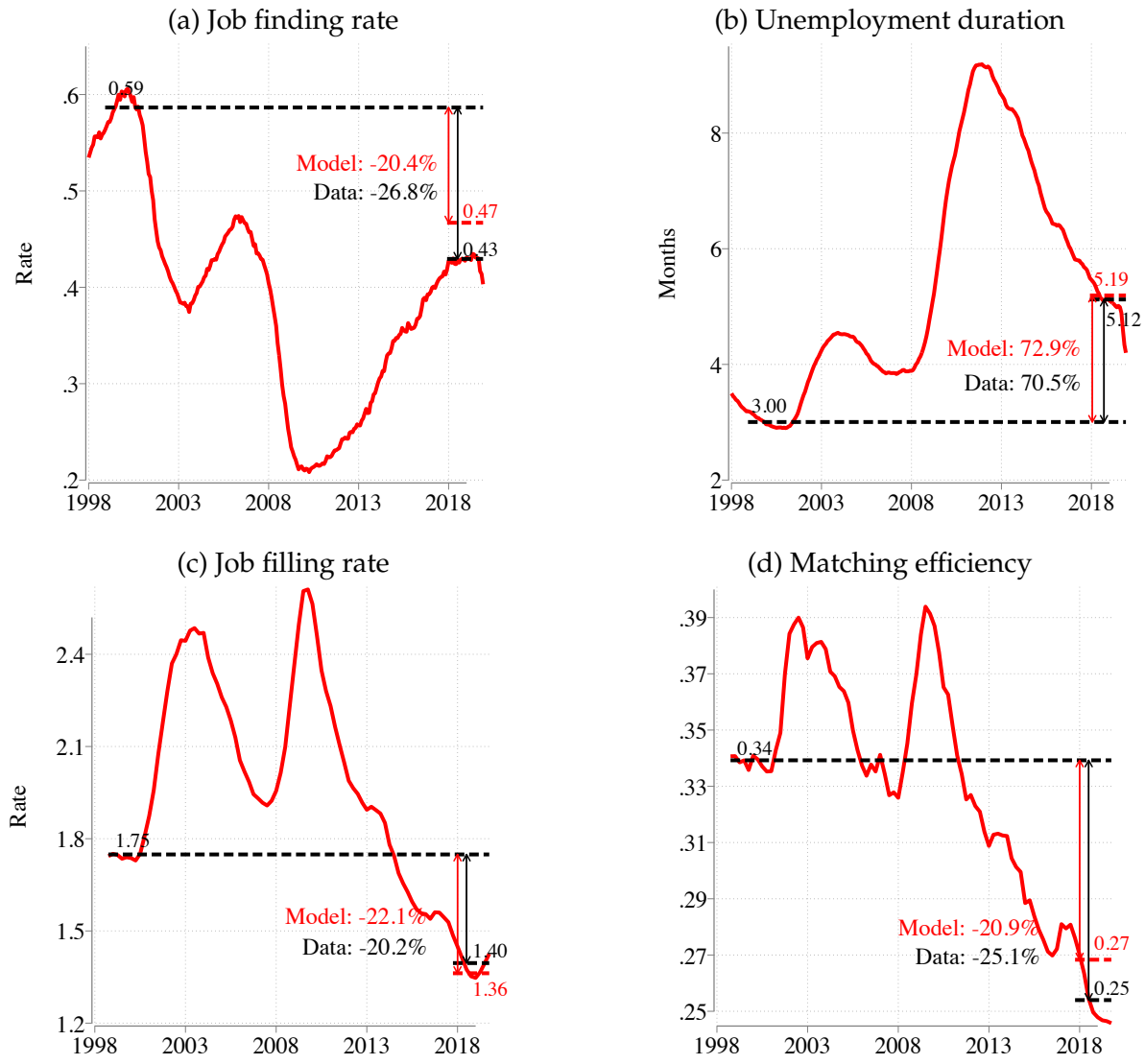


Figure 11: Comparison of aggregates across the two steady states

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

show in Appendix B.4, the job filling rate in our model equals

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

The rate at which jobs are filled depends on the encounter rate and the probability that an encounter converts into a hire. In equilibrium, the encounter rate is determined by the return on signals (Δ^f) through a standard free entry channel, discussed in Appendix B.4. In our quantification, the encounter rate rises from the shock and thus exerts upward pressure on the job filling rate as the return on purchasing a signal declines. However, two opposing effects reverse the overall direction and lengthen vacancy durations: first, a substantive decline in the admission probabilities of workers without a signal; second, a decline in the conversion probability of signaling events.

Fourth, matching efficiency declines by 21% in the model versus 25% in the data.[26] Conditional on the number of market participants, the number of hires in the economy decreases as firms become more selective.

4.2 Worker-side implications

We now explore and test the implications of the model for the worker side. We begin by describing the micro-level effects on workers induced by the shock. Table 2 shows model-predicted changes in worker-side moments.

The rate of applications rises for both types of workers. Overall, it doubles, which is consistent with the data and a targeted moment. However, this increase in the application rate does not translate into higher job finding rates because it is offset by a corresponding increase in the rate of rejections. L-type workers especially struggle to find work: even though they send about twice as many applications, they experience a four times lower success rate per application, halving their job finding rate. For H-type workers, on the other hand, the increase in the application rate and the rise in rejections roughly offset one another. Thus, the job finding rate of L-type workers (f_L) declines substantially, and their unemployment duration increases, while the labor market prospects of H-type workers remain roughly unchanged

[26] Since the model matches the decline in both the job finding and job filling rate, it must mechanically also match the decline in matching efficiency. This follows from the insight that

$$\text{ME} \propto \frac{\text{hires}}{v^{\omega} u^{1-\omega}} = \left(\underbrace{\frac{\text{hires}}{v}}_{\text{job filling rate}} \right)^{\omega} \cdot \left(\underbrace{\frac{\text{hires}}{u}}_{\text{job finding rate}} \right)^{1-\omega}.$$

Variable	Level (1990s)	Level (2019)	Change
Application rate (L)	4.410	8.728	+97.9%
Application rate (H)	9.225	19.310	+109.3%
Application rate	6.820	13.010	+90.8%
Job finding rate (L)	0.201	0.125	-37.9%
Job finding rate (H)	0.972	0.971	-0.1%
Job finding rate	0.587	0.467	-20.4%
Unemployment duration (L)	4.983	8.021	+61.0%
Unemployment duration (H)	1.029	1.030	+0.1%
Unemployment duration	3.004	5.192	+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.

