WORKING PAPER NO. 17-40
THE AGGREGATE EFFECTS OF LABOR MARKET FRICTIONS

Michael W. L. Elsby
University of Edinburgh

Ryan Michaels
Research Department
Federal Reserve Bank of Philadelphia

David Ratner
Board of Governors of the Federal Reserve System

November 8, 2017

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA
Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

The Aggregate Effects of Labor Market Frictions

Michael W. L. Elsby Ryan Michaels David Ratner[1]

November 8, 2017

Abstract

Labor market frictions are able to induce sluggish aggregate employment dynamics. However, these frictions have strong implications for the *source* of this propagation: They distort the path of aggregate employment by impeding the flow of labor across firms. For a canonical class of frictions, we show how observable measures of such flows can be used to assess the effect of frictions on aggregate employment dynamics. Application of this approach to establishment microdata for the United States reveals that the empirical flow of labor across firms deviates markedly from the predictions of canonical labor market frictions. Despite their ability to induce persistence in aggregate employment, firm-size flows in these models are predicted to respond aggressively to aggregate shocks but react sluggishly in the data. This paper therefore concludes that the propagation mechanism embodied in standard models of labor market frictions fails to account for the sources of observed employment dynamics.

JEL codes: E32, J63, J64

Keywords: Labor market frictions, firm dynamics, adjustment costs

[1] Elsby: University of Edinburgh (mike.elsby@ed.ac.uk). Michaels: Federal Reserve Bank of Philadelphia (ryan.michaels@phil.frb.org). Ratner: Federal Reserve Board (david.d.ratner@frb.gov).

We are grateful to Philipp Kircher, John Moore, Giuseppe Moscarini, Victor Ríos Rull, and anonymous referees, as well as seminar participants at numerous institutions, for helpful comments. All errors are our own. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. We thank Jess Helfand and Mike LoBue for their excellent support at the BLS. The views expressed here do not necessarily reflect the views of the BLS, the Federal Reserve Bank of Philadelphia, the staff and members of the Federal Reserve Board, or the Federal Reserve System as a whole.

Elsby and Michaels gratefully acknowledge financial support from the UK Economic and Social Research Council (ESRC), Award reference ES/L009633/1.

This paper is available free of charge at www.philadelphiafed.org/research-and-data/publications/working-papers/.

What are the effects of labor market frictions on aggregate employment dynamics? In this paper, we provide a new approach to this question for a canonical class of frictions. This class encompasses influential models of fixed adjustment costs that induce intermittent, discrete adjustments (Caballero, Engel, and Haltiwanger, 1997); per-worker hiring and firing costs that induce further distortions to the magnitude of adjustments (Bentolila and Bertola, 1990; Hopenhayn and Rogerson, 1993); and search and matching frictions (Pissarides, 1985; Mortensen and Pissarides, 1994).

These models of labor market frictions are a compelling class to study, from both micro- and macroeconomic perspectives. First, they are able to capture a key stylized fact of microeconomic establishment dynamics, namely, the empirical prevalence of inaction in employment adjustment. Second, we show that some models in this class, especially those including search frictions and other per-worker costs of hiring and firing, are also able to propagate aggregate shocks and induce sluggishness in aggregate employment dynamics, thereby contributing to a key stylized fact of macroeconomic adjustment. Thus, models in this class provide potentially fertile ground for an explanation of the microfoundations of aggregate employment dynamics. And, any successful explanation in this class will imply a prominent aggregate role for labor market frictions. Perhaps for these reasons, such models inform a large body of modern research on aggregate labor markets.[2]

Our contribution in this paper is to inspect the *channel* through which canonical labor market frictions distort the path of aggregate employment and to confront it with novel empirical evidence. We show that the fundamental channel through which models of frictions in this class are able to propagate aggregate employment dynamics is by restricting the incidence and/or size of employment adjustments, thereby retarding the flow of labor across the firm-size distribution. Aggregate employment dynamics and firm-size flows are thus inextricably linked in these models. And, crucially, these firm-size flows can be measured in establishment panel data, opening up the possibility of a new empirical evaluation of the propagation mechanism embodied in a large class of canonical models.

Our findings suggest that standard labor market frictions provide a poor account of the dynamics of firm-size flows. Under these models, we show that the flows are predicted to respond aggressively to aggregate shocks. Intuitively, since the frictions retard the flow of labor, there is a "pent-up" demand for adjusting, which implies that the flows (though dampened in levels) are very *elastic* to shifts in the aggregate state. In the data, however, firm-size flows evolve sluggishly following macroeconomic disturbances. Since the behavior of these flows lies at the heart of the propagation mechanism inherent in all models in

[2] An exhaustive list of models in this class is too numerous to cite. Additional examples include Hamermesh (1989), Caballero and Engel (1993) and Bachmann (2012) for fixed costs; Oi (1962) and Nickell (1978) for linear frictions; and Elsby and Michaels (2013) and Acemoglu and Hawkins (2014) for "large-firm" extensions of search frictions. Further prominent studies that consider hybrids of these frictions include Bertola and Caballero (1990), Mortensen and Nagypal (2007), Bloom (2009), Pissarides (2009), and Cooper, Haltiwanger, and Willis (2007, 2015).

this canonical class, this failure suggests that standard frictions also provide a poor account of the empirical sources of aggregate employment persistence. While they may account for microeconomic inaction and aggregate persistence, they do so at the cost of predicting counterfactual microdata on the firm-size flows through which these observations are predicted to be linked.

It is important to distinguish this contribution from prior related work. Several papers have explored the extent to which some of the models we consider are able to induce sluggish aggregate employment. Our contribution contrasts with, and builds on, this literature in two ways.

First, the majority of this literature explores the aggregate implications of *fixed* adjustment costs only (for example, Caballero and Engel, 2007; Elsby and Michaels, 2017; Khan and Thomas, 2008). By contrast, our analysis further admits analysis of linear and search costs. This is important since, as we will show, the latter generate much greater, and more realistic, propagation of aggregate employment dynamics relative to models of fixed costs.

Second, we further show that the propagation mechanism in this class of models can be evaluated empirically by inspecting the behavior of firm-size flows. This too is important, as prior literature has broadly neglected these flows. Yet, we show that these frictions impinge on aggregate employment outcomes *only* by distorting these flows. That the dynamics of the model-implied flows deviate substantially from their empirical counterparts therefore calls into question the credibility of the models' propagation mechanism.[3] In other words, even if some of these models are able to produce realistic sluggishness in aggregate employment, the means by which they do so can be, and is, refuted by the data.[4]

We begin in section 1 by establishing the theoretical results that will inform our later empirical analysis. Here, we show that intermittent adjustment implies that only a fraction of desired, frictionless adjustments are implemented, retarding flows of labor to and from each firm size relative to an economy without frictions. In addition, distortions to the magnitude of adjustments induced by per-worker or search frictions further divert inflows away from their frictionless destination. By obstructing these firm-size flows, labor market frictions distort aggregate employment, since the latter is proportional to the mean of the firm-size distribution.

In general, however, the flows to and from each position in the firm-size distribution are functions of the employment level at each position and are thus complicated objects to distill. We show in section 1 how it is possible to devise a single summary statistic for

[3] In this sense, our work is similar in spirit to Bils, Klenow, and Malin (2012) who use microdata on price adjustments to assess the propagation mechanism in monetary DSGE models.

[4] Moreover, to the extent that our results call into question the microfoundations of canonical labor market frictions, they suggest caution in carrying out policy analysis using these theories. Understanding the nature of the underlying frictions is a key task for future research.

the behavior of the firm-size flows that, in theory, provides a diagnostic for their aggregate effects. This summary statistic is the mean of a notional firm-size distribution associated with flow balance—that is, the distribution that equates inflows to outflows at each employment level. We show that a robust implication of canonical models is that aggregate flow-balance employment exhibits an overshooting property relative to its frictionless counterpart, rising more than frictionless employment in aggregate expansions, and declining more in recessions. This behavior of flow-balance employment reflects the fast-moving dynamics of the firm-size flows.[5]

This overshooting property is quite general and is shaped by two economic forces: a partial equilibrium effect that holds in the absence of adjustment of wages; and a further equilibrium effect induced by such wage adjustment. In partial equilibrium, the response of aggregate flow-balance employment to a positive aggregate shock captures a rightward shift in the distribution of desired employment, just as aggregate frictionless employment does. In addition, it reflects an increased propensity of firms to adjust *to* versus *from* high employment levels; the elasticity of these cross-sectional flows is a critical component of the model's dynamics. Consequently, mean flow-balance employment responds at least as much as its frictionless counterpart to aggregate shocks.

Equilibrium wage adjustment reinforces this property. Consider a rise in aggregate labor productivity. To the extent that labor market frictions attenuate the response of labor demand, equilibrium wages will rise less in the presence of frictions than in their absence. Hence, aggregate flow-balance employment is conditional on a smaller increase in wages. Equivalently, the rise in equilibrium frictionless employment is choked off to a greater extent by rises in wages. For this reason, the equilibrium response of aggregate flow-balance employment is further amplified relative to its frictionless counterpart.

We confirm these properties of canonical models in two sets of complementary results. The first establishes analytical results for popular special cases of the models in which frictionless labor demand evolves within each firm according to a random walk, and aggregate disturbances are unanticipated and permanent. The second explores numerical simulations that relax these assumptions. These theoretical results reveal that models in this class, especially variants with linear and search frictions, can induce significant propagation in aggregate employment dynamics. However, at the same time, all such models imply considerable overshooting of flow-balance employment relative to frictionless employment. Importantly, this overshooting property holds quite generally for a wide array of parameterizations of the persistence and volatility of idiosyncratic shocks, and of

[5] These fast-moving dynamics of the firm-size flows are reminiscent of earlier findings in related literature on price and capital frictions. For example, Calvo models of price setting, in which the adjustment probability is an exogenous constant, fail to capture the sluggishness of average price changes—i.e., aggregate inflation (Fuhrer and Moore, 1995; Mankiw and Reis, 2002). Similarly, Veracierto's (2002) early study of the special case of irreversible investment found numerically that the model failed to capture the sluggishness of average capital changes—aggregate investment. (See also Christiano and Todd, 1996.) Our results show analytically that the origins of such findings lie in the behavior of firm-size flows, can be generalized to a much wider class of frictions, and can be tested using microdata on firm dynamics.

the magnitude of adjustment frictions. Since there remains some uncertainty in the literature over these parameters, and how they might separately be identified, the generality of the result is especially useful.

The upshot of section 1, then, is that frictions in this class may distort the path of aggregate employment, but only by virtue of their ability to restrain the flow of labor across firms. However, while such frictions dampen the level of the cross-sectional flows, these flows are predicted to be highly elastic to aggregate shocks. A consequence is that employment under flow balance responds to shocks even more aggressively than its frictionless counterpart. A natural question is whether available data are consistent with such a stark response of firm-size flows, as summarized by aggregate flow-balance employment.

In section 2, we confront these implications of canonical models with empirical counterparts measured using rich establishment microdata. The data we use are derived from the U.S. Quarterly Census of Employment and Wages for the period 1992Q1 through 2014Q2. Being a natural establishment panel, these data enable us to observe the outflows from, and inflows to, each employment level in the employer-size distribution. Accordingly, we can derive an empirical measure of aggregate employment implied by flow balance along the lines suggested by the theoretical work of section 1.

Using this measure, we present the results of several exercises that assess the empirical relevance of the propagation mechanism in this class of models. An initial, revealing finding is that the empirical time series for aggregate flow-balance employment tracks very closely the time series for actual, observed aggregate employment. Intuitively, it is hard to reconcile such an observation with the prediction of this class of models that flow-balance employment must overshoot its frictionless (let alone its observed) counterpart.

We formalize this intuition in three further empirical exercises. For all of them, we begin by selecting a parameterization of the adjustment frictions that replicates the sluggishness of observed aggregate employment. We find that a relatively large linear friction is needed to achieve this.

The first exercise then finds a sequence of aggregate shocks to match the empirical time series of observed aggregate employment in our data and compares the model-implied series for flow-balance employment with its analogue in the data. Consistent with the above intuition, the model-implied series for flow-balance employment is much more volatile than its empirical counterpart, exhibiting around 50 percent more peak-to-trough variation around recessions.

The second exercise provides a further illustration of this result by comparing the dynamic correlations between aggregate flow-balance employment and labor productivity in the model and the data. By construction, the parameterized model generates an impulse response of actual observed employment to labor productivity that resembles its sluggish, hump-shaped analogue in the data. However, while the empirical impulse response of flow-balance employment is only modestly less persistent and hump shaped than that for actual

employment, the model-implied response exhibits very volatile, jump dynamics with respect to labor productivity.

In a final exercise, we directly compare impulse responses of measures of the inflows to, and outflows from, three employment-size classes in both the model and the data. Qualitatively consistent with models that feature canonical frictions, positive innovations to output-per-worker in the data are associated with an increase in the share of firms adjusting to, rather than from, higher employment levels. But, in stark contrast to the predictions of such models, the empirical impulse responses of firm-size flows are sluggish, hump shaped, and an order of magnitude smaller than their model-implied counterparts. This finding confirms that the differences between model-implied and observed flow-balance employment can be traced to the fast-moving dynamics of the firm-size flows under canonical frictions.

The results of these exercises form the basis of our conclusion that canonical models provide a poor account of the propagation mechanism underlying observed employment persistence. In the concluding section of the paper, we speculate on potential resolutions of this failure. A particularly satisfying resolution would be one that acknowledges the prominent microeconomic observation of inaction in employment adjustment and explores its interactions with other frictions that can account for our observation of sluggishness in the flow of labor across firms. We suggest one example in which the costs of adjusting employment interact with information frictions, thereby building on and borrowing from applications of related ideas in the price setting literature, among others.[6] A distinctive feature of canonical labor market frictions is that they render employment decisions partially irreversible. Consequently, information frictions induce a natural signal extraction problem whereby firms adjust to aggregate disturbances to the extent that they are perceived to be permanent, and render desired employment flows sluggish, as we observe in establishment microdata.

1. Labor market frictions and firm-size dynamics

In this section, we first formalize the observation that canonical labor market frictions affect aggregate employment by impeding the flow of firms across different firm sizes. We then use the implied structure of these firm-size dynamics to motivate a summary statistic for their behavior, which enables us to characterize tractably key properties of canonical models. Another virtue of this measure, which we discuss in later sections, is that it can be measured directly from establishment microdata.

[6] Gorodnichenko (2010) and Alvarez, Lippi, and Paciello (2011) are two recent contributions to the literature that integrates menu costs of price adjustment and information frictions.

1.1 Fixed costs

A leading model of labor market frictions postulates the presence of a fixed cost of adjusting employment, independent of the scale of adjustment. The early work of Hamermesh (1989) suggested that such a friction could account for important features of establishment employment dynamics, an observation that informed the later influential empirical analyses of Caballero and Engel (1993) and Caballero, Engel, and Haltiwanger (1997).[7]

The case of a fixed cost is a natural starting place, not only in view of its prominence in the literature, but also because it provides a setting in which to convey our approach, and the intuition behind our results, most easily. As we shall see, the key insights will carry over to other canonical models. In what follows we first review the well-understood distortions of firms' labor demand policies induced by this friction. More important for our purposes, we use this to infer the implications for firm-size flows, and thereby aggregate employment.

With regard to the structure of labor demand, the key implication of a fixed cost is that employment will be adjusted only intermittently and, upon adjustment, discretely—adjustment will be "lumpy." Thus, labor demand takes the form of a threshold "*Ss*" policy, as illustrated in Figure 1A:

$$n = \begin{cases} n^* & \text{if} \quad n^* > U(n_{-1}), \\ n_{-1} & \text{if} \quad n^* \in [L(n_{-1}), U(n_{-1})], \\ n^* & \text{if} \quad n^* < L(n_{-1}). \end{cases} \tag{1}$$

Here n^* is the level of employment that a firm chooses if it adjusts. Under the *Ss* policy, a firm's current employment n is adjusted away from its past level n_{-1} whenever n^* deviates sufficiently from n_{-1}, as dictated by the triggers $L(n_{-1}) < n_{-1} < U(n_{-1})$.

Caballero, Engel, and Haltiwanger (1995, 1997) refer to n^* as *mandated* employment, interpreted as the level of employment the firm would choose if the friction were suspended for the current period. In principle, the latter is distinct from *frictionless* employment, which emerges if the fixed cost is suspended *indefinitely*. For reasonably calibrated models within this canonical class, however, the dynamics of mandated and frictionless employment are very similar.[8] Henceforth, then, we shall refer to n^* as frictionless, or desired, employment.

