Destabilizing Search Technology*

Tristan Potter†
Drexel University

Click here for the latest version.

February 5, 2024

Abstract

Modern job search technologies enable job seekers to monitor the arrival of newly posted vacancies. This paper conceptualizes search as a monitoring decision and shows that monitoring technologies give rise to a novel source of strategic complementarities in search and can thus lead to potentially destabilizing multiplicity of equilibria. The model provides a theory of belief-driven fluctuations in labor supply that can permanently shift the path of the economy, and offers an explanation for persistently weak wage growth despite low unemployment during the recovery from the Great Recession.

Keywords: Search and matching; online job search; hysteresis; beliefs
JEL Classification: E24, J64, E71, O33

*Thanks to Marco Airaudo, Jim Albrecht, Rich Barnett, Ryan Chahrour, Ian King, Ioannis Kospentaris, André Kurmann, Guido Menzio, Edouard Schaal, Ricardo Serrano-Padial, Susan Vroman, Ronald Wolthoff, and various conference and seminar participants for helpful comments on earlier versions of this paper. Any remaining errors are my own.

†School of Economics, LeBow College of Business, Drexel University. Email: tristan.l.potter@drexel.edu. Phone: (215) 895-2540.

1 Introduction

Since the turn of the century, online job search has become ubiquitous in the U.S. labor market. Between 2000 and 2010, the share of unemployed workers in the U.S. using the internet to search for work increased from 25% to 75% (Faberman and Kudlyak, 2016; Kuhn and Mansour, 2013). Further, as of 2013, 70% of all job openings in the U.S. were posted online (Carnevale et al., 2014). Understanding the consequences of this shift to online search requires understanding how various online search technologies affect the matching process and thus equilibrium labor market dynamics. Motivated by these questions, a growing literature has begun studying the channels through which such technologies affect the labor market. Recent examples include work considering how declining search costs can render workers more selective (Menzio and Martellini, 2020); how online recruiting can improve screening of applicants (Pries and Rogerson, 2022); and how online matching contributes to the existence of phantom vacancies (Cheron and Decreuse, 2017; Albrecht et al., 2017).[1]

This paper studies a feature of online search technologies not previously explored in the literature: Monitoring technologies that enable job seekers to find and apply to jobs soon after they are posted. A number of technologies facilitate monitoring: For instance, personal computers and broadband connections enable job seekers to easily and frequently check for new listings; job search engines sort listings by the date on which they were posted, thus facilitating finding the most recent; and matching platforms offer "job alerts" to notify job seekers of new listings that match their profiles.[2] An important consequence of such technologies is that they allow workers who actively monitor new vacancies to see and apply for these jobs *before* those who do not, a dimension of the search decision not present in traditional models. This paper conceptualizes search as a monitoring decision—that is, as a decision governing how quickly a worker is able to see and apply to newly posted vacancies. There are two main results: (i) Monitoring decisions are characterized by strategic complementarities that give rise to multiple equilibria, and (ii) the resulting multiplicity can destabilize the economy by exposing the labor market to self-fulfilling fluctuations in beliefs among jobless workers.

The first main result—that monitoring technologies can lead to multiple equilibria—can be understood in a stylized one-period game in which u unemployed workers compete for $v < u$ vacant jobs. I operationalize the notion of search as a monitoring decision by assuming that workers choose between either matching in the morning (when jobs are first posted) or enjoying a day of leisure before matching with remaining vacancies in the evening. This is a simple way to capture the fact that a decision to monitor new vacancies confers a first-mover advantage to workers who do so. Suppose, for now, that matching is frictionless in both the morning and the evening, and let w denote the value of finding a job, $\theta \equiv v/u$ the probability of finding a job if a worker applies to jobs at the same time as other workers (i.e., in either the morning by monitoring or the evening by waiting), and κ the value of leisure. Table 1 depicts this game, with a particular worker's choice represented by the rows and the (symmetric) choices of all other workers depicted by the columns.

[1]See the discussion of related literature below for a more complete list of contributions to this literature.

[2]As early as 2003, Kuhn (2003) identified the potential importance of job alerts, which he refers to as "electronic job search agents." Job alerts are now offered by virtually all major job search platforms. For example, Monster.com tells searchers that they can "Get relevant jobs matching your profile and criteria straight in your inbox with our Free Job Alerts," and visitors to Indeed.com are greeted by a pop-up inviting them to enter their email address to "Be the first to see new jobs in [your location]."

Potential equilibria are represented by diagonal entries in the matrix.

Table 1: Monitoring game

Worker i	Everybody else: Monitor	Everybody else: Wait
Monitor	θw	w
Wait	κ	$\kappa + \theta w$

Suppose all other workers wait to search in the evening. If an individual worker does the same, she enjoys a day of leisure and competes with other workers for jobs in the evening, yielding an expected payoff of $\kappa + \theta w$. On the other hand, if she decides to apply for jobs as soon as they are available in the morning, she gets a first-mover advantage and is thus assured to find a job, but forgoes leisure, giving her a payoff of w. Instead, suppose all other workers monitor and thus apply for jobs as soon as they become available in the morning. If our unemployed worker does the same, she now must compete with these workers in the morning and receives no leisure, yielding an expected payoff of θw. If instead she enjoys leisure and waits, there are no remaining jobs left in the evening, so she only receives κ. It is straightforward to show that this game has two symmetric equilibria for $\kappa \in [(1-\theta)w, \theta w]$. The possibility of multiplicity arises because of strategic complementarities in monitoring decisions: If a worker believes that others are waiting until the evening to search, then forgoing leisure to actively monitor the arrival of jobs in the morning is unnecessary because jobs will still be available in the evening; the worker should simply wait like everyone else. By contrast, if a worker believes that others are waking up early to monitor new postings, then doing the same becomes necessary to avoid falling to the back of the queue.

Section 2 formalizes this monitoring game in a simple one-period model featuring endogenous wages and entry by firms.[3] First, I show that monitoring technology leads to an aggregate match function with increasing returns, implying an upward-sloping best-response function for workers' monitoring decisions. This is true even with endogenous wages and entry, which will tend to militate against multiplicity: As more workers choose to monitor, unemployment becomes less palatable, which depresses wages. Falling wages reduce the incentive to monitor both directly by reducing the value of employment but also indirectly by stimulating job creation that alleviates the congestion caused by monitoring. This is an important result, as it implies that multiplicity is possible in an equilibrium environment that shares key features with the quantitative model I study later in the paper. Second, I show that multiplicity cannot occur in the free-entry limit, implying that inelastic entry is a necessary condition for multiplicity. Finally, I provide intuitive sufficient conditions under which multiplicity can occur related to the elasticity of entry and dispersion in monitoring costs.

The second main result of the paper is that the multiplicity introduced by the availability of monitoring technologies can destabilize the economy. To demonstrate this point, Section 3 embeds a version of the monitoring game described above in a quantitative equilibrium model of the labor market. Workers who choose to monitor observe the arrival of vacancies while other unemployed workers only have access to a standard frictional matching process that operates at the end of each period. Monitoring technologies in the quantitative model thus function to both enable workers who monitor to match quickly (as highlighted in the model and discussion above) and to reduce

[3]The game in Table 1 is a limiting case of the model in Section 2.

matching frictions. The model also features dynamic processes for unemployment and vacancies, bargained wages, and an entrepreneurial sector that gives rise to a flexible model of vacancy creation that nests various entry processes considered in the literature as special cases (e.g., free entry and "Diamond entry"). When no workers monitor the arrival of new vacancies, the model reduces to a standard search and matching model of the labor market in the Diamond-Mortensen-Pissarides (DMP) tradition,[4] augmented with a generalized vacancy-creation process.

Section 4 calibrates the quantitative model. The calibrated model has three steady states such that, across the high- and low-monitoring steady states, wages and unemployment are positively correlated. This observation reflects differences in labor supply sustained by self-confirming beliefs among unemployed workers about how dire their job-finding prospects will be if they do not actively monitor the arrival of vacancies and are thus relegated to the back of the queue for jobs. Different beliefs induce different monitoring decisions that, in equilibrium, confirm the beliefs that led to them. I show that social welfare is higher in the low-monitoring, high-unemployment steady state than in the high-monitoring, low-unemployment steady state, reflecting strong negative congestion externalities from monitoring that are not fully offset by the matching efficiency and job creation gains that monitoring brings. The model thus reverses the Pareto-ranking in the canonical treatment of multiplicity in Diamond (1982a).

Section 5 shows that monitoring can destabilize the labor market. First, I show that the model features global indeterminacy, in the sense that from a range of initial states, the economy can converge to different steady states. This can destabilize the economy in two ways. First, it implies that changes in beliefs among workers can permanently change the path of the economy. In this way, the model provides a novel theory of endogenous labor supply shocks. Second, under a simple equilibrium selection criterion, the model also implies that when the unemployment rate is sufficiently high, there is only one equilibrium path—the path leading to the high-monitoring, low-unemployment steady state—while multiple paths remain for the range of unemployment rates associated with normal business cycle fluctuations. This implies that a sufficiently large, but transitory, adverse demand shock can force the economy onto a path with permanently higher monitoring than the path it began on, thus providing a novel theory of hysteresis.

These observations can potentially offer insight into the recovery from the Great Recession. After unemployment peaked at 10% in late 2009, it ultimately fell substantially *below* its pre-recession level, while the vacancy rate rose substantially *above* its pre-recession level. Meanwhile, as was noted at the time,[5] even as late as 2019, wage growth remained weak relative to what would have been expected based on vacancy and unemployment data. These observations are difficult to square with both traditional models of the labor market and models of multiple steady states generated by demand-side mechanisms. In the present model, a demand shock that drives the economy to 10% unemployment forces the economy to the high-monitoring steady state. This implies not only a slow recovery, but also one that overshoots the original levels of unemployment and vacancies from which the economy started, and in which wages never fully recover.

In Section 6, I discuss evidence in support of the mechanism. While direct evidence on monitoring is virtually non-existent, I review some recent indirect evidence suggesting that there is a first-mover advantage in search, and then suggest three approaches that could be taken to bring more direct evidence to bear on the mechanisms in the paper in future research.

[4] Diamond (1982b); Mortensen (1982); and Pissarides (1985).

[5] For example, see John Robertson's discussion in a post on the Atlanta Fed's Macroblog on November 25, 2019.

Related literature. This paper lies at the intersection of the macroeconomic literature that examines the possibility of multiple equilibria and self-fulfilling fluctuations arising in the labor market, the growing literature on how the internet affects job search and the matching process, and the macroeconomic literature that considers the role of labor supply shocks in aggregate fluctuations.

The paper is most closely related to the macroeconomic literature interested in mechanisms through which multiplicity can arise in models of the labor market. The seminal contribution in this literature is Diamond (1982a), with important subsequent work including Howitt and McAfee (1992); Pissarides (1992); Schaal and Taschereau-Dumouchel (2016); Kaplan and Menzio (2016); Sterk (2016); Eeckhout and Lindenlaub (2018); and Acharya et al. (2018). While the work since Diamond (1982a) considers various mechanisms, much of this work focuses on the role of firms' hiring or employment decisions in generating multiplicity. In this paper, strategic complementarities arise from the search decisions of unemployed workers rather than from firms' hiring or employment decisions.[6] The fact that multiplicity arises through a labor supply channel allows my model to account for several recent empirical phenomena that are otherwise difficult to explain.

The paper also relates to a large and growing literature regarding how the internet affects matching and the labor market. In addition to the papers mentioned previously, contributions to this literature include Autor (2001); Kuhn (2003); Kuhn and Skuterud (2004); Stevenson (2009); Kuhn and Mansour (2013); Kroft and Pope (2014); Faberman and Kudlyak (2016); Faberman and Kudlyak (2019); Belot et al. (2019); and Davis and Samaniego de la Parra (2020). Two of these papers are particularly relevant. First, Kuhn (2003) is among the first to explicitly identify what he refers to as "electronic job search agents"—that is, a job board feature that "emails a worker when a new vacancy satisfying worker-supplied criteria is posted." As discussed above, this feature of job boards is essential for enabling workers to monitor vacancies. Second, Davis and Samaniego de la Parra (2020) provide direct empirical evidence of "application bunching" in online matching platforms—an empirical phenomenon in which new vacancies receive a torrent of applications immediately after posting, consistent with workers monitoring vacancies.[7] Also related are models of stock-flow matching (Coles and Smith, 1998). In the present paper, I model search as a decision that allows a worker to have immediate access to the inflow of new vacancies.

Finally, the paper relates to an empirical macroeconomic literature concerned with identifying labor supply shocks. Blanchard and Diamond (1989); Chang and Schorfheide (2003); Peersman and Straub (2009); and Foroni et al. (2018), all find an important role for such shocks. This paper provides a novel theory of endogenous labor supply shocks. I also show that the belief shocks in my model would be identified empirically as supply shocks in Foroni et al. (2018), and that, given the explosive growth in online job search, this could explain their evidence that labor supply shocks appear to have become more salient in recent data.

Outline of paper. The paper proceeds as follows. Section 2 studies a stylized model of monitoring to illustrate the mechanism. Section 3 embeds the game in a quantitative model of the labor market. Section 4 calibrates the model and considers its steady states. Section 5 turns to dynamics, focusing on how belief-driven fluctuations can emerge, and the model's implications for the Great Recession. Section 6 considers evidence and suggests paths for future empirical work.

[6] In Eeckhout and Lindenlaub (2018), multiplicity arises due to strategic complementarities between firms' vacancy posting decisions and workers' on-the-job search decisions, thus highlighting an important role for job search but through a fundamentally different mechanism.

[7] I discuss this evidence in greater detail in Section 6.

2 Simple Model

The central insight of this paper is that monitoring technologies can lead to strategic complementarities in search decisions and thus multiple equilibria. In order to provide intuition for the key forces giving rise to (and precluding) multiplicity, in this section I study a stylized one-period general-equilibrium model in which searching workers have access to a monitoring technology. There are three main results: First, monitoring leads to an aggregate matching function with increasing returns.Second, the economy must be away from the free-entry limit for multiplicity to occur. Third, if entry is sufficiently inelastic and monitoring costs not too dispersed, multiplicity will exist.

2.1 Environment

Consider a single period in which u unemployed workers compete to match with v vacancies that are posted sequentially at random times during the period. Workers who are matched by the end of the period produce y units of output and receive wage w; all others receive payoff $b < w$.

2.1.1 Monitoring and matching

I first study workers' monitoring decisions and the matching process, taking w and v as given.

Monitoring decision. At the start of the period, each worker i draws flow utility of leisure (or home production) κ_i from continuous and differentiable distribution F with support on $(0, \infty)$.[8] After observing κ_i, workers decide how to spend their time: They can either actively monitor the arrival of vacancies throughout the period, thus gaining a first-mover advantage in submitting applications, or enjoy leisure/home production and wait until the end of the period to try to find work. Workers choosing to monitor match with probability p^m but forgo leisure/home production, while workers choosing to wait match with probability $p^w \leq p^m$ but enjoy flow utility κ_i. The values of monitoring and waiting for worker i, respectively, are thus

$$U^m \quad = p^m w + (1 - p^m) b \tag{1}$$
$$U^w(\kappa_i) = p^w w + (1 - p^w) b + \kappa_i. \tag{2}$$

Workers monitor if and only if the value of monitoring exceeds the value of waiting: $U^m > U^w(\kappa_i)$. Equations (1) and (2) reveal that monitoring decisions are characterized by a cutoff rule such that a worker drawing $\kappa_i < \kappa_i^*$ will choose to monitor while a worker drawing $\kappa_i \geq \kappa_i^*$ will choose to wait, where κ_i^* is defined by $\kappa_i^* \equiv \{\kappa_i | U^m = U^w(\kappa_i)\}$. Using (1) and (2),

$$\kappa_i^* = \big(p^m - p^w\big)(w - b) \tag{3}$$

where, implicitly, the match rates p^m and p^w depend on the equilibrium behavior of other workers (and the number of vacancies) in the economy. I next derive expressions for these two match rates.

Matching. The preceding description implies that matching effectively occurs in two phases: A monitoring phase (in which monitoring workers match with arriving vacancies) and an aftermarket (in which all vacancies and workers that did not match in the monitoring phase are matched).

[8] The restriction on the support of F is not necessary, but serves to simplify exposition. Note that the Lognormal distribution used in Section 3 satisfies these properties.

Because vacancies arrive sequentially throughout the period and monitoring allows workers to immediately observe and apply for new jobs, matching in the monitoring phase is frictionless. Letting κ^* denote the (symmetric) equilibrium monitoring cutoff, $F(\kappa^*)u$ is the number of monitoring workers, and the number of matches in the monitoring phase is

$$\mathring{m} = \min\left\{v, F(\kappa^*)u\right\} \tag{4}$$

implying that the match rate for workers in the monitoring phase is $\mathring{p} \equiv \frac{\mathring{m}}{F(\kappa^*)u}$.

For simplicity, I assume that the aftermarket is likewise frictionless.[9] Thus, the number of matches in the aftermarket is given by

$$\hat{m} = \min\{v - \mathring{m}, u - \mathring{m}\} \tag{5}$$

$$= v - \mathring{m} \tag{6}$$

implying that the match rate for workers in the aftermarket is $\hat{p} \equiv \frac{\hat{m}}{u - \mathring{m}}$.

Workers who choose to monitor are able to match during the monitoring phase and, failing that, in the aftermarket.[10] Workers who choose to wait are only able to match in the aftermarket. This implies that the match rates for monitoring workers and waiting workers that appear in (3) are, respectively, given by $p^m = \mathring{p} + (1 - \mathring{p})\hat{p}$ and $p^w = \hat{p}$. Using these together with (4) and (6), and defining market tightness as $\theta \equiv v/u$,[11] it is straightforward to show that the match rates for monitoring workers and waiting workers, respectively, can be written as

$$p^m = \begin{cases} 1 & \text{if} \quad F(\kappa^*) \leq \theta \\ \frac{\theta}{F(\kappa^*)} & \text{if} \quad F(\kappa^*) > \theta \end{cases} \tag{7}$$

$$p^w = \begin{cases} \frac{\theta - F(\kappa^*)}{1 - F(\kappa^*)} & \text{if} \quad F(\kappa^*) \leq \theta \\ 0 & \text{if} \quad F(\kappa^*) > \theta. \end{cases} \tag{8}$$

Search decisions are fully described by (3), (7), and (8). Figure 1 depicts a stylized representation of the matching process in the presence of monitoring.

Increasing returns. Before endogenizing w and v, it is useful to reflect on the match function. Notice that the match rates in (7) and (8) imply that (for fixed v), the return to monitoring is

$$p^m - p^w = \begin{cases} \frac{1-\theta}{1-F(\kappa^*)} & \text{if} \quad F(\kappa^*) \leq \theta \\ \frac{\theta}{F(\kappa^*)} & \text{if} \quad F(\kappa^*) > \theta. \end{cases} \tag{9}$$

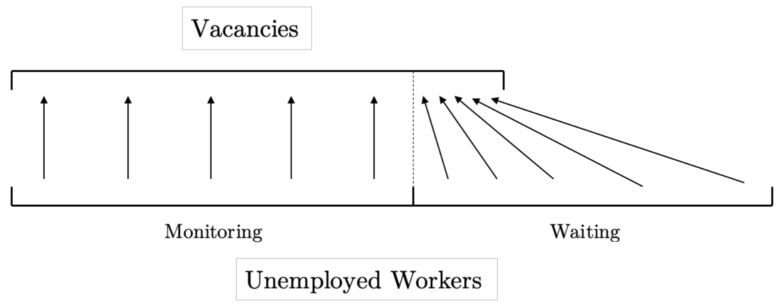
Equation (9) reveals that, when the labor market is tight ($F(\kappa^*) \leq \theta$), the matching function exhibits increasing returns in the sense that the individual benefit from monitoring relative to waiting, $p^m - p^w$, is an increasing function of the equilibrium monitoring cutoff, κ^*. As has been well known since the seminal work of Diamond (1982a), this property can expose the economy to multiple equilibria. Monitoring technologies lead to increasing returns because they confer a first-mover advantage to searchers, who in turn directly congest the matching process for those who wait, all of whom must compete for a diminished stock of remaining jobs.