Decomposing unemployment duration. This divergence of type-specific job finding rates is important for the behavior of unemployment duration. This becomes clear from the following decomposition, derived in 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 of unemployment is determined by two factors. First, the inverse aggregate job finding rate, which increases from the shock. Second, heterogeneity in job finding rates, as captured by the second term, which likewise increases from the shock. The model thus predicts that the increase in unemployment duration should exceed the fall in the job finding rate. This is consistent with the data, as Appendix Figure A.3 illustrates.

The decomposition of equation (15) is not limited to the environment of our model: in Appendix B.6, we generalize the set of assumptions under which a decomposition of this type holds true. Specifically, we show that even with job finding rates that vary by unemployment duration, and with an unlimited number of worker types, the following

decomposition can be derived:

$$\text{unemployment duration} = (\text{job finding rate})^{-1} \cdot \underbrace{\frac{1}{2}(1+\text{CV}^2)}_{\text{duration dependence}} \cdot \underbrace{\left(1+\text{Cov}_i(f_i, -f_i^{-1})\right)}_{\text{type heterogeneity}}. \tag{16}$$

Here, the covariance operator is taken with respect to the type composition of the pool of unemployed and CV is the coefficient of variation of unemployment spell lengths, which is a sufficient statistic for the strength of duration dependence in job finding rates. This decomposition reveals that changes in duration dependence are an alternative explanation for the increasing wedge between unemployment duration and the job finding rate. However, in Appendix B.7, we analyze such changes in micro-data from the National Longitudinal Survey of Youth (NLSY) and find no evidence for trends in duration dependence.

In what follows, we do, however, show evidence that type heterogeneity has increased in micro-data, as implied by the model. 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 (16) are invariant to replacing f_i with f_i^{-1}. We focus on duration for the purposes of the coming section, but note that we get similar results when studying job finding rates.

Evidence for duration heterogeneity in the NLSY. We document time variation in the heterogeneity in unemployment duration by analyzing unemployment spells in micro-data from the NLSY. We rely on both the NLSY79 and NLSY97 cohorts, which follow roughly 6,000 and 8,000 individuals, respectively, from the time they first enter the labor market until the present day. The NLSY records employment status at a weekly frequency, allowing us to measure unemployment spells at a granular level. We restrict the sample to individuals aged 25 and older and exclude the minority and military over-samples.

We classify workers into two groups based on the mean unemployment spell duration observed over their lifetime. Through the lens of the model, such a classification is a suitable proxy for their underlying type. Figure 12 plots the unemployment duration of each group over time.

We observe that the duration gap between groups widens over time for both cohorts. By comparing both cohorts' trajectories, we can see that this is in large part due to time rather than age effects. Figure 12 thus provides direct evidence that the dispersion in

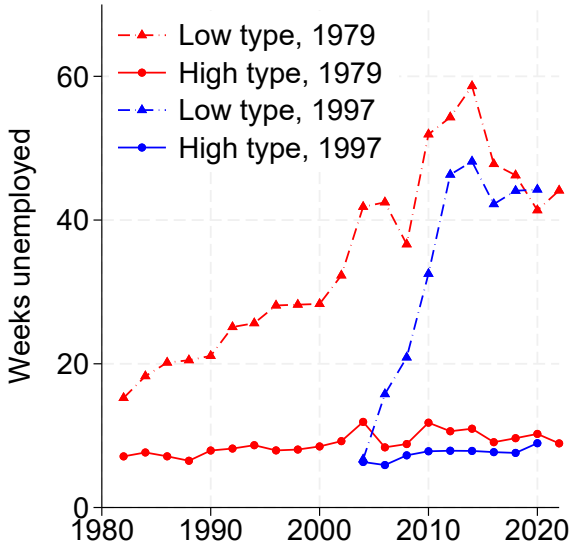


Figure 12: Type-specific heterogeneity in the NLSY

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.

average unemployment duration by group has widened over time, as implied by the model.

4.3 Firm-side implications

We now evaluate and validate the model's implications for firm side variables. We begin by analyzing the composition of hires and then show that the model is consistent with the evidence we presented in Section 2.3 that suggests that firms commit more resources to screening candidates.

Compositional effects. We first explore evidence pertaining to a central and testable firm-side implication of the model: Firms should become more successful at screening out less desired worker types.

The firm-side moments listed in Table 3 illustrate the mechanism: as a result of the shock, firms increase R, which raises the probability that a hire is someone on whom the firm has received a positive signal. Thus, even though the application pool shifts towards

Variable	Level (1990s)	Level (2019)	Change
P(signal | hire)	0.471	0.602	+27.9%
Prob. applicant is type L	0.323	0.399	+23.7%
Prob. hire is type L	0.171	0.159	-7.0%
Duration rank of av. hire	0.442	0.424	-4.0%

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

Notes: See Table 2. "Duration rank of av. hire" corresponds to $\mathcal{R}$ in the model.

type-L workers, the share of type-L workers among hires decreases. An implication of this shift is that the flow of hires should tilt towards workers with comparatively low unemployment durations, as measured by the quantile that these workers occupy in the duration distribution of unemployed workers at the moment of their hire. As Table 3 shows, the average unemployment duration rank of hired workers does indeed decrease across steady states in the model. In what follows, we test both implications in the data.