The dynamics of *aggregate* employment implied by the behavior of firms in equation (1) can be inferred from its implications for firm-size flows. Imagine that the economy enters the period with a density of past employment, $h_{-1}(\cdot)$, and that realizations of

[7] See also King and Thomas (2006), Cooper, Haltiwanger, and Willis (2007, 2015), and Bachmann (2012).

[8] This has been proved analytically for the case of a plausibly small fixed adjustment cost (Gertler and Leahy, 2008; Elsby and Michaels, 2017). In the Appendix (see Figure A), we also verify numerically that the distinction between frictionless and mandated employment is quantitatively inconsequential for the results we report below.

idiosyncratic and aggregate shocks induce a density of desired employment $h^*(\cdot)$. Our strategy is to infer a law of motion for the current-period density $h(\cdot)$ implied by equation (1). This in turn will imply a path for aggregate employment in the economy, which we denote by N, since the latter is captured by the mean of the density, $N \equiv \int mh(m)dm$.

The adjustment policy in Figure 1A suggests a straightforward approach to constructing a law of motion for the firm-size density $h(\cdot)$. Consider first the outflow of mass from some employment level m. Among the $h_{-1}(m)$ mass of firms that enter the period with m workers, only the fraction whose desired employment n^* lies outside the inaction region $[L(m), U(m)]$ will choose to incur the adjustment cost and leave the mass. Symmetrically, now consider the inflow of mass to employment level m. Among the $h^*(m)$ mass of firms whose desired employment is equal to m, only the fraction whose inherited employment n_{-1} lies outside the inverse inaction region $[U^{-1}(m), L^{-1}(m)]$ will choose to incur the adjustment cost and flow to m. Thus, the change in the mass at employment level m follows the law of motion

$$\Delta h(m) = \tau(m)h^*(m) - \phi(m)h_{-1}(m), \tag{2}$$

where $\tau(m)$ and $\phi(m)$ are, respectively, the probabilities of adjusting to and from an employment level m,

$$\begin{aligned} \tau(m) &= \Pr(n_{-1} \notin [U^{-1}(m), L^{-1}(m)] | n^* = m)\text{, and} \\ \phi(m) &= \Pr(n^* \notin [L(m), U(m)] | n_{-1} = m). \end{aligned} \tag{3}$$

Formal derivations of equations (2) and (3) are provided in the Appendix.

The role of frictions in shaping the evolution of aggregate employment is evident in equations (2) and (3). In the absence of frictions, the probabilities of adjusting to and from m are given by $\tau(m) = 1 = \phi(m)$. Hence, (2) collapses to $\Delta h(m) = h^*(m) - h_{-1}(m)$: Any gap between the initial and frictionless densities is closed immediately. Thus, frictions distort the path of the firm-size density, and thereby aggregate employment, by impeding the flows of labor across firms, in the sense that $\tau(m)$, $\phi(m) \in (0,1)$.

1.2 An empirical diagnostic

With this theoretical law of motion in hand, our next step is to consider which of its components can be *measured* empirically using available data. As we shall see, establishment-level panel data allow one to observe much of equation (2): One can measure the mass at each employment level at each point in time, $h_{-1}(m)$ and $h(m)$; one can also observe the fraction of establishments at each employment level that adjusts away, $\phi(m)$, as well as the total inflow, $\tau(m)h^*(m)$.[9]

[9] That we can observe only the *total* inflow, $\tau(m)h^*(m)$, rather than its constituent parts, is of course a perennial identification problem in this literature. If one could measure both $\tau(m)$ and $h^*(m)$, the latter would allow one to infer a measure of aggregate *frictionless* employment $N^* \equiv \int mh^*(m)dm$. Comparison of

Our point of departure is to note that, for fixed adjustment rates $\tau(m)$ and $\phi(m)$, the firm-size density will converge to a position where the inflow of mass to each m is balanced by outflows from that point. This *flow balance* condition implies a density

$$\hat{h}(m) \equiv \frac{\tau(m)}{\phi(m)} h^*(m). \tag{4}$$

$\hat{h}(m)$ is useful for several reasons. First, it can be measured straightforwardly, since it requires knowledge only of the total inflow, $\tau(m)h^*(m)$, and the probability of outflow $\phi(m)$, both of which are observed in establishment panel data.

Second, we argue in what follows that the mean of the flow-balance density offers a single summary statistic that conveys the effects of canonical frictions on the dynamics of firm-size flows, and thereby on the dynamics of aggregate employment. Specifically, note that, using (4), the aggregate employment level implied by flow balance, $\widehat{N} \equiv \int m\hat{h}(m)dm$, can be written as

$$\widehat{N} = N^* + cov_{h^*}\left(m, \frac{\tau(m)}{\phi(m)}\right), \tag{5}$$

where cov_{h^*} denotes a covariance taken with respect to the distribution of frictionless employment, $h^*(m)$.

Equation (5) reveals that aggregate employment under flow balance $\widehat{N}$ will *overshoot* the path of aggregate frictionless employment N^* under a monotonicity condition—namely, that firms on average are more likely to adjust to, versus from, high (low) employment levels following positive (negative) innovations to aggregate frictionless employment. This implies that, after a positive innovation, $\tau(m)/\phi(m)$ will decline for low m (since fewer firms adjust to, versus from, low m) and rise for high m (since more firms adjust to, versus from, high m). Thus, $\tau(m)/\phi(m)$ "tilts up" with respect to m, raising the covariance term in (5). Under this condition, $\widehat{N}$ will rise more than N^* when the latter rises and fall more than N^* when it falls.

The monotonicity condition that underlies this intuition is closely related to the *selection effect* that has been emphasized in the literature on adjustment frictions (Caballero and Engel, 2007; Golosov and Lucas, 2007). This refers to a property shared by state-dependent models of adjustment whereby the firms that adjust tend to be those with the greatest desired adjustment. By the same token, firms in these models also will adjust in the *direction* of the desired adjustment.

The forgoing intuition can be formalized tractably in standard models of fixed adjustment frictions, such as that set out in Caballero and Engel (1999). In this

N^* with the observed path of actual aggregate employment N would then indicate the wedge between these two induced by the adjustment friction.

environment, firms face an isoelastic production function $y = pxn^{\alpha}$ that is subject to idiosyncratic shocks x. Firms thus face the following decision problem

$$\Pi(n_{-1}, x) \equiv \max_{n}\{pxn^{\alpha} - wn - C^{+}\mathbb{I}[n > n_{-1}] - C^{-}\mathbb{I}[n < n_{-1}] + \beta\mathbb{E}[\Pi(n, x')|x]\}, \quad (6)$$

where p denotes (for now, fixed) aggregate productivity, w the wage, and $C^{+/-}$ the fixed costs of adjusting employment up and down.

Caballero and Engel (1999) show that, if idiosyncratic shocks follow a geometric random walk, $\ln x' = \ln x + \varepsilon'_x$, and the adjustment costs $C^{+/-}$ are scaled to be proportional to the firm's frictionless labor costs, the labor demand problem has a tractable homogeneity property. This has two useful implications: First, the adjustment triggers in (1) are linear and time invariant, $L(n_{-1}) = L \cdot n_{-1}$ and $U(n_{-1}) = U \cdot n_{-1}$ for constants $L < 1 < U$. Second, desired (log) employment adjustments, $\ln(n^*/n_{-1})$, are independent of initial firm size n_{-1}.[10]

Proposition 1 uses these properties of the canonical model to formalize the heuristic claim above that changes in aggregate employment under flow balance overshoot changes in aggregate frictionless employment. It assumes that firms perceive aggregate productivity p as fixed, and characterizes comparative statics with respect to a (one-time) change in p. Because of the model's loglinear structure, the result is most simply derived in terms of aggregate *log* frictionless employment, which we shall denote by $\mathcal{N}^*$, and its counterpart under flow balance, $\widehat{\mathcal{N}}$.

Proposition 1 *Consider the model of fixed adjustment costs (6). To a first-order approximation around a small change in aggregate log frictionless employment $\Delta\mathcal{N}^*$, the change in aggregate log employment under flow balance, relative to a prior constant-$\mathcal{N}^*$ steady state, is*

$$\Delta\widehat{\mathcal{N}} \approx \frac{1 - \epsilon_w}{1 - \epsilon_{w^*}} \cdot (1 + \psi) \cdot \Delta\mathcal{N}^*, \quad (7)$$

where $\psi > 0$, and ϵ_w and ϵ_{w^*} are the elasticities with and without frictions, respectively, of equilibrium wages to aggregate productivity p.

In Proposition 1, the response of $\widehat{\mathcal{N}}$ overshoots the frictionless response of $\mathcal{N}^*$ for two reasons. The first is a partial equilibrium response: Even if $\epsilon_w = \epsilon_{w^*} = 0$, Proposition 1 indicates that the change in aggregate log employment under flow balance strictly overshoots its frictionless counterpart This reflects the intuition conveyed by equation (5) that increases in desired employment $\mathcal{N}^*$ are augmented in $\widehat{\mathcal{N}}$ by increases in the propensity to adjust toward higher employment levels. Put another way, frictions induce

[10] The Appendix provides a formal statement and proof of this result in Lemma 1.

a "pent-up" demand for adjusting, such that the propensity to adjust reacts sharply after aggregate shocks and leads $\widehat{\mathcal{N}}$ to overshoot $\mathcal{N}^*$.

In addition, Proposition 1 reveals how differential equilibrium wage responses reinforce this overshooting property still further. To the extent that adjustment frictions restrict the response of labor demand to an aggregate shock, they also will restrict the response of equilibrium wages for a given labor supply schedule, $\epsilon_w < \epsilon_{w^*}$. It follows that $(1-\epsilon_w)/(1-\epsilon_{w^*}) > 1$, thereby further amplifying the equilibrium employment response under flow balance.

While Proposition 1 has a number of virtues—it holds irrespective of whether adjustment is symmetric ($C^+ = C^-$) or asymmetric ($C^+ \neq C^-$), for example—it also has limitations. It relies on the homogeneity of the canonical model implied by the assumption that idiosyncratic productivity, x, follows a random walk. It is also a comparative statics result, describing the response of the economy to a change in aggregate labor demand, indexed by p, that is expected to occur with zero probability from the firms' perspectives. For these reasons, in the next subsection, we explore the robustness of the overshooting result in numerical simulations that relax these assumptions.

1.3 Quantitative illustrations

We illustrate the dynamics of fixed costs models that resemble the canonical model described above, but with two differences. First, we relax the random walk assumption on idiosyncratic shocks, which we allow to follow a geometric first-order autoregression (AR(1)),

$$\ln x' = \rho_x \ln x + \varepsilon_x', \text{ where } \varepsilon_x' \sim N(0, \sigma_x^2). \tag{8}$$

Second, we allow for the presence of aggregate productivity shocks, and for their stochastic process to be known to firms in the model. The evolution of these aggregate shocks also is assumed to follow a geometric AR(1),

$$\ln p' = \rho_p \ln p + \varepsilon_p', \text{ where } \varepsilon_p' \sim N\left(0, \sigma_p^2\right). \tag{9}$$

To mirror the timing of the data we use later in the paper, a period is taken to be one quarter. Based on this, we set the discount factor β to 0.99, consistent with an annual interest rate of around 4 percent. To parameterize the remainder of the model, we appeal to the empirical literature that estimates closely related models of firm dynamics.

The returns to scale parameter α is set to 0.64, as in the estimates of Cooper, Haltiwanger, and Willis (2007, 2015).

The choice of parameters of the idiosyncratic productivity shock process (8) is informed by the estimates of Abraham and White (2006). They estimate a quarterly persistence parameter ρ_x of approximately 0.7, which we implement. Our choice of the standard deviation of the idiosyncratic innovation ε_x' of $\sigma_x = 0.15$ is set a little higher than Abraham and White's estimate of 0.10, since the latter lies at the lower end of the

range of estimates in the literature. The Appendix derives these quarterly parameters from Abraham and White's annual estimates and contrasts them with other estimates of ρ_x and σ_x reported in related literature. The results are very similar to those described in what follows (see Figure B in the Appendix).

The parameters of the process for aggregate technology in (9) are chosen so that aggregate frictionless employment in the model exhibits a persistence and volatility comparable with aggregate employment in U.S. data. This yields $\rho_p = 0.95$ and $\sigma_p = 0.018$. Although frictions augment persistence and dampen volatility, the intent is for the model environment to resemble broadly the U.S. labor market with respect to these *un*conditional moments. Importantly, the approach does not build in any persistence in employment *conditional* on technology.

Finally, with respect to the adjustment cost, here we report results for the case of symmetric frictions, $C^+ = C^-$, the most common choice in the literature (see, for example, Bloom 2009). We explore three parameterizations that successively raise the friction to replicate a range of inaction rates. In the data used later in the paper, the observed fraction of firms that do not adjust employment from quarter to quarter averages 52.5 percent. We find that a fixed cost equal to 1.3 percent of quarterly revenue replicates this inaction rate. However, for two reasons, we also consider fixed costs that induce higher inaction rates. First, the latter calibration lies at the lower end of available estimates of fixed costs (Bloom 2009; Cooper, Haltiwanger, and Willis, 2007, 2015). Second, consistent with this, inaction rates measured at a year-to-year frequency lie closer to 40 percent, much higher than implied by a naïve extrapolation of the quarterly inaction rate. A natural explanation for this fact is that some quarter-to-quarter shifts in employment reflect quits, which are subsequently replaced, rather than "active" employment adjustments that are subject to frictions and are the focus of canonical models. For these reasons, we also explore larger fixed costs that imply quarterly inaction rates of 67 percent and 80 percent. These correspond to adjustment costs of 2.7 percent and 5.8 percent of quarterly revenue, respectively, which also lie in the range of estimates in the literature.

We solve the labor demand problem via value function iteration on an integer-valued employment grid, $n \in \{1,2,3 \ldots\}$. The latter mirrors the integer constraint in the data, allowing one to construct the density $\hat{h}(\cdot)$ in the simulated data in the same way as we later implement in the real data.

To simulate equilibrium wage responses, we impose an aggregate labor supply schedule. Based on the estimates of Chetty (2012) and Chetty et al. (2012), we parameterize the labor supply function to have a (constant) Frisch elasticity of 0.5.[11] We maintain the same elasticity in the frictionless model. Chetty has argued that longer-run

[11] Using survey questions about the long-run response to hypothetical wealth windfalls, Kimball and Shapiro (2010) estimate a median Frisch elasticity of 0.6 and a mean of 1. Consistent with Proposition 1, we have verified that aggregate employment under flow balance overshoots its frictionless counterpart even in the latter parameterization. Results are available on request.

labor supply responses (e.g., Hicksian elasticities), which are arguably less influenced by frictions, imply a Frisch elasticity that is still no more than 0.5.

To solve the model, we implement the bounded rationality algorithm of Krusell and Smith (1998), whereby firms condition their labor demands on a linear forecast rule that relates the log aggregate employment to its lag and aggregate productivity. We then iterate on the coefficients of this forecast rule until the firms' simulated choices are consistent with the rule.

Figure 2 plots simulated impulse responses of aggregate employment N, together with its frictionless and flow-balance counterparts, N^* and $\widehat{N}$, respectively. The overshooting result anticipated in Proposition 1 is clearly visible in the model dynamics. For all three parameterizations of the adjustment cost, our proposed diagnostic, $\widehat{N}$, responds more aggressively to the aggregate shock than frictionless employment N^*. Moreover, the magnitude of the overshooting of $\widehat{N}$ relative to N^* is substantial in the model, responding on impact around twice as much to the impulse.

These results provide a first example of how canonical frictions have clear predictions on the dynamics of firm-size flows, as summarized by the dynamics of $\widehat{N}$—namely, that they respond aggressively to aggregate shocks. Since these firm-size flows reflect the channel through which frictions distort the path of aggregate employment, observable measures of such flows can be used to assess the empirical relevance of the propagation mechanism implied by canonical frictions. The next subsections extend this insight to two other popular models of labor market frictions.