[9] The frictionless property of the monitoring phase is a consequence of monitoring assuming vacancies are posted randomly throughout the period and firms can quickly make offers, whereas the frictionless property of the aftermarket is an assumption that I relax in Section 3.

[10] This assumption is inessential.

[11] It can be shown that, even with entry, it must be that $v < u$.

Figure 1: Matching with monitoring



Notes: The figure depicts the matching process in the monitoring game.

2.1.2 Wages and vacancy creation

I next endogenize w and v to study the model's general equilibrium properties.

Wages. I endogenize the wage, w, by positing a simple variation on the alternating-offer game of Hall and Milgrom (2008), adapted to a one-period environment and designed to reflect key features of the model of wages in Section 3. The game proceeds in two stages: When a worker and a vacancy are matched via the process described above (regardless of whether or not the worker was monitoring), the firm makes an offer to the worker. If the worker rejects the offer, with probability $1-\chi$ she is able to make a take-it-or-leave-it counter-offer (after which the opportunity for production is assumed to have passed), and with probability χ bargaining breaks down and she enters the aftermarket (where she can potentially match again). Solving this backwards, the worker's optimal wage offer (w^W) conditional on the game reaching its second and final round is $w^W = y$. Given this, the firm's optimal offer (w^F) leaves the worker indifferent between accepting and delaying, i.e. $w^F = \chi\left[p^w w + (1-p^w)b\right] + (1-\chi)w^W$. Substituting $w^W = y$ into the firm's optimal offer and observing that the firm's offer will be the equilibrium wage, $w = w^F$, we can solve for the equilibrium wage, yielding:

$$w = \frac{\chi(1-p^w)b + (1-\chi)y}{1-\chi p^w}. \tag{10}$$

Notice that the wage is increasing in workers' match rate in the aftermarket, p^w, and when the aftermarket is entirely crowded out, the wage reduces to $w = \chi b + (1-\chi)y$.

Vacancy creation. I endogenize the number of vacancies, v, by assuming that the economy consists of a measure $\eta \leq u$ of entrepreneurs indexed by j, each of whom can create a vacancy for a single job at cost ξ_j which is assumed to be drawn from continuous and differentiable distribution G with support on $[0, \bar{\xi}]$. Because the measure of entrepreneurs is assumed to be weakly less than the measure of unemployed workers,[12] and the matching process is frictionless, entrepreneurs who create vacancies are guaranteed to match, so the value of a vacancy is given by $y - w - c$, where $c > 0$ denotes the usual cost of maintaining a vacancy.[13] This description implies that the number

[12]This is convenient but not essential. It is straightforward to show that $\theta < 1$ even if $\eta > u$.

[13]I assume throughout that $c < (1-\chi)(y-b)$, which ensures that at least some jobs exist when no workers monitor.

of vacancies in the economy is:

$$v = \eta G(y - w - c). \tag{11}$$

To simplify notation, I assume $\eta = u$ so that this can be written as $\theta = G(y - w - c)$.

2.2 Equilibria

The economy's equilibria are characterized by equations (3), (7), (8), (10) and (11), together with the equilibrium condition $\kappa_i^* = \kappa^*$.

2.2.1 Key properties

The partial-equilibrium analysis of the matching process in Section 2.1.1 revealed that, for fixed wages and market tightness, monitoring technologies lead to increasing returns in the match function, and hence an upward sloping best-response function, as is necessary for multiplicity. However, understanding whether monitoring technologies can give rise to multiplicity in *general equilibrium* requires accounting for the behavior of wages and job creation, both of which affect incentives to monitor. Proposition 1 summarizes the model's implications for these variables.

Proposition 1 (Wages and job creation). *Wages fall (weakly) as more workers monitor. Job creation rises (weakly) as more workers monitor.*

Proof. See Appendix A. □

Thus, the general-equilibrium forces introduced in Section 2.1.2 both militate against an upward-sloping best-response function and thus the scope for multiplicity: If wages fall as more workers monitor, this will directly reduce the value of employment, and so tend to flatten out the best-response function via (3). Furthermore, falling wages will tend to stimulate entry, which will alleviate the strong congestion effects that lead to increasing returns when $F(\kappa^*) < \theta$ via (7) and (8). Nevertheless, Proposition 2 demonstrates a striking result: Despite the effect of endogenous wages and entry, when $F(\kappa^*) < \theta$ the best-response function in (3) continues to be upward-sloping so long as there is not free entry.

Proposition 2 (Properties of the best-response function). *When G is non-degenerate (i.e., away from the free-entry limit), the best-response function in* (3) *has the following properties:*

1. *There is a unique cutoff $\hat{\kappa} \equiv F^{-1}\big(G\big(\chi(y - b) - c\big)\big)$ such that $F(\kappa^*) < \theta$ for $\kappa^* < \hat{\kappa}$ and $F(\kappa^*) > \theta$ for $\kappa^* > \hat{\kappa}$.*
2. *When monitoring is limited ($\kappa^* < \hat{\kappa}$), the best-response function is increasing.*
3. *When monitoring is pervasive ($\kappa^* > \hat{\kappa}$), the best-response function is decreasing.*

Proof. See Appendix A. □

The result that the best-response function continues to be upward-sloping (for $\kappa^* < \hat{\kappa}$) even with endogenous wages and entry of firms is important: It implies that the source of multiplicity generated by the presence of monitoring technologies is sufficiently potent to expose the economy to multiplicity even when realistic general-equilibrium forces are present. The characteristics of the best-response function summarized in Proposition 2 continue to be present in the dynamic model studied in the next section—see, for example, Figure 2a in the next section.

2.2.2 A necessary condition for multiplicity

Much of the macroeconomics literature has focused on the limiting case of the general entry process described above, namely free entry (which occurs when G is degenerate at zero, or equivalently $\xi \to 0$). Proposition 3 shows that the free-entry limit of the model described above is indeed a special case—it is the only case in which the best-response function is flat, a fact which immediately precludes multiplicity.

Proposition 3 (Uniqueness under free entry/Necessity of inelastic entry). *Under free entry, there is at most one equilibrium. Thus, the economy must be away from the free-entry limit in order for multiplicity to occur.*

Proof. When $\xi \to 0$, entry occurs until $y - w = c$. If $\kappa^* > \hat{\kappa}$, because wages and entry are invariant to κ^* in this case, the right-hand side of (3) is constant. If $\kappa^* < \hat{\kappa}$, the free-entry condition uniquely pins down the wage. Furthermore, notice that (10) can be written in terms of the return to monitoring, $p^m - p^w$: $w = \frac{\chi b(p^m - p^w) + (1-\chi)y}{1-\chi+\chi(p^m-p^w)}$. Thus, the free-entry condition also uniquely pins down the return to monitoring. These two observations imply that the right-hand side of (3) is constant, and can intersect the 45-degree line in at most one point. □

The intuition is that entrepreneurs instantaneously arbitrage away any reductions in the wage, which serves to keep the wage from falling as more workers monitor. But because workers' wage bargaining position is fully determined by the return to monitoring (up to the model's parameters), this means that any congestion caused by additional monitoring is immediately exactly offset by entry, such that the return to monitoring is also constant. This result is important because it implies that the standard free-entry assumption is not innocuous, and can obscure rich dynamics (such as multiple equilibria) that might otherwise occur. Accordingly, in the quantitative model in Section 3, I attend to the entry process carefully when evaluating whether multiplicity obtains in a calibrated model of the U.S. economy.

2.2.3 Sufficient conditions for multiplicity

The fact that the model features an upward-sloping best-response function even when wages and entry are endogenous suggests that the influence of monitoring on the matching process is sufficiently strong to give rise to multiplicity even in a general equilibrium environment. As a final result before turning to the quantitative model, Proposition 4 provides a set of sufficient conditions under which multiplicity occurs in the model.

Proposition 4 (Multiplicity). *The model features multiplicity if the following conditions hold:*

1. *There is sufficient entry when monitoring is low (and thus wages are high):* $\underline{\theta} > \frac{(1-\chi)(y-b)-\hat{\kappa}}{(1-\chi)(y-b)-\hat{\kappa}\chi}$.
2. *There is not too much entry when monitoring is high (and thus wages are low):* $\bar{\theta} < F((1-\chi)(y-b))$.
3. *Monitoring costs are not too dispersed, i.e.* $Var(\kappa_i)$ *is sufficiently small.*

where $\underline{\theta} \equiv G\left((y-b)\frac{\chi(1-\underline{\theta})}{1-\chi\underline{\theta}} - c\right)$ *implicitly defines the level of entry when no workers monitor,*[14] *and* $\bar{\theta} \equiv G(\chi(y-b)-c)$ *defines the level of entry when* $\kappa^* \geq \hat{\kappa}$.

[14]Note that assumptions on c imply that $\underline{\theta}$ is unique and lies in $(0, 1)$.

Proof. See Appendix A. □

Because abundant jobs reduce congestion and thus monitoring incentives, the first condition ensures that individual workers are not too incentivized to monitor when no other workers do so, thereby helping to sustain a low-monitoring equilibrium. For the same reason, the second condition ensures that individual workers are *strongly* incentivized to monitor when other workers are monitoring, thereby helping to sustain a high-monitoring equilibrium. The intuition for the third condition, relating multiplicity to the dispersion in monitoring costs, follows from the fact that when monitoring costs are tightly centered around their mean, small changes in the equilibrium monitoring cutoff can lead to rapid congestion of the aftermarket and thus rapidly increasing incentives to monitor, which manifests as a high degree of curvature in the best-response function in (3). This condition therefore ensures that *both* a high and low monitoring equilibrium—and thus, by continuity of F, an intermediate-monitoring equilibrium—exist.[15] This condition is also related to the simple game analyzed in the Introduction, which corresponds to the limiting case of degenerate F (with fixed w and θ). I next turn to a quantitative dynamic model of monitoring.

3 Dynamic Model

To study the implications of monitoring for equilibrium labor market dynamics, I embed a variation on the static game from Section 2 within a quantitative macroeconomic model of the labor market. The model is fully dynamic and features a generalized version of the matching process in Section 2, bargained wages, and a flexible model of vacancy creation that nests various special cases considered in the literature. In the absence of monitoring, the model reduces to a standard DMP-style model of the labor market with a generalized job-creation process.

3.1 Environment

Time is discrete and runs forever. The economy is populated by a unit measure of ex ante identical workers who are either employed or unemployed and a fixed measure of entrepreneurs who create vacancies. Workers and entrepreneurs all discount the future with discount factor β, have perfect foresight with respect to the evolution of aggregate variables,[16] and seek to maximize the present discounted value of lifetime utility.

3.1.1 Accounting

Let u_t denote the total number of unemployed workers in a period and v_t the total number of vacancies in a period.[17] Unemployed workers either actively monitor the arrival of vacancies throughout the period (there are u_t^m such workers) or wait until the end of the period to match (there are u_t^w such workers). Vacancies are either newly posted by an entrepreneur (there are v_t^n such vacancies)

[15] Note that, except in the case of degenerate F, there must be an odd number of equilibria.

[16] In Section 5, I consider the effects of unanticipated "MIT shocks" to the economy.

[17] While I use the term "unemployment" throughout the paper, the model should be understood as referring more broadly to any non-employed workers who have some attachment to the labor market.

or were posted in a previous period and are old (there are v_t^o such vacancies). Thus, we have:

$$u_t = u_t^m + u_t^w \tag{12}$$
$$v_t = v_t^n + v_t^o. \tag{13}$$

Unemployed workers and vacancies are matched through a frictional matching process that I describe in further detail below. Matches become productive in the period after they are formed and are destroyed with probability δ at the end of each period (including matches that have formed but have not yet become productive). Letting m_t denote the total number of matches in a period, unemployment and vacancies evolve according to the following laws of motion:

$$u_{t+1} = u_t + \delta(1 - u_t) - (1 - \delta)m_t \tag{14}$$
$$v_{t+1} = (1 - \delta)(v_t - m_t) + v_{t+1}^n. \tag{15}$$

Equations (14) and (15) are standard and describe the aggregate dynamics of unemployment and vacancies in the model.

3.1.2 Matching

As in Section 2, unemployed workers choosing to monitor new job postings are able to observe (and apply for) new jobs as soon as they become available—and thus before other unemployed workers who are *not* actively monitoring new postings.[18] Monitoring thus results in a temporally segmented matching process that can be thought of as taking place in two phases: A monitoring phase (which lasts for the duration of the period) followed by an aftermarket phase (at the end of the period).

Monitoring. The monitoring phase proceeds exactly as in the static model in Section 2. Specifically, the total number of matches is given by[19]

$$\mathring{m}_t = \min\{v_t^n, u_t^m\}. \tag{16}$$

The corresponding match rates for workers and vacancies in the monitoring phase are, respectively,

$$\mathring{p}_t = \mathring{m}_t / u_t^m \tag{17}$$
$$\mathring{q}_t = \mathring{m}_t / v_t^n. \tag{18}$$

Aftermarket. At the end of each period, after all new vacancies have been posted, the aftermarket opens. In the aftermarket, all remaining unmatched workers (monitoring workers who failed to match while monitoring and workers who did not monitor) match with all unmatched vacancies (newly posted vacancies that failed to match with monitoring workers and old vacancies that were posted in previous periods and thus were not observed by monitoring workers). Thus, the total number of matches in the aftermarket is given by[20]

$$\hat{m}_t = \mu\big(v_t - \mathring{m}_t, u_t - \mathring{m}_t\big). \tag{19}$$

[18] My formulation of the matching process thus assumes that monitoring technologies afford workers a first-mover advantage in applying to *new* vacancies, which can intuitively be understood as reflecting, e.g., the functionality of job alerts (which notify workers when a new job has been posted). While it is possible to imagine other search technologies that might afford workers a first-mover advantage in applying to *all* vacancies, such a variation on the matching process described above would be unlikely to qualitatively affect the results since, in the calibrated version of the model, the significant majority of vacancies (roughly 85%) are new vacancies.

[19] I denote variables corresponding to the monitoring phase with circles, e.g. $\mathring{x}$.

[20] I denote variables corresponding to the aftermarket with hats, e.g. $\hat{x}$.

where μ is a matching function. The corresponding match rates for workers and vacancies in the aftermarket are, respectively,

$$\hat{p}_t = \hat{m}_t/(u_t - \mathring{m}_t) \tag{20}$$
$$\hat{q}_t = \hat{m}_t/(v_t - \mathring{m}_t). \tag{21}$$

The preceding implies that the total number of matches in the economy is given by

$$m_t = \hat{m}_t + \mathring{m}_t \tag{22}$$

with corresponding average match rates $p_t = m_t/u_t$ and $q_t = m_t/v_t$. Notice that the matching process described above reduces to a standard matching model in the absence of monitoring technology: That is, if $\mathring{m}_t = 0$, then the total number of matches is just $m_t = \mu(v_t, u_t)$. Thus, the model is simply a generalization of a traditional matching model with no search decision, in which monitoring technology affords searchers a first-mover advantage. I next consider workers' optimal choice to use this monitoring technology.

3.1.3 Workers

Employment and unemployment. Workers are either employed or unemployed. Employed workers receive wage w_t at the beginning of each period. Unemployed workers receive income $b_t = bw_t$ and flow value of leisure/home production $\bar{l}$ at the beginning of each period and decide whether or not to monitor new job postings. Note that $\bar{l}$ is leisure/home production enjoyed by all unemployed workers, whether or not they monitor, reflecting a baseline flow of utility from sources other than transfers that is independent of monitoring activity. This parameter will be important for calibrating the size of monitoring costs below.

Workers who do not monitor new postings draw i.i.d. flow utility κ_{it} (e.g., additional leisure or home production) from distribution F and are only able to match in the aftermarket at the end of the period.[21] Workers who monitor new postings, on the other hand, do not get to draw flow utility κ_{it}, but have the opportunity to match during the monitoring phase with newly posted vacancies in addition to being able to match in the aftermarket if they fail to match while monitoring. Thus, once again letting p_t^w denote the match rate for workers who wait until the aftermarket rather than monitoring and p_t^m the match rate for workers who monitor, the preceding implies

$$p_t^m = \mathring{p}_t + (1 - \mathring{p}_t)\hat{p}_t \tag{23}$$
$$p_t^w = \hat{p}_t. \tag{24}$$

Monitoring decision. Let W_t denote the value of entering period t employed, U_{it} the value of entering period t unemployed with draw κ_{it}, U_t^m the value of being unemployed and monitoring new postings throughout the period, and U_{it}^w the value of being unemployed and waiting until the

[21] Thus, κ_{it} can be understood as the cost of monitoring in terms of foregone leisure/home production. I provide further discussion of the interpretation of monitoring costs in Section 4 below.

aftermarket to match. Then the description of the environment above implies

$$W_t = w_t + \Big[\delta\beta\mathbb{E}U_{it+1} + (1-\delta)\beta W_{t+1}\Big] \tag{25}$$

$$U_{it} = b_t + \bar{l} + \max\Big\{U_t^m, U_{it}^w\Big\} \tag{26}$$

$$U_t^m = p_t^m\Big[\delta\beta\mathbb{E}U_{it+1} + (1-\delta)\beta W_{t+1}\Big] + (1-p_t^m)\Big[\beta\mathbb{E}U_{it+1}\Big] \tag{27}$$

$$U_{it}^w = p_t^w\Big[\delta\beta\mathbb{E}U_{it+1} + (1-\delta)\beta W_{t+1}\Big] + (1-p_t^w)\Big[\beta\mathbb{E}U_{it+1}\Big] + \kappa_{it} \tag{28}$$

where the expectation operator reflects uncertainty regarding future (i.i.d.) draws of κ_{it}. As may be seen in (27) and (28), an unemployed worker's decision about whether or not to monitor the arrival of new postings entails a tradeoff between a higher probability of matching if she chooses to monitor ($p_t^m \geq p_t^w$), and higher flow utility if she chooses not to monitor ($b_t + \bar{l} + \kappa_{it} \geq b_t + \bar{l}$).

Unemployed workers choose to monitor the arrival of new postings if and only if the value of doing so exceeds the value of not doing so, i.e. $U_t^m > U_{it}^w$. Because κ_{it} is i.i.d., (27) and (28) imply that the monitoring decision takes the form of a cutoff rule such that workers choose to monitor if and only if the value of leisure/home production exceeds a threshold defined by the value of κ_{it} such that $U_t^m = U_{it}^w$. Using (25)-(28), it is straightforward to show that this cutoff, κ_{it}^*, is given by

$$\kappa_{it}^* = \beta(1-\delta)(p_t^m - p_t^w)\mathbb{E}\Big[W_{t+1} - U_{it+1}\Big]. \tag{29}$$

In a symmetric equilibrium $\kappa_{it}^* = \kappa_t^*$ for all i. Thus, the number of monitoring and waiting workers in period t are given, respectively, by

$$u_t^m = F(\kappa_t^*)u_t \tag{30}$$

$$u_t^w = (1 - F(\kappa_t^*))u_t. \tag{31}$$

3.1.4 Entrepreneurs and job creation

I consider a generalized job-creation process that parameterizes the elasticity of job creation and thus nests free entry and inelastic job-creation processes as limiting cases.[22] Specifically, there is a fixed number η of entrepreneurs in the economy, each of whom can invest in the creation of a new job (i.e., create a vacancy). During each period t, each such entrepreneur j receives an opportunity to draw i.i.d. sunk investment cost ξ_{jt} from distribution G. If an entrepreneur decides to invest, it pays the sunk cost and creates a new vacancy with value V_t^n. Entrepreneurs undertake the investment if the expected value of creating a new vacancy exceeds the sunk investment cost, i.e., if $V_t^n > \xi_{jt}$. Thus, the number of newly created vacancies in the economy, v_t^n, is determined by

$$v_t^n = \eta G\big(V_t^n\big). \tag{32}$$

[22] See Fonseca et al. (2000), Fujita and Ramey (2005), Beaudry et al. (2018), and Coles and Moghaddasi Kelishomi (2018) for examples of similar processes.