We begin by constructing proxies for a worker's type in data from the Current Population Survey (CPS). The CPS is an overlapping panel of individuals who are interviewed for four months, then left unobserved for eight months, and finally interviewed again for another four months. Workers' wages are observed in interview month 4, conditional on employment. To construct type measures that capture worker quality conditional on a worker's labor market, we residualize these wage measures with respect to their occupation and education by estimating

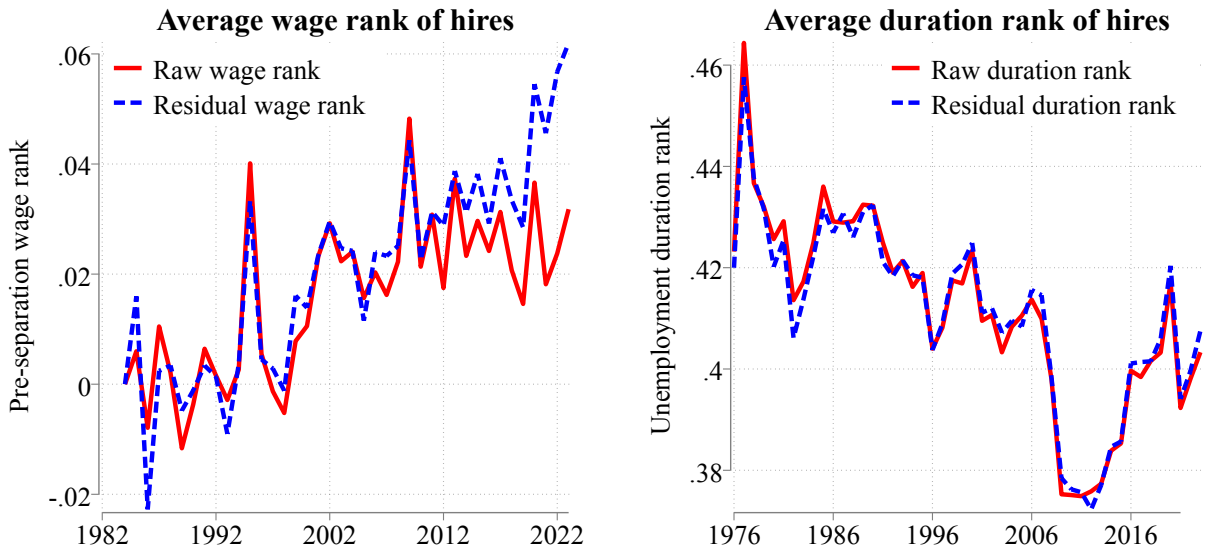
$$w_{it} = \beta_{\text{occ}(i,t)} + \beta_{\text{edu}(i)} + \tilde{w}_{it}.$$

Within each period, we then rank workers by the residual wage $\tilde{w}_{it}$ relative to other workers in this month, and interpret this residual wage rank as a proxy for a worker's type.

Next, we document how the composition of hires as measured by this proxy shifts over time. As a basis for this exercise we use the universe of all hiring events from unemployment or non-participation observed during interview months 5-8 (9-12 months after observing their wage). Figure 13a documents the average wage rank and residual wage rank among such hires.

The average wage rank measure of workers among hires rises steadily after the 1990s, which establishes that hires increasingly shift towards workers with higher previous wages, as implied by the model.

A drawback of this exercise is that the selection on month-4 employment introduces



(a) Wage rank of hires in the CPS

(b) Duration 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 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. The right 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.

selection bias that is hard to control for. Even if our rank measure perfectly corresponded to worker types, an ideal version of this exercise would require us to measure (potential) wages for every worker in interview month 4, not only those who are observed as employed.

We therefore complement this analysis with a separate one that studies a more easily measurable aspect of the compositional shift in hires: as the composition of hires shifts towards H-types, workers who are hired should occupy lower duration ranks among the unemployed. Formally, defining $\mathcal{R}$ as the average duration rank of hired workers, we show in Appendix B.8 that, in the model, $\mathcal{R}$ is characterized by the following formula:

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

For $\mathcal{R}$, too, heterogeneity in job finding rates among the unemployed plays a central role. More heterogeneity in job finding rates among unemployed workers lowers the average

Variable	Level (1990s)	Level (2019)	Change
Monthly applications	20.334	37.956	+86.7%
Acceptance probability	0.075	0.029	-61.8%
Vacancy duration	0.572	0.734	+28.4%
Screening resources per hire	0.383	0.492	+28.4%

Table 4: 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.

duration rank of hires. In the model, $\mathcal{R}$ thus shifts from a value of 0.44 to 0.42.

Figure 13b plots the corresponding measure of hires in the CPS. It declines steadily from 0.42 in the 1990s to 0.40 in the more recent period, reflecting both a level and a decline that quantitatively match the model closely.

Resources committed to screening. Last, we validate that the model predictions are consistent with the evidence presented in Figure 6; that is, firms commit more resources towards screening activities.

To establish this claim, we first note that, in the model, the resource consumption of screening operations per unit of time is invariant and depends only on k, r, δ, and η_s:[27]

$$\text{screening resource consumption rate} = k(r+\delta)\eta_s. \tag{17}$$

Thus, the amount of resources that the firm commits to filling one job is exactly proportional to the time it takes to fill that job.

Table 4 shows the model-predicted vacancy-level changes arising as a result of the shock. Vacancies receive more applications, but lower their acceptance rates more than proportionally, leading to longer vacancy duration overall, consistent with the data. Through the lens of our model, online job search, which raises the time it takes to fill vacancies, thus also raises the amount of resources committed to screening operations by firms. This implication of the model is thus qualitatively consistent with the evidence on growing HR employment presented in Section 2.3.