1.4 Linear costs

Prominent alternative models of labor market frictions appeal instead to linear costs of adjustment in which the friction is discrete *at the margin* and rises with the scale of adjustment. This class encompasses models of per-worker hiring and firing costs, including the contributions of Oi (1962); Nickell (1978); Bentolila and Bertola (1990); Hopenhayn and Rogerson (1993); and Veracierto (2008). The case of linear adjustment costs is especially important to examine since, as we shall see, these costs can induce more realistic sluggishness in aggregate employment than fixed costs.

Relative to the fixed costs case, linear frictions alter the structure of both labor demand and firm-size dynamics. Although labor demand will continue to feature intermittent adjustment, a key difference is that, conditional on adjusting, firms will no longer discretely set employment to their frictionless target n^*. Rather, they will reduce the magnitude of hires and separations, shedding fewer workers when they shrink, and hiring fewer workers when they expand. Formally, the policy rule for separations, which we shall denote by $l(\cdot)$, will differ from the policy rule used for hiring, denoted by $u(\cdot)$, inducing the *continuous Ss* policy as illustrated in Figure 1B,

$$n = \begin{cases} u^{-1}(n^*) & \text{if} \quad n^* > u(n_{-1}), \\ n_{-1} & \text{if} \quad n^* \in [l(n_{-1}), u(n_{-1})], \\ l^{-1}(n^*) & \text{if} \quad n^* < l(n_{-1}), \end{cases} \tag{10}$$

where $l(n_{-1}) < n_{-1} < u(n_{-1})$ for all n_{-1}.

The key distinction, that the *direction* of adjustment must be taken into account in the presence of linear costs, also leaves its imprint on the law of motion for the firm-size distribution. As before, the labor demand policy in Figure 1B motivates the form of this law of motion. This reveals that the structure of outflows is qualitatively unchanged—of the $h_{-1}(m)$ density of firms currently at employment level m, only those with frictionless employment outside the inaction region $[l(m), u(m)]$ will adjust away. But inflows are now differentiated by the direction of adjustment. The inflow of mass adjusting down to m is composed of firms whose past employment n_{-1} is greater than m, and whose frictionless employment n^* is equal to $l(m) < m$. Likewise, the inflow of mass flowing up to m consists of firms with $n_{-1} < m$ and $n^* = u(m) > m$.

Piecing this logic together yields the following law of motion for the firm size density:

$$\Delta h(m) = \tau_l(m) h_l^*(m) + \tau_u(m) h_u^*(m) - \phi(m) h_{-1}(m). \tag{11}$$

Extending the interpretation of the fixed costs case above, here $h_l^*(m) = l'(m) h^*\big(l(m)\big)$ and $h_u^*(m) = u'(m) h^*\big(u(m)\big)$ are the densities of employment that would emerge if all firms adjusted, respectively, according to the separation rule, $l(m)$, and hiring rule, $u(m)$. However, only a fraction of firms will in fact adjust. The adjustment probabilities take the form

$$\begin{aligned} \tau_l(m) &= \Pr\big(n_{-1} > m | n^* = l(m)\big), \\ \tau_u(m) &= \Pr\big(n_{-1} < m | n^* = u(m)\big), \text{ and} \\ \phi(m) &= \Pr(n^* \notin [l(m), u(m)] | n_{-1} = m), \end{aligned} \tag{12}$$

where $\tau_l(m)$ is the probability that a firm adjusts *down to* m, while $\tau_u(m)$ is the probability that a firm adjusts *up to* m.

To construct the density under flow balance for the linear costs case, note that, for fixed adjustment rates $\tau_l(m)$, $\tau_u(m)$, and $\phi(m)$, the law of motion (11) implies that the firm-size density will converge to

$$\hat{h}(m) \equiv \frac{\tau_l(m)}{\phi(m)} h_l^*(m) + \frac{\tau_u(m)}{\phi(m)} h_u^*(m). \tag{13}$$

Like its counterpart (4) in the case of fixed costs, equation (13) offers a glimpse into the behavior of aggregate employment under flow balance, $\widehat{N} \equiv \int m \hat{h}(m) dm$. The flow-balance density $\hat{h}(m)$ is again related both to the propensities to adjust, $\tau_l(m)$, $\tau_u(m)$, and $\phi(m)$, and to the densities of "desired" employment conditional on adjusting, $h_l^*(m)$ and $h_u^*(m)$. As in the case of fixed costs, changes in the propensities to adjust following

an aggregate shock will tend to amplify the response of employment under flow balance relative to the frictionless benchmark that lacks intermittent adjustment. However, what is new is that the presence of a *linear* cost implies that, conditional on adjusting, employment responds less aggressively than the frictionless benchmark. Under certain conditions, we can characterize the relative strength of these two opposing forces.

Once again, further insight can be gained if we consider a canonical linear cost model in which firms face isoelastic production $y = pxn^{\alpha}$, and idiosyncratic shocks that follow a geometric random walk.[12] The key difference is that the adjustment friction is now scaled by the magnitude of adjustment, so that firms face the decision problem:[13]

$$\Pi(n_{-1}, x) \equiv \max_{n}\{pxn^{\alpha} - wn - c^{+}\Delta n^{+} + c^{-}\Delta n^{-} + \beta\mathbb{E}[\Pi(n, x')|x]\}. \tag{14}$$

A simple extension of Caballero and Engel's (1999) homogeneity results for the fixed cost model can be used to show that if idiosyncratic shocks follow a geometric random walk, and if per-worker hiring and firing costs are proportional to wages, the adjustment triggers in (10) are linear and time invariant, $l(n) = l \cdot n$ and $u(n) = u \cdot n$ for constants $l < 1 < u$, and that desired (log) employment adjustments, $\ln(n^{*}/n_{-1})$, are independent of initial firm size n_{-1}.[14]

As in Proposition 1 above for the fixed costs case, the latter properties allow one to relate the response of aggregate flow-balance log employment $\widehat{\mathcal{N}}$ to the response of aggregate frictionless log employment $\mathcal{N}^{*}$ following a change in aggregate productivity.

Proposition 2 *Consider the model of linear adjustment costs (14). To a first-order approximation around a small change in aggregate log frictionless employment $\Delta\mathcal{N}^{*}$, the change in aggregate log employment under flow balance, relative to a prior constant-$\mathcal{N}^{*}$ steady state, is*

$$\Delta\widehat{\mathcal{N}} \approx \frac{1 - \epsilon_{w}}{1 - \epsilon_{w^{*}}} \cdot \Delta\mathcal{N}^{*}, \tag{15}$$

where ϵ_{w} and $\epsilon_{w^{*}}$ are the elasticities with and without frictions, respectively, of equilibrium wages to aggregate productivity p.

Just as in the model of fixed costs, the response of $\widehat{\mathcal{N}}$ relative to $\mathcal{N}^{*}$ is shown to be mediated by the wage elasticities ϵ_{w} and $\epsilon_{w^{*}}$, and is qualitatively independent of any asymmetries in the frictions $c^{+} \neq c^{-}$. In contrast to the fixed costs case, though, the

[12] Nickell (1978, 1986) first formalized the linear cost model in the context of a labor demand model. Bentolila and Bertola (1990) introduced uncertainty into Nickell's continuous-time formulation. Equation (14) is a discrete-time analogue to Bentolila and Bertola's model (although the shocks need not be Gaussian, as in their paper).

[13] We use Δn^{+} and Δn^{-} as shorthand for $\Delta n\mathbb{I}[n > n_{-1}]$ and $\Delta n\mathbb{I}[n < n_{-1}]$, respectively.

[14] Again, the Appendix provides a formal statement and proof of this result in Lemma 1.

extent to which $\widehat{\mathcal{N}}$ overshoots the frictionless response of $\mathcal{N}^*$ now depends entirely on the response of equilibrium wages.

For fixed wages, the response of $\widehat{\mathcal{N}}$ no longer overshoots that of $\mathcal{N}^*$ but is approximately *equal* to it. The key difference is that firms adjust only *partially* toward their frictionless employment under linear frictions. A rise in $\mathcal{N}^*$ places more firms on the hiring margin, where employment is set below its frictionless counterpart, and fewer firms on the separation margin, where employment exceeds its frictionless level. Both forces serve to attenuate the response of $\widehat{\mathcal{N}}$ relative to the fixed costs case. Proposition 2 shows that, to a first order, this attenuation offsets exactly the partial equilibrium overshooting of the diagnostic $\widehat{\mathcal{N}}$ in the fixed costs case.

The effects of differential equilibrium wage responses remain as before, however. Sluggish frictional responses of labor demand to an aggregate shock will induce sluggish equilibrium wage responses under frictions, such that $\epsilon_w < \epsilon_{w^*}$. This again gives rise to overshooting, as shown in Proposition 2.

Figures 3 and 4 show that the result of Proposition 2 is mirrored in numerical simulations of models that incorporate a general stationary process for idiosyncratic productivity, x, and a fully stochastic process for aggregate productivity, p. We again present results for three parameterizations of the friction, each of which induces a different inaction rate. As with the fixed costs case above, the Appendix provides further results that vary the persistence and volatility of idiosyncratic shocks, ρ_x and σ_x. The results are again very similar to those described here (see Figure C in the Appendix). The numerical methods and the details of the calibration strategy are as described in section 1.3.

Figure 3 illustrates impulse responses of actual, frictionless, and flow-balance aggregate employment in the presence of *symmetric* linear frictions where $c^+ = c^-$. As before, each panel of Figure 3 successively raises the friction to produce increasingly higher average rates of inaction in employment adjustment. Note that the response of actual employment becomes progressively more sluggish as the friction rises, which dampens the response of the wage. As foreshadowed by Proposition 2, the response of flow-balance employment therefore increasingly overshoots the frictionless path.

Figure 4 in turn reveals that this result is unimpaired by the presence of *asymmetric* frictions, as suggested by Proposition 2. Its first three panels report results for successively higher hiring costs, $c^+ > 0$ and $c^- = 0$; the latter three panels do the same for firing costs, $c^- > 0$ and $c^+ = 0$. Strikingly, it is hard to discern differences between the impulse responses in the cases with a hiring (but no firing) cost and a firing (but no hiring) cost, and between these and the impulse response for the symmetric case in Figure 3.

The message of Figures 3 and 4, then, is that the insight of Proposition 2 is robust to empirically reasonable parameterizations of canonical models of linear frictions. This reinforces the message of section 1.3 that flow-balance employment is indeed a useful summary statistic for the impact of canonical frictions on firm-size dynamics, and thereby the effects of such frictions on aggregate employment dynamics.

However, Proposition 2 does not allow the adjustment triggers to vary, since these are independent of $\Delta\mathcal{N}^*$ under the time-invariant linear frictions we have considered thus far. This is a key distinction with respect to models of search frictions, to which we now turn.

1.5 Search costs

The canonical Diamond-Mortensen-Pissarides (DMP) model of search frictions, in which a single firm matches with a single worker, can be extended to a setting with "large" firms that operate a production technology with decreasing returns to scale (Acemoglu and Hawkins, 2014; Elsby and Michaels, 2013). The presence of search frictions implies two modifications to the canonical linear cost model studied above.

First, search frictions induce a *time-varying* per-worker hiring cost. Hiring is mediated through vacancies, each of which is subject to a flow cost c and is filled with a probability q that depends on the aggregate state of the labor market. Under a law of large numbers, the effective per-worker hiring cost is thus c/q, which varies over time with the vacancy-filling rate q. The typical firm's problem, therefore, takes the form:

$$\Pi(n_{-1}, x) \equiv \max_{n} \left\{ pxn^{\alpha} - w(n,x)n - \frac{c}{q}\Delta n^{+} + \beta\mathbb{E}[\Pi(n,x')|x] \right\}. \tag{16}$$

Second, search frictions induce *ex post* rents to employment relationships over which a firm and its workers may bargain. In an extension of the bilateral Nash sharing rule invoked in standard one-worker-one-firm search models, Elsby and Michaels (2013) show that a marginal surplus-sharing rule proposed by Stole and Zwiebel (1996) implies a wage equation of the form

$$w(n,x) = \eta \frac{px\alpha n^{\alpha-1}}{1-\eta(1-\alpha)} + (1-\eta)\omega. \tag{17}$$

Here $\eta \in [0,1]$ indexes worker bargaining power, and ω is the annuitized value of the workers' threat point. Bruegemann, Gautier, and Menzio (2015) show that the marginal surplus-sharing rule underlying (17) can be derived from an alternating-offers bargaining game between a firm and its many workers in which the strategic position of each worker in the firm is symmetric.

As before, we consider first a version of the search model with a tractable homogeneity property. Specifically, we study the case in which the friction, embodied in the vacancy cost, is proportional to the workers' outside option, $c \propto \gamma\omega$.[15] Under these assumptions, the Appendix shows that the homogeneity properties used for the models discussed in

[15] This can be motivated through the presence of a dual labor market in which recruitment is performed by workers hired in a competitive market, who are paid according to the annuitized value of unemployment ω.

previous subsections continue to hold, with one exception: Although the adjustment triggers remain linear, they no longer are invariant to shifts in aggregate productivity, for the simple reason that the friction, c/q, varies with the aggregate state.

Proposition 3 reveals that the result of Proposition 2 extends to search frictions, under a few restrictions.

Proposition 3 *Consider the model of search costs in (16) and (17). Assume (i) firms are patient, $\beta \approx 1$; (ii) frictions are small, $\gamma^2 \approx 0$; and (iii) the distribution of ε_x is symmetric. Then, to a first-order approximation around a small change in aggregate log frictionless employment $\Delta \mathcal{N}^*$, the change in aggregate log employment under flow balance, relative to a prior constant-$\mathcal{N}^*$ steady state, is*

$$\Delta \widehat{\mathcal{N}} \approx \frac{1 - \epsilon_\omega}{1 - \epsilon_{w^*}} \Delta \mathcal{N}^*, \tag{18}$$

where ϵ_ω and ϵ_{w^*} are the elasticities of ω and frictionless wages w^* to aggregate productivity p.

As in earlier results, Proposition 3 suggests that the responses of $\widehat{\mathcal{N}}$ and $\mathcal{N}^*$ are shaped by both partial equilibrium and equilibrium forces, which we consider in turn.

In partial equilibrium, Proposition 3 shows that the response of aggregate employment under flow balance $\widehat{\mathcal{N}}$ still approximates the response of aggregate log frictionless employment $\mathcal{N}^*$, but under a few additional restrictions. We argue in what follows that these restrictions are plausible.

The first two restrictions—that firms are patient, and that frictions are small—are quantitative. We address their plausibility by examining results from a numerical model that does not impose these restrictions. This model sets the discount factor β to match an annual interest rate of 4 percent and sets c to match evidence on recruitment costs. The numerical results will thus address the extent to which β is close enough to one, and the friction sufficiently small, for the insight of Proposition 3 to hold.

The third restriction concerns the symmetry of the distribution of idiosyncratic shocks. This can be justified along two grounds. First, it is conventional to implement shock processes with symmetrically distributed—typically normal—innovations. Second, it is also consistent with the observed pattern of employment adjustment, which is close to symmetric (see Davis and Haltiwanger, 1992; Elsby and Michaels, 2013; among others).

These three restrictions aid the proof of Proposition 3, which is based on symmetry. If the firm is sufficiently patient ($\beta \approx 1$), the cost of hiring in the current period implies an equal cost of firing in the subsequent period. As a result, one can show that the optimal policy is symmetric, to a first-order approximation around $\gamma = 0$, as long as the driving force ε_x is symmetric. In terms of the notation of the policy rules, this means the upper and lower adjustment triggers, $u(n_{-1}) = u \cdot n_{-1}$ and $l(n_{-1}) = l \cdot n_{-1}$, satisfy $\ln u \approx -\ln l$,

and move by approximately the same amount in response to a shift in aggregate productivity.

As in preceding sections, we explore the robustness of the conclusion of our theoretical analysis by solving a numerical version of the model that relaxes the restrictions used in deriving the proposition. The numerical model extends (16) slightly by including a per-worker cost of hiring k (akin to c^+ in (14)) that is *independent* of the aggregate state of the labor market. Numerous authors have noted that a time-invariant cost of hiring aids the ability of search and matching models to generate realistic degrees of amplitude and persistence in employment (Mortensen and Nagypal, 2007; Pissarides, 2009; and Moscarini and Postel-Vinay, 2016).

We again present results for three parameterizations, each one targeting a different inaction rate. Details of our calibration strategy, as well as values of all structural parameters, can be found in the Appendix. Here, we describe the more salient structural parameters that underlie the elasticities, ϵ_{w^*} and ϵ_{ω}, highlighted by Proposition 3. These elasticities measure, respectively, the flexibility of frictionless wages w^*, and the workers' outside option in the presence of frictions ω, to aggregate productivity p.