Letting q_t^n denote the match rate for new vacancies and q_t^o the match rate for old vacancies, the value of a new vacancy, an old vacancy, and a filled job, respectively, are given by

$$V_t^n = -c + \beta(1-\delta)\Big[q_t^n J_{t+1} + (1-q_t^n)V_{t+1}^o\Big] \tag{33}$$

$$V_t^o = -c + \beta(1-\delta)\Big[q_t^o J_{t+1} + (1-q_t^o)V_{t+1}^o\Big] \tag{34}$$

$$J_t = y_t - w_t - T_t + \beta(1-\delta)\Big[J_{t+1}\Big] \tag{35}$$

where y_t is match output, T_t is a lump-sum tax used to finance UI payments b_t to unemployed workers, and c is the cost of maintaining a vacancy.[23] Note that the description of the matching process above implies that q_t^n and q_t^o are given by

$$q_t^n = \mathring{q}_t + (1-\mathring{q}_t)\hat{q}_t \tag{36}$$

$$q_t^o = \hat{q}_t. \tag{37}$$

Considering a generalized job-creation process, such as that described in (32), is important for the existence of multiple equilibria. In particular, in the free-entry limit, vacancy creation is highly responsive to workers' monitoring decisions, and because the incentive to monitor diminishes with the number of firms in the economy (all else equal), a highly elastic job creation process precludes multiplicity. In Section 4, I calibrate the elasticity of the entry process to be consistent with empirical estimates of the wage elasticity of labor demand in the literature and also explore the implications of more elastic entry processes.

3.1.5 Wages

Wages are determined via an alternating offers game as described in Hall and Milgrom (2008). Letting γ denote the cost to the firm of making a counteroffer to a worker and δ^B denote the probability that bargaining breaks down between periods, the wage may be written as

$$\begin{aligned} w_t =& (1-\beta(1-\delta))(\mathbb{E}[\kappa_{it}] + \bar{l}) \\ &+ \frac{\beta(1-\delta^B)}{1+\tau_{t+1}}\left[y_{t+1} + \gamma(1-\beta(1-\delta)) - \beta(1-\delta^B)(y_{t+2} - w_{t+2}(1+\tau_{t+2}))\right] \\ &+ \beta(\delta^B - \delta)\mathbb{E}[U_{it+1} - \beta U_{it+2}]. \end{aligned} \tag{38}$$

Notice that as $\delta^B \to \delta$, workers' outside option has no influence on the wage. Hence, δ^B has important implications for the degree to which wages respond to fluctuations in workers' outside option—and thus equilibrium monitoring intensity—in the model.[24]

[23] The value functions and structure of the entry process are as in Coles and Moghaddasi Kelishomi (2018).

[24] Under this wage protocol, bargaining operates in the same way for workers who monitor and for those who do not because of the usual assumption in models calibrated to monthly data that matches do not become productive—and thus wages are not bargained—until the period *after* the match is formed. Because monitoring costs are i.i.d. across time and workers, whether a worker found a job by monitoring or in the aftermarket in the previous period is irrelevant to their outside option at the time of bargaining, and thus all workers receive the same wage.

3.1.6 Taxes and transfers

The government is assumed to run a balanced budget, so that total transfers to unemployed workers are paid for by a lump-sum tax on firms T_t:

$$T_t = \frac{b w_t u_t}{1 - u_t} = \tau_t w_t \tag{39}$$

where it will be convenient to define $\tau_t \equiv b u_t / (1 - u_t)$.

3.2 Equilibrium

Definition 1 defines a perfect foresight equilibrium of the dynamic model described above.

DEFINITION 1. *A (symmetric) perfect foresight equilibrium is a sequence* $\{u_t, v_t, \kappa^*_{it}, v^n_t\}$ *such that* $\kappa^*_{it} = \kappa^*_t$ *for all i, and satisfying* (14),(15), (29), *and* (32) *for all* $t \geq 1$, *given:*

1. *The definitions and equilibrium conditions in* (12)-(39)*;*
2. *Initial conditions:* $\{u_0, v_0\}$*;*
3. *Transversality conditions:* $\lim_{t\to\infty} v^n_t < \infty$ *and* $\lim_{t\to\infty} \kappa^*_t < \infty$.

4 Calibration and Steady States

With the complete dynamic model in hand, I turn to calibrating the model's structural parameters.

4.1 Calibration

I directly calibrate the model's standard parameters to match values from the literature. To calibrate parameters related to monitoring and the matching process, I assume that the economy was in the low-monitoring steady state and choose parameters to match values from the early years of online job search. See Appendix B for complete details.

4.1.1 Functional forms

Following Beaudry et al. (2018) and Coles and Moghaddasi Kelishomi (2018), I adopt a flexible form for the entrepreneurial investment cost function,

$$G(\xi_{jt}) = \left(\xi_{jt}/\bar{\xi}\right)^{\nu}. \tag{40}$$

The parameter ν governs the elasticity of job creation, allowing (40) to nest both free-entry and an inelastic vacancy creation process as special cases.[25] I assume that workers and vacancies that remain unmatched at the end of each period are matched via an urn-ball matching function,

$$\mu\big(v_t - \mathring{m}_t, u_t - \mathring{m}_t\big) = (v_t - \mathring{m}_t)\left[1 - e^{-\psi \frac{u_t - \mathring{m}_t}{v_t - \mathring{m}_t}}\right]. \tag{41}$$

Finally, I assume that the distribution of monitoring costs, F, is Log-Normal with parameters μ_κ and σ_κ, both of which I calibrate to match moments in the data below.

[25] The parameter $\bar{\xi} \equiv \max\{\xi_{jt}\}$ ensures that this function can be interpreted as a proper distribution function. However, from (32) it can be seen that $v^n_t = \eta G(V^n_t) = \eta (V^n_t/\bar{\xi})^\nu = \eta \bar{\xi}^{-\nu} (V^n_t)^\nu$, implying that this parameter is not separately identified from η. Thus, in Table 2, I report the calibrated value of $\tilde{\eta} \equiv \eta \bar{\xi}^{-\nu}$ rather than η.

4.1.2 Standard parameters

Direct calibration. I directly calibrate a subset of the model's parameters to match values commonly used in the literature. The period length is assumed to be one month.[26]

Labor productivity is normalized to $y = 1$. I choose a discount factor of $\beta = 0.997$ consistent with a 4% annual steady-state interest rate. I choose a UI replacement rate of $b = 0.25$ following Hall and Milgrom (2008). I choose a vacancy posting cost of $c = 0.17$ following Barron et al. (1997).

Indirect calibration. I calibrate the remaining standard parameters to match a set of moments from the data and other studies. When possible, I use data from periods prior to the Great Recession when monitoring technologies existed but were not yet in widespread use.[27] I choose the match efficiency parameter, ψ, to match the average monthly job-finding probability between 2000Q1 and 2007Q3 of 0.37. I choose the monthly separation rate, δ, to match the average unemployment rate for the same period of 5.0%. I choose the parameter scaling the entry process, $\tilde{\eta}$, to match the average online job posting duration of 15.7 days identified by Marinescu and Wolthoff (2020). Turning to parameters related to wage determination, I interpret the firm's cost of making a counteroffer to a worker, γ, as the cost of re-evaluating the worker's application for employment. In light of this interpretation, and the fact that the cost of maintaining a vacancy, c, entails a similar evaluation of *all applicants* (that is presumably at least as costly on a per-applicant basis), I choose γ to match the per-applicant cost of maintaining a vacancy, $c/10$ (where 10 is the average number of applications to a posting in the data used to compute $c = 0.17$). Finally, I choose the probability of breakdown during bargaining, δ^B, such that at the low-monitoring steady state, the model-implied elasticity of labor market tightness with respect to labor productivity is equal to 19, the value computed in Shimer (2005), thereby allowing the model to generate plausibly large responses to productivity shocks.[28]

4.1.3 Non-standard parameters

Job creation elasticity. Because the elasticity parameter of the job-creation process, ν, is not entirely standard in DMP-style models (that have traditionally focused on the limiting case of free entry) and because of its importance for the existence of multiplicity (a point that I return to below

[26]Because of the stock-flow nature of matching in the monitoring phase, the matching process that I study will generally be sensitive to the choice of period length (or taking the continuous-time limit). This implication can be avoided by assuming that all vacancies (new and old) are reposted in each period and are thus able to be matched with monitoring workers in the monitoring phase.

[27]Because the paper studies a model of job search in which workers have access to monitoring technologies associated with the internet, I calibrate the model to a period during which such technologies existed—hence my focus on post-2000 data for calibration where possible. I focus the calibration on the *early* years of online job search (2000-2007) in particular for two related reasons: (i) This period contains a full business cycle that is not influenced by the Great Recession and thus averages over this period will approximate the economy's steady state reasonably well, and (ii) including post-Great Recession data in the calibration would preclude using the recovery from the Great Recession as an out-of-sample test of the model's ability to explain, e.g., the observed coexistence of low unemployment and tepid wage growth, as in Section 5.

[28]Formally, $\delta_B = \left\{\tilde{\delta}_B \middle| \frac{d\ln\theta}{d\ln y}(\tilde{\delta}_B)\big|_{@\text{ low-}\kappa^*\text{ SS}} = 19\right\}$ where (i) I have written $\frac{d\ln\theta}{d\ln y}(\tilde{\delta}_B)$ to emphasize that this elasticity is an endogenous object that depends on δ_B and (ii) I have indicated explicitly that the derivative is taken at the low-monitoring steady state. Because the model features multiple steady states, this elasticity can be computed at any of those steady states. For this reason, calibrating the model requires assuming the economy is at one of those steady states throughout the period over which the empirical targets are computed. As with the other parts of the calibration, I assume the economy is at the low-monitoring steady state.

in Section 4.3), it bears briefly discussing how this parameter is calibrated. A natural approach is to target the wage elasticity of labor demand, an object that has been estimated by a large and diverse literature. Hamermesh (1993) provides an early survey of this literature, finding values in a range around −0.3. Dube (2019) surveys the minimum wage literature, which is also interested in this elasticity, and finds a median estimate of the own-wage elasticity of employment of −0.17. Yet another approach is that of Beaudry et al. (2013, 2018), who estimate city level and city-industry level wage elasticities of labor demand of −0.3 and −1, respectively.

A common issue in such studies is that, in order to identify a labor demand or job creation elasticity, it is necessary to find an exogenous source of variation in wages. For example, Beaudry et al. (2018) use a Bartik-style instrument for the wage in their empirical specification.[29] This implies that, in order to calibrate the model to match empirical estimates of the labor demand or job creation elasticity, the underlying source of variation in the model must come only through the model's wage equation. Accordingly, I calibrate ν such that a small perturbation to δ_B—a parameter that only shows up in the model's wage equation—yields a wage elasticity of employment of $\bar{\epsilon}_{e,w} = -0.3$:

$$\nu = \left\{ \tilde{\nu} \Big| \frac{d \ln e / d\delta_B}{d \ln w / d\delta_B}(\tilde{\nu}) \Big|_{@ \text{ low-}\kappa^* \text{ SS}} = -0.3 \right\}. \tag{42}$$

In Appendix B.3, I elaborate on the use of δ_B as the source of variation to generate the model-implied elasticity to match the empirical estimates surveyed above. Interestingly, the value that I find for ν, 0.39, is not too different from the value of 0.24 identified by Coles and Moghaddasi Kelishomi (2018). In Section 4.3 below, I explore the sensitivity of the results to the empirical target of $\bar{\epsilon}_{e,w} = -0.3$.

4.1.4 Monitoring

The cost of monitoring in the model corresponds to the opportunity cost in terms of leisure/home production and is characterized by two parameters, μ_κ and σ_κ. Calibrating these parameters is difficult because monitoring decisions are difficult to observe directly. I thus proceed by choosing μ_κ and σ_κ indirectly to match two moments in the data with implications for monitoring behavior in the model. In what follows, I denote empirical analogs to model variables with bars over variables.[30]

Share monitoring. The first moment I target is the share of workers using monitoring technologies, $F(\kappa^*)$. I target the average share of workers using online job boards in 1998, 2001 and 2003, and compute this based on data from the CPS Computer and Internet Use Supplements as

[29] Discussing their results from estimating equation (6) in their paper, in Section IV.A the authors write: "...the employment equation derived from the model implies that OLS estimation of this equation should not provide consistent estimates. The fact that productivity shocks...enter the employment equation's error terms, and that wages are likely positively related to productivity, explains why the OLS regression coefficient on wages is positive."

[30] These two moments—the share of monitoring workers and the flow value of unemployment associated with leisure from not monitoring—together enable identification of μ_κ and σ_κ. Identification comes from the fact that, given κ^*, for $F < \frac{1}{2}$ (the relevant range given $F = 0.027$), F is decreasing in μ_κ and increasing in σ_κ, whereas $\int_{\kappa^*}^{\infty} \kappa dF(\kappa)$ is increasing in μ_κ and either (approximately) invariant to—or increasing in—σ_κ. It is therefore possible to numerically find a unique combination $\{\mu_\kappa, \sigma_\kappa\}$ that achieves the targeted moments.

tabulated by Stevenson (2009), which yields a value of 0.027.[31] Precisely, the first restriction is:

$$F(\kappa^*) = \bar{F}. \tag{43}$$

Flow values. The second moment I target is the residual average flow value of unemployment after netting out UI transfers (bw_t) and the baseline leisure/home production time enjoyed by all workers in the model ($\bar{l}$). Formally, letting $\mathbb{E}[z]$ denote the average flow value of unemployment, in the model we have $\mathbb{E}[z] \equiv bw_t + \bar{l} + \int_{\kappa^*}^{\infty} \kappa dF(\kappa)$, where the final term is the partial expectation representing additional leisure/home production enjoyed by workers who do not monitor. The restriction I impose involves taking a standard value for average flow value ($\mathbb{E}[z]$) from the literature and computing a value for baseline leisure/home production $\bar{l}$ from the data, which, together with the value for b above and the model wage equation in (38), implies a restriction on the partial expectation of the monitoring cost distribution:

$$\int_{\kappa^*}^{\infty} \kappa dF(\kappa) = \bar{\mathbb{E}[z]} - bw_t - \bar{l}. \tag{44}$$

It remains to compute $\bar{\mathbb{E}[z]}$ and $\bar{l}$. For the former, I follow Hall and Milgrom (2008) and set $\bar{\mathbb{E}[z]} = 0.71$. For the latter, I begin by noting that (i) the model implies that $\bar{l}$ is the consumption value imputed to leisure/home production among workers who engage in monitoring, and (ii) because the flow value of employment in (25) is normalized to be the wage, $\bar{l}$ corresponds to leisure/home production of workers who monitor relative to those who are working. Thus, $\bar{l}$ can be measured as the additional leisure/home production time enjoyed by workers who monitor relative to those who work. Accordingly, I link the October 2003 Current Population Survey with the 2003 American Time Use Survey and compute the difference in the average fraction of a 16 hour day spent on leisure/home production between an employed worker and a worker who reports having used the internet for job search in the CPS. As a baseline, I measure $\bar{l}$ as time not explicitly devoted to work, implying $\bar{l} = 0.28$. I also consider defining $\bar{l}$ more narrowly to only include "Socializing, Relaxing, and Leisure" and "Household Activities," which yields $\bar{l} = 0.14$.[32] Given the importance of the size of monitoring costs implied by $\bar{l}$, while I focus on the more expansive definition as a baseline, in the next section I show how the results depend on this choice.[33] See Appendix B.2 for additional details on the data used to construct this value.

From Proposition 4, we know that the value of σ_κ is critical for the existence of multiplicity. Because σ_κ and μ_κ are jointly determined by the empirical targets for $\bar{l}$ and $\bar{F}$, choosing reasonable values

[31] Stevenson (2009) finds that, averaging across 1998, 2001 and 2003, roughly 27% of unemployed workers and 3% of non-participants searched for jobs online, of whom roughly half used online job boards, corresponding to approximately 2.7% of such workers using online job boards to find work in the late 1990s/early 2000s. See Table 1 in Stevenson (2009) for details.

[32] Specifically, I compute $\bar{l} = 0.28$ as the difference between the fraction of a 16-hour day that unemployed searchers allocate to leisure (95%) and the fraction of a 16-hour day that employed workers allocate to leisure (67%), under a baseline assumption that a hypothetical full 16-hour day of leisure would yield a consumption value equal to labor's marginal product, y. Hence, $\bar{l} = y \cdot (0.95 - 0.67) = 1 \cdot 0.28 = 0.28$. See Appendix B.2, for details of this calculation, including how I define leisure in the baseline case as well as an alternative definition. Because the assumption that the consumption value of leisure is equal to y is somewhat arbitrary, Figure 2b in Section 4.3 illustrates the set of equilibria with the smaller value of $\bar{l} = 0.14$ (since presumably y is an upper bound) resulting from a narrower definition of leisure.

[33] I take as a baseline the more expansive definition of leisure and home production because, given my calibration strategy, high values of $\bar{l}$ imply smaller monitoring costs, so the results associated with this calibration represent a conservative view on the size, and thus importance, of monitoring costs in the economy.

for both of these moments is important. Accordingly, in Figure 2b below, I explore alternative targets for both of these moments and show that multiplicity continues to obtain.

4.1.5 Summary

Table 2 summarizes parameter values. See Appendix B for complete details and discussion.

Table 2: Calibration

Concept	Parameter	Value
Discount factor	β	0.997
Separation rate	δ	0.02
UI replacement rate	b	0.25
Mean of $\ln(\kappa_i)$	μ_κ	-1.48
SD of $\ln(\kappa_i)$	σ_κ	0.20
Baseline leisure/home production	$\bar{l}$	0.28
Productivity	y	1
Cost of counter-offer	γ	0.017
Bargaining breakdown probability	δ^B	0.032
Vacancy posting cost	c	0.17
Job creation (scale)	$\tilde{\eta} \equiv \eta\bar{\xi}^{-\nu}$	0.01
Job creation (elasticity)	ν	0.39
Matching efficiency	ψ	0.79

Notes: Monthly frequency. Moments computed by the author for calibration based on data between 2000 and 2007.