[27] The rate of screening resource consumption is $\tilde{q} \cdot \mathbb{E}\left[\kappa \cdot \mathbb{I}\{\kappa \leq \Delta^f\}\right]$, 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 (29) in Appendix B.4 yields equation (17).

Variable	Level (1990s)	Level (2019)	Change
Worker welfare (avg)	1803.9	1801.8	-0.1%
Worker welfare (L)	207.5	196.8	-5.2%
Worker welfare (H)	2176.8	2176.8	-0.0%
Wage (L)	0.93	0.91	-2.5%
Wage (H)	9.24	9.24	-0.0%

Table 5: Welfare 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

Having validated the micro- and macro-economic predictions of our model, we are now in a position to assess welfare consequences of the online job search shock through the lens of our model. We preface this analysis by acknowledging that it comes with an asterisk: Our results can be viewed as a lower bound given that screening on match quality, rather than worker characteristics, would introduce a force that raises welfare when search frictions diminish (Martellini and Menzio, 2020). We argue that our analysis is nonetheless insightful for two reasons: first, it highlights that in a world where firms screen primarily on worker characteristics, diminishing search frictions can lead to a Pareto deterioration that makes every participant of the economy weakly worse off;[28] second, it shows that such welfare losses can be quantitatively large for some groups of workers.

We define overall and type-specific worker welfare[29] as

$$\mathcal{W} = \alpha^L_{\text{pop}} \mathcal{W}(L) + (1 - \alpha^L_{\text{pop}}) \mathcal{W}(H), \qquad \mathcal{W}(i) = \frac{\delta}{\delta + f_i} U_i + \frac{f_i}{\delta + f_i} W_i. \tag{18}$$

Table 5 shows the changes in worker welfare between the 1990s and 2019 steady state. Welfare declines only moderately overall and stays nearly constant for H-type workers, but declines by over 5% for L-type workers.[30] This substantial welfare loss comes from

[28] Auster et al. (2025) reach a similar result in a directed search model with adverse selection. One way to read our paper is that it formalizes the connection between this theoretical result and long-run trends in aggregate labor market variables.

[29] Free entry guarantees that firms make zero profits in equilibrium and can be omitted from the welfare calculation.

[30] The welfare level of L-type workers is an order of magnitude smaller in our calibration compared to

two sources. First, wages of L-type workers decrease as their bargaining position diminishes. Second, workers lose labor market access as their job finding rates plummet.

Overall, a counter-intuitive welfare result arises in this economy: diminishing search frictions endogenously lead to longer screening times and more selectivity, which can congest the labor market enough to make every participant of the economy 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.

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 beyond γ_a that may characterize the job search or screening technology. 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 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

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

that of H-type workers, which reduces their overall impact on aggregate welfare.

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

Proof. See Appendix B.9. □

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.9. □

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. In what follows, we now make a best effort attempt to quantify them.

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.[31] 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.[32] 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.10. 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 its 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. 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 6, 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

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

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

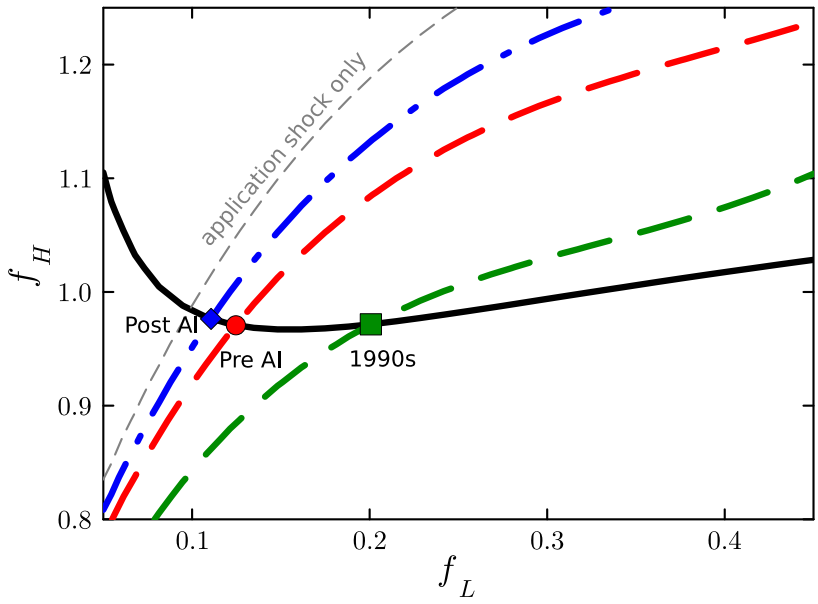


Figure 14: Shifts in the screening curve resulting from AI

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 γ_a^{AI}**	**Effect of s^{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 6: 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).

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

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*.

Denderski, P. and F. Sniekers (2024). Declining search frictions, unemployment and self-employment. *The Economic Journal 134*(659), 1100–1145.

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.

Modestino, A. S., D. Shoag, and J. Ballance (2020). Upskilling: Do employers demand greater skill when workers are plentiful? *Review of Economics and Statistics 102*(4), 793–805.

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.