As before, in the frictionless case, ϵ_{w^*} is related to the Frisch elasticity of labor supply, which we again set to 0.5. This implies $\epsilon_{w^*} = 2/(3-\alpha) \approx 0.848$ when α is set to equal 0.64.[16]

The counterpart to ϵ_{w^*} in the search model, ϵ_{ω}, depends on the structure of the workers' threat point ω, which in turn is shaped by the hiring costs faced by firms. These include c, the vacancy cost, as well as k. The vacancy cost is set such that the average cost of recruiting, c/q, equals 14 percent of the quarterly wage, following Hall and Milgrom (2008) and Elsby and Michaels (2013). We then select the value of k to match the three inaction rates studied in the preceding sections.

Given this structure, a simple extension of the "large-firm" wage bargain implemented in Elsby and Michaels (2013) to this environment implies that

$$\omega = \frac{\eta}{1-\eta}\big(c\theta + kf(\theta)\big) + b, \tag{19}$$

where θ is labor market tightness, the ratio of aggregate vacancies to unemployment, $f(\theta)$ is the job-finding rate, and b is the flow payoff to unemployment. Intuitively, since firms would have to pay both vacancy and hiring costs to replace a worker, both frictions act as a lever to raise his wage, and so both c and k enter into ω.

It remains to choose worker bargaining power, η. We pin this down based on evidence from microdata on wages. Taking account of the shifting composition of employment over

[16] Strictly speaking, labor supply is inelastic in the canonical search model, and so the elasticity that would emerge absent frictions is zero. In principle, though, it is possible to compare the behavior of flow-balance employment to the dynamics of any frictionless model. Accordingly, we benchmark against a more compelling frictionless alternative that uses a Frisch elasticity of 0.5.

the business cycle, microdata-based estimates are broadly consistent with a rule of thumb that real wages are about as cyclical as employment (Solon, Barsky, and Parker, 1994; Elsby, Shin, and Solon, 2016). Accordingly, we set η to match an elasticity of average real wages with respect to aggregate employment approximately equal to one. This choice, in turn, implies the elasticity of the workers' threat point to aggregate productivity, ϵ_ω.

The implied magnitudes for ϵ_ω are measured by the response of ω *on impact* of a shock to p, consistent with the interpretation of Proposition 3. The results vary somewhat across the different parameterizations of the search friction. We find that ϵ_ω lies between 0.6 (when the frictions are set to induce an inaction rate of 52.5 percent per quarter) and 0.35 (when the frictions induce an inaction rate of 80 percent per quarter).

Proposition 3 implies that the response of aggregate employment under flow balance should overshoot that of frictionless employment under these parameterizations, since $(1-\epsilon_\omega)/(1-\epsilon_{w^*})$ lies between 2.7 (in the case of a 52.5 percent inaction rate) and 4.3 (in the case of an 80 percent inaction rate). Figure 5 shows that this prediction of Proposition 3 is visible in numerical simulations of the model. As before, these are based on the methods and baseline parameterization described in section 1.3—that is, with stationary idiosyncratic shocks x and fully stochastic aggregate shocks p. The impulse responses in Figure 5 suggest that aggregate employment under flow balance reacts on impact of the aggregate shock considerably more than does its frictionless counterpart.

2. Empirical implementation

The previous section gives a theoretical rationale for how the aggregate effects of a class of canonical frictions are mediated through their effects on the dynamics of firm-size flows, and how a summary statistic for these dynamics is provided by aggregate flow-balance employment $\widehat{N}$. A key virtue of $\widehat{N}$ is that it can be measured with access to establishment panel data on employment. In this section, we apply these results to a rich source of microdata from the United States.

2.1 Data

The data we use are taken from the Quarterly Census of Employment and Wages (QCEW), which is compiled by the Bureau of Labor Statistics (BLS) in concert with State Employment Security Agencies. The latter collect data from all employers that are subject to their state's Unemployment Insurance (UI) laws. Firms file with the state agency quarterly UI Contribution Reports, which provide payroll counts of employment in each month; the BLS further disaggregates to the establishment level as necessary for multi-establishment firms. These are then aggregated by the BLS, which defines employment as the total number of workers on the establishment's payroll during the pay

period that includes the 12th day of each month. Following BLS procedure, we define quarterly employment as the level of employment in the third month of each quarter.[17]

From the cross-sectional QCEW data, the BLS constructs the Longitudinal Database of Establishments (LDE), which we use in what follows. Although data are available for the period 1990Q1 to 2014Q2, we restrict attention to data from 1992Q1 due to difficulty in matching establishments in the first two years of the sample.[18]

Sample restrictions. The QCEW data are a near-complete census of workers in the United States, covering approximately 98 percent of employees on nonfarm payrolls. The dotted line in Figure 6 plots the time series of log aggregate employment in private establishments in the full QCEW sample. Relative to this full sample, we apply three further sample restrictions, illustrated by the successive lines in Figure 6.

First, our access to QCEW/LDE microdata is restricted to a subset of forty states that approved access onsite at the BLS for this project. As a result, our sample excludes data for Florida, Illinois, Massachusetts, Michigan, Mississippi, New Hampshire, New York, Oregon, Pennsylvania, Wisconsin, and Wyoming.

Second, we restrict our sample to continuing establishments with positive employment in consecutive quarters. Specifically, we construct a set of overlapping quarter-to-quarter balanced panels that exclude births and deaths of establishments *within* the quarter. Note that we do not balance *across* quarters, so births in a given panel will appear as incumbents in the subsequent panel (if they survive). We focus on continuing establishments because the canonical models of adjustment frictions analyzed above are intended to describe adjustment patterns among incumbent firms.[19]

Our final sample restriction is to exclude establishments with more than 1000 employees in consecutive quarters. We do this for practical reasons. To measure the flow-balance employment distribution in equations (4) and (13), and hence the diagnostic suggested by the theory, we require measures of establishment flows between points in the firm-size distribution—specifically, inflows of mass to each employment level, and the probability of outflow. To measure the latter with sufficient precision requires sufficient sample sizes at all points in the distribution. Since establishments with more than 1000 employees comprise a very small fraction of U.S. establishments—for example, less than 0.1 percent in 2014Q2—sample sizes become impracticably thin beyond 1000 employees, inducing substantial noise in implied estimates of our diagnostic.

[17] The count of workers includes all those receiving any pay during the pay period, including part-time workers and those on paid leave.

[18] Although the underlying microdata are available from 1990 on, the BLS does not publish data based on longitudinally matched data for 1990-1991 due to changes in administrative procedures for how firms reported their data over that period.

[19] In constructing our sample of continuers, we also exclude the small subset of establishments that are flagged as undergoing a potential change of ownership, since their employment adjustment may be subject to measurement error. Those establishments, which the BLS attempts to link with their predecessor or successor, constitute only 0.1 percent of our total sample in 2014Q2.

Though the foregoing sample restrictions reduce the level of employment relative to the U.S. total, fluctuations in employment in our sample closely mimic the behavior of the published aggregate. Figure 6 reveals that, in terms of levels, the largest loss of sample size occurs because we are unable to access data for all states, accounting for around 30 percent of total employment in the United States. The further exclusion of noncontinuing establishments and large establishments accounts for around 2 percent and 10 percent of employment, respectively. However, Figure 6 shows that the path of aggregate employment in our sample resembles, in both trend and cycle, the path of aggregate employment in the full QCEW sample. The correlation between log aggregate employment in the published QCEW series for all states and that in our final microdata sample is 0.99.

Measurement. To estimate our diagnostic, we require first an estimate of the distribution of employment under flow balance, $\hat{h}(m)$. Substituting equations (4) and (13), respectively, into the laws of motion (2) and (11), we can write the density under flow balance as

$$\hat{h}_t(m) = h_{t-1}(m) + \frac{\Delta h_t(m)}{\phi_t(m)}, \tag{20}$$

where t indexes quarters, $h_{t-1}(m)$ is the previous quarter's mass of establishments with employment m, $\Delta h_t(m) \equiv h_t(m) - h_{t-1}(m)$ is the quarterly change in that mass, and $\phi_t(m)$ is the fraction of establishments that adjusts away from an employment level of m in quarter t. Thus, estimation of $\hat{h}_t(m)$ requires only an estimate of the outflow adjustment probability $\phi_t(m)$, in addition to measures of the evolution of the firm-size distribution $h_t(m)$.

The simplest approach to measuring $\phi_t(m)$ is to use our microdata to compute the fraction of establishments with m workers in quarter t that reports employment different from m in quarter $t+1$. As alluded to above in motivating our sample restrictions, however, a practical issue that arises is that sample sizes become small as m gets large, inducing sampling variation in estimates of $\phi_t(m)$.

We further address this issue by discretizing the employment distribution at large m. An advantage of the substantial sample sizes in the QCEW/LDE microdata is that we can be relatively conservative in this regard. In particular, we allow individual bins for each integer employment level up to 250 workers. In excess of 99 percent of the establishments lie in this range; therefore, sample sizes in each bin are large, between about 100 and 1.3 million establishments. For establishment sizes of 250 through 500 workers, we use bins of length five, allowing us to maintain sample sizes of at least 80 establishments in each quarter. Further up the distribution, of course, sample sizes get smaller, so we extend our bin length to ten for employment levels between 500 and 999 workers. In this range, sample sizes are at least 15 establishments in each quarter.

Denoting an individual bin by b, we estimate the firm-size mass and the outflow probability as

$$h_t(b) = \sum_i \mathbb{I}[n_{it} \in b], \text{ and } \phi_t(b) = \frac{\sum_i \mathbb{I}[n_{it} \notin b | n_{it-1} \in b]}{\sum_i \mathbb{I}[n_{it-1} \in b]}, \tag{21}$$

where i indexes establishments. We use these measures to compute the flow-balance mass in each bin according to equation (20) as $\hat{h}_t(b) = h_{t-1}(b) + [\Delta h_t(b)/\phi_t(b)]$. Finally, we compute aggregate employment and its flow-balance counterpart by taking the inner product of h_t and $\hat{h}_t$ with the midpoints of each bin, denoted m_b,

$$N_t = \sum_b m_b h_t(b), \text{ and } \widehat{N}_t = \sum_b m_b \hat{h}_t(b). \tag{22}$$

2.2 Inferring the aggregate effects of frictions

With this estimate of flow-balance aggregate employment $\widehat{N}_t$ in hand, we can now contrast its dynamics with the predictions of the canonical models summarized in section 1, and in Figures 2 through 5.

A first look at the data. Figure 7 plots the time series of N_t and $\widehat{N}_t$ derived from application of equation (22) to the QCEW/LDE microdata. Both series are expressed in log deviations from a quadratic trend.[20] Figure 7 reveals that $\widehat{N}_t$ is a leading indicator of actual employment N_t and is also more volatile. Specifically, the standard deviation of N_t is 0.025, whereas the standard deviation of $\widehat{N}_t$ is 0.031.

On the whole, however, the differences between the two series are modest. The median (mean) absolute difference between the series is just 0.5 (0.8) log points. Indeed, there is remarkably little daylight between the two series from 1992 to 2008. Even in the 2001 recession, flow-balance employment very closely tracks the drop in actual employment. The only substantial difference between the series emerges in the Great Recession. For instance, in the five quarters that bracket the trough of the recession, 2008Q4 to 2009Q4, the mean difference between the series is about 3 log points. However, this difference is short-lived. Since 2010, the two series have moved in tandem: Employment has increased 11.6 log points, whereas flow-balance employment has increased 11.9 log points.

By contrast, recall from the theoretical results in section 1 that canonical models share the prediction that flow-balance employment jumps aggressively in response to aggregate shocks. Together these observations give a first suggestion that the propagation

[20] Throughout our empirical analysis, we use quadratic time trends, rather than an HP filter, as the latter is well known to suffer from end point problems, and the end of our sample is dominated by recovery from the most recent recession. The aggregate time series, as well as the impulse responses we show later, are nonetheless qualitatively similar when an HP filter is applied to the data instead.

mechanism embodied in canonical models fails to capture the source of sluggishness in empirical employment dynamics.

Time series matching. To contrast the data with the models' predictions more precisely, we undertake a simulation exercise devised by King and Rebelo (1999) and Bachmann (2012). They show that it is possible to find a sequence of aggregate shocks that generates a path for aggregate model-generated outcomes—in our case employment—that matches an empirical analogue. In what follows, we use this technique to contrast the time series of flow-balance employment in the model and the data when the path of aggregate employment in each is constructed to be the same.

The procedure relies on the ability to summarize the dynamics of aggregate employment implied by the model using a simple aggregate law of motion. In a related adjustment cost model, Bachmann shows that an AR(1) specification that relates log aggregate employment to its own lag and current labor productivity does an excellent job of summarizing these dynamics. We find that the same property holds for our model.

Figures 2 through 5 suggest that linear cost models are especially capable of generating persistence in actual aggregate employment. We therefore initiate an algorithm with a variant of the (symmetric) linear cost model that is calibrated to replicate the amplitude and persistence of the empirical dynamics of actual employment.[21] We find that a model with fixed wages and a linear cost that generates a quarterly inaction rate of 86 percent achieves this goal. Note that this procedure is being generous to the model by enabling it to match observed employment at the expense of violating the inaction rate and the flexibility of real wages observed in the data. Further, recalling Proposition 2, by suppressing movements in the real wage, we are dampening the volatility of flow-balance employment implied by the model. Accordingly, we shall see that we obtain a lower bound on the discrepancy between the model and the data.

In a first step, we use this model to generate 85 quarters of simulated data (the same time span as in the data). We then estimate via OLS the following AR(1) process that relates model-generated log aggregate employment to its lag and current aggregate productivity p_t,

$$\ln N_t = \hat{\nu}_0 + \hat{\nu}_1 \ln N_{t-1} + \hat{\nu}_2 \ln p_t. \tag{23}$$

[21] Specifically, we choose the flexibility of wages and the linear cost to minimize the (sum of squares) distance between the impulse responses of observed employment N and those implied by an equivalent specification run on model-generated data (the latter results are presented later in Figure 10A). Since this step measures the response of employment to a *given* increase in productivity, it can be implemented before we back out the realized sequence of aggregate shocks. We do not pursue the effects of asymmetries in adjustment costs here because the results of sections 1.3 and 1.4 suggest that any such asymmetries affect neither the dynamics of aggregate employment, nor its flow-balance counterpart. We do not use the search model, since its implications mirror those of the model we simulate (see Figures 3 to 5), but come at the expense of greater computational burden (due to the additional fixed-point problem over market tightness).

With estimates of equation (23) in hand, we check whether the law of motion matches the *empirical* path of aggregate employment by substituting the latter time series into (23) and solving for the implied series of productivity. If the resultant sequence $\{p_t\}$ is consistent with the assumed data-generating process, we stop. Otherwise, in a second step, we re-parameterize the productivity process and re-initialize the model with this updated process. These steps are repeated until the moments of the productivity series implied by (23) are consistent with the parameterization assumed. In practice, the AR(1) specification in (23) fits the data closely (the R-squared of the regression is 0.9985), and so the algorithm converges quite quickly, after just a few iterations.[22]

Figures 8 and 9 illustrate the results. To smooth out high frequency noise, we apply the above algorithm to match a four-quarter moving average of log aggregate employment in the data. The standard deviation of the resulting time series for actual employment $\ln N$, shown in Figure 8, is 0.023. The model yields a notably more variable path for aggregate flow-balance employment, $\widehat{N}$. The model-implied standard deviation of $\ln \widehat{N}$ is 0.038, 36 percent larger than its empirical counterpart of 0.028.

The deviations between model-implied and observed flow-balance employment are thrown into even starker relief in and around recessions, as shown in Figure 9. When the model-implied series is near its nadir, it lies 5-6 log points below its empirical counterpart. Aggregate flow-balance employment also recovers significantly quicker in the wake of these downturns. In the eight quarters after the Great Recession, for instance, the model's flow-balance employment rises 12 log points. Its empirical counterpart increases by half that amount over the same period.