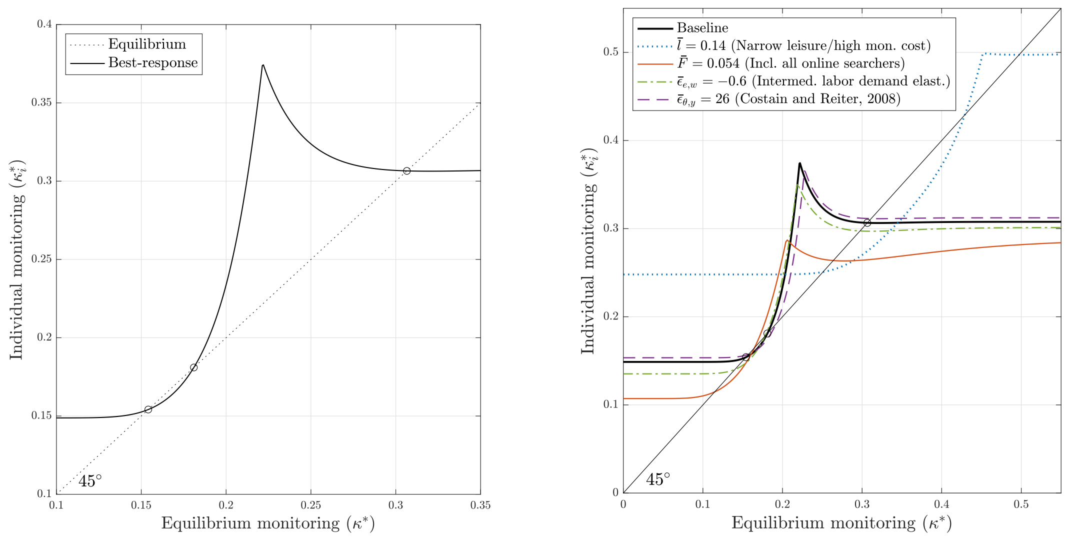
4.2 Steady-state equilibria

Figure 2a plots the (steady-state) best-response function for worker i in equation (29) for the baseline calibration described above. As may be seen in Figure 2a, the best-response function is increasing when $F(\kappa^*) < \theta$ (as anticipated in Proposition 2 from Section 2) and there are three steady states. Table 3 reports the implied values of the targeted moments for all three steady states and several other moments of interest.[34]

The steady states with higher levels of monitoring feature (i) higher average job-finding rates, (ii) lower unemployment, and (iii) more job creation. Furthermore, there is a non-monotonic relationship between unemployment and wages across steady states, such that the wage increases from the low-κ^* to the intermediate-κ^* steady state, but then decreases relative to both in the high-κ^* steady state. These observations reflect endogenously different levels of monitoring across steady states, sustained by workers' self-confirming beliefs about the actions of other workers in the economy. Specifically, when workers believe that many other workers are monitoring, they know that there will be relatively few jobs available in the aftermarket. This, in turn, makes forgoing leisure/home production in order to monitor necessary to avoid falling to the back of the queue for jobs and remaining unemployed. Because monitoring is frictionless whereas the aftermarket is not,

[34] In additional analysis in the appendix, I study the calibrated model's implications for welfare and show that an identical model without monitoring ceases to admit multiple steady states.

Figure 2: Best-response function



(a) Baseline calibration

(b) Alternative calibration targets

Notes: Panel (a) depicts the best-response function in the baseline model. Panel (b) depicts the best-response function in the baseline model (solid line) and various alternatives calibration targets. Steady-state equilibria occur where best response functions intersect of the 45-degree line.

more monitoring leads to higher average job-finding rates which reduces unemployment and exerts upward pressure on wages. However, because monitoring requires forgoing leisure, when a large fraction of workers monitor, unemployment looks less attractive in expectation for a given match rate (because it is likely a worker will find monitoring necessary in the event of job loss), which in turn exerts downward pressure on wages, thus stimulating entry and job creation. The effect of higher average match rates dominates when the economy moves from the low-κ^* steady state to the intermediate-κ^* steady state, such that unemployment falls but wages rise. By contrast, the effect of higher expected search costs dominates when the economy shifts from the low-κ^* steady state to the high-κ^* steady state, implying a substantial drop in unemployment and a modest drop in wages.[35] I return to this latter observation in the context of the Great Recession in Section 5.

4.3 Robustness

Several of the model's key parameters are non-standard in macroeconomic models, leading to uncertainty about how best to calibrate them. To address this, I study the implications of using various plausible alternative calibration targets in place of the targets described above.[36]

[35]It bears noting that if search technologies also serve to render the aftermarket frictionless (in addition to introducing the possibility of monitoring), then there are no match efficiency gains from more monitoring, implying an unambiguously positive correlation between wages and unemployment across all three steady states.

[36]In Appendix B.4, I carry out additional sensitivity analysis by studying how key moments of the model vary with the structural parameters δ_B, ν, μ_κ, and σ_κ. However, because it is not obvious what alternative values we should focus on for many of these parameters (most of which are non-standard), whereas there are natural alternative calibration targets, I leave that analysis for the appendix and focus on an analysis of various alternative calibration

Table 3: Description of equilibria

	Data	Monitoring model		
		Low-κ^*	*Int.-*κ^*	*High-*κ^*
Targeted moments				
Unemployment rate (u)	5.0%	5.0%	4.7%	4.0%
Job-finding prob. (p)	0.37	0.370	0.399	0.473
Share monitoring ($F(\kappa^*)$)	2.7%	2.7%	12.8%	92.6%
Untargeted moments				
Wage (w)		0.804	0.804	0.803
Job creation (v^n/v)		0.864	0.930	1
Job-finding prob. (p^m)		1	1	0.511
Job-finding prob. (p^w)		0.352	0.311	0
Social welfare (Appendix D)				
Output net of costs (Ω)		0.931	0.932	0.927

Notes: See Section 4 and Appendix B for further details.

Figure 2b illustrates that the existence of multiple steady states is robust to various plausible alternative calibration targets. Specifically, the figure depicts the best-response function in the baseline calibration (solid black line) alongside the best-response function implied by four plausible alternative calibrations: (i) the dotted blue line uses the alternative value of $\bar{l} = 0.14$ corresponding to a narrower definition of leisure in the ATUS and thus implying larger monitoring costs (see Appendix B.2 for details of how this is computed); (ii) the solid red line uses an alternative value of $\bar{F} = 0.054$ corresponding to defining the share of monitoring workers as including *all* workers engaging in online job search (rather than the share of online searchers who report using online job boards specifically, as in the baseline calibration); (iii) the dot-dashed green doubles (in magnitude) the value of the labor demand elasticity target, yielding $\bar{\epsilon}_{e,w} = -0.6$, bringing the target closer to the upper bound of the estimates surveyed above; and (iv) the dashed purple line uses a value for the elasticity of tightness with respect to productivity of $\bar{\epsilon}_{\theta,y} = 26$ taken from Costain and Reiter (2008) (rather than $\bar{\epsilon}_{\theta,y} = 19$ as in Shimer (2005) and the baseline calibration).[37] Importantly, three steady states arise in all four of these alternative calibrations. One circumstance in which multiplicity does disappear, however, is when labor demand is highly elastic. For example, setting a target of $\bar{\epsilon}_{e,w} = -1$, corresponding to the upper bound of the estimates surveyed above, implies that only the low-monitoring steady state exists. The basic reason for this is that when labor demand is highly elastic with respect to the wage, as more workers monitor and average wages fall, there is a large response in vacancy creation, which in turn increases the likelihood of a worker matching in the aftermarket and obviates monitoring, leading to a flat best response function.

targets in the main text.

[37] Hornstein et al. (2005) obtain a similar value to 26.

5 Dynamics and Implications

I next study the model's global dynamics and explore some of its central implications. In Section 5.1, I use a limiting case of the model ($\psi \to \infty$) to show that the model has multiple equilibria converging to different steady states, implying the existence of a novel source of endogenous (belief-driven) fluctuations in labor supply. I also show how purely temporary exogenous shocks can permanently alter the trajectory of the economy, thus giving rise to hysteresis effects emanating from online search technologies. In Section 5.2, I return to the full quantitative model and show how it can help to explain several features of the recovery from the Great Recession that are difficult to explain with traditional models featuring unique steady states or models of multiple steady states driven by strategic complementarities in labor demand.

5.1 ψ-limit economy

To facilitate exposition of the model's dynamics, I begin by focusing on the limiting case as $\psi \to \infty$. This corresponds to a frictionless aftermarket, implying $\hat{q}_t = q_t^n = q_t^o = q_t = 1$ and thus $v_t = v_t^n$. In Appendix E.1, I show that the ψ-limit model's dynamics are characterized by four equations and that the model has a single endogenous state variable (u_t), which significantly facilitates both analysis and graphical exposition of the model's global dynamics.[38] In Appendix E.2, I show that the model continues to have three steady states, and that these steady states are qualitatively similar to those reported in Table 3 for the full quantitative model described in Section 3.

5.1.1 Global dynamics

Constructing equilibrium paths. As a prelude to studying the model's global dynamics, it is useful to first note that a first-order approximation of the solution to the ψ-limit model in (E.14)-(E.17) around each of the three steady states implies that all three steady states feature unique local dynamics.

Knowledge of the model's local dynamics, in turn, can be used to construct global stable manifolds and thus to study the model's global dynamics. Specifically, following Judd (1999) and Brunner and Strulik (2002), I numerically construct the global stable manifolds converging to the three steady states using backward integration.[39] The global dynamics obtained in this way imply that the economy features *global* indeterminacy despite locally unique dynamics: For a wide range of initial values of the state variable u_t, the equilibrium conditions of the model do not uniquely pin down the path of the economy nor do they pin down the steady state towards which the economy will converge. As I discuss in Sections 5.1.2 and 5.2 below, this result has important implications for the sources of fluctuations in the economy and how we understand the recovery from the Great

[38] Computationally, when computing global non-linear saddle paths around the three steady states via backward integration, none of the numerical difficulties associated with multiple state variables (such a those described in Atolia and Buffie (2009)) arise when there is a single state variable. Expositionally, a single state variable makes it particularly easy to graphically illustrate the intuition for global indeterminacy.

[39] This method takes advantage of the fact that, when the model is solved in reverse time, the stable manifolds in forward time become the unstable manifolds. Paths attracted to the unstable manifold thus feature exponentially decreasing deviations from the stable manifold as the system is run in reverse time, implying that the global stable manifolds will be well-approximated by beginning from a point in a neighborhood of each steady state that need not lie precisely on the (zero measure) global stable manifold and running the system in reverse time. See Atolia and Buffie (2009) for a good discussion of this point.

Recession. Before discussing those implications, however, I first consider an empirical challenge posed by such models.

Equilibrium selection. While the existence of multiple equilibria leads to interesting and potentially important theoretical insights, it also presents an empirical challenge since the model does not make unique predictions about the trajectory of the economy following, e.g., a disturbance to fundamentals. Obtaining such predictions, that can then be compared with the data, requires a systematic way of reducing the set of candidate equilibria. I therefore propose a simple selection criterion based on a restriction on workers' beliefs along the equilibrium path that can help to alleviate this problem.

Definition 2 (Non-reversing return to monitoring criterion). *An equilibrium satisfies the non-reversing return to monitoring criterion if workers believe that, in the absence of fundamental shocks, along the equilibrium path:*

(i) The economy will stay on a single global stable manifold;

(ii) The return to monitoring ($p_t^m - p_t^w$) will never switch between being increasing and decreasing. That is, either $\mathbb{E}_t\left[\frac{\partial(p_{t+\tau}^m - p_{t+\tau}^w)}{\partial\kappa_{t+\tau}^*}\right] > 0$ *for all* $\tau > 0$ *or* $\mathbb{E}_t\left[\frac{\partial(p_{t+\tau}^m - p_{t+\tau}^w)}{\partial\kappa_{t+\tau}^*}\right] < 0$ *for all* $\tau > 0$.

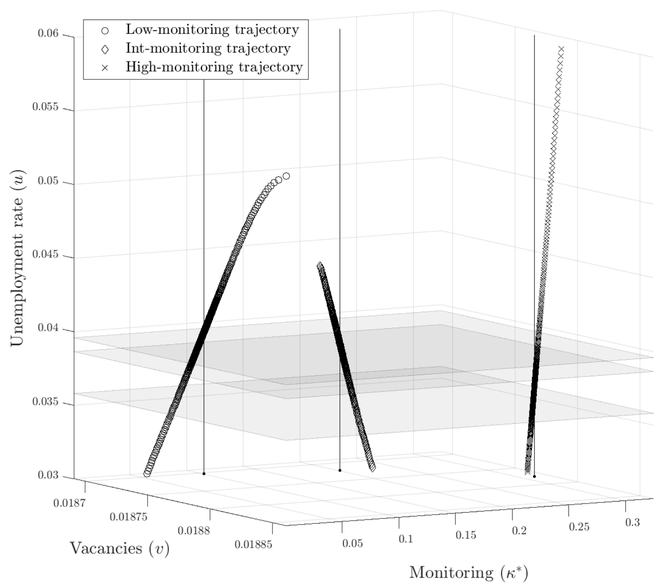
The first restriction simply rules out workers anticipating shifts between the trajectories converging to the different steady states. The second restriction provides a selection criterion among those paths: It rules out workers anticipating the equilibrium transiting between one in which the return to monitoring effort is positive and one in which it is negative (or vice versa). Importantly, this does not rule out the possibility of fundamental shocks shifting the economy between the global stable manifolds; rather it states that the set of global stable manifolds that the economy can reach from any initial position is restricted based on workers' beliefs about the evolution of the economy.

The criterion proposed in the second part of Definition 2 is appealing for two reasons. First, it can be understood as a restriction on the "sophistication" of workers' beliefs: Trajectories along which the qualitative nature of a choice fundamentally changes such as would be implied by a reversal of the return to monitoring are inherently more complicated than trajectories along which such a change does not occur. Second, it is relatively weak: In particular, it does not rule out multiplicity in large regions of the state space, which implies that it does not rule out the possibility of endogenous shocks to beliefs such as those I consider below. Importantly, however, it does function to *eliminate* multiplicity when a shock raises unemployment sufficiently. The basic reason for this is that a sufficiently high unemployment rate forces the economy into a position in which newly posted vacancies are the short end of the market in the monitoring phase—a situation in which the return to monitoring is *negative*. When this is the case, in order for the economy to converge to either the low- or intermediate-monitoring steady states, it must transit into a position in which workers eventually become the short end of the market in the monitoring phase—a transition which necessitates a reversal such that the return to monitoring eventually becomes *positive*. Thus, when a shock is sufficiently large to drive the economy to a high level of unemployment, the only remaining equilibrium is the one that leads to the high-monitoring steady state. I return to this result in Section 5.2, where I explore the implications of multiplicity for understanding the Great Recession.

Figure 3 plots the three global stable manifolds that satisfy the selection criterion in Definition 2. The grey planes depict the three steady-state levels of unemployment. The vertical lines depict the three steady-state pairs of values $\{\kappa^*, v\}$. Note that, when a recession drives the unemployment rate above roughly 6%, the only remaining equilibrium converges to the high-monitoring steady-state.

For smaller shocks, the economy's trajectory is indeterminate.

Figure 3: Global dynamics



Notes: The figure depicts the global stable manifolds converging towards the three steady states. The shaded planes correspond to the steady-state levels of unemployment. The vertical lines depict the three steady-state pairs of values $\{\kappa^*, v\}$. Global dynamics are non-unique at any value of u_t through which multiple stable manifolds pass.

5.1.2 Belief-driven labor supply fluctuations

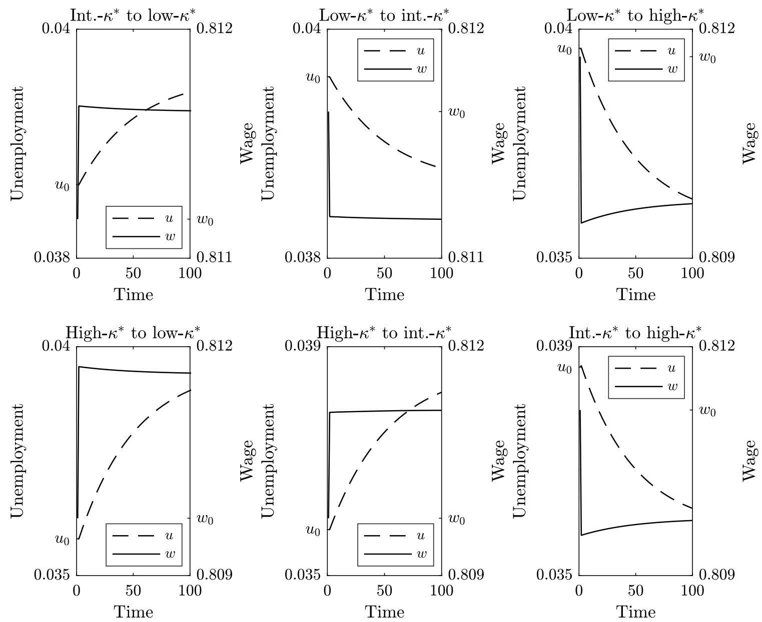
The multiplicity of dynamic equilibria observed in Figure 3 implies that the economy is susceptible to autonomous shocks to workers' beliefs that can permanently alter the trajectory of the economy by shifting the economy from a path converging to one steady state to a path converging to another. Because such changes in beliefs manifest through unemployed workers' decisions to actively monitor the arrival of vacancies, the model can be understood as providing a theory of endogenous belief-induced labor supply fluctuations.[40] Specifically, Figure 4 depicts the responses of wages (solid lines) and unemployment (dashed lines) to the six possible belief shocks shifting the economy between the trajectories depicted in Figure 3.[41] In each case, I assume that the economy begins at a particular steady state and then workers' beliefs suddenly (and unexpectedly) change such that the economy is expected to converge to a different steady state.

This figure is important for two reasons. First, it illustrates the rich dynamics that can be induced

[40] As I have described it above, the model can accommodate *unanticipated* shocks to beliefs without any additional analysis. Such shocks can be thought of as simply shifting the economy between the various global stable manifolds. This is the approach taken by Eeckhout and Lindenlaub (2018). Incorporating *anticipated* shocks to beliefs would require modeling beliefs as following, e.g., a Markov process, as in Kaplan and Menzio (2016).

[41] There are six possible shocks because there are three steady states and from each steady state belief shocks can shift the economy to a path converging towards either of the two other steady states.

Figure 4: Belief-driven labor supply shocks



Notes: The panels depict impulse responses of unemployment (dashed lines) and wages (solid lines) associated with the six possible belief shocks. Initial steady state values of unemployment and wages are indicated by u_0 and w_0, respectively. For each shock, the correlation between wages and unemployment over short horizons following the shocks are positive, consistent with the interpretation of the belief shocks as labor supply shocks.

by shocks to workers' beliefs. Second, it illustrates why these shocks can be interpreted as labor supply shocks: In all six cases, the effect of the shock in the first several periods induces a positive correlation between wages and unemployment. The observation that belief shocks necessarily induce a positive short-run correlation between wages and unemployment is important because it suggests that, in VAR analyses of the sources of business cycles that exploit short-run restrictions on the relationship between unemployment and wages, the belief shocks considered in this model will be identified as labor supply shocks. Using this type of short-run restriction, recent work by Foroni et al. (2018) finds evidence in support of an important role for labor supply shocks in driving aggregate fluctuations, and particularly so in more recent years, consistent with the advent of online job search exposing the economy to belief-driven labor supply shocks.

5.2 The Great Recession

I next return to the full quantitative model (i.e., with a frictional aftermarket) to study the model's implications for the recovery from the Great Recession. I show that a shock to the financial sector sufficiently large to replicate observed levels of unemployment forces the economy onto the high-monitoring trajectory, along which unemployment and labor market tightness both overshoot their pre-recession levels and wages undershoot their pre-recession level—features that are qualitatively consistent with the economy's observed recovery.

5.2.1 Financial shocks

To study the model's implications for the Great Recession and the economy's subsequent recovery, I study a shock to the cost of financing investments in job creation among entrepreneurs, $\tilde{\eta}$.[42] While this is not intended to be a quantitative exercise, it is a simple way to qualitatively explain several facts about the recovery from the Great Recession following an impulse that can be understood as emanating from the financial sector.

Formally, in the context of the model described above, I assume that $\tilde{\eta}$ follows an AR(1) process,

$$\tilde{\eta}_{t+1} = \rho\tilde{\eta}_t + (1-\rho)\tilde{\eta} \tag{45}$$

where $\tilde{\eta}$ is the value calibrated in Section 4. Agents in the model are assumed to know the AR(1) process but do not anticipate shocks to $\tilde{\eta}_t$. I calibrate the initial state of the economy (i.e., $\tilde{\eta}_t$ and u_t) and the persistence of the financial shock (ρ) to minimize the distance between the model-implied and observed unemployment rates in January 2008 (shortly before the unemployment rate began to rise), October 2009 (when the unemployment rate peaked at 10%), November 2019 (the date of John Robertson's Atlanta Fed blog post cited in the Introduction), and February 2020 (the beginning of Covid-19).[43] Importantly, a shock sufficiently large to generate 10% unemployment eliminates the two equilibrium paths converging to the low- and intermediate-monitoring steady states under the selection criterion suggested above.[44] Thus, beginning from the low-monitoring steady state, the shock forces the economy onto a path converging to the high-monitoring steady state. In this way, purely transitory shocks can induce a permanent shift in workers' labor supply decisions. Below, I argue that this fact can help to account for several anomalous features of the recovery from the Great Recession.