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

A Empirical appendix

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

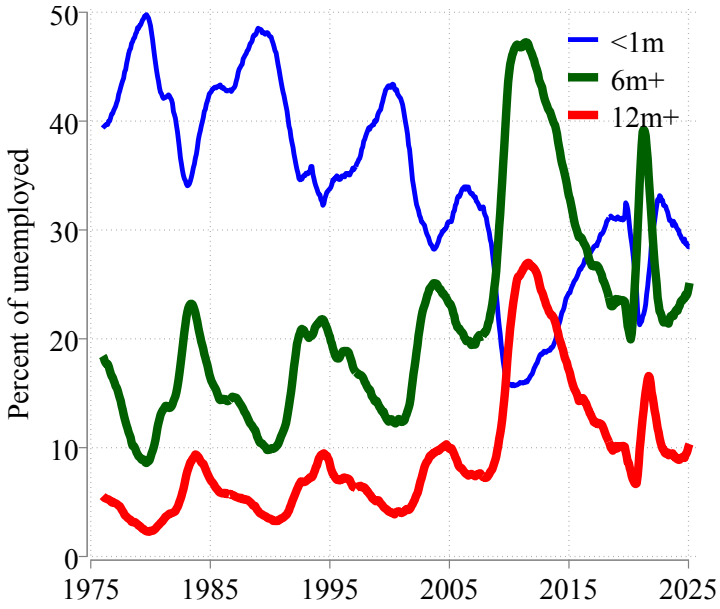


Figure A.1: Long-term vs. short-term unemployed

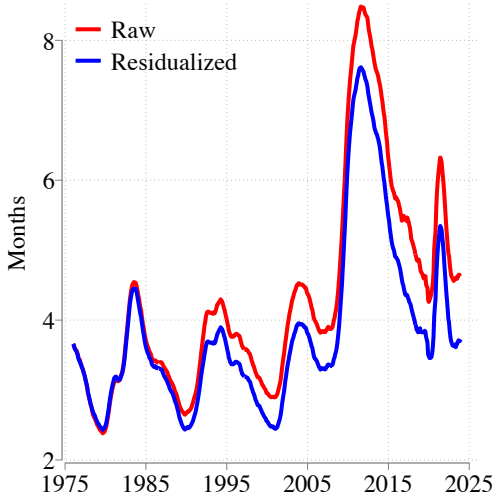
Notes: CPS and authors' calculations. Duration groups are defined by elapsed weeks of unemployment reported in the CPS, converted to months. 12 month centered moving averages are reported.

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 (16)), 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

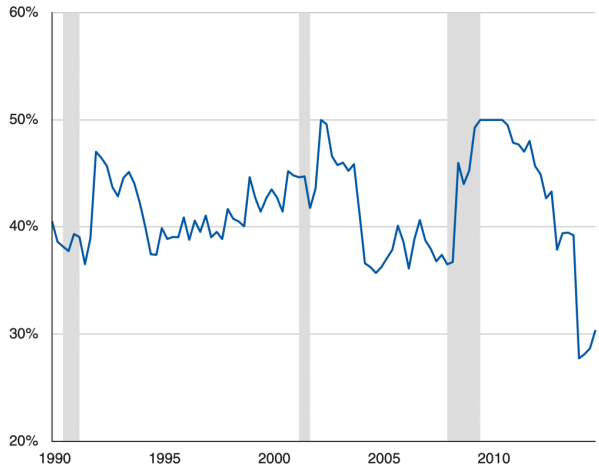


Figure A.2: Demographics-adjusted duration and UI replacement rate

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.

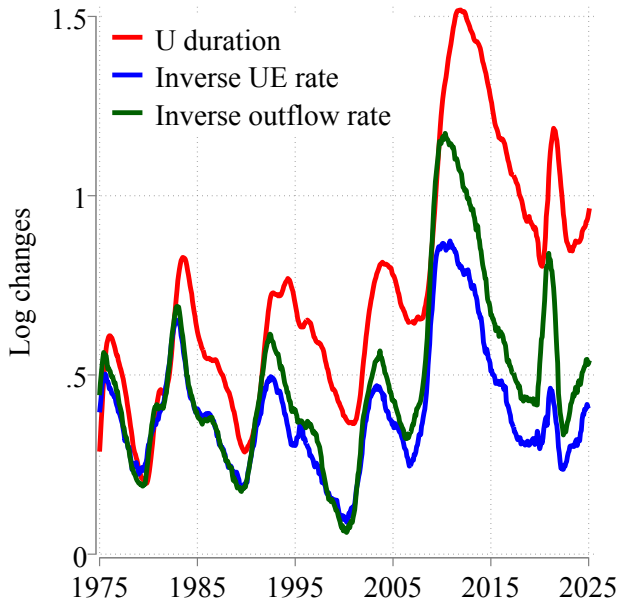


Figure A.3: Unemployment duration and inverse job finding rate

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.