Measuring persistence. A final way of visualizing the difference between the data and the models' predictions is to contrast the response of flow-balance employment to estimated shifts in the aggregate driving force. Rather than attempting to use the data to identify structural shocks, which is prone to controversy, we instead undertake a descriptive analysis of the dynamic properties of aggregate employment. A commonly used gauge for the latter is a comparison of the dynamics of employment relative to output-per-worker. In what follows, we interpret unforecastable movements in output-per-worker as being indicative of innovations to the (latent) driving force, and estimate the reaction of flow-balance employment, in the model and the data, to these forecast errors. This serves as a simple way of summarizing the persistence of flow-balance employment.

Formally, we proceed as follows. Denote log output-per-worker by y_t. In a first stage, we estimate innovations in y_t that are unforecastable conditional on lags of y, and lags of log aggregate employment $\ln N$. Specifically, we use quarterly data on output-per-worker

[22] The implied process for output-per-worker in the model generated data shares roughly the same statistical properties as a similarly smoothed output-per-worker series taken from the Bureau of Labor Statistics Productivity and Costs data. An estimated AR(1) through model-implied output-per-worker data gives a persistence parameter of about 0.94 and a standard deviation of residuals of about 0.004, comparable with estimates from the data.

in the nonfarm business sector from the BLS Productivity and Costs release and our measure of actual employment from the QCEW to estimate the following AR(L) specification:

$$y_t = \alpha^y + \sum_{s=1}^{L} \beta_s^y y_{t-s} + \sum_{s=1}^{L} \gamma_s^y \ln N_{t-s} + \delta_1^y t + \delta_2^y t^2 + \varepsilon_t^y. \tag{24}$$

Within the context of the models considered in section 1, lags of output-per-worker y can be interpreted as proxies for lags of aggregate technology p, conditional on lags of N, as in (24). More broadly, they can be viewed as proxies for past realizations of business cycle driving forces. Note that secular trends are captured using a quadratic time trend.

The estimated residuals from this first-stage regression, $\hat{\varepsilon}_t^y$, are then used as the innovations to output-per-worker from which we derive impulse responses of actual and flow-balance employment in a second stage,

$$\begin{aligned} \ln N_t &= \alpha^N + \sum_{s=0}^{L-1} \beta_s^N \hat{\varepsilon}_{t-s}^y + \sum_{s=1}^{L} \gamma_s^N \ln N_{t-s} + \delta_1^N t + \delta_2^N t^2 + \varepsilon_t^N, \text{ and} \\ \ln \widehat{N}_t &= \alpha^{\widehat{N}} + \sum_{s=0}^{L-1} \beta_s^{\widehat{N}} \hat{\varepsilon}_{t-s}^y + \sum_{s=1}^{L} \gamma_s^{\widehat{N}} \ln \widehat{N}_{t-s} + \delta_1^{\widehat{N}} t + \delta_2^{\widehat{N}} t^2 + \varepsilon_t^{\widehat{N}}. \end{aligned} \tag{25}$$

Note that the timing in the lag structure of innovations to output-per-worker permits a *contemporaneous* relationship between these innovations and employment, as suggested by the model-based impulse responses described in section 1.

The estimates from the regressions in equations (24) and (25) allow us to trace out the dynamic relationship between each measure of log aggregate employment and a one-log-point innovation in output-per-worker. In practice, we use a lag order of $L = 4$ in both stages, (24) and (25).[23] Given the availability of our QCEW data, we estimate these regressions over the period, 1992Q2 to 2014Q2.

Panel A of Figure 10 plots the results. The dynamic response of aggregate employment takes a familiar shape, rising slowly after the innovation with a peak response of around 1 log point after five quarters. These hump-shaped dynamics mirror similar results found using different methods elsewhere in the literature (Blanchard and Diamond, 1989; Fujita and Ramey, 2007; Hagedorn and Manovskii, 2011). This is one representation of the persistence of aggregate employment.

As suggested by the time series in Figure 7, the dynamics of the flow-balance diagnostic $\widehat{N}$ share many of these properties. Although its peak response occurs earlier—after three quarters—reinforcing the impression of Figure 7 that $\widehat{N}$ is a leading indicator of the path of N, it exhibits a similar volatility and a clear hump shape.

[23] Experiments with different lag orders suggest that, although the peak of the hump-shaped impulse responses varies slightly across different lag lengths, Figure 10 is representative of results across these specifications.

To contrast the empirical dynamics illustrated in Figure 10A with those implied by canonical models of frictions, we rerun the regressions in equations (24) and (25) using model-generated data. Following our preceding discussion, we use the model with symmetric linear costs, chosen to minimize the distance between the empirical dynamics of actual employment N in Figure 10A and those implied by the model.

Panel B of Figure 10 reveals that this parameterization of the model is able to generate a dynamic relationship between *actual* employment and output-per-worker that is comparable with the data. Although the model overstates the impact response, the amplitude and persistence of employment are similar to their empirical counterparts.

A key result of Figure 10B, however, is that the model-implied dynamics of flow-balance employment are profoundly different from those seen in the data. Confirming the impression of the theoretical impulse responses in Figure 3, $\widehat{N}$ *jumps* in response to innovations in output-per-worker in the model, with an initial response *five times* larger than that of actual employment N. In marked contrast, the empirical dynamics of $\widehat{N}$ in Figure 10A are much more sluggish, bearing a closer resemblance to the empirical path of actual employment than its model-implied counterpart.

The substantial discrepancy between the implied and observed dynamics of flow-balance employment is an important failure of canonical models of frictions, in the sense that the models do not capture a key aspect of how shocks are propagated through the labor market.

2.3 Understanding the failure of canonical models

To examine the origins of this failure of canonical models, recall that the link between our diagnostic flow-balance employment $\widehat{N}$ and frictionless employment N^* is mediated through the behavior of firm-size flows—the τs and ϕs of equations (4) and (13)—and that canonical frictions have strong predictions regarding the dynamics of these flows by establishment size.

As we have emphasized, a key benefit of the data is that we can measure aspects of these flows using the longitudinal dimension of the QCEW microdata—specifically, the total inflow to, and the probability of outflow from, each employment level. Our next exercise, therefore, is to contrast the dynamics of the firm-size distribution in the data to those implied by canonical models of frictions.

To do this, we first split establishments in the data into three size classes. We choose these to correspond to the lower quartile (fewer than 15 employees), interquartile range (16 to 170 employees), and upper quartile (171 or more employees) of establishment sizes. We then estimate descriptive impulse responses that mirror equations (24) and (25) for the total inflow to, and probability of outflow from, each size class.[24] As in our previous

[24] To aggregate within a quartile range, we take a weighted average across establishment sizes, where the weight is the size's share of all establishments in the range.

analysis of the dynamics of aggregate employment, we repeat these same steps using data simulated from the model underlying Figure 10B that is calibrated to match as closely as possible the empirical dynamics of aggregate employment.

Panels C through F of Figure 10 illustrate the results of this exercise. The empirical and model-implied dynamics share a qualitative property, namely, that positive aggregate shocks render small (large) establishments more (less) likely to adjust away from their current employment, and induce fewer (more) establishments to adjust to low (high) employment levels.

Aside from this broad qualitative similarity, the quantitative dynamics reveal striking contrasts. The empirical behavior of firm-size flows exhibits an inertia not only in the sense that their *levels* are retarded relative to a frictionless environment, but also in the sluggishness of their *responses* to aggregate disturbances.

We highlight three manifestations of this general observation. First, note that the empirical responses of the firm-size flows in Figures 10C and 10E are an order of magnitude smaller than their theoretical counterparts in Figures 10D and 10F. Second, the dynamics of the flows in the data are much more sluggish than implied by canonical frictions. Firm-size dynamics in the model respond aggressively on impact of the aggregate shock. In the data, the response is mild and delayed. Third, the empirical dynamics reveal an establishment size gradient in the magnitude of the response of firm-size flows: Flows to and from smaller establishments respond less than their counterparts for larger establishments.

The upshot of this exercise is that canonical models of labor market frictions do a poor job of capturing the empirical dynamics of the firm-size distribution. Since the latter is the key channel through which canonical frictions are supposed to impede aggregate employment dynamics, this is an important limitation of this class of model.

3. Summary and discussion

In this paper, we have explored the propagation mechanism embodied in a canonical class of labor market frictions. In postulating several forms of nonconvex adjustment frictions, this class has the virtue of being able to reproduce the conspicuous degree of inaction observed in establishment employment dynamics. We further show that (some of) these labor market frictions are in turn able to generate at least part of the observed sluggishness in aggregate employment dynamics.

However, canonical frictions have strong implications for the source of this propagation, for which we do not find empirical support. In this class of models, deviations of aggregate employment from its frictionless path arise because frictions retard the flow of labor across firms. But since the latter induces pent-up demand for adjustment, these firm-size flows are predicted to respond rapidly to aggregate shocks. We use this to motivate a summary statistic for these flows, which we have labeled flow-balance employment that can be measured with access to establishment panel data.

We find that empirical measures of flow-balance employment display only mild departures from the path of actual employment, exhibit much more sluggish dynamics than implied by canonical frictions, and that the source of this tension can be traced to a failure of canonical models to capture the empirical persistence of firm-size flows.

We highlight two possible conclusions in light of these findings. The first is that labor market frictions that induce inaction are indeed unimportant for aggregate employment dynamics. This suggests a return to an older literature on convex adjustment costs (as in, for example, Sargent (1978) and Shapiro (1986)). The latter can induce sluggishness in firms' choice of employment *conditional* on adjustment and may thus be able to attenuate the elasticity of the firm-size flows.

A drawback of such a conclusion, however, is that the presence of inaction is perhaps the most prominent stylized fact of microeconomic employment adjustment. In acknowledgment of this fact, a second, alternative conclusion is that future work should explore the possibility that such inaction might *interact* with other frictions to induce the observed sluggishness in firm-size flows.

We provide one example of this possibility, based on an interaction of labor market and information frictions. Intuitively, if firms do not have full information on aggregate disturbances, they may attenuate their hiring and firing, dampening the response of flows of labor across firms.

To illustrate, suppose aggregate productivity is the sum of transitory and permanent components. Firms observe aggregate productivity but not its constituent parts.[25] In the absence of labor market frictions, firms' labor demand is the outcome of a simple static optimization problem, for which only knowledge of aggregate productivity is required. Thus, absent labor frictions, the information friction has no bite.

In the presence of employment adjustment frictions, however, firms must forecast the path of the aggregate state, which requires a judgment of the degree to which an aggregate disturbance is permanent. Hiring and firing decisions are thus based on perceptions of the persistent component of productivity. Standard signal extraction arguments will imply that such perceptions are a slow-moving state variable. Accordingly, hiring and firing decisions respond less aggressively to aggregate shocks on impact. This can lead, qualitatively, to the drawn-out dynamics of the labor market flows we observe in the data. Critically, this persistence in hiring and firing policies will in turn contribute to persistent aggregate employment dynamics.

The quantitative success of such a model will hinge on the rate at which firms update their assessments of the persistent component of aggregate productivity as well as the extent to which such persistence can be reconciled with the large cyclical volatility of employment. Nonetheless, we suspect that an interaction of labor market frictions with a

[25] For early applications of this information structure in macroeconomics, see Brunner, Cukierman, and Meltzer (1980) and Gertler (1982). More recently, see Erceg and Levin (2003).

notion of imperfect information provides a promising avenue of further research that seeks to understand aggregate employment persistence.[26]

References

Abraham, Arpad, and Kirk White. 2006. "The Dynamics of Plant-Level Productivity in U.S. Manufacturing." Center for Economic Studies WP 06-20.

Acemoglu, Daron, and William Hawkins. 2014. "Search with Multi-worker Firms." *Theoretical Economics* 9(3): 583-628.

Alvarez, Fernando, Francesco Lippi, and Luigi Paciello. 2011. "Optimal Price Setting with Observation and Menu Costs." *Quarterly Journal of Economics* 126(4): 1909-1960.

Bachmann, Ruediger. 2012. "Understanding Jobless Recoveries After 1991 and 2001." Mimeo, University of Notre Dame.

Bentolila, Samuel, and Giuseppe Bertola. 1990. "Firing Costs and Labour Demand: How Bad Is Eurosclerosis?" *Review of Economic Studies* 57(3): 381-402.

Bertola, Giuseppe, and Ricardo J. Caballero. 1990. "Kinked Adjustment Costs and Aggregate Dynamics." *NBER Macroeconomics Annual 1990* Cambridge, MA: MIT Press.

Bils, Mark, Peter J. Klenow and Benjamin A. Malin. 2012. "Reset Price Inflation and the Impact of Monetary Policy Shocks." *American Economic Review* 102(6): 2798-2825.

Blanchard, Olivier Jean, and Peter Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity* 1: 1–60.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77(3): 623-685.

Bruegemann, Bjoern, Pieter Gautier and Guido Menzio. 2015. "Intra Firm Bargaining and Shapley Values." Mimeo, Vrije Universiteit Amsterdam.

Brunner, Karl, Alex Cukierman, and Allan Meltzer. 1980. "Stagflation, Persistent Unemployment, and the Permanence of Economic Shocks." *Journal of Monetary Economics* 6: 467-492.

Caballero, Ricardo, and Eduardo M. R. A. Engel. 1993. "Microeconomic Adjustment Hazards and Aggregate Dynamics." *Quarterly Journal of Economics* 433(2): 359-383.

Caballero, Ricardo, and Eduardo M. R. A. Engel. 1999. "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized *(S,s)* Approach." *Econometrica* 67(4): 783-826.

Caballero, Ricardo, and Eduardo M. R. A. Engel. 2007. "Price Stickiness in Ss Models: New Interpretations of Old Results." *Journal of Monetary Economics* 54(S1): 100-121.

[26] Interestingly, though information frictions in macro have been revived in recent literature in monetary economics (see Mankiw and Reis's 2011 survey), they have been used much more sparingly in understanding of labor dynamics. (For an exception, see Venkateswaran 2013.)

Caballero, Ricardo, Eduardo M. R. A. Engel, and John Haltiwanger. 1995. "Plant Level Adjustment and Aggregate Investment Dynamics." *Brookings Papers on Economic Activity* 2*:* 1-54.

Caballero, Ricardo, Eduardo M. R. A. Engel, and John Haltiwanger. 1997. "Aggregate Employment Dynamics: Building from Microeconomics." *American Economic Review*, 87(1): 115-137.

Chetty, Raj. 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80(3): 969-1018.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2012. "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." In *NBER Macroeconomics Annual 2012*, edited by D. Acemoglu, J. Parker, and M. Woodford, University of Chicago Press, 1-56.

Christiano, Lawrence, and Richard Todd. 1996. "Time to Plan and Aggregate Fluctuations." Federal Reserve Bank of Minneapolis *Quarterly Review* Winter: 14-27.

Cooper, Russell, John Haltiwanger, and Jonathan L. Willis. 2007. "Search Frictions: Matching Aggregate and Establishment Observations." *Journal of Monetary Economics* 54(1): 56-78.

Cooper, Russell, John Haltiwanger, and Jonathan L. Willis. 2015. "Dynamics of Labor Demand: Evidence from Plant-Level Observations and Aggregate Implications." *Research in Economics* 6(1): 37-50.

Davis, Steven J., and John Haltiwanger. 1992. "Gross Job Creation, Gross Job Destruction, and Employment Reallocation." *Quarterly Journal of Economics* 107(3): 819-864.

Elsby, Michael W. L., and Ryan Michaels. 2013. "Marginal Jobs, Heterogeneous Firms, and Unemployment Flows." *American Economic Journal: Macroeconomics* 5(1): 1-48.

Elsby, Michael W. L., and Ryan Michaels. 2017. "Fixed Adjustment Costs and Aggregate Fluctuations." Mimeo, University of Edinburgh.

Elsby, Michael W. L., Donggyun Shin and Gary Solon. 2016. "Wage Adjustment in the Great Recession and Other Downturns: Evidence from the United States and Great Britain." *Journal of Labor Economics* 34 (S1): S249-S291.

Erceg, Christopher, and Andrew Levin. 2003. "Imperfect Credibility and Inflation Persistence." *Journal of Monetary Economics* 90: 915-944.

Fuhrer, Jeff, and George Moore. 1995. "Inflation Persistence." *Quarterly Journal of Economics* 110(1): 127-159.

Fujita, Shigeru, and Garey Ramey. 2007. "Job Matching and Propagation." *Journal of Economic Dynamics and Control* 31(11): 3671–98.

Gertler, Mark. 1982. "Imperfect information and wage inertia in the business cycle." *Journal of Political Economy* 90(5): 967-987.

Gertler, Mark, and John Leahy. 2008. "A Phillips Curve with an *Ss* Foundation." *Journal of Political Economy* 116(3): 533-572.