5.2.2 Implications

Unemployment. Figure 5 depicts the effect of the entrepreneurial shock on the unemployment rate in the model, and compares the model's response with the data. The dashed horizontal line corresponds to the pre-shock (low-monitoring) steady state. The solid horizontal line depicts the high-monitoring steady state toward which the economy converges. While the model's response to the shock decays more slowly than the data, the figure is intended to illustrate a simple qualitative point: The model provides a natural explanation for why unemployment overshot its pre-recession value, namely that a temporarily high level of unemployment forced the economy onto a trajectory

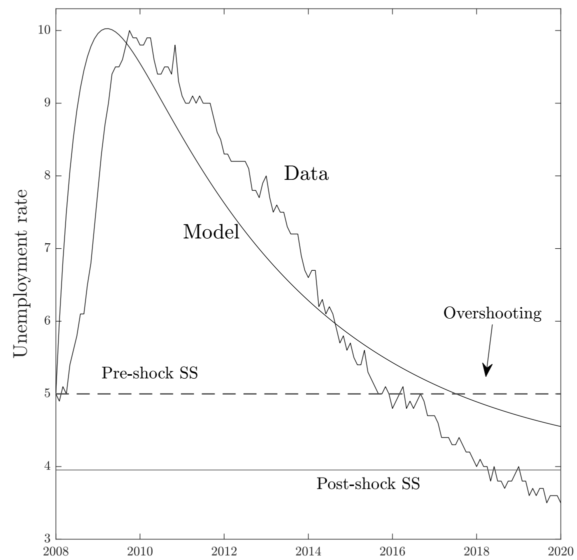
[42] Recall from Section 3 that, in order to undertake an investment project and thus create a job, an entrepreneur must pay a sunk investment cost drawn from distribution G with support $[0, \bar{\xi}]$. Alternatively, this can simply be interpreted as a shock to the measure of entrepreneurs in the economy emanating from some other source.

[43] This procedure implies that the economy begins close to the low-κ^* steady-state level of unemployment and is hit by a large (roughly 50%) contraction in $\tilde{\eta}$ that decays according to $\rho = 0.9$.

[44] For a given monitoring cutoff κ^* and thus a given fraction of the unemployed population choosing to monitor $F(\kappa^*)$, a sufficiently high level of unemployment will lead to a large number of workers choosing to monitor, thereby inducing a negative return to monitoring across all three equilibria. However, along the trajectories leading to the low- and intermediate-monitoring steady states, the return to monitoring will eventually switch and become positive as unemployment falls. The restriction I impose on beliefs rules out workers looking into the future and anticipating this type of switch on the grounds that such anticipation reflects implausible sophistication, and thus rules out trajectories leading to these two steady states from the moment the shock occurs. As a consequence, following a large shock, there is only one trajectory for the economy to move to, and that is the trajectory converging to the high-monitoring steady state. Moreover, the economy must stay on this trajectory, since anything else would necessitate an eventual reversal of beliefs regarding the return to monitoring.

converging towards the high-monitoring, low-unemployment steady state. Standard models with unique steady states will be unable to account for this feature of the data.[45]

Figure 5: Unemployment overshooting



Notes: Pre-shock SS corresponds to the (targeted) low-monitoring steady state. Post-shock SS corresponds to the (untargeted) high-monitoring steady state.

Wages and tightness. On November 5, 2019—almost exactly ten years after unemployment reached 10% in October 2009—John Robertson posted on the Atlanta Fed's MacroBlog about the state of the labor market ten years into the recovery. Robertson writes:

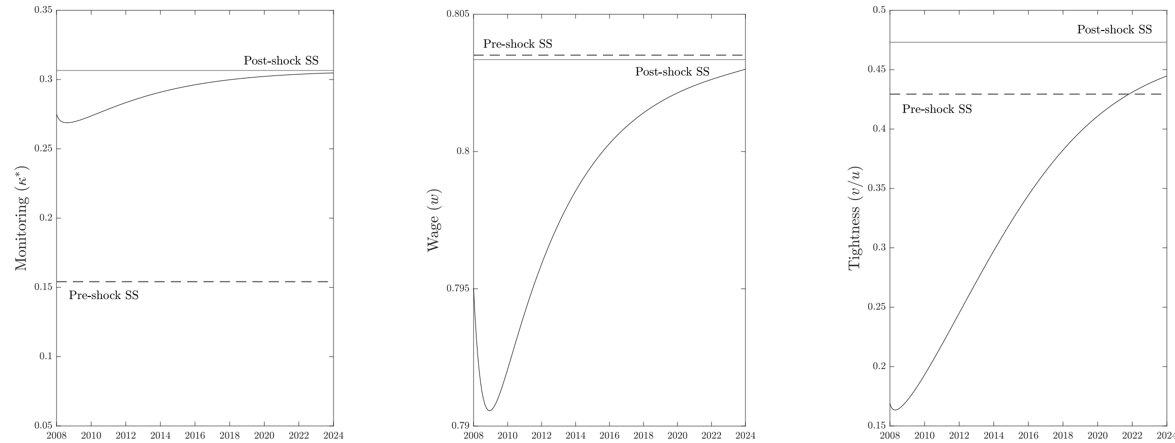
> *Here's a puzzle. Unemployment is at a historically low level, yet nominal wage growth is not even back to prerecession levels (see, for example, the Atlanta Fed's own Wage Growth Tracker). Why is wage growth not higher if the labor market is so tight?*

This quote captures two additional features of the recovery from the Great Recession that are perplexing from the standpoint of traditional models: The labor market (ultimately) was quite tight, but wage growth was persistently tepid. In the presence of a large initial shock such as the one described above, the model considered in this paper can qualitatively account for both of these observations. Figures 6b and 6c illustrate this point, plotting the model-implied trajectories of wages and tightness.

Following the shock, the economy is forced onto the high-monitoring trajectory, which *permanently* depresses wages (albeit modestly, as in the data) by reducing the expected value of unemployment as workers know they will have to compete aggressively for jobs, and thus forgo leisure/home production, if they return to unemployment. Wages eventually begin to recover, but only to the high-monitoring steady state at which they are marginally lower than they were prior to the shock, as may be seen in Figure 6b, and also lower than what would be implied by a pure demand

[45]Figure 8 in Appendix B.4 shows that, under three of the four alternative calibration targets considered in Figure 2b, the model fits the data from the recovery as well as the baseline calibration in Figure 5.

Figure 6: The recovery from the Great Recession



(a) Monitoring (b) Wages (c) Tightness

Notes: Panel (a) depicts the response of equilibrium monitoring intensity (κ^*) to the shock. Panel (b) depicts the response of wages (w). Panel (c) depicts the response of labor market tightness ($\theta \equiv v/u$). In each case, "Pre-shock SS" refers to the low-monitoring steady state targeted in calibration and "Post-shock SS" refers to the high-monitoring steady state towards which the economy is forced following a sufficiently large impulse.

shock calibrated to match the same unemployment data. In turn, this weak recovery in wages, coupled with the high level of monitoring, gives rise to a strong recovery in job creation—new vacancy creation in the model rises to a permanently elevated level. Together with low (and falling) unemployment, this implies that the economy was on a trajectory converging to a historically tight labor market, as may be seen in Figure 6c.

These observations would be difficult to understand through the lens of a traditional model with a unique steady state. Moreover, they would also be difficult to understand in the context of a model of multiple equilibria driven by strategic complementarities in labor demand, since in this case a plausibly tight labor market will tend to be associated with *elevated* wages.

6 Evidence on the mechanism

A central assumption in this paper is that there is an important first-mover advantage in job search, i.e. that workers who apply first typically get the job. In this section, I first discuss some suggestive evidence from vacancy durations and application flows in support of this assumption, and then suggest several possible approaches for more directly testing the assumption in future research.

Evidence that completed vacancy/posting durations tend to be short suggests that job seekers who do not apply for jobs quickly after they are posted will have access to substantially fewer jobs than those who do and thus supports the hypothesis of a first-mover advantage. For example, Davis and Samaniego de la Parra (2020) use second-by-second matched applicant-vacancy-employer data from the Dice.com platform to show that the average posting duration is 9 days. More to the point, they show that over 20% of vacancies are posted for under two days, and on the basis of this evidence argue that the "meeting phase" of the matching process is very short. Relatedly, Faberman and

Kudlyak (2019) study matched data from SnagAJob, an online private job search platform, and find that 22% of vacancies are posted for under one week.[46]

Evidence of "application bunching"—the fact that new vacancies tend to receive an immediate influx of applications—also supports the hypothesis of a first-mover advantage, since absent such an advantage there would be no reason to expect applicants to rush to apply to jobs. For example, Davis and Samaniego de la Parra (2020) document that a third of all applications are submitted to jobs posted within the past 24 hours and nearly half to jobs posted within the past 48 hours. Likewise, Faberman and Kudlyak (2019) show that 14% of applications are sent to newly posted vacancies and that the average number of applications per week received by a vacancy is negatively correlated with average posting duration. Indeed, many major platforms actively advise job seekers that applying first is essential in order to get the job.[47] For example, Monster informs job seekers that "as soon as recruiters post jobs, they're ready to start reviewing incoming resumes—meaning the sooner you apply, the better your chances are" while Indeed advises that "when you're one of the first to apply, you may be able to increase your chances of being interviewed."

Future research could evaluate the hypothesis of a first-mover advantage in search in several ways. One approach would be to apply new measures of "clumpiness" in incidence data to any longitudinal and high-frequency job search data to measure whether an individual's search tends to occur all at once (and hence be "clumpy") or be more spread out (consistent with monitoring).[48] A second approach would be to use *matched* applicant-application-vacancy data, such as that used by Davis and Samaniego de la Parra (2020), to construct direct time-varying applicant-level measures of monitoring—for example, the average number of minutes an applicant takes to apply to new job openings—that could be evaluated in terms of their ability to predict a match being formed. A third approach would be to exploit a resume/hiring audit study design to randomize over the speed at which a (fictitious) application is submitted to a newly posted job, controlling for other characteristics that are observable on a resume, and to measure how this affects the callback rate.

In Appendix F, I explore the first possibility: Using the Survey of Unemployed Workers in New Jersey (SUWNJ), I show that (i) online job search tends to be less clumpy, and (ii) less clumpy search is associated with more job offers. While more research is needed, this evidence supports the central hypothesis of this paper that there is a first-mover advantage in job search.

7 Conclusion

This paper conceptualizes search as a monitoring decision and studies the implications for equilibrium labor market dynamics. I first show that monitoring leads to a novel source of strategic complementarities in search and thus multiple equilibria. I then embed this mechanism in a quantitative model of the labor market and show that the model has multiple steady-state equilibria and features global indeterminacy. This implies that endogenous changes in workers' beliefs can permanently alter the trajectory of the economy. Moreover, the model offers a parsimonious account of the recovery from the Great Recession.

[46]Consistent with these studies, Davis et al. (2013) find that 42% of hires are at establishments with no reported vacancies at the start of the month, which they argue "suggests that average vacancy durations are very short."

[47]For example, see Monster.com, LinkedIn.com and Indeed.com.

[48]Recent developments in measuring "clumpiness" in incidence data have begun to be applied in other social/natural sciences. See Zhang et al. (2013) for a discussion of several such measures.

References

Acharya, S., J. Bengui, K. Dogra, and S. L. Wee (2018). Slow recoveries and unemployment traps: Monetary policy in a time of hysteresis. *Working Paper*.

Albrecht, J., B. Decreuse, and S. Vroman (2017). Directed search with phantom vacancies. *Unpubished manuscript*.

Atolia, M. and E. Buffie (2009). Reverse shooting made easy: Automating the search for the global nonlinear saddle path. *Manuscript*. Manuscript.

Autor, D. (2001). Wiring the labor market. *Journal of Economic Perspectives*.

Barron, J., M. Berger, and D. Black (1997). Employer search, training, and vacancy duration. *Economic Inquiry*.

Beaudry, P., D. Green, and B. Sand (2013). How elastic is the job creation curve? *Unpublished manuscript*.

Beaudry, P., D. Green, and B. Sand (2018). In search of labor demand. *American Economic Review*.

Belot, M., P. Kircher, and P. Muller (2019). Providing advice to job seekers at low cost: An experimental study on online advice. *Review of Economic Studies*.

Blanchard, O. and P. Diamond (1989). The beveridge curve. *Brookings Papers on Economic Activity*.

Brunner, M. and H. Strulik (2002). Solution of perfect foresight saddle point problems: a simple method and applications. *Journal of Economic Dynamics and Control*.

Cairo, I., S. Fujita, and C. Morales-Jimenez (2021). The cyclicality of labor force participation flows: The role of labor supply elasticities and wage rigidity. *Review of Economic Dynamics*.

Carnevale, A., T. Jayasundera, and D. Repnikov (2014). Understanding online job ads data. Technical report, Georgetown University.

Chang, Y. and F. Schorfheide (2003). Labor-supply shifts and economic fluctuations. *Journal of Monetary Economics*.

Cheron, A. and B. Decreuse (2017). Matching with phantoms. *Review of Economic Studies*.

Coles, M. and A. Moghaddasi Kelishomi (2018). Do job destruction shocks matter in the theory of unemployment? *American Economic Journal: Macroeconomics*.

Coles, M. and E. Smith (1998). Marketplaces and matching. *International Economic Review*.

Costain, J. and M. Reiter (2008). Business cycles, unemployment insurance, and the calibration of matching models. *Journal of Economic Dynamics and Control*.

Davis, S., J. Faberman, and J. Haltiwanger (2013). The establishment-level behavior of vacancies and hiring. *Quarterly Journal of Economics*.

Davis, S. and B. Samaniego de la Parra (2020). Application flows. *Working paper*.

Diamond, P. (1982a). Aggregate demand management in search equilibrium. *Journal of Political Economy 90*(5).

Diamond, P. (1982b). Wage determination and efficiency in search equilibrium. *Review of Economic Studies*.

Dube, A. (2019). Impacts of minimum wages: Review of the international evidence. Technical report, HM Treasury.

Eeckhout, J. and I. Lindenlaub (2018). Unemployment cycles. *American Economic Journal: Macroeconomics*.

Faberman, J. and M. Kudlyak (2016). What does online job search tell us about the labor market? *Federal Reserve Bank of Chicago Economic Perspectives*.

Faberman, J. and M. Kudlyak (2019). The intensity of job search and search duration. *American Economic Journal: Macroeconomics*.

Fonseca, R., P. Lopez-Garcia, and C. Pissarides (2000). Entrepreneurship, start-up costs and employment. *European Economic Review*.

Foroni, C., F. Furlanetto, and A. Lepetit (2018). Labor supply factors and economic fluctuations. *International Economic Review*.

Fujita, S. and G. Ramey (2005). The dynamic beveridge curve. *Federal Reserve Bank of Philadelphia Working Papers*.

Hall, R. and P. Milgrom (2008). The limited influence of unemployment on the wage bargain. *American Economic Review*.

Hamermesh, D. (1993). *Labor Demand*. Princeton University Press.

Hornstein, A., P. Krusell, and G. Violante (2005). Unemployment and vacancy fluctuations in the matching model: Inspecting the mechanism. *Federal Reserve Bank of Richmond: Economic Quarterly*.

Howitt, P. and P. McAfee (1992). Animal spirits. *American Economic Review*.

Judd, K. (1999). *Numerical Methods in Economics*. MIT Press.

Kaplan, G. and G. Menzio (2016). Shopping externalities and self-fulfilling unemployment fluctuations. *Journal of Political Economy*.

Kroft, K. and D. Pope (2014). Does online search crowd out traditional search and improve matching efficiency? Evidence from Craigslist. *Journal of Labor Economics*.

Krueger, A. and A. Mueller (2011). Job search, emotional well-being and job finding in a period of mass unemployment: Evidence from high-frequency longitudinal data. *Brookings Papers on Economic Activity 42*, 1–81.

Krusell, P., T. Mukoyama, R. Rogerson, and A. Sahin (2017). Gross worker flows over the business cycle. *American Economic Review*.

Kuhn, P. (2003). The internet and matching in labor markets. In *New Economy Handbook*.

Kuhn, P. and H. Mansour (2013). Is internet job search still ineffective? *The Economic Journal*.

Kuhn, P. and M. Skuterud (2004). Internet job search and unemployment durations. *American Economic Review*.

Marinescu, I. and R. Wolthoff (2020). Opening the black box of the matching function: The power of words. *Journal of Labor Economics*.

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

Mortensen, D. (1982). The matching process as a noncooperative bargaining game. In U. of Chicago Press (Ed.), *The Economics of Information and Uncertainty*.

Peersman, G. and R. Straub (2009). Technology shocks and robust sign restrictions in a euro area svar. *International Economic Review*.

Pissarides, C. (1985). Short-run dynamics of unemployment, vacancies, and real wages. *American Economic Review*.

Pissarides, C. (1992). Loss of skill during unemployment and the persistence of employment shocks. *Quarterly Journal of Economics*.

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

Schaal, E. and M. Taschereau-Dumouchel (2016). Aggregate demand and the dynamics of unemployment. *Working Paper*.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*.

Sterk, V. (2016). The dark corners of the labor market. *Discussion Paper*.

Stevenson, B. (2009). The internet and job search. In *Labor Market Intermediation*. University of Chicago Press.

Zhang, Y., E. Bradlow, and D. Small (2013). New measures of clumpiness for incidence data. *Unpubished manuscript*.

Appendices

A Simple model

This appendix provides further details on the simple model in Section 2.

A.1 Proofs

Throughout the analysis below it will be useful to define $p \equiv p^m - p^w$. Furthermore, using (10), note that when $F(\kappa^*) < \theta$, the wage can be written as

$$w = \frac{\chi b p + (1-\chi)y}{1-\chi+\chi p} \tag{A.1}$$

and when $F(\kappa^*) > \theta$, the wage is constant

$$w = \chi b + (1-\chi)y. \tag{A.2}$$

A.1.1 Proposition 1: Wages and job creation

Wages. Wages only depend on the monitoring cutoff through p, thus we can write

$$\frac{dw}{d\kappa^*} = \frac{dw}{dp}\frac{dp}{d\kappa^*}. \tag{A.3}$$

The first derivative on the right-hand side is obtained from (A.1)

$$\frac{dw}{dp} = -\frac{\chi(1-\chi)(y-b)}{(1-\chi+\chi p)^2} < 0. \tag{A.4}$$

For the second, we must consider two cases. If $F(\kappa^*) < \theta$, then $p = \frac{1-\theta}{1-F(\kappa^*)}$. Because θ depends on w which in turn depends on κ^*, we implicitly differentiate:

$$dp = \frac{g\frac{dw}{dp}}{1-F}dp + \frac{(1-\theta)f}{(1-F)^2}d\kappa^* \tag{A.5}$$

$$\implies \frac{dp}{d\kappa^*} = \frac{\frac{f}{1-F}\frac{1-\theta}{1-F}}{1-\frac{g\frac{dw}{dp}}{1-F}}. \tag{A.6}$$

Because $\frac{dw}{dp} < 0$, $\frac{dp}{d\kappa^*} > 0$ and it follows that $\frac{dw}{d\kappa^*} = \frac{dw}{dp}\frac{dp}{d\kappa^*} < 0$. If $F(\kappa^*) > \theta$, then $\frac{dw}{dp} = 0$ so it follows immediately that $\frac{dw}{d\kappa^*} = 0$.