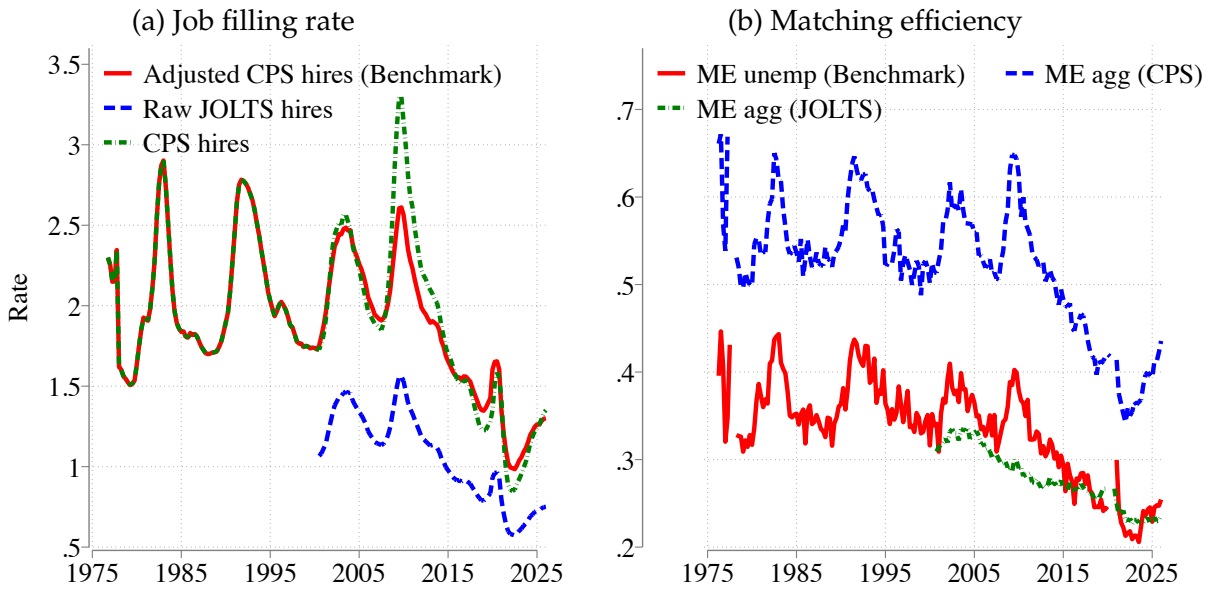


Figure A.4: Alternative measures of job filling rate and matching efficiency

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

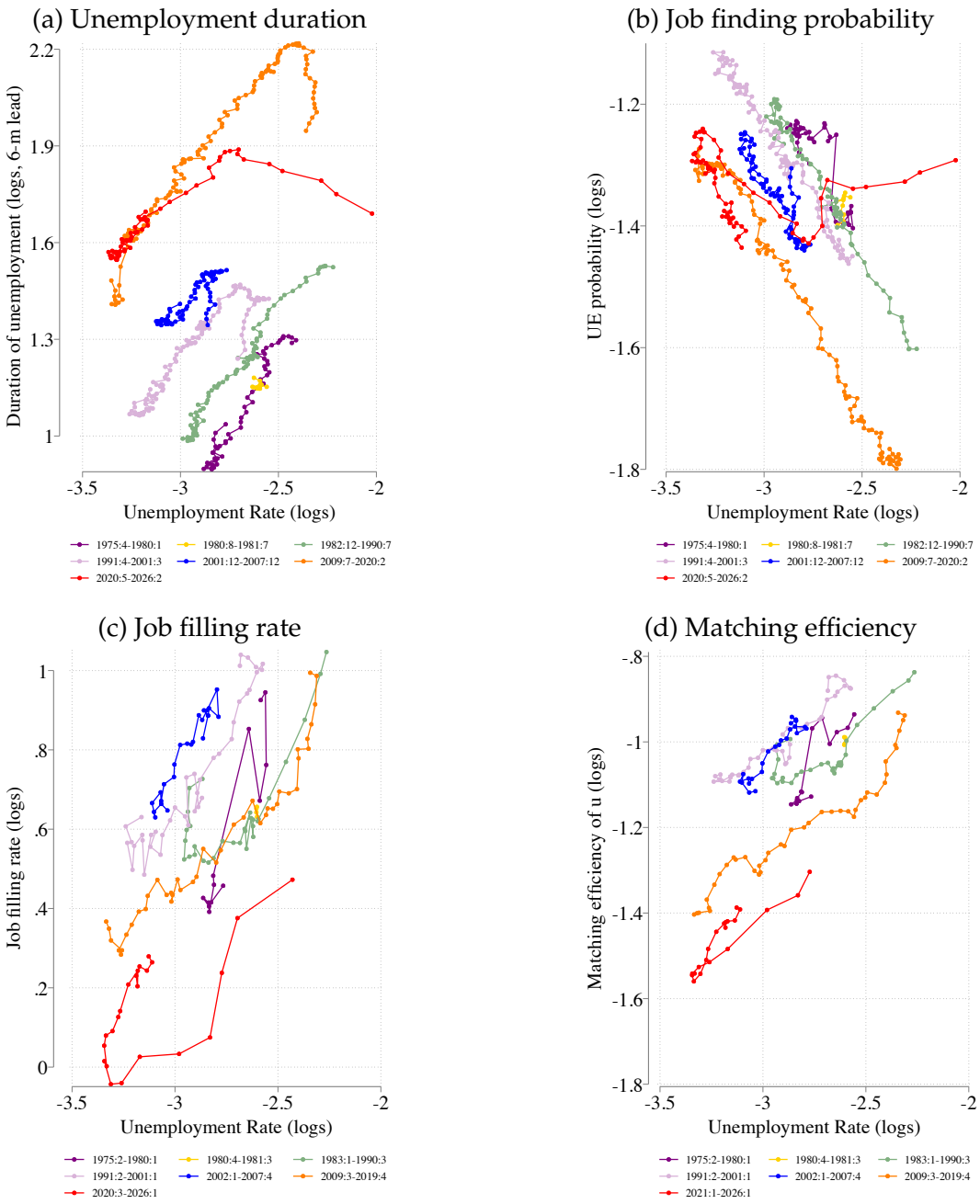


Figure A.5: Covariation of aggregate labor market variables with unemployment rate

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.