Golosov, Mikhail, and Robert E. Lucas. 2007. "Menu Costs and Phillips Curves." *Journal of Political Economy* 115(2): 171-199.

Gorodnichenko, Yuriy. 2010. "Endogenous Information, Menu Costs and Inflation Persistence." Mimeo, University of California, Berkeley.

Hagedorn, Marcus, and Iourii Manovskii. 2011. "Productivity and the Labor Market: Comovement over the Business Cycle." *International Economic Review* 52 (3): 603–19.

Hall, Robert, and Paul Milgrom. 2008. "The Limited Influence of Unemployment on the Wage Bargain," *American Economic Review*, 98(4): 1653-1674.

Hamermesh, Daniel. 1989. "Labor Demand and the Structure of Adjustment Costs." *American Economic Review* 79(4): 674-689.

Hopenhayn, Hugo, and Richard Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy* 101(5): 915-938.

Khan, Aubhik, and Julia Thomas (2008). "Idiosyncratic Shocks and the Role of Nonconvexities in Plant & Aggregate Investment Dynamics," *Econometrica*, 76(2): 395-436.

Kimball, Miles, and Matthew Shapiro. 2010. "Labor Supply: Are Income and Substitution Effects Both Large or Both Small?" Mimeo, University of Michigan.

King, Robert, and Sergio Rebelo. 1999. "Resuscitating Real Business Cycles." *Handbook of Macroeconomics*, volume 1, chapter 14: 927-1007.

King, Robert, and Julia Thomas. 2006. "Partial Adjustment Without Apology." *International Economic Review* 47(3): 779-809.

Krusell, Per, and Anthony Smith. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106(5): 867-896.

Mankiw, Gregory, and Ricardo Reis. 2002. "Sticky Information v. Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *Quarterly Journal of Economics* 117(4): 1295-1328.

Mankiw, Gregory, and Ricardo Reis. 2011. "Imperfect Information and Aggregate Supply," in Friedman, B., and Woodford, M. (eds.), *Handbook of Monetary Economics*, North Holland: 183-230.

Mortensen, Dale T., and Eva Nagypal. 2007. "More on Unemployment and Vacancy Fluctuations." *Review of Economic Dynamics* 10(3): 327-347.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Review of Economic Studies* 61(3): 397–415.

Moscarini, Giuseppe, and Fabien Postel-Vinay. 2016. "Wage Posting and Business Cycles: A Quantitative Exploration." *Review of Economic Dynamics* 19: 135-160.

Nickell, Stephen. 1978. "Fixed Costs, Employment and Labour Demand Over the Cycle." *Economica* 45(180): 329-345.

Nickell, Stephen. 1986. "Dynamic Models of Labour Demand," in Ashenfelter, O., and Layard, R. (eds.), *Handbook of Labor Economics*, North Holland.

Oi, Walter Y. 1962. "Labor as a Quasi-Fixed Factor." *Journal of Political Economy* 70(6): 538-555.

Petrongolo, Barbara, and Christopher A. Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature* 39(2): 390–431.

Pissarides, Christopher A. 1985. "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages." *American Economic Review* 75(4): 676–90.

Pissarides, Christopher A. 2009. "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica* 77: 1339-1369.

Sargent, Thomas J. 1978. "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations." *Journal of Political Economy* 86(6): 1009–1044.

Shapiro, Matthew D. 1986. "The Dynamic Demand for Capital and Labor." *Quarterly Journal of Economics* 101(3): 513-542.

Solon, Gary, Robert Barsky, and Jonathan Parker. 1994. "Measuring the Cyclicality of Real Wages: How Important Is Composition Bias?" *Quarterly Journal of Economics* 109(1): 1-25.

Stole, Lars A., and Jeffrey Zwiebel. 1996. "Intra-Firm Bargaining Under Non-Binding Contracts." *Review of Economic Studies* 63(3): 375-410.

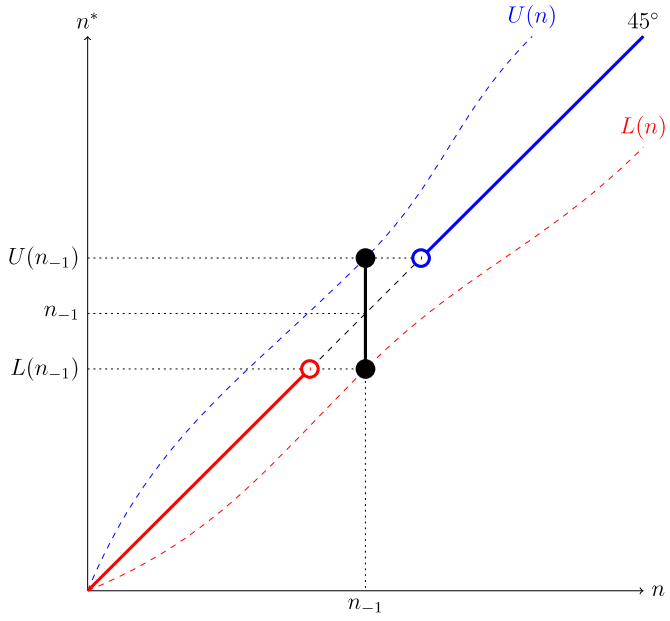
Venkateswaran, Venky. 2013. "Heterogeneous Information and Labor Market Fluctuations." Mimeo, New York University.

Veracierto, Marcelo. 2002. "Plant-Level Irreversible Investment and Equilibrium Business Cycles." *American Economic Review* 92(1): 181-197.

Veracierto, Marcelo. 2008. "Firing Costs and Business Cycle Fluctuations." *International Economic Review* 49(1): 1-39.

Figure 1. *Ss* policies in the presence of fixed and linear adjustment frictions

A. Fixed costs



B. Linear costs

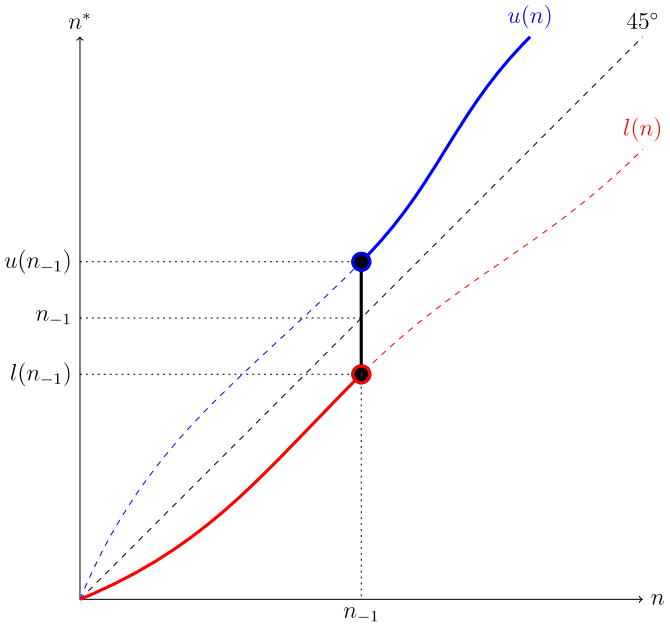


Figure 2. Impulse responses of aggregate employment: Fixed costs

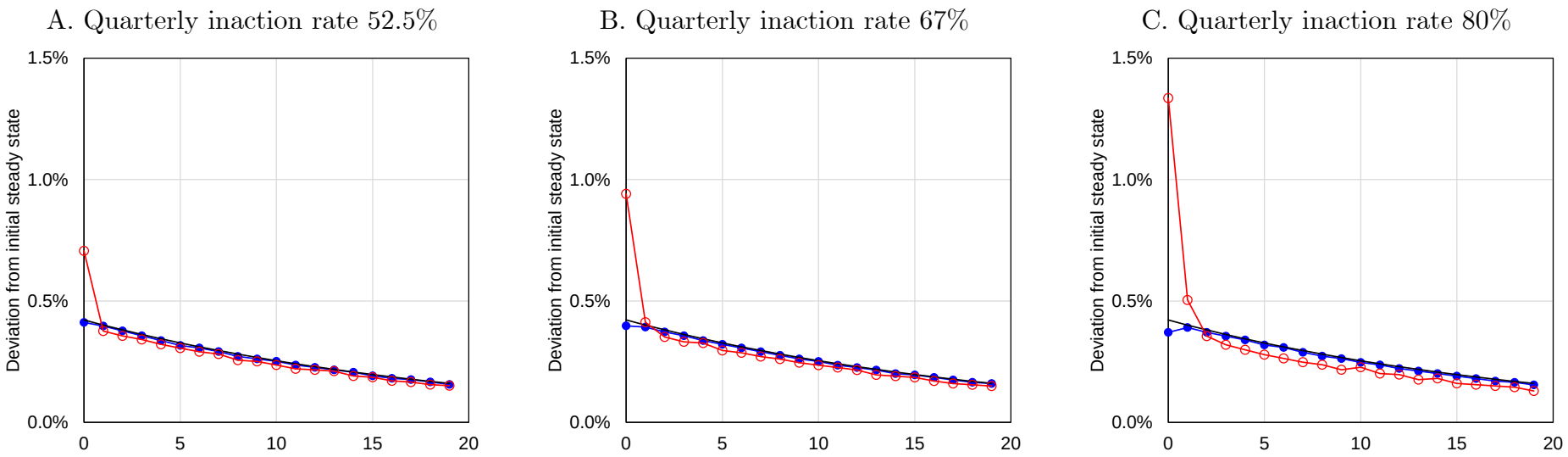


Figure 3. Impulse responses of aggregate employment: Linear costs

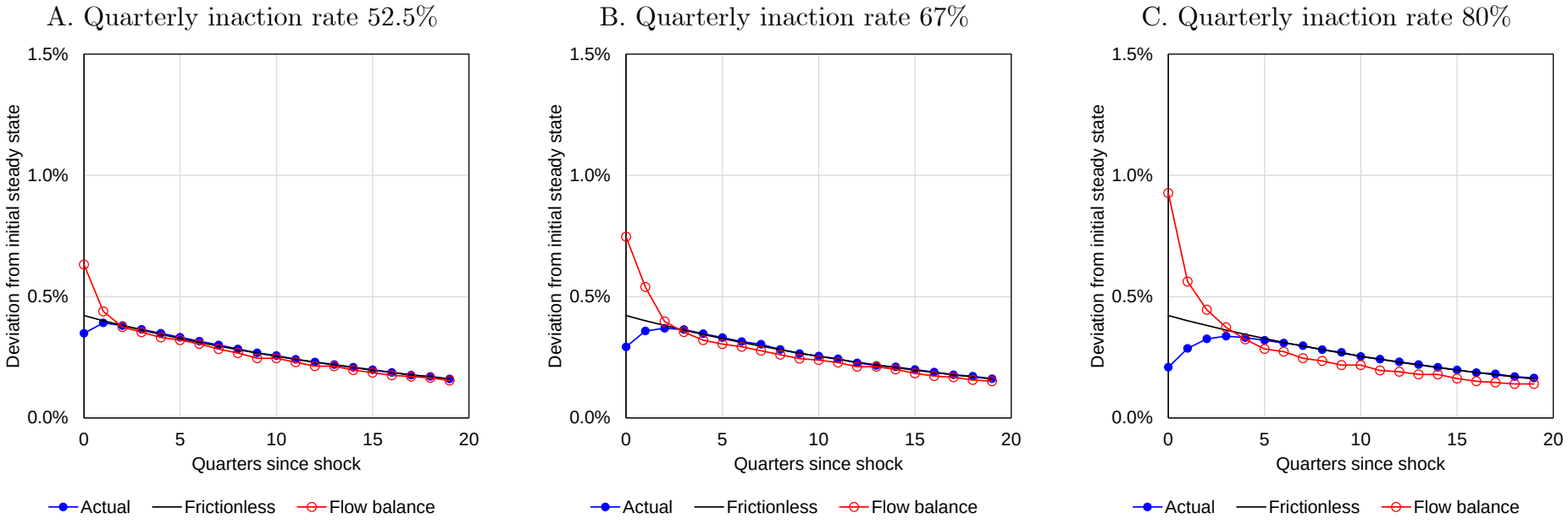


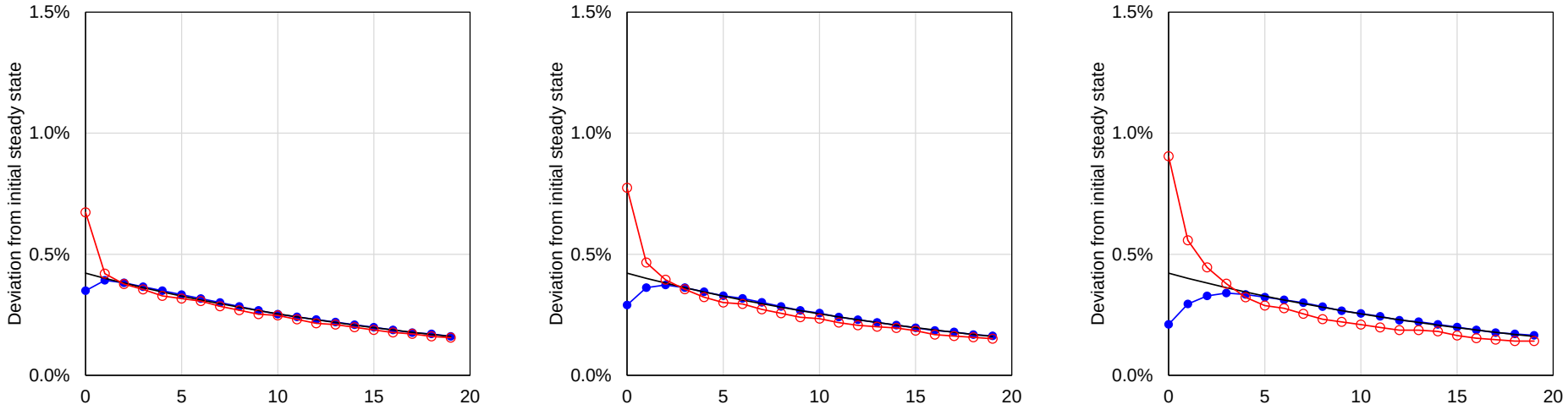
Figure 4. Impulse responses of aggregate employment: Asymmetric linear costs