Entry. Entry is determined by (**??**): $\theta = G(y - w - c)$. Inspection of this equation reveals immediately that entry is a decreasing function of the wage, implying (following from the analysis above) that entry is increasing in κ^* for $F(\kappa^*) < \theta$ and constant for $F(\kappa^*) > \theta$.

A.1.2 Proposition 2: Properties of the best-response function

Existence/uniqueness of $\hat{\kappa}$. I first show that $\hat{\kappa} \equiv F^{-1}(G(\chi(y-b)-c))$ exists and is unique. First, observe that $c < (1-\chi)(y-b)$ implies that $\theta > 0$. Thus, θ must lie strictly above F for sufficiently small κ^*. Next, suppose that for some κ^*, $\theta = 1$. Because in this case it must be that $F < \theta$, the wage is $w = y$ (reflecting $\theta = 1$ in (A.1)), implying $y - w - c = -c < 0$, which contradicts $\theta = 1$. Thus, $\theta < 1$, so θ must lie strictly below F for sufficiently large κ^*. This, together with continuity, implies that there must be at least one crossing. Suppose there are several. Then it must be that, after the first crossing, $\theta < F$. But in this case $w = \chi b + (1-\chi)y$ which is constant, implying that there is no further entry so θ is constant, which in turn implies that there can be no more crossings (since F must be increasing). Note also that the crossing occurs where $F(\kappa^*) = \theta$, and at this point $w = \chi b + (1-\chi)y$, so $\theta = G(\chi(y-b)-c)$. Thus,

$$\hat{\kappa} \equiv F^{-1}(G(\chi(y-b)-c)) \tag{A.7}$$

as in the main text.

$\kappa^* < \hat{\kappa}$. In this case, the best response function in (3) may be written as

$$k_i^* = p\left(\frac{(1-\chi)(y-b)}{1-\chi+\chi p}\right) \tag{A.8}$$

so the slope of the best response function is

$$\frac{d\kappa_i^*}{d\kappa^*} = \frac{d\kappa_i^*}{dp}\frac{dp}{d\kappa^*}. \tag{A.9}$$

We have already established that $\frac{dp}{d\kappa^*} > 0$. Differentiating the best response function above with respect to p yields

$$\frac{d\kappa_i^*}{dp} = \frac{(1-\chi)^2(y-b)}{(1-\chi+\chi p)^2} > 0. \tag{A.10}$$

Thus, for $\kappa^* < \hat{\kappa}$, the best-response function is upward-sloping.

$\kappa^* > \hat{\kappa}$. In this case, we have already established that w is constant, implying that θ is constant. Thus, the slope of the best-response function is determined by the slope of $p = \theta/F(\kappa^*)$ which we immediately see is negative.

A.1.3 Proposition 4: Sufficient conditions for multiplicity.

The strategy is to use the properties of the limit economy as F becomes degenerate to show that multiplicity must obtain when monitoring costs are sufficiently concentrated around their mean. To keep the notation clear, I index variables associated with economies away from the limit by σ.

Fix $\mu = E[\kappa_i]$ and consider the limit economy in which the monitoring cost distribution collapses around μ (i.e., $\sigma \to 0$). In this case, F is a delta function at μ such that

$$F(\kappa_i) = \begin{cases} 0 & \text{if} \quad \kappa_i < \mu \\ 1 & \text{if} \quad \kappa_i \geq \mu \end{cases} \tag{A.11}$$

implying that the best response function $\kappa^*_{i,\sigma}(\kappa^*)$ in (3) collapses to

$$\kappa^*_i(\kappa^*) = \begin{cases} (1-\underline{\theta})(\underline{w}-b) & \text{if} \quad \kappa^* < \mu \\ \bar{\theta}(\bar{w}-b) & \text{if} \quad \kappa^* \geq \mu \end{cases} \tag{A.12}$$

where $\underline{\theta} \equiv G\left((y-b)\frac{\chi(1-\theta)}{1-\chi\underline{\theta}} - c\right)$ implicitly defines the level of entry when no workers monitor,[49] and $\bar{\theta} \equiv G(\chi(y-b)-c)$ defines the level of entry when all workers monitor. First, note that $\hat{\kappa}_\sigma$ approaches μ in the limit, so using the condition that $(1-\underline{\theta})(\underline{w}-b) < \hat{\kappa}_\sigma$, for sufficiently small σ, $(1-\underline{\theta})(\underline{w}-b) < \mu$. Now, fix $\kappa^*_0 \in ((1-\underline{\theta})(\underline{w}-b), \min\{\mu, \hat{\kappa}_\sigma\})$ and observe that, again for sufficiently small σ, it must be that $\kappa^*_0 < \hat{\kappa}_\sigma$ and we can make $\kappa^*_{i,\sigma}(\kappa^*_0)$ arbitrarily close to $\kappa^*_i(\kappa^*_0) = (1-\underline{\theta})(\underline{w}-b)$. Choose σ such that $\kappa^*_{i,\sigma}(\kappa^*_0) \in ((1-\underline{\theta})(\underline{w}-b), \kappa^*_0)$. Then, we have the following: (i) $\kappa^*_{i,\sigma}(0) = (1-\underline{\theta})(\underline{w}-b) \geq 0$ (by the first part of the proposition), (ii) $\kappa^*_{i,\sigma}(\kappa^*_0) < \kappa^*_0$ (as shown above), and (iii) $\kappa^*_{i,\sigma}(\hat{\kappa}_\sigma) > \hat{\kappa}_\sigma$ (also by the first part of the proposition), so the best response function crosses the 45-degree line at least twice and there are at least two equilibria below $\hat{\kappa}_\sigma$. Furthermore, because $\kappa^*_{i,\sigma}$ is weakly decreasing for $\kappa^* > \hat{\kappa}_\sigma$, there must be a third equilibrium for some $\kappa^* > \hat{\kappa}_\sigma$.

B Calibration

This appendix provides further details on the calibration of ν, δ^B, ψ, δ, $\tilde{\eta}$, μ_κ, and σ_κ. All other parameters are calibrated directly to match values in the data and existing literature as described in Section 4. The seven moments used to pin down these remaining seven parameters are: $\bar{d}$ (average vacancy duration), $\bar{p}$ (average match rate of workers), $\bar{u}$ (unemployment rate), $\mathbb{E}[z] \equiv bw + \bar{l} + \int_{\kappa^*}^{\infty} \kappa dF(\kappa)$ (average flow value of unemployment), $\bar{F}$ (share of workers monitoring), $\bar{\epsilon}_{e,w} = \frac{d\ln(e)}{d\ln(w)}$ (elasticity of employment with respect to the wage), and $\bar{\epsilon}_{\theta,p} = \frac{d\ln(\theta)}{d\ln(y)}$ (elasticity of tightness with respect to productivity).

B.1 Calibrating the model

The steady state of the model cannot be solved analytically, so I proceed numerically. My procedure calibrates the model's parameters via an inner and an outer loop, summarized below:

1. Inner loop: For given values of $\{\nu, \delta^B\}$, use the model's steady-state conditions to calibrate the five remaining parameters: $\{\psi, \delta, \tilde{\eta}, \mu_\kappa, \sigma_\kappa\}$.
2. Outer loop: Search over values of $\{\nu, \delta^B\}$ to match two steady-state elasticities from the data:
   - Wage elasticity of employment: $\bar{\epsilon}_{e,w} = \frac{d\ln e}{d\ln w}$
   - Productivity elasticity of tightness: $\bar{\epsilon}_{\theta,y} = \frac{d\ln\theta}{d\ln y}$

   where $\bar{\epsilon}_{e,w} = -0.3$ and $\bar{\epsilon}_{\theta,p} = 19.1$ as discussed in the main text.

Below, I elaborate on the inner loop of the calibration procedure.

Match efficiency (ψ). Because the economy is assumed to be in the low-monitoring steady state in which workers are the short-end of the monitoring market, $\bar{F}$ and $\bar{u}$ can be used to pin down the

[49]Note that assumptions on c imply that $\underline{\theta}$ is unique and lies in $(0, 1)$.

number of matches during the monitoring phase:

$$\mathring{m} = \bar{F}\bar{u}. \tag{B.1}$$

Additionally, average posting duration $\bar{d}$ can be used to pin down the average match rate for firms,

$$q = 1 - e^{-31/\bar{d}} \tag{B.2}$$

which, in turn, can be combined the with $\bar{p}$ to solve for labor market tightness,

$$\theta = \bar{p}/q. \tag{B.3}$$

Using the match functions in (16) and (19), q can be written as

$$q = m/v = \frac{\mathring{m} + \hat{m}}{v} = \bar{F}/\theta + (1 - \bar{F}/\theta)\left(1 - e^{-\psi\frac{1-\bar{F}}{\theta-\bar{F}}}\right) \tag{B.4}$$

which in turn can be solved for ψ,

$$\psi = -\left(\frac{\theta - \bar{F}}{1 - \bar{F}}\right)\ln\left(\frac{1-q}{1-\bar{F}/\theta}\right). \tag{B.5}$$

Separation rate (δ). Given $\bar{p}$ and $\bar{u}$, equation (14) can be solved for the separation rate, δ:

$$\delta = \frac{\bar{p}}{\bar{p} + 1/\bar{u} - 1}. \tag{B.6}$$

Tax rate (τ). The calibration of $\bar{u}$ together with the value of b pins down the tax rate from (39),

$$\tau = b\bar{u}/(1 - \bar{u}). \tag{B.7}$$

Remainder of the model (μ_κ, σ_κ, $\tilde{\eta}$). Using $\theta \equiv v/\bar{u} = \bar{p}/q$, it is possible to pin down the steady-state level of vacancies,

$$v = \frac{\bar{p}}{q}\bar{u}. \tag{B.8}$$

Given v and q, the law of motion for vacancies in (15) pins down the number of new vacancies via

$$v^n = v(1 - (1-\delta)(1-q)). \tag{B.9}$$

From here, it is possible to solve for all of the relevant match rates. Using (18) and (21) and recalling that $\mathring{m} = \bar{F}\bar{u}$ and $\hat{m} = \bar{u}\bar{p} - \mathring{m}$, the steady-state match rates for vacancies in the monitoring phase and aftermarket, respectively, are

$$\mathring{q} = \mathring{m}/v^n \tag{B.10}$$

$$\hat{q} = \hat{m}/(v - \mathring{m}). \tag{B.11}$$

From (36) and (37), the steady-state match rates for new and old vacancies, respectively, are

$$q^n = \mathring{q} + (1 - \mathring{q})\hat{q} \tag{B.12}$$

$$q^o = \hat{q}. \tag{B.13}$$

Likewise, from equations (17) and (20), the steady-state match rates for workers in the monitoring phase and aftermarket, respectively, are

$$\mathring{p} = \mathring{m}/u^m = \mathring{m}/(\bar{F}\bar{u}) \tag{B.14}$$

$$\hat{p} = \hat{m}/(\bar{u} - \mathring{m}). \tag{B.15}$$

From (23) and (24), the steady-state match rates for workers who choose to monitor and who choose not to monitor, respectively, are

$$p^m = \mathring{p} + (1 - \mathring{p})\hat{p} \tag{B.16}$$

$$p^w = \hat{p}. \tag{B.17}$$

Next, we can use the two restrictions from Section 4.1.4 to solve numerically for $\{\mu_\kappa, \sigma_\kappa\}$,

$$F(\kappa^*(\mu_\kappa); \mu_\kappa, \sigma_\kappa) = \bar{F} \tag{B.18}$$

$$\int_{\kappa^*(\mu_\kappa)}^{\infty} \kappa dF(\kappa; \mu_\kappa, \sigma_\kappa) = \mathbb{E}[\bar{z}] - bw(\mu_\kappa) - \bar{l} \tag{B.19}$$

where the dependence of κ^* and w on μ_κ exploits the fact that, given the calibration of the model to this point, the steady-state version of the wage equation in (38) and the first-order condition for monitoring in (29) can be used to write κ^* and w as explicit functions of μ_κ:

$$w(\mu_\kappa) = \frac{(1-\beta(1-\delta))(\mu_\kappa+\bar{l}) + \frac{\beta(1-\delta)}{1+\tau}\left(y(1-\beta(1-\delta^B)) + \gamma(1-\beta(1-\delta))\right) + \frac{\beta(\delta^B-\delta)(1-\beta(1-\delta))}{1-\beta(1-\delta)(1-\bar{p})}\mathbb{E}[\bar{z}]}{1-\left((\beta(1-\delta^B))^2 + \frac{\beta^2(\delta^B-\delta)\bar{p}(1-\delta)}{1-\beta(1-\delta)(1-\bar{p})}\right)} \tag{B.20}$$

$$\kappa^*(\mu_\kappa) = \beta(1-\delta)\left(\frac{1-\theta}{1-\bar{F}}\right)\left(\frac{w(\mu_\kappa) - \mathbb{E}[\bar{z}]}{1-\beta(1-\delta)(1-\bar{p})}\right). \tag{B.21}$$

Next, given the solutions for $\{\mu_\kappa, \sigma_\kappa\}$ and thus for w and κ^* implied by the preceding four equations, we can solve for the value of a filled job using (35),

$$J = \frac{y - w - \tau}{1 - \beta(1-\delta)} \tag{B.22}$$

which allows us to solve for the value of an old vacancy and the value of a new vacancy, respectively:

$$V^o = \frac{-c + \beta(1-\delta)q^o J}{1 - \beta(1-\delta)(1-q^o)} \tag{B.23}$$

$$V^n = -c + \beta(1-\delta)\left[q^n J + (1-q^n)V^o\right]. \tag{B.24}$$

Finally, I calibrate $\tilde{\eta}$ by solving (32) for $\tilde{\eta}$ given the value of ν parameterizing the inner loop:

$$\tilde{\eta} = v^n/(V^n)^\nu. \tag{B.25}$$

B.2 Leisure time and monitoring costs

In the calibration described in Section 4 of the main text, I describe how I use data on leisure time to calibrate $\bar{l}$ and thus the implied monitoring cost distribution. In this appendix, I first provide further details on the data used in that calculation, and then suggest an alternative calibration that reflects the fact that stress from job search tends to spill over into non-search activities (Krueger and Mueller, 2011), implying that frequent search (i.e., monitoring) will be uniquely costly.

Leisure time in the ATUS. The data used in the calibration described in the main text comes from two sources: (i) the October 2003 Current Population Survey (CPS) and (ii) the 2003 American Time Use Survey (ATUS). Specifically, I merge the datasets using common respondent IDs and restrict the sample to respondents who, at the time of the ATUS survey, were between the ages of 20 and 65. I then compute total hours of leisure/home production (defined below), convert the value to a fraction of a 16 hour day, and compare the mean leisure time among unemployed workers who report that they engaged in online job search in the October 2003 CPS Computer and Internet Use Supplement (the same data source used for calibrating the share of monitoring workers) with workers who report being employed. If online job search is interpreted as monitoring, this comparison corresponds to $\bar{l}$ in the model, since workers who monitor in the model only receive $\bar{l}$ units of leisure/home production and the flow value of employment is normalized to be the wage (and hence leisure time of employed workers needs to be netted out of the flow value of unemployment).

As a baseline, I consider a broad measure of leisure/home production: The sum of all time spent on any activity *not* in the ATUS "Work and Work-Related Activities" (050000) category. I choose a broad measure as my baseline because doing so will tend to yield a conservative estimate of monitoring costs given my calibration strategy of fixing the flow value of unemployment to a value in the literature and then measuring monitoring costs as the residual of the flow value after netting out benefits and leisure/home production time. The results from such a calibration will therefore represent a lower bound on the influence of monitoring on equilibrium outcomes. The calculation described above implies that the fraction of time allocated to leisure/home production is 67% for employed workers and 95% for unemployed workers who used the internet for job search, implying a difference of 28%, or $\bar{l} = 0.28$ as in the main text. The less conservative approach I also discuss in the text instead defines leisure/home production more narrowly to include "Household Activities" (020000) and "Socializing, Relaxing, and Leisure" (120000), in which case the leisure/home production differential would be 14%, implying $\bar{l} = 0.14$. As may be seen in Figure 2a, the model features multiple equilibria in both cases.

Monitoring and spillovers from frequent search. As discussed in the main paper, the defining feature of monitoring is that it entails *frequent* search. One reason monitoring may be particularly costly as a form of search—even if the total amount of time spent on search is not very high—is that, as documented by Krueger and Mueller (2011), the negative feelings associated with job search tend to spill over into other (non-search) activities on days when search occurs. Specifically, Krueger and Mueller (2011) document that an unemployed worker who searches on a particular day is likely to be more stressed during non-search activities on the same day, and that reported life satisfaction is lower for workers who searched on the previous day, even after controlling for total time spent searching over the previous week and person fixed effects. As the authors write, these results suggest that "the experience of job search has a substantial and lingering effect on reported life satisfaction." These observations imply that spreading out a fixed amount of search time over multiple episodes of search, as I show in Appendix F tends to happen with online job search, may lead to more disutility than simply confining search to a single long episode.

A simple back-of-envelope calculation can be used to compute a value for $\bar{l}$ based on such stress spillovers from search. First, note that regressions similar to those in Appendix F imply that the probability of a worker engaging in search (for any amount of time) on a particular weekday, conditional on searching online, is approximately $p^{M\text{-}F} = 0.66$, and the corresponding figure for weekends is $p^{SS} = 0.38$. Suppose that weekdays entail half of a day of work for employed workers (an 8-hour workday out of 16 hours awake) and that weekends entail no work for employed workers, and let $s \in [0, 1]$ denote the spillover stress from search—that is, the fraction of a day's leisure/home production that is contaminated by an episode of search. Then $\bar{l}$ can be computed as:

$$\bar{l} = \frac{5}{7}\left(p^{M\text{-}F}\frac{16}{16}(1-s) + (1-p^{M\text{-}F})\frac{16}{16} - \frac{8}{16}\right) + \frac{2}{7}\left(p^{SS}\frac{16}{16}(1-s) + (1-p^{SS})\frac{16}{16} - \frac{16}{16}\right).$$

Substituting in the probabilities above and assuming that the consumption value of leisure/home production on a day on which a worker engages in any amount of search is 90% of what it would otherwise be (i.e., $s = 0.1$), the preceding implies a value of $\bar{l} = 0.3$, similar to the value of $\bar{l} = 0.28$ implied by the calibration in the paper.

Finally, it is interesting to note that, across the three steady states implied by the baseline calibration, the realized monitoring costs (that is, average monitoring costs conditional on monitoring) lie between 0.15 (in the low-monitoring steady state to which the model is calibrated) and 0.22 (in the high-monitoring steady state). These values are close to the marginal effect of online job search on the probability of a worker searching on a given day as calculated in Appendix F (which lie between 0.11 and 0.16), which can be interpreted as an alternative measure of monitoring costs. While these values are not an ideal target for calibration since realized monitoring costs in the model depend on the number of workers who monitor in equilibrium, and the SUWNJ was implemented during a tumultuous period when, as I argue in Section 5, the economy was moving towards a new steady state, it is nevertheless reassuring that the values fall within a similar range.