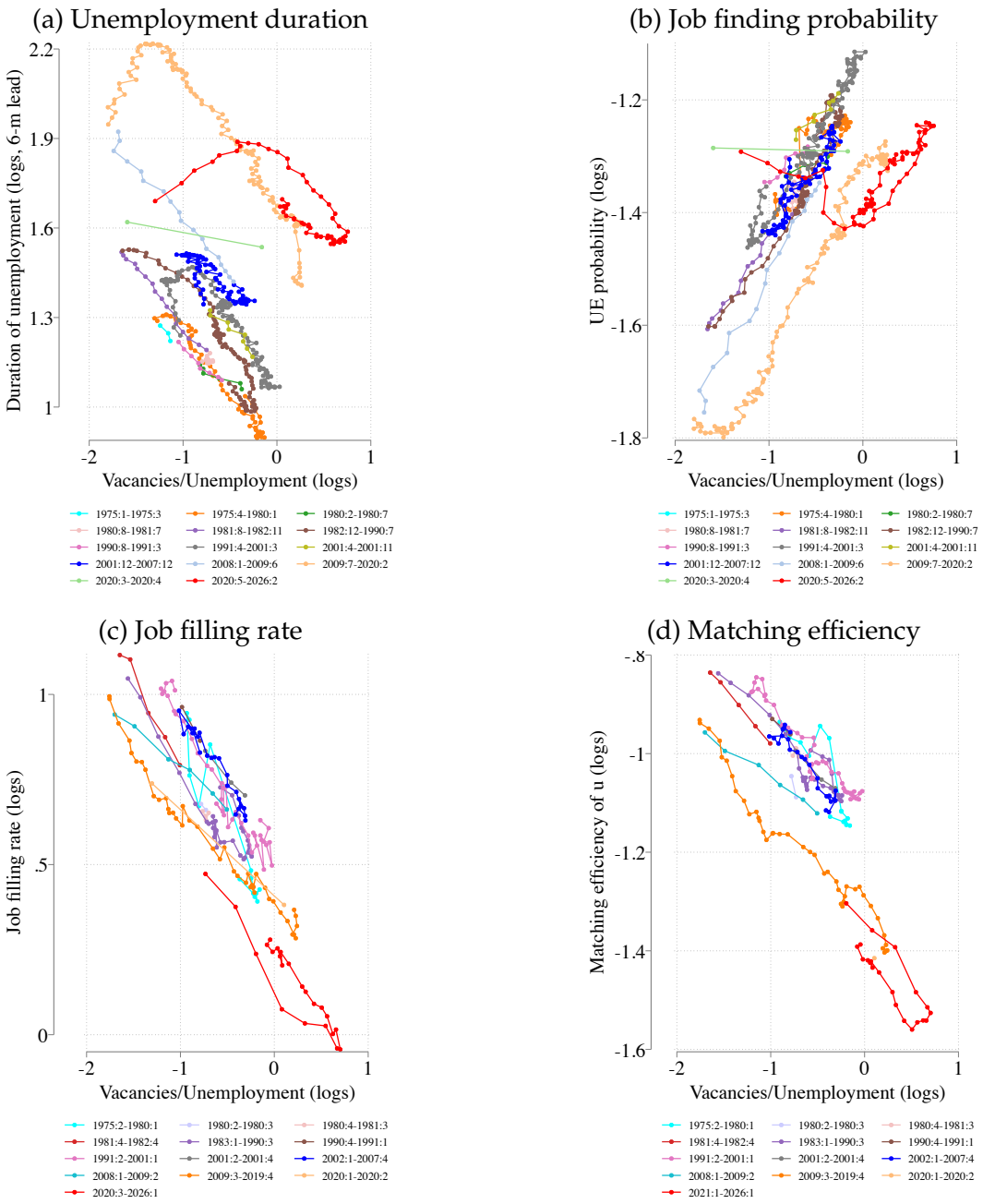


Figure A.6: Covariation of aggregate variables with labor market tightness

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.

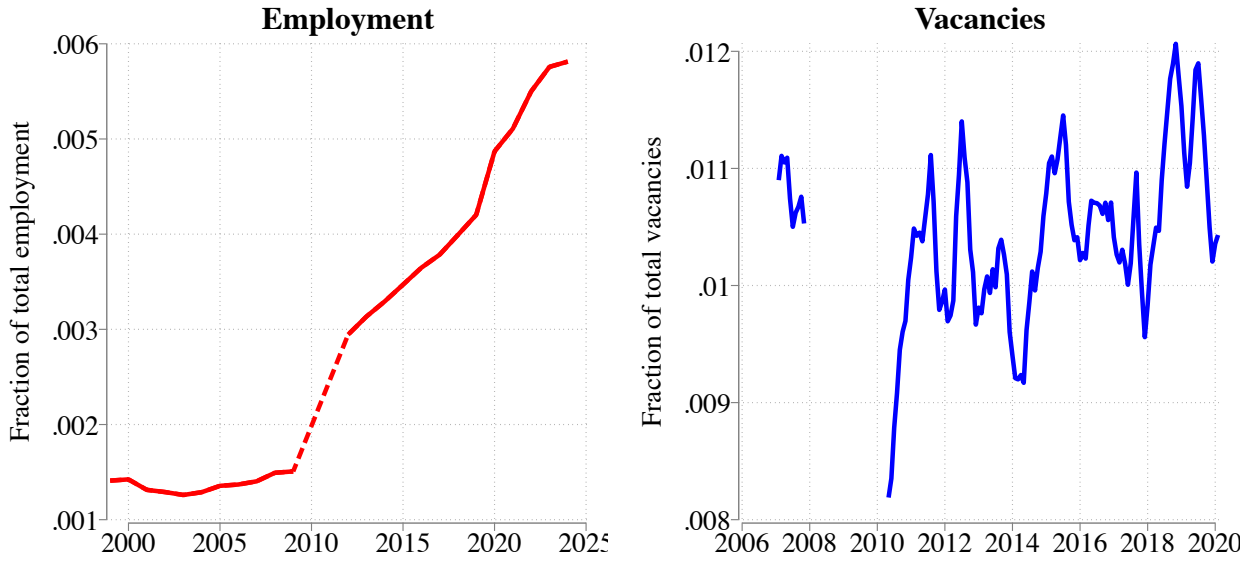


Figure A.7: HR specialists as a fraction of all occupations

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 (19)-(27) 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:

$$J_i(f_i) = \left(\frac{(1-\beta)\rho y_i}{r+\delta+\beta(1+\eta_a)^{-1} f_i} \right) \tag{19}$$

$$a_L^u(\vec{f}) = \frac{\alpha_L^{\text{pop}}(f_H+\delta)}{\alpha_L^{\text{pop}} f_H + (1-\alpha_L^{\text{pop}}) f_L + \delta} \tag{20}$$

$$a_L^{\text{app}}(\vec{f}) = a_L^u(\vec{f}) \left(a_L^u(\vec{f}) + (1-a_L^u(\vec{f})) \left(\frac{f_H}{f_L} \right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{r+\delta+\beta(1+\eta_a)^{-1} f_L}{r+\delta+\beta(1+\eta_a)^{-1} f_H} \right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{y_H}{y_L} \right)^{\frac{\eta_a}{1+\eta_a}} \right)^{-1} \tag{21}$$