A. Quarterly inaction rate 52.5%

B. Quarterly inaction rate 67%

C. Quarterly inaction rate 80%

i. Pure hiring cost



ii. Pure firing cost

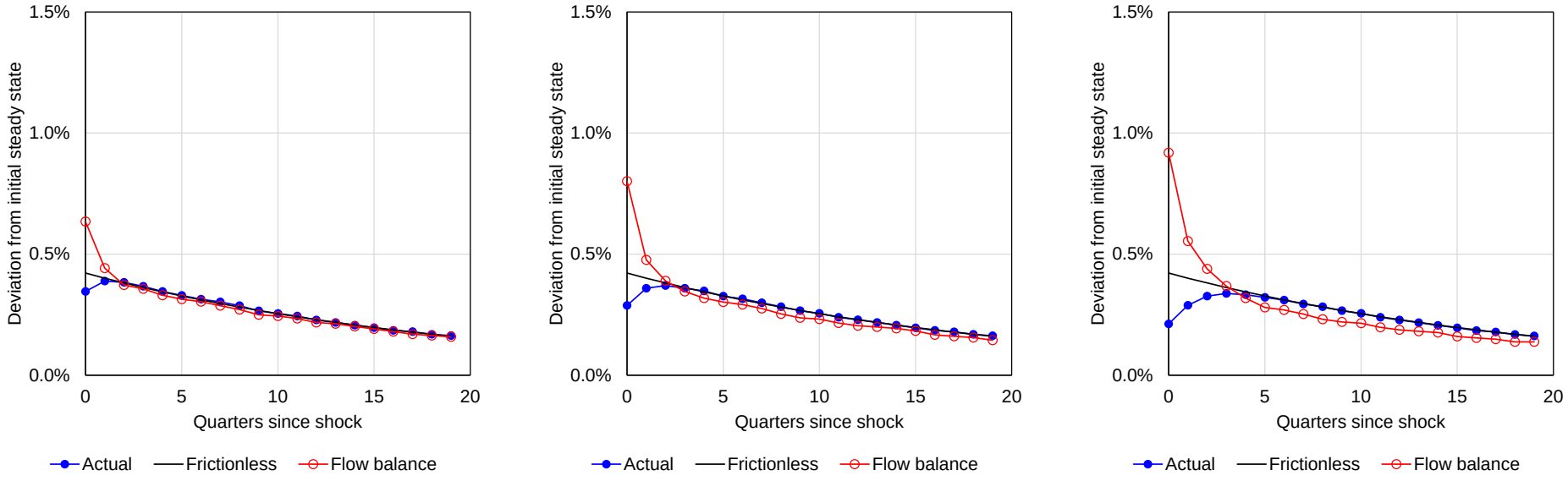


Figure 5. Impulse responses of aggregate employment: Search costs

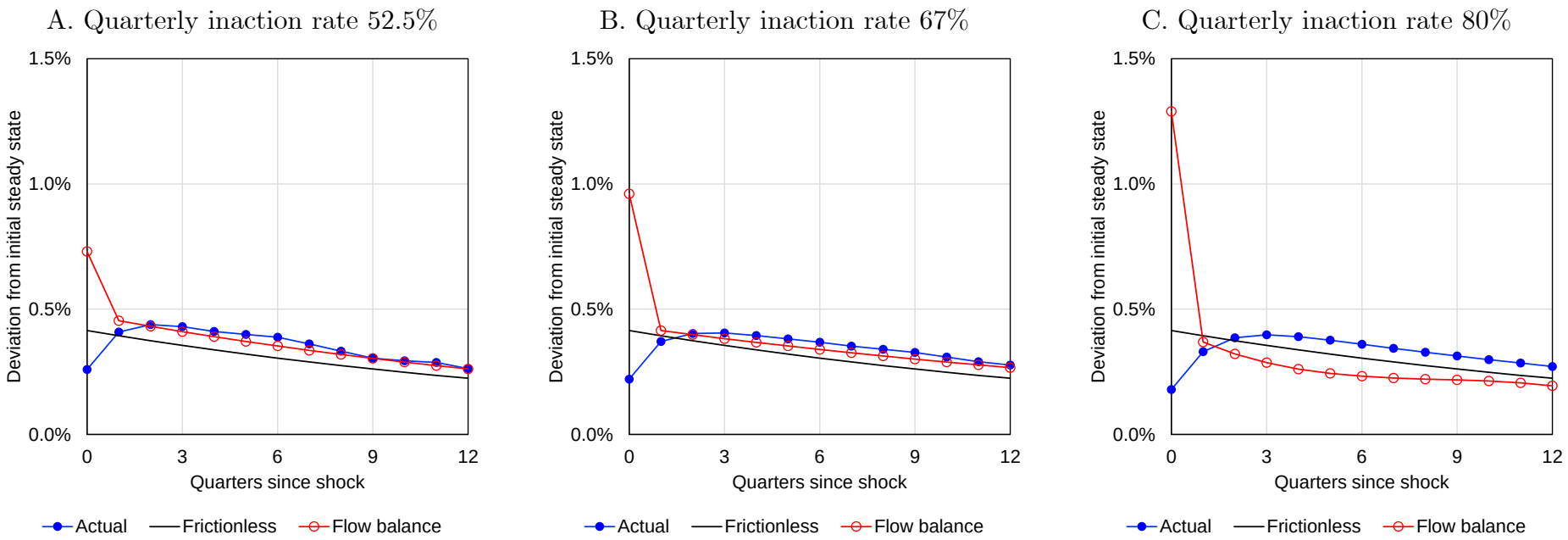


Figure 6. Aggregate employment in the QCEW by sample restriction

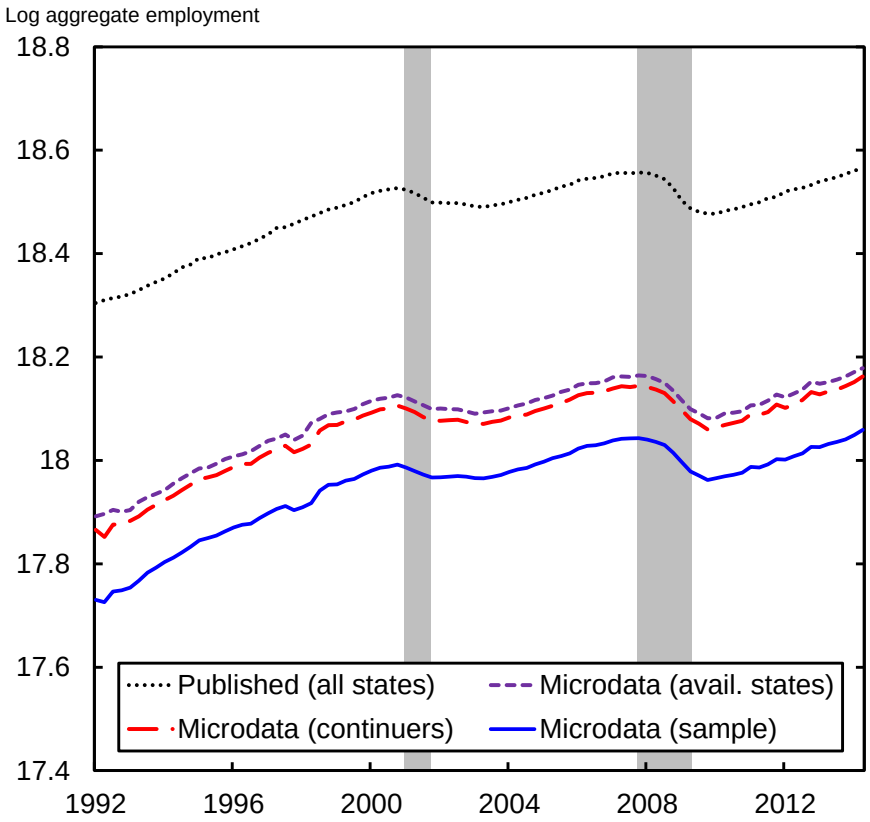


Figure 7. Actual and flow-balance log aggregate employment

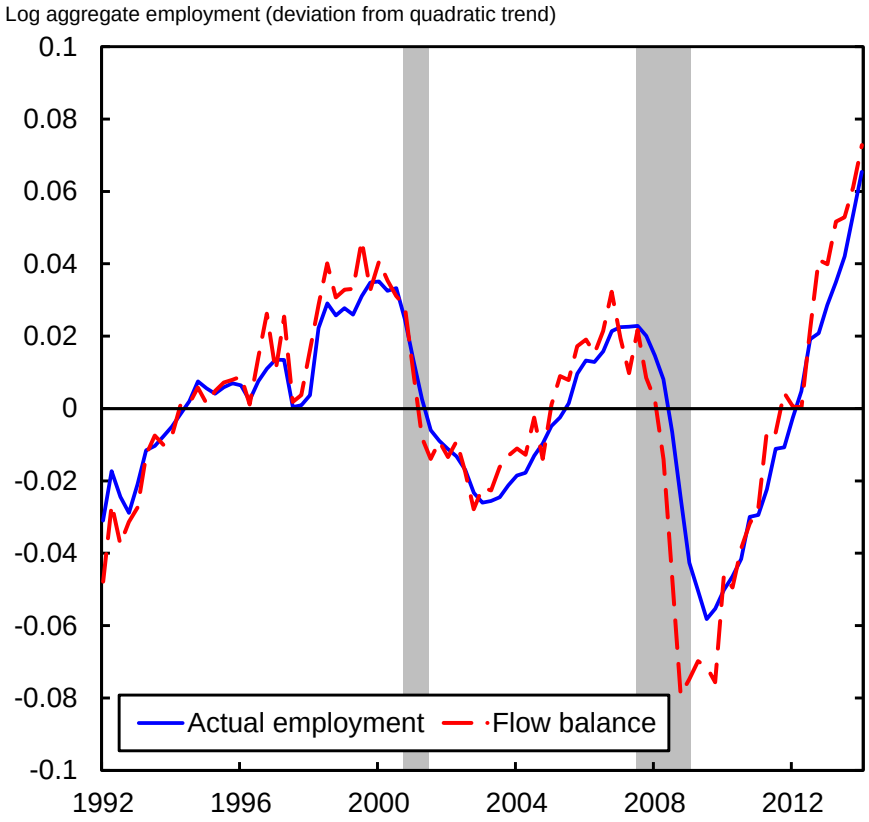


Figure 8. Model-implied flow-balance log aggregate employment: Time series

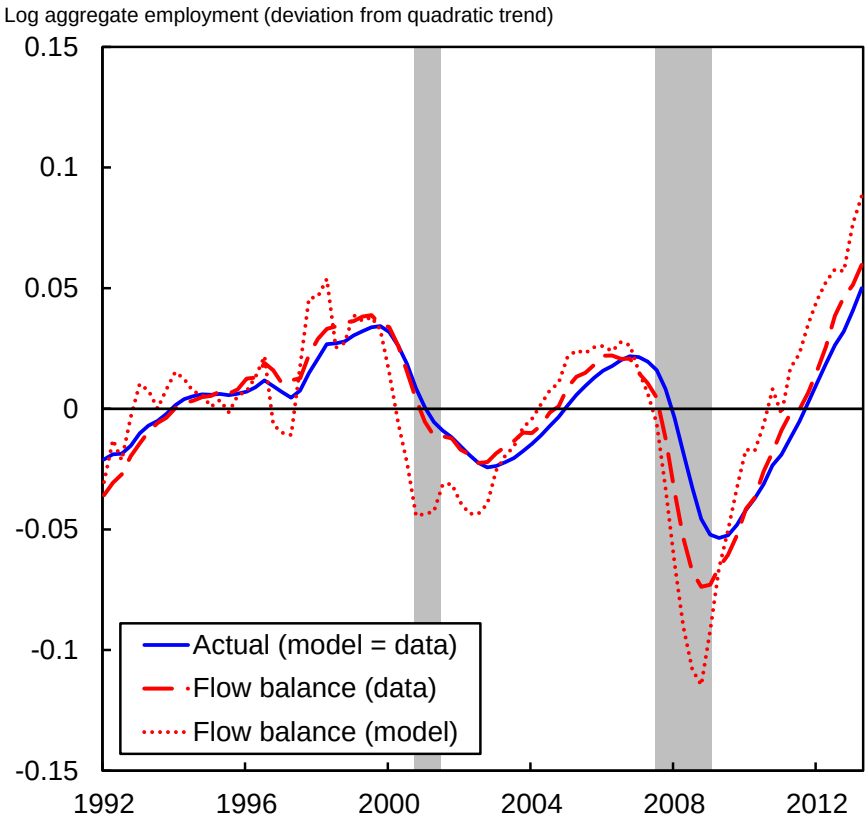
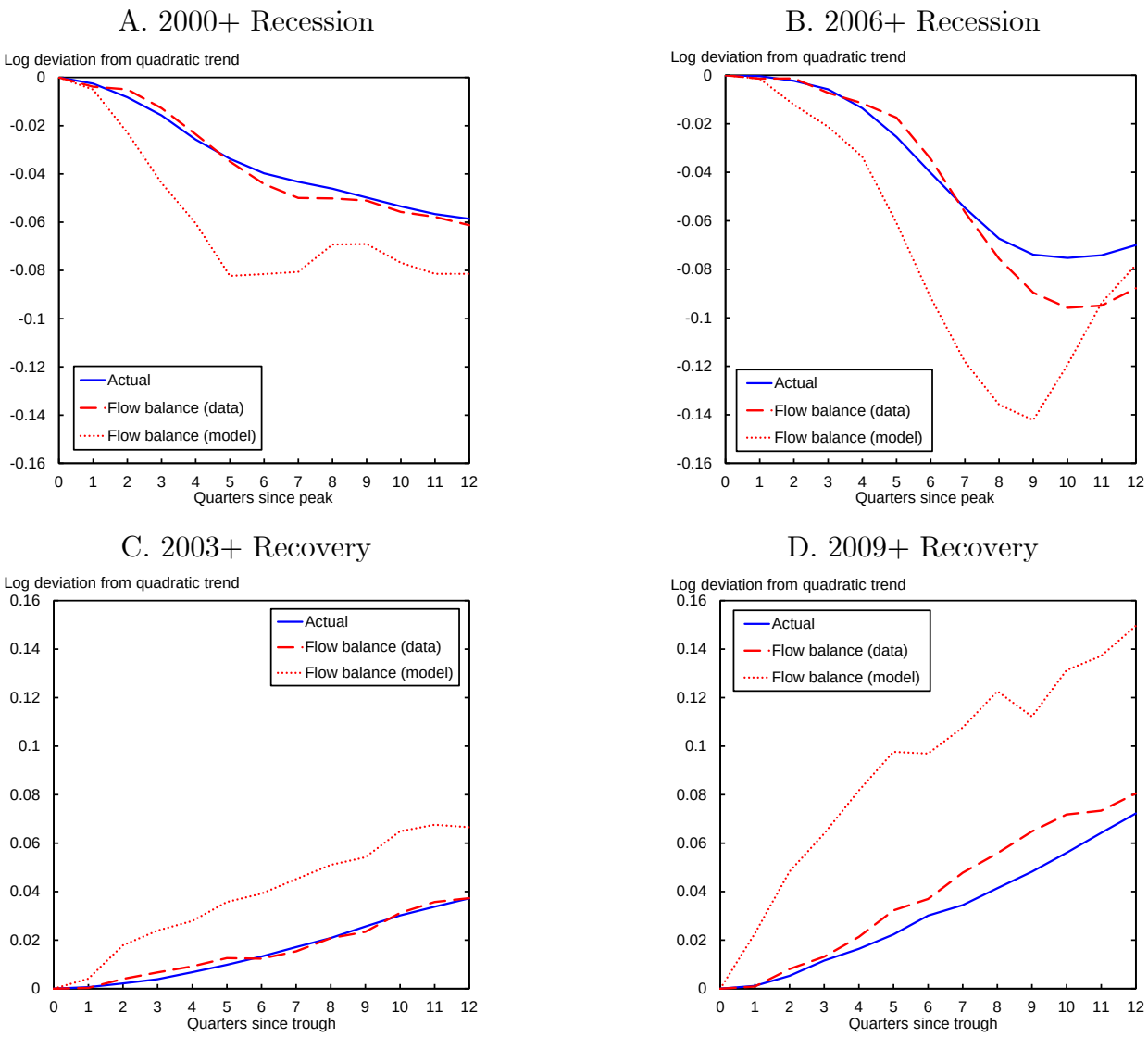


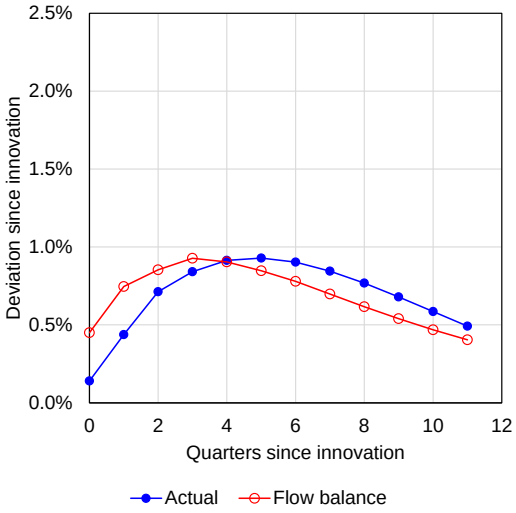
Figure 9. Model-implied flow-balance log aggregate employment: Recession and recovery episodes



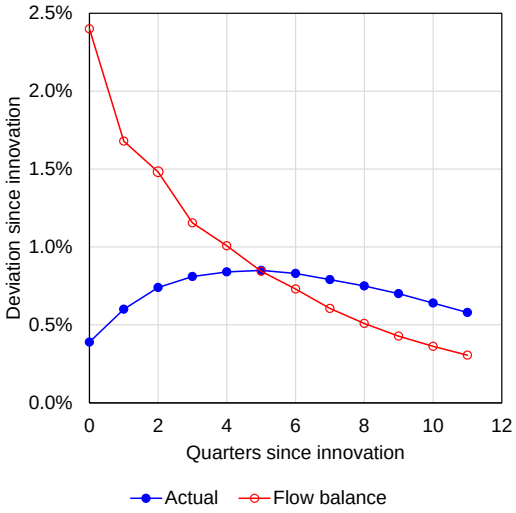
Notes: Each series is plotted relative to its own cyclical peak (panels A and B) or trough (panels C and D) since the timing of the cycle can differ across series (although in practice they only differ at most by two quarters).