B.3 Job creation elasticity

The theoretical analog to the empirical elasticities in the studies I survey in the main text is the elasticity of employment with respect to the wage generated by either (i) a shock to a parameter of the model that only appears in the model's wage equation (e.g., δ_B) or (ii) a shock directly to the wage itself when treated as a parameter. Furthermore, these two approaches yield approximately the same results in a neighborhood around the steady state (i.e., for small shocks). To understand these claims, first note that the goal of these studies is to understand how exogenous changes in the wage affect employment, and hence to measure the slope of the job creation or labor demand curve. Exactly how exogenous variation in the wage is obtained varies depending on the paper—Beaudry et al. (2018) use a Bartik-style instrument for the wage, whereas minimum wage studies appeal to, e.g., border-county discontinuity designs—but regardless, the relevant observation is that, by only using exogenous variation in the wage, these studies are attempting to discard any source of variation in the wage that also directly impacts employment (rather than doing so only through the wage). This rules out variation driven by productivity shocks, separation shocks, etc. because such shocks appear in both the model's wage equation but also elsewhere in the model. Thus, the only remaining candidate sources of variation are shocks to parameters that only appear in the wage equation—for example, δ_B. Furthermore, because the model is approximately linear in a neighborhood around the steady state, the change in employment following such any such shock (provided it is relatively small) will be proportional to the induced change in the wage, implying that any such shock will generate approximately the same value for the elasticity of employment with respect to the wage as the value generated by simply treating the wage as parametric and perturbing it directly.[50] I have verified this claim numerically: The former approach (as described in the main text) yields $\nu = 0.3857543$, whereas the latter (when the variation is instead driven by a shock of the same magnitude to w), yields $\nu = 0.3857541$, a difference of 0.0000002 that vanishes as the shock becomes arbitrarily small.

B.4 Sensitivity and robustness

Sensitivity analysis. Figure 7 reports the values of five moments from the model at each of the three (potential) steady states as δ_B, ν, μ_{κ^*}, and σ_{κ^*} vary from their baseline values (indicated with vertical lines). Multiplicity of steady states for a given parameter value is indicated by the presence of multiple plot lines (solid gold; dashed red; dot-dashed blue) passing through the corresponding value on the horizontal axis.[51]

From the figure, we observe that larger values of δ_B, smaller values of ν, or an increase in μ_κ will tend to eliminate the low- and intermediate-monitoring steady states, whereas substantial increases in ν or reductions in μ_κ will tend to eliminate the intermediate- and high-monitoring steady states. The multiplicity of steady states that arises in the baseline calibration therefore appears to be somewhat sensitive to changes in parameter values.

Notwithstanding this observation, the figure also highlights that the mere *possibility* of multiplicity is important even if the economy happens to only be nearby, but not in, a region where multiplicity obtains: If, as is the case in several of the plots in Figure 7, a somewhat higher value of a particular parameter results in only the high-κ^* steady state existing, whereas a somewhat lower value of the same parameter results in only the low-κ^* steady state existing (as is the case with μ_κ, for example), then small shifts in the parameter value can move the economy across *unique but qualitatively different* steady states, implying potentially dramatically different dynamics and optimal policy prescriptions.[52] Thus, the central message of the paper—that monitoring technologies can be destabilizing—continues to hold whether the economy is in a region of multiplicity, or just nearby one.

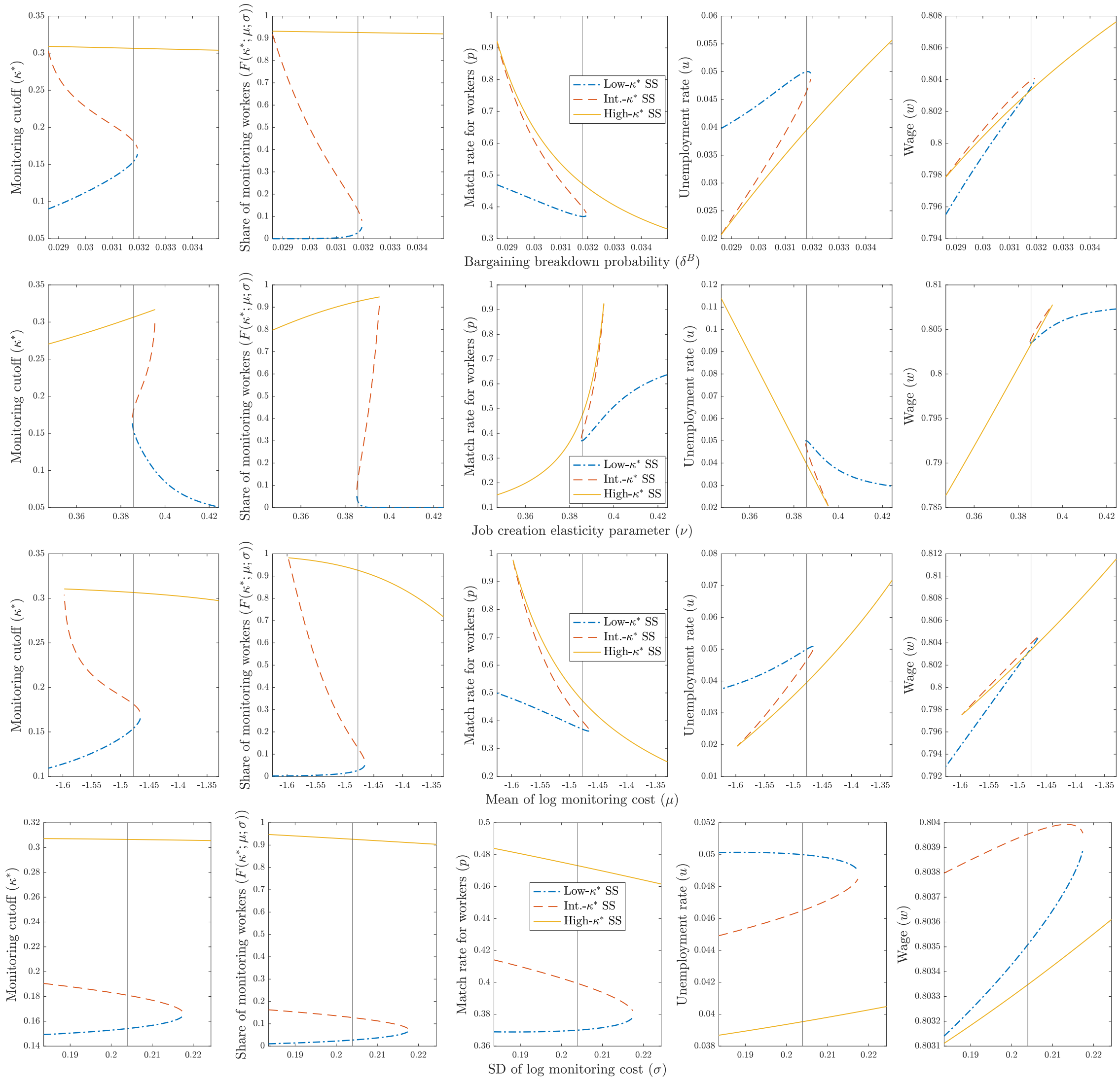
Robustness to alternative moments. Figure 8 depicts results from the Great Recession simulation under four alternative target moments. In three of the four scenarios—all of which generate multiple steady states—the model is able to provide a strong out-of-sample fit for the dynamics of unemployment following the Great Recession. This is important evidence that the particular moments used for calibration do not appear to be driving the main results.

[50] One can think of this latter approach as adding a "wage shock" to the model.

[51] For example, at a value of $\delta_B = 0.03$, we observe that there would still be multiple steady states since a hypothetical vertical line emanating from this value in the plots in the first row would pass through all three steady state curves.

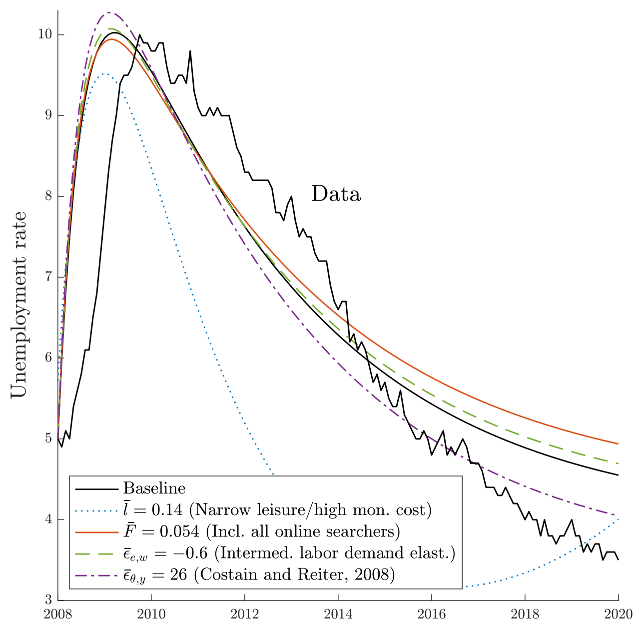
[52] As an extreme example, consider ν in Figure 7: A small reduction in ν will tend to eliminate the intermediate- and low-κ^* steady states, whereas a small increase will tend to eliminate the intermediate- and high-κ^* steady states. But because the low- and high-κ^* steady states that are selected by these changes are dramatically different in terms of the moments depicted in the figure—to say nothing of the dynamics around these steady states—the fact that the economy can exhibit multiplicity in a region of the parameter space can still be destabilizing in a meaningful sense.

Figure 7: Parameter sensitivity



Notes: Rows vary structural parameters (solid vertical lines indicate the value from the baseline calibration); columns report different model-implied moments. Row 1: δ_B (+/-10%); Row 2: ν (+/-10%); Row 3: μ_κ (+/-10%); Row 4: σ_κ (+/-10%). Column 1: Monitoring cutoff (κ^*); Column 2: Share of monitoring workers (F); Column 3: Match rate for workers (p); Column 4: Unemployment rate (u); Column 5: Wage (w). Dot-dashed blue line corresponds to the low-κ^* SS, dashed red line corresponds to the intermediate-κ^* SS, and the solid gold line corresponds to the high-κ^* SS.

Figure 8: Robustness to alternative moments



Notes: Dotted blue line: Uses the alternative value of $\bar{l} = 0.14$ corresponding to a narrower definition of leisure in the ATUS and thus implying larger monitoring costs (the precise definition of leisure that gives rise to this value is discussed in the main text); Solid red line: Uses an alternative value of $\bar{F} = 0.054$ corresponding to defining the share of monitoring workers as including *all* workers engaging in online job search (rather than the share of online searchers who report using online job boards specifically, as in the baseline calibration); Dot-dashed green line: Uses a value for the labor demand elasticity that falls between point estimates surveyed in the main text; Dashed purple line: Uses a value for the elasticity of tightness with respect to productivity taken from Costain and Reiter (2008) (rather than Shimer (2005), as in the baseline calibration).

C Traditional model

This appendix describes a traditional model of matching reflecting how search is typically modeled in the literature—as a decision concerning the intensity of effort rather than as a decision concerning whether to monitor new jobs.

C.1 Search as variable effort

The traditional approach to modeling job search decisions in quantitative macroeconomic models is to assume that searchers and non-searchers are perfect substitutes in the match function. In this case, non-searchers are simply ineffective searchers who match at a rate proportional to that of searchers. Recent examples of such models include Krusell et al. (2017) and Cairo et al. (2021). The description of search and matching below reflects this interpretation.

Consider a model in which workers choosing to search exert one unit of effort, whereas workers choosing not to search exert $z < 1$ units of effort. "Effort" here can be interpreted as the probability of participating in the market. Taking this interpretation and dropping time subscripts for ease of notation, the effective measure of searchers in the economy, s, can be written as

$$s = F(\kappa^*)u + (1 - F(\kappa^*))zu \tag{C.1}$$

$$= \Big[F(\kappa^*) + (1 - F(\kappa^*))z\Big]u \tag{C.2}$$

$$= \tilde{s}u. \tag{C.3}$$

where $\tilde{s}$ can be interpreted as average search effort in the economy.

C.2 Dynamic model

Adopting the functional form of the match function from Section 3, the total number of matches in this economy is now given by

$$m_t = \mu(v_t, s_t) = v_t \left[1 - e^{-\psi \frac{s_t}{v_t}}\right]. \tag{C.4}$$

The implied match rates for searchers (high-effort) and non-searchers (low-effort), respectively, are thus

$$p_t^h = \frac{v_t}{s_t}\left[1 - e^{-\psi \frac{s_t}{v_t}}\right] \tag{C.5}$$

$$p_t^l = z\frac{v_t}{s_t}\left[1 - e^{-\psi \frac{s_t}{v_t}}\right]. \tag{C.6}$$

The average match rate for workers, in turn, is

$$p_t = F(\kappa_t^*)p_t^h + (1 - F(\kappa_t^*))p_t^l \tag{C.7}$$

$$= \theta_t \left[1 - e^{-\psi \frac{\tilde{s}_t}{\theta_t}}\right] \tag{C.8}$$

where the second line uses the expression for s_t in (C.3), (C.5), (C.6), and $\theta_t \equiv v_t/u_t$. Thus, changes in average search effort are isomorphic to changes in ψ, the parameter governing frictions in the matching process. Note that this implies that it is possible to write the difference between the match rate associated with searching and the match rate associated with not searching (which governs the search decision) as

$$p_t^h - p_t^l = \frac{1-z}{\tilde{s}_t} p_t. \tag{C.9}$$

Finally, the average match rate for firms in the traditional model is just

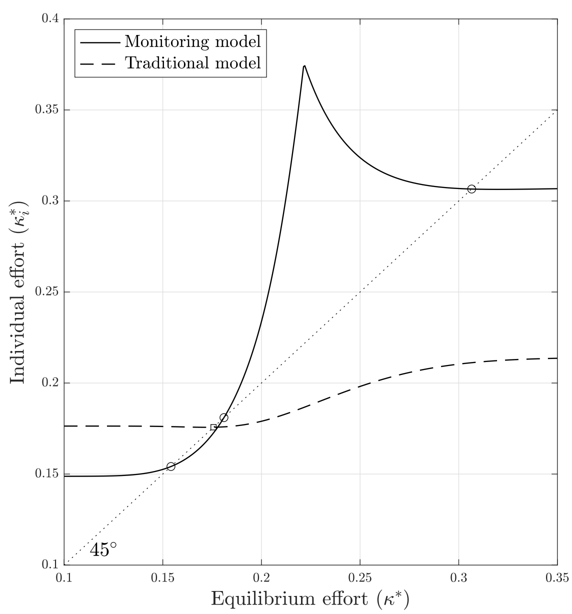
$$q_t = p_t / \theta_t = 1 - e^{-\psi \frac{\tilde{s}_t}{\theta_t}}. \tag{C.10}$$

The foregoing can be used to calibrate the model using the exact same moments that are used to calibrate the model in Section 4 as described in Appendix B.

C.3 Calibrated model

The dashed line in Figure 9 depicts the best-response function implied by embedding this more traditional model of search within an otherwise identical framework and calibration procedure to the one I describe above. The more traditional model has a unique steady state, and important observation because it highlights that a key feature of the model giving rise to multiple equilibria is the conceptualization of search as a monitoring decision.

Figure 9: Traditional model



Notes: The figure depicts the best-response function in the baseline model (solid line) and the best-response function in a model with a traditional interpretation of search (dashed line).

D Social welfare

To understand the implications of monitoring for social welfare, it is useful to express steady state social welfare (i.e., per-period consumption/leisure net of costs) as a function of the monitoring cutoff κ^* (that is, treating κ^* as exogenous and not imposing (29)):

$$\Omega(\kappa^*) = \underbrace{(1-u(\kappa^*))y}_{\text{Production }(\Omega_y)} + \underbrace{u(\kappa^*)\left[\bar{l} + \int_{\kappa^*}^{\infty} \kappa dF(\kappa)\right]}_{\text{Leisure/home prod. }(\Omega_l)} - \underbrace{\left[cv(\kappa^*) + m\int_0^{V^n(\kappa^*)} \xi dG(\xi)\right]}_{\text{Vacancy costs + Entry costs }(\Omega_v)}. \quad \text{(D.1)}$$

The first term is increasing in κ^* because unemployment is decreasing in κ^*. The second term is decreasing in κ^* because more workers choosing to monitor implies both (i) fewer workers unemployed who can enjoy leisure/home production and (ii) a smaller *share* of unemployed workers choosing to enjoy leisure/home production. The final term is less clear: More workers monitoring means that there are fewer vacancies that are not matched immediately, and thus fewer firms paying costs of maintaining a vacancy (c), but it also depresses wages by reducing the value of unemployment, which in turn increases the value to an entrepreneur of posting a new vacancy, thus resulting in more entry from entrepreneurs with higher costs. The final row of Table 3 shows that the net effect of these forces is a non-monotonic social welfare function in κ^*: At low levels of monitoring, the efficiency gains from reallocating workers to the frictional aftermarket to the frictionless monitoring process and the associated savings on vacancy posting costs are large, and dwarf the additional search costs, leading to higher social welfare in the intermediate-κ^* equilibrium than in the low-κ^* equilibrium. However, these efficiency gains eventually become exhausted, and the additional search costs associated from more monitoring dominate the social welfare function, leading to a roughly four tenths of a percentage point reduction in social welfare at the high-κ^* steady state relative to the low-κ^* steady state. Interestingly, the results for the low-κ^* and high-κ^* steady states in Table 3 reverse the Pareto ranking obtained in Diamond (1982a), in which the "high activity" equilibrium necessarily dominates the "low activity" one. This important difference reflects the strong congestion externalities associated with monitoring technology that are not internalized by job seekers who are trying to avoid being shut out of the queue for jobs combined with the inelastic job creation process.

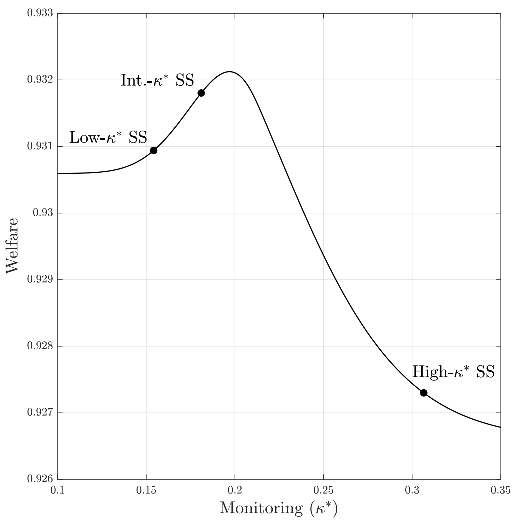
Quantitatively, the important tradeoff for social welfare is between the first two terms, Ω_y and Ω_l in the social welfare function in (D.1). Moreover, the first term can be decomposed into two components, reflecting the channels through which more monitoring increases total production: A component due to reduced matching frictions associated with monitoring (recall that the monitoring phase is frictionless whereas the aftermarket is not, so as more workers decide to monitor, the matching process becomes more efficient), and a component due to more entry/job creation resulting from monitoring. Formally, making explicit the dependence of Ω_y on κ^* via unemployment, which in turn depends on κ^* both directly and via entry of new firms, Ω_y can be written as

$$\Omega_y(\kappa^*) = \Big(1 - u(\kappa^*, v^n(\kappa^*))\Big)y \quad \text{(D.2)}$$

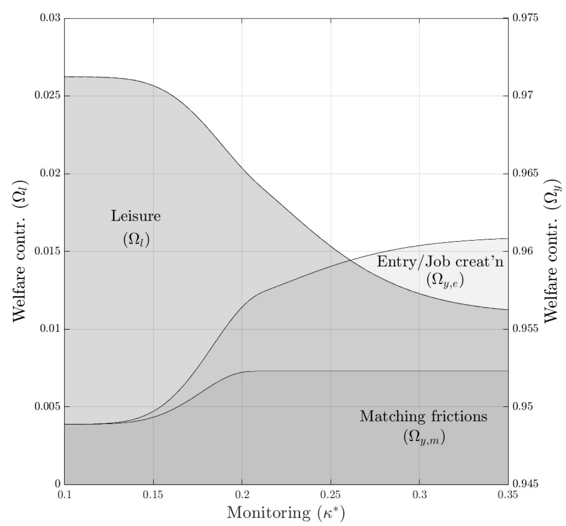
$$= \underbrace{\Big(1 - u(\kappa^*, \underline{v}^n)\Big)y}_{\text{Matching frictions }(\Omega_{y,m})} + \underbrace{\Big(u(\kappa^*, \underline{v}^n) - u(\kappa^*, v^n(\kappa^*))\Big)y}_{\text{Entry/job creation }(\Omega_{y,e})} \quad \text{(D.3)}$$

where $\underline{v}^n$ is some fixed level of job creation. Fixing $\underline{v}^n = 0.25$, Figure 10a plots total social welfare Ω as a function of κ^*, indicating the levels of monitoring associated with the three steady states. Figure 10b plots the three most important components of social welfare, $\Omega_{y,m}$, $\Omega_{y,e}$, and Ω_l, also as functions of κ^*, with the left axis measuring the contribution of leisure/home production and the right axis measuring the contribution of the components of output. Figure 10 reveals several

Figure 10: Social welfare



(a) Total



(b) Components

Notes: Panel (a) depicts social welfare (Ω) as a function of the equilibrium monitoring cutoff (κ^*). Panel (b) depicts various components of social welfare (Ω_l, $\Omega_{y,e}$, and $\Omega_{y,m}$) as functions of the equilibrium monitoring cutoff (κ^*).

interesting features of how monitoring influences social welfare. First, as can be seen in Figure 10a, social welfare is non-monotone with respect to the monitoring cutoff κ^*. This reflects the competing forces of leisure and output described above. Second, the low-monitoring steady state Pareto dominates the high-monitoring steady state. As can be seen in Figure 10b, this results from the fact that, at sufficiently high levels of monitoring, the increased matching efficiency and increased job creation associated with more monitoring contribute an increasingly small amount to social welfare, whereas the leisure/home production costs of more monitoring are continuing to substantially detract from social welfare. Finally, the model implies that the gains from increased monitoring are largely due to entry rather than reduced matching frictions.