$$\Delta^f(\vec{f}) = a_L^{\text{app}}(\vec{f})(1-a_L^{\text{app}}(\vec{f}))(1-\beta)(2s-1)\cdot \left(\frac{1}{r+\delta+\beta\,(1+\eta_a)^{-1} f_H}(\rho y_H) - \frac{1}{r+\delta+\beta(1+\eta_a)^{-1} f_L}(\rho y_L) \right) \tag{22}$$

$$\lambda(\vec{f}) = \left(\frac{Z}{((r+\delta)k)^\omega} \right)^{\frac{(1+\eta_a)}{1+\eta_a-\omega}} \left[\gamma_a^{\frac{1}{1+\eta_a}} (1+\eta_a)^{\frac{\eta_a}{1+\eta_a}} \left(a_L^u(\vec{f}) \left(\frac{\beta(1+\eta_a)^{-1} f_L}{r+\delta+\beta(1+\eta_a)^{-1} f_L}(\rho y_L) \right)^{\frac{\eta_a}{1+\eta_a}} + (1-a_L^u(\vec{f})) \left(\frac{\beta(1+\eta_a)^{-1} f_H}{r+\delta+\beta(1+\eta_a)^{-1} f_H}(\rho y_H) \right)^{\frac{\eta_a}{1+\eta_a}} \right) (1+\eta_s)^{-1} \gamma_s \Delta^f(\vec{f})^{(1+\eta_s)} \right]^{\frac{\omega(1+\eta_a)}{1+\eta_a-\omega}} \tag{23}$$

$$\Delta_L^w(\vec{f}) = \left(\gamma_a^{-1}(1+\eta_a)\lambda(\vec{f})^{-1} \left(\frac{\beta(1+\eta_a)^{-1} f_L}{r+\delta+\beta(1+\eta_a)^{-1} f_L} \right) (\rho y_L) \right)^{\frac{1}{1+\eta_a}} \tag{24}$$

$$\Delta_H^w(\vec{f}) = \left(\gamma_a^{-1}(1+\eta_a)\lambda(\vec{f})^{-1} \left(\frac{\beta(1+\eta_a)^{-1} f_H}{r+\delta+\beta(1+\eta_a)^{-1} f_H} \right) (\rho y_H) \right)^{\frac{1}{1+\eta_a}} \tag{25}$$

$$p_i^{\text{app}}(\vec{f}) = F(\Delta_i^w(\vec{f})) \tag{26}$$

$$p^{\text{sig}}(\vec{f}) = G(\Delta^f(\vec{f})) \tag{27}$$

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 (24), (25) and (26). We can immediately use this result to derive

$$\frac{F(\Delta_H^w)}{F(\Delta_L^w)} = \left(\frac{f_H}{f_L}\right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{r+\delta+\beta(1+\eta_a)^{-1}f_L}{r+\delta+\beta(1+\eta_a)^{-1}f_H}\right)^{\frac{\eta_a}{1+\eta_a}} \left(\frac{y_H}{y_L}\right)^{\frac{\eta_a}{1+\eta_a}}. \tag{28}$$

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 (19). 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 (22) and (27).

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 (23).

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 (20). 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 (28), establishes equation (21).

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,

$$k(r+\delta) = (1+\eta_s)^{-1}\tilde{q}G(\Delta^f)\Delta^f. \tag{29}$$

Equation (29) 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:

$$\begin{aligned}\text{job filling rate} &= \tilde{q}\left(\left(1-G(\Delta^f)\right)(1-R) + G(\Delta^f)[a_L^{\text{app}}(1-s) + (1-a_L^{\text{app}})s]\right) \\ &= \underbrace{k(r+\delta)(1+\eta_s)(G(\Delta^f)\Delta^f)^{-1}}_{\text{encounter rate}}\left(\underbrace{\left(1-G(\Delta^f)\right)(1-R)}_{\text{admissions without signal}} + \underbrace{G(\Delta^f)}_{=p^{\text{sig}}}\underbrace{[a_L^{\text{app}}(1-s) + (1-a_L^{\text{app}})s]}_{\text{conversion prob. of signals}}\right).\end{aligned} \tag{30}$$

Equation (30) 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,

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

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 each 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.[33] (A2) imposes the assumption

[33] 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

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

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*

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

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

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

□

Intuitively, heterogeneity matters for the duration of unemployment because the type-

specific job finding rate and unemployment duration are inversely related. 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 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.

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.

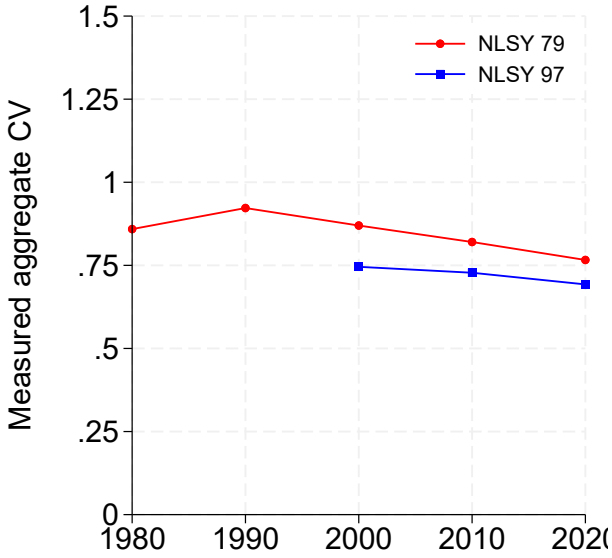


Figure B.1: Estimated CV in the NLSY

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.

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 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

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

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,

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

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

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

and

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

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

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

We now simplify this expression by using the identity

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

so

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

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 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

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

and write

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

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 (28). 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.10 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

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

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