E ψ-limit economy

This appendix derives equations (E.14)-(E.17) that characterize the limit of the model in Section 3 as $\psi \to \infty$.[53] When the aftermarket is frictionless, provided $\theta_t < 1$, all firms match immediately, and thus $\hat{q}_t = q_t^n = q_t^o = q_t = 1$.[54]

E.1 Model

Unemployment

Because firms are guaranteed to match by the end of the period, $m_t = v_t$, which implies that the law of motion for unemployment in (14) can be written as

$$u_{t+1} = u_t + \delta(1-u_t) - (1-\delta)v_t. \tag{E.1}$$

Vacancies

Combining the vacancy value functions in (33), (34), and (35) with the entry condition in (32) and the fact that $q_t = 1$ implies $v_t^n = v_t$ via (15), so that

$$v_t = \eta_t G\left(-c + \beta(1-\delta)\left[y_{t+1} - w_{t+1}(1+\tau_{t+1}) + c + G^{-1}\left(\frac{v_{t+1}}{\eta_{t+1}}\right)\right]\right). \tag{E.2}$$

Monitoring

From (29), the optimality condition for monitoring is

$$\kappa_t^* = \beta(1-\delta)(p_t^m - p_t^w)\mathbb{E}\left[W_{t+1} - U_{it+1}\right]. \tag{E.3}$$

Using (25)-(28), we have

$$\mathbb{E}[W_t - U_{it}] = w_t(1-b) - \bar{l} - \textstyle\int_{\kappa_t^*}^{\infty} \kappa dF(\kappa) + \beta(1-\delta)(1-\frac{v_t}{u_t})\mathbb{E}[W_{t+1} - U_{it+1}]. \tag{E.4}$$

Solving the time-t and time-$t+1$ versions of (E.3) for the employment rent (i.e., $\mathbb{E}\left[W_{t+1} - U_{it+1}\right]$ and $\mathbb{E}\left[W_{t+2} - U_{it+2}\right]$), and substituting the resulting expressions into the period-ahead version of (E.4), we obtain

$$\kappa_t^* = \beta(1-\delta)\frac{1-v_t/u_t}{1-F(\kappa_t^*)}\left[w_{t+1}(1-b) - \bar{l} - \int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + (1-\frac{v_{t+1}}{u_{t+1}})\frac{\kappa_{t+1}^*}{p_{t+1}^m - p_{t+1}^w}\right]. \tag{E.5}$$

[53]Note that I index η by t in order to indicate that I will treat this as an exogenous variable in the quantitative analysis of Section 5.

[54]I also assume throughout the analysis below that there are no taxes to facilitate exposition.

Finally, when $F(\kappa_t^*)u_t < v_t$,

$$p_t^m - p_t^w = \frac{1 - v_t/u_t}{1 - F(\kappa_t^*)} \tag{E.6}$$

and when $F(\kappa_t^*)u_t > v_t$,

$$p_t^m - p_t^w = \frac{v_t/u_t}{F(\kappa_t^*)}. \tag{E.7}$$

Combining the preceding, the optimal monitoring condition can be written as

$$\kappa_t^* = \begin{cases} \beta(1-\delta)\frac{1-v_t/u_t}{1-F(\kappa_t^*)}\left[w_{t+1}(1-b) - \bar{l} - \int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + (1-F(\kappa_{t+1}^*))\kappa_{t+1}^*\right] & \text{if} \quad F(\kappa^*) \leq \frac{v_t}{u_t} \\ \beta(1-\delta)\frac{v_t/u_t}{F(\kappa_t^*)}\left[w_{t+1}(1-b) - \bar{l} - \int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + (1-\frac{v_{t+1}}{u_{t+1}})\frac{v_{t+1}}{u_{t+1}F(\kappa_{t+1}^*)}\kappa_{t+1}^*\right] & \text{if} \quad F(\kappa^*) > \frac{v_t}{u_t}. \end{cases} \tag{E.8}$$

Wages

Recall that the model's wage equation in (38) is

$$\begin{aligned} w_t =& (1-\beta(1-\delta))(\mathbb{E}[\kappa_{it}] + \bar{l}) \\ &+ \frac{\beta(1-\delta^B)}{1+\tau_{t+1}}\left[y_{t+1} + \gamma(1-\beta(1-\delta)) - \beta(1-\delta^B)(y_{t+2} - w_{t+2}(1+\tau_{t+2}))\right] \\ &+ \beta(\delta^B - \delta)\mathbb{E}[U_{it+1} - \beta U_{it+2}]. \end{aligned} \tag{E.9}$$

To operationalize this equation we need to eliminate the value functions in the last term on the right-hand side. Using (26)-(28), after some manipulation we can write

$$\mathbb{E}[U_{it+1} - \beta U_{it+2}] = \int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + bw_{t+1} + \bar{l} + \frac{v_{t+1}}{u_{t+1}}\beta(1-\delta)\mathbb{E}[W_{t+1} - U_{it+2}]. \tag{E.10}$$

Using the period-ahead version of (E.3), this can be written as

$$\mathbb{E}[U_{it+1} - \beta U_{it+2}] = \int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + bw_{t+1} + \bar{l} + \frac{v_{t+1}}{u_{t+1}}\left(\frac{\kappa_{t+1}^*}{p_{t+1}^m - p_{t+1}^w}\right). \tag{E.11}$$

Substituting (E.11) into (E.9),

$$\begin{aligned} w_t =& (1-\beta(1-\delta))(\mathbb{E}[\kappa_{it}] + \bar{l}) \\ &+ \frac{\beta(1-\delta^B)}{1+\tau_{t+1}}\left[y_{t+1} + \gamma(1-\beta(1-\delta)) - \beta(1-\delta^B)(y_{t+2} - w_{t+2}(1+\tau_{t+2}))\right] \\ &+ \beta(\delta^B - \delta)\left[\int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + bw_{t+1} + \bar{l} + \frac{v_{t+1}}{u_{t+1}}\left(\frac{\kappa_{t+1}^*}{p_{t+1}^m - p_{t+1}^w}\right)\right]. \end{aligned} \tag{E.12}$$

Using (E.6) and (E.7), assuming for simplicity that $\tau_t = 0 \ \forall t$, and defining $\chi \equiv (1-\beta(1-\delta))[\mu + \bar{l}] + \beta(1-\delta^B)[y(1-\beta(1-\delta^B)) + \gamma(1-\beta(1-\delta))]$, (E.12) can be written as

$$w_t = \begin{cases} \chi + (\beta(1-\delta^B))^2 w_{t+2} + \beta(\delta^B - \delta)\left[\int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + bw_{t+1} + \bar{l} + \frac{v_{t+1}}{u_{t+1}}\left(\frac{1-\frac{v_{t+1}}{u_{t+1}}}{1-F(\kappa_{t+1}^*)}\right)\kappa_{t+1}^*\right] & \text{if} \quad F(\kappa^*) \leq \frac{v_t}{u_t} \\ \chi + (\beta(1-\delta^B))^2 w_{t+2} + \beta(\delta^B - \delta)\left[\int_{\kappa_{t+1}^*}^{\infty} \kappa dF(\kappa) + bw_{t+1} + \bar{l} + \frac{v_{t+1}}{u_{t+1}}\left(\frac{\frac{v_{t+1}}{u_{t+1}}}{F(\kappa_{t+1}^*)}\right)\kappa_{t+1}^*\right] & \text{if} \quad F(\kappa^*) > \frac{v_t}{u_t}. \end{cases} \tag{E.13}$$

Collecting equations, the ψ-limit economy can be be summarized by the following four equations:

$$u_{t+1} = u_t + \delta(1-u_t) - (1-\delta)v_t \tag{E.14}$$

$$v_t = \eta G\left(-c + \beta(1-\delta)\left[y_{t+1} - w_{t+1} + c + G^{-1}\left(\frac{v_{t+1}}{\eta}\right)\right]\right) \tag{E.15}$$

$$\kappa_t^* = \begin{cases} \beta(1-\delta)\frac{1-v_t/u_t}{1-F(\kappa_t^*)}\left[w_{t+1}(1-b) - \bar{l} - \int_{\kappa_{t+1}^*}^{\infty}\kappa dF(\kappa) + (1-F(\kappa_{t+1}^*))\kappa_{t+1}^*\right] & \text{if } F(\kappa^*) \leq \frac{v_t}{u_t} \\ \beta(1-\delta)\frac{v_t/u_t}{F(\kappa_t^*)}\left[w_{t+1}(1-b) - \bar{l} - \int_{\kappa_{t+1}^*}^{\infty}\kappa dF(\kappa) + (1-\frac{v_{t+1}}{u_{t+1}})\frac{v_{t+1}}{u_{t+1}F(\kappa_{t+1}^*)}\kappa_{t+1}^*\right] & \text{if } F(\kappa^*) > \frac{v_t}{u_t} \end{cases} \tag{E.16}$$

$$w_t = \begin{cases} \chi + (\beta(1-\delta^B))^2 w_{t+2} + \beta(\delta^B - \delta)\left[\int_{\kappa_{t+1}^*}^{\infty}\kappa dF(\kappa) + bw_{t+1} + \bar{l} + \frac{v_{t+1}}{u_{t+1}}\left(\frac{1-\frac{v_{t+1}}{u_{t+1}}}{1-F(\kappa_{t+1}^*)}\right)\kappa_{t+1}^*\right] & \text{if } F(\kappa^*) \leq \frac{v_t}{u_t} \\ \chi + (\beta(1-\delta^B))^2 w_{t+2} + \beta(\delta^B - \delta)\left[\int_{\kappa_{t+1}^*}^{\infty}\kappa dF(\kappa) + bw_{t+1} + \bar{l} + \frac{v_{t+1}}{u_{t+1}}\left(\frac{\frac{v_{t+1}}{u_{t+1}}}{F(\kappa_{t+1}^*)}\right)\kappa_{t+1}^*\right] & \text{if } F(\kappa^*) > \frac{v_t}{u_t} \end{cases} \tag{E.17}$$

where $\chi \equiv (1-\beta(1-\delta))[\mu + \bar{l}] + \beta(1-\delta^B)[y(1-\beta(1-\delta^B)) + \gamma(1-\beta(1-\delta))]$.

E.2 Moments

Table 4 reports the moments from the ψ-limit economy alongside (in parentheses) the corresponding moments from the baseline economy (identical to those in Table 3 in the main text). As is clear

Table 4: Moments from ψ-limit economy

	Data	Monitoring model		
		Low-κ^*	*Int.-κ^**	*High-κ^**
Targeted moments				
Unemployment rate (u)	5.0%	4.4% (5.0%)	4.3% (4.7%)	4.0% (4.0%)
Job-finding prob. (p)	0.37	0.421 (0.370)	0.429 (0.399)	0.473 (0.473)
Share monitoring ($F(\kappa^*)$)	2.7%	0.0% (2.7%)	18.8% (12.8%)	92.6% (92.6%)
Untargeted moments				
Wage (w)		0.805 (0.804)	0.804 (0.804)	0.803 (0.803)
Job creation (v^n/v)		1 (0.864)	1 (0.930)	1 (1)
Job-finding prob. (p^m)		1 (1)	1 (1)	0.511 (0.511)
Job-finding prob. (p^w)		0.421 (0.352)	0.297 (0.311)	0 (0)
Social welfare (Appendix D)				
Output net of costs (Ω)		0.934 (0.931)	0.933 (0.932)	0.927 (0.927)

Notes: Moments for the frictional economy (identical to those in the corresponding table in the main text) are in parentheses following moments from the ψ-limit economy.

from the table, both economies feature three equilibria that are qualitatively similar. The high-monitoring steady state is identical in the baseline economy and in the ψ-limit economy because in the high-monitoring steady state all vacancies match with monitoring workers at the start of the period in the (frictionless) monitoring stage and thus the (frictional) aftermarket ceases to be operative. Social welfare is monotonically decreasing in the amount of monitoring because one of the main benefits of monitoring in the frictional economy—the reallocation of workers from the frictional aftermarket to frictionless monitoring—is no longer operative as $\psi \to \infty$.

F Evidence on Monitoring and Job Offers

The analysis in this appendix leverages high-frequency panel data from the Survey of Unemployed Workers in New Jersey (SUWNJ) to shed light on two central assumptions in this paper: (i) that online search technologies reduce the fixed costs of search, enabling more frequent search that is less "clumpy",[55] and (ii) that more frequent search improves job-finding prospects.

The SUWNJ is a weekly longitudinal survey (lasting for up to 26 weeks) of 6,025 unemployment insurance benefit recipients in New Jersey, conducted between 2009 and 2010.[56] Importantly, the survey contains data on job search behavior, both in the form of weekly recall questions about activity over the entire previous week, as well as a once-weekly time diary. Most relevantly for the purposes of this paper, the SUWNJ contains a question asking respondents whether they used the internet to search for work in the past week, and information from time diaries on whether someone searched on the previous day and how that search was distributed over the course of the day. I first show that workers who use the internet for job search are more likely to search on any given day and also engage in more episodes of search within a day than those who do not, conditional on time spent searching. I then show that workers who engage in more episodes of search within a day are more likely to receive a job offer in the subsequent week than those who do not.

F.1 Online job search is less clumpy

If online job search is less clumpy than traditional job search, then the probability that a respondent searches on any given day—and the number of distinct episodes of search within a day—should be higher for a person who uses the internet for job search than for a person who does not, all else equal. Following this logic, I estimate two regressions. First, I estimate a linear probability model in which I regress an indicator for whether or not a respondent reports engaging in any search on the previous day on an indicator for whether or not the respondent reports having used the internet for job search in the past week, as well as the total amount of time reported searching in the previous week, unemployment duration, calendar-week, day-of-week, and respondent fixed effects. Second, I estimate a Poisson model in which I regress the number of distinct episodes of search on the previous day on the same controls plus total reported search time on the previous day.[57] In both cases, I am interested in the coefficient on the online search term, indicating whether online search is associated with an increased likelihood of search on a given day and more episodes of search on a given day. In all specifications, I restrict attention to respondents between the ages of 20 and 60 who have not accepted a job offer. Moreover, I follow Krueger and Mueller (2011) and drop time diaries in which fewer than four distinct activities are reported or in which three or more hour-long episodes are left incomplete.

Table 5 reports the main results. The first two columns correspond to the linear probability model, while the second two columns correspond to the Poisson model. Within each model, the first

[55] See Zhang et al. (2013) for a discussion of various ways of measuring "clumpiness" in incidence data.

[56] Complete survey data and documentation can be obtained from https://dss.princeton.edu/catalog/resource1350.

[57] I measure the number of distinct episodes of search as the number of continuous episodes of search within a day (i.e., adjacent hours all containing some search) separated by at least one period without search.

column ("Spec. 1") corresponds to the baseline regression described above and the second column ("Spec. 2") adds a control for the total number of applications that a respondent reports submitting during the week prior to the survey and 12 dummies for whether or not a respondent used various specific search methods during the week prior to the survey.[58]

Table 5: Online job search and search frequency

	Pr(Daily search)		*Distinct episodes*	
	Spec. 1	Spec. 2	Spec. 1	Spec. 2
Online search	0.16***	0.11***	0.56***	0.48***
	(0.03)	(0.03)	(0.10)	(0.11)
Weekly search	0.03***	0.01**	0.01	0.00
	(0.00)	(0.00)	(0.01)	(0.01)
Weekly app'ns		0.04***		-0.00
		(0.01)		(0.02)
Fixed effects:				
Respondent:	×	×	×	×
Calendar time:	×	×	×	×
Day of week:	×	×	×	×
Search methods:		×		×
Observations	17,024	16,108	15,743	14,895

Robust standard errors in parentheses.
Notes: Standard errors clustered at the individual level. Sample restricted to respondents who have never accepted a job offer and who are between the ages of 20 and 60. All specifications use survey weights.

The results in the first two columns indicate that engaging in online search is associated with an increased probability of searching on the previous day of at least 11 percentage points, an effect that is highly significant across specifications. The results in the second two columns indicate that online search is associated with a nearly 50% increase in the average number of episodes of search on the previous day, which is also highly significant. Both results are consistent with online search allowing job seekers to smooth out their search across days rather than devoting, e.g., a single day of the week to search to economize on fixed costs. Moreover, the results are robust to various alternative specifications, sample selection criteria, and inclusion of additional controls.

F.2 Less clumpy search is associated with more job offers

The second key assumption of the model, as discussed in Section 6, is that there is a first-mover advantage to search. If this is so, then more frequent search should increase the likelihood of a job offer, conditional on total time spent on search. To test this in the SUWNJ, I regress an indicator

[58] The 12 options are contacting an employer directly, contacting a public employment agency, contacting a private employment agency, contacting friends or relatives, contacting a school/university employment center, checking union/professional registers, attending a job training program/course, placing or answering ads, going to an interview, sending out resumes/completing applications, looking at ads, and a category for all other methods.

for whether or not a worker reports receiving a job offer in the following week on the number of distinct episodes of search on the preceding day and all of the other controls considered in the regressions described above, including respondent fixed effects.

Table 6: Search frequency and job offers

	Spec. 1	Spec. 2
Episodes	0.010**	0.011**
	(0.005)	(0.005)
Search yesterday	-0.011**	-0.012**
	(0.005)	(0.005)
Weekly search	-0.003	-0.004
	(0.002)	(0.002)
Weekly app'ns		0.010**
		(0.005)
Fixed effects:		
Respondent:	×	×
Calendar time:	×	×
Day of week:	×	×
Search methods:		×
Observations	14,886	14,127

Robust standard errors in parentheses.
Notes: Standard errors clustered at the individual level. Sample restricted to respondents who have never accepted a job offer and who are between the ages of 20 and 60. All specifications use survey weights.

Table 6 reports the main results. The important observation corresponds to the first row: Even after controlling for various measures of weekly and daily search time and respondent fixed effects, the frequency of search appears to be positively related to the probability of receiving an offer. The point estimates indicate that an additional episode of search is associated with a one percentage point increase in the probability of an offer in the following week, which is significant at the 5% level in both specifications. I view this as suggestive additional evidence of the mechanism analyzed in the paper.