Figure 10. Descriptive impulse responses of employment and firm-size flows: Data versus model

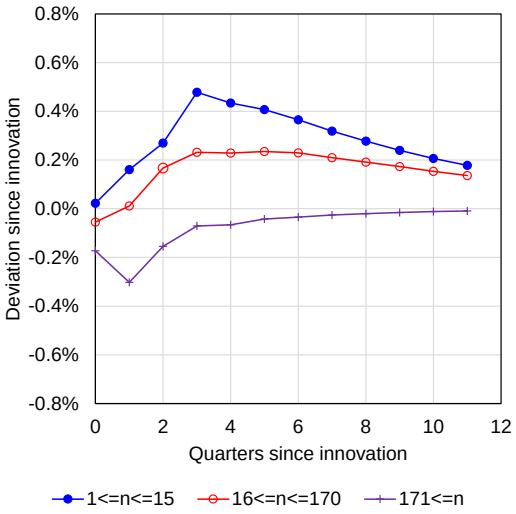
A. Employment: Data



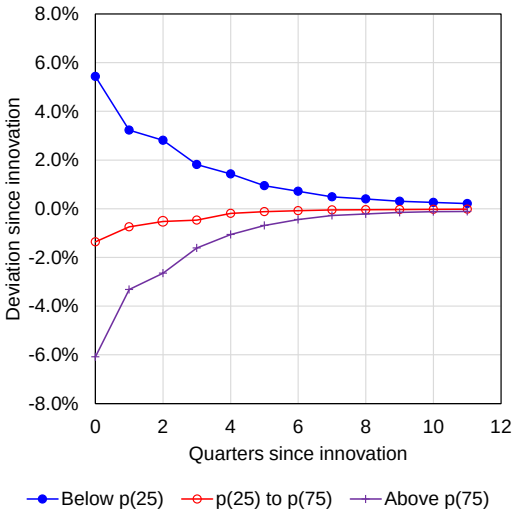
B. Employment: Model



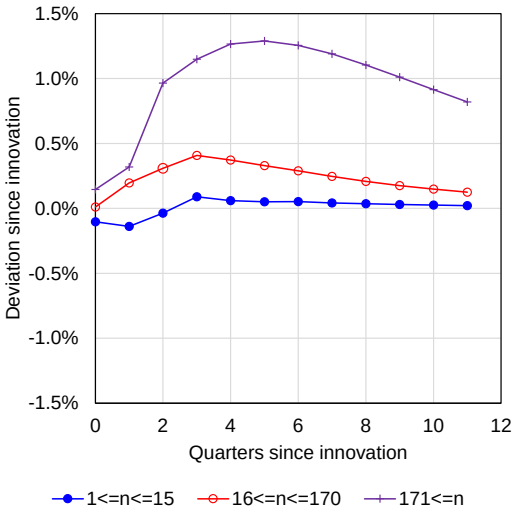
C. Outflow probability: Data



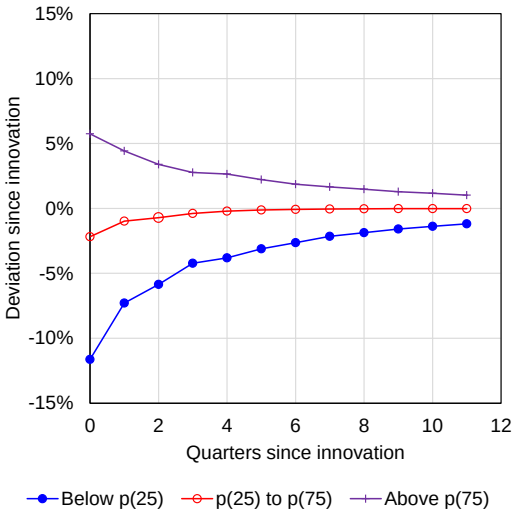
D. Outflow probability: Model



E. Total inflow: Data



F. Total inflow: Model



Appendix

A. Laws of motion for the firm-size distribution

To derive the laws of motion in equations (2) and (11) in the main text, we require notation for several distributions. As in the main text, we denote the densities of employment, lagged employment, and frictionless employment by h, h_{-1}, and h^*, respectively, and will refer to their distribution functions by analogous uppercase letters, H, H_{-1}, and H^*. In addition, however, we require notation for the distributions of frictionless employment conditional on lagged employment, which we denote by $\mathcal{H}^*(\xi|\nu) = \Pr(n^* < \xi | n_{-1} = \nu)$, and the distribution of lagged employment conditional on frictionless employment, denoted by $\mathcal{H}(\nu|\xi) = \Pr(n_{-1} < \nu | n^* = \xi)$. The latter are related by Bayes' rule, $\mathcal{h}(\nu|\xi)h^*(\xi) = \mathcal{h}^*(\xi|\nu)h_{-1}(\nu)$, where lowercase script letters denote associated density functions. But we preserve separate notation to aid clarity.

We can now use the labor demand policy rules—(1) for the fixed costs case, (10) for the linear costs case—to construct laws of motion for the distribution function of actual employment $H(n)$ implied by each type of friction. We then show how these imply the laws of motion for the density $h(n)$ stated in equations (2) and (11) in the main text.

Fixed costs. Consider a point m in the domain of the employment distribution. We wish to derive the flows in and out of $H(m)$. To do so, we first derive flows for a given lagged employment level n_{-1}. Then inflows into $H(m)$ are summarized as follows:

1) If $m < L(n_{-1})$, or equivalently $n_{-1} > L^{-1}(m)$, then the inflow is equal to $\mathcal{H}^*(m|n_{-1})$.
2) If $m \in [L(n_{-1}), n_{-1})$, or equivalently $n_{-1} \in (m, L^{-1}(m)]$, then the inflow is equal to $\mathcal{H}^*(L(n_{-1})|n_{-1})$.

Likewise, the outflows from $H(m)$ for a given n_{-1} can be evaluated as:

3) If $m \in (n_{-1}, U(n_{-1})]$, or equivalently $n_{-1} \in [U^{-1}(m), m)$, then the outflow is equal to $1 - \mathcal{H}^*(U(n_{-1})|n_{-1})$.
4) If $m > U(n_{-1})$, or equivalently $n_{-1} < U^{-1}(m)$, then the outflow is equal to $1 - \mathcal{H}^*(m|n_{-1})$.

Integrating the latter with respect to the distribution of lagged employment $H_{-1}(n_{-1})$ recovers the aggregate flows and thereby the law of motion for $H(m)$,

$$\begin{aligned}\Delta H(m) = &\int_{L^{-1}(m)} \mathcal{H}^*(m|n_{-1}) dH_{-1}(n_{-1}) + \int_{m}^{L^{-1}(m)} \mathcal{H}^*(L(n_{-1})|n_{-1}) dH_{-1}(n_{-1}) \\ &- \int_{U^{-1}(m)}^{m} [1 - \mathcal{H}^*(U(n_{-1})|n_{-1})] dH_{-1}(n_{-1}) \\ &- \int^{U^{-1}(m)} [1 - \mathcal{H}^*(m|n_{-1})] dH_{-1}(n_{-1}).\end{aligned} \tag{26}$$

Linear costs. Likewise, one can use the adjustment rule for the linear costs case, (10), to construct an analogous law of motion under linear costs. Again, we first fix a given level of lagged employment, n_{-1}, and evaluate inflows to, and outflows from, $H(m)$. These flows are simpler in the linear costs case. Inflows are given by:

1) If $m < n_{-1}$, or equivalently $n_{-1} > m$, then the inflow is equal to $\mathcal{H}^*(l(m)|n_{-1})$.

Similarly, outflows are given by:

2) If $m > n_{-1}$, or equivalently $n_{-1} < m$, then the outflow is equal to $1 - \mathcal{H}^*(u(m)|n_{-1})$.

Following the same logic as above, the law of motion for $H(m)$ is thus given by

$$\Delta H(m) = \int_m \mathcal{H}^*(l(m)|n_{-1}) dH_{-1}(n_{-1}) - \int^m [1 - \mathcal{H}^*(u(m)|n_{-1})] dH_{-1}(n_{-1}). \quad (27)$$

Laws of motion for $h(n)$. Differentiating (26) and (27) with respect to m, cancelling terms, and using Bayes' rule to note that $\int_0^\nu \hbar^*(\xi|n_{-1}) h_{-1}(n_{-1}) dn_{-1} = \int_0^\nu \hbar(n_{-1}|\xi) h^*(\xi) dn_{-1}$ yields the simpler laws of motion for the density of employment $h(n)$, equations (2) and (11) in the main text.

B. Proofs of Propositions 1 and 2

To establish Propositions 1 and 2 in the main text, it is convenient first to define a notion of *quasi-frictionless* employment, defined as the employment level implied by *frictionless* labor demand, evaluated at the *frictional* wage, w. Lemma 1 shows that the firm's problem can be normalized with respect to quasi-frictionless employment to establish some useful homogeneity properties. Using this homogeneous problem, we can relate the change in aggregate log flow steady-state employment $\Delta\widehat{\mathcal{N}}$ to the change in aggregate log quasi-frictionless employment. In a final step, we link the latter to the change in aggregate log *frictionless* employment.

Definition *(i) Quasi-frictionless employment $n^\star$ solves $pxan^{\star\alpha-1} \equiv w$, where w is the frictional equilibrium wage; and (ii) frictionless employment n^* solves $pxan^{*\alpha-1} \equiv w^*$, where w^* is the frictionless equilibrium wage.*

Remark *The change in aggregate log quasi-frictionless employment $\Delta\mathcal{N}^\star$ induced by a change in aggregate productivity $\Delta \ln p$ is related to the change in aggregate log frictionless employment $\Delta\mathcal{N}^*$ according to*

$$\Delta\mathcal{N}^\star = \frac{1 - \epsilon_w}{1 - \epsilon_{w^*}} \Delta\mathcal{N}^*, \quad (28)$$

where ϵ_w and ϵ_{w^*} denote the elasticities of the equilibrium wage to aggregate productivity p, respectively, with and without frictions.

Lemma 1 (Caballero and Engel, 1999) *Consider the firm's problem,*

$$\Pi(n_{-1},x) \equiv \max_{n}\{pxn^{\alpha} - wn - C^{+}\mathbb{I}[n > n_{-1}] - C^{-}\mathbb{I}[n < n_{-1}] - c^{+}\Delta n^{+} + c^{-}\Delta n^{-} + \beta\mathbb{E}[\Pi(n,x')|x]\}. \tag{29}$$

If (i) $\ln x' = \ln x + \varepsilon'_x$ *with* ε'_x *i.i.d., and (ii)* $C^{-/+} = \Gamma^{-/+}wn^{\star}$ *and* $c^{-/+} = \gamma^{-/+}w$*, then (a) the adjustment policy takes the form*

$$n = \begin{cases} n^{\star}/u & \text{if} \quad n^{\star} > U \cdot n_{-1}, \\ n_{-1} & \text{if} \quad n^{\star} \in [L \cdot n_{-1}, U \cdot n_{-1}], \\ n^{\star}/l & \text{if} \quad n^{\star} < L \cdot n_{-1} \end{cases} \tag{30}$$

for constants $L \leq l < 1 < u \leq U$ *; and (b) desired (log) employment adjustments,* $\ln(n^{\star}/n_{-1})$*, are independent of initial firm size* n_{-1}*.*

Proof of Lemma 1. Since idiosyncratic shocks follow a geometric random walk, $\ln x' = \ln x + \varepsilon'_x$, so does (quasi-) frictionless employment, $\ln n^{\star\prime} = \ln n^{\star} + \varepsilon'_{n^{\star}}$ *where* $\varepsilon'_{n^{\star}} = \varepsilon'_x/(1-\alpha)$. Defining $z = n/n^{\star}$ and $\zeta = n_{-1}/n^{\star}$, a conjecture that $\Pi(n_{-1},x) = wn^{\star}\widetilde{\Pi}(\zeta)$ implies

$$\widetilde{\Pi}(\zeta) = \max_{z}\left\{\frac{z^{\alpha}}{\alpha} - z - \Gamma^{+}\mathbb{I}[z > \zeta] - \Gamma^{-}\mathbb{I}[z < \zeta] - \gamma^{+}(z-\zeta)^{+} + \gamma^{-}(z-\zeta)^{-} + \beta\mathbb{E}\left[e^{\varepsilon'_{n^{\star}}}\widetilde{\Pi}\left(e^{-\varepsilon'_{n^{\star}}}z\right)\right]\right\}. \tag{31}$$

We highlight two aspects of (31). First, the expectation over the forward value is no longer conditional, since it is taken over $\varepsilon'_{n^{\star}}$, which is i.i.d. Second, the firm's problem is simplified to the choice of a number $z = n/n^{\star}$ for each realization of the single state variable $\zeta = n_{-1}/n^{\star}$.

An *Ss* policy will thus stipulate that $z = \zeta$ for intermediate values of $\zeta \in [1/U, 1/L]$ and will set $z = 1/u$ whenever $\zeta < 1/U$, and $z = 1/l$ whenever $\zeta > 1/L$. Mapping back into employment terms implies the adjustment policy in (30), establishing part a) of the result. Note that the case of pure fixed costs implies $u = l$, while pure linear costs imply $l = L < U = u$.

To establish part b), note that the probability of a desired log employment adjustment of size less than δ can be written, in general, as

$$\Pr\left(\ln\left(n^{\star\prime}/n\right) < \delta|n\right) = \Pr(\varepsilon'_{n^{\star}} < \delta + \ln z\,|n) = \int \Pr(\varepsilon'_{n^{\star}} < \delta + \ln z\,|n,z)\,dZ(z|n), \tag{32}$$

where $Z(z|n)$ denotes the distribution function of z given n. In the context of the canonical model, however, (32) simplifies. First, $\varepsilon'_{n^{\star}}$ is independent of n since the former is i.i.d. Second, z is also independent of n. To see this, note first that if a firm adjusts this period, its choice of z is uninformed by n—it sets $z = 1/u$ or $z = 1/l$. If the firm sets $n = n_{-1}$ but adjusted *last* period, then it sets $\ln z = \ln n_{-1} - \ln n^{\star} = \ln z_{-1} - \varepsilon_{n^{\star}}$ and z_{-1} is $1/u$ or $1/l$. Thus, z is again independent of n. More generally, suppose the firm last adjusted T periods ago, that is, $n = n_{-1} = \cdots = n_{-T}$ and $z_{-T} = 1/u$ or $1/l$. Then, $\ln z =$

$\ln n_{-T} - \ln n^\star = \ln z_{-T} - \sum_{t=0}^{T-1} \varepsilon_{n^\star_{-t}}$. Each term here is independent of $n = n_{-T}$. Equation (32) therefore collapses to

$$\Pr\big(\ln(n^{\star\prime}/n) < \delta | n\big) = \int \Pr(\varepsilon'_{n^\star} < \delta + \ln z \,|z)\, dZ(z), \tag{33}$$

which does not depend on n.

Proof of Proposition 1. Denoting log employment by n, the adjustment rules take the form $L(n) = n - \lambda$ and $U(n) = n + \upsilon$ for $\lambda > 0$ and $\upsilon > 0$. The density of log employment in flow balance is then defined by

$$\hat{h}(n) \equiv \frac{1 - \mathcal{H}(n+\lambda|n) + \mathcal{H}(n-\upsilon|n)}{1 - \mathcal{H}^\star(n+\upsilon|n) + \mathcal{H}^\star(n-\lambda|n)} h^\star(n), \tag{34}$$

where $\mathcal{H}^\star(\xi|\nu) \equiv \Pr(n^\star < \xi | n_{-1} = \nu)$ and $\mathcal{H}(\nu|\xi) \equiv \Pr(n_{-1} < \nu | n^\star = \xi)$. The property of the canonical model noted in result b) of Lemma 1, that $n^\star - n_{-1}$ is independent of n_{-1}, implies that

$$\mathcal{H}^\star(\xi|\nu) = \Pr(n^\star - n_{-1} < \xi - \nu) \equiv \widetilde{\mathcal{H}}^\star(\xi - \nu). \tag{35}$$

This implies that the probability of adjusting away from n is independent of n,

$$1 - \mathcal{H}^\star(n+\upsilon|n) + \mathcal{H}^\star(n-\lambda|n) = 1 - \int_{-\lambda}^{\upsilon} \tilde{\hbar}^\star(z)dz \equiv \phi. \tag{36}$$

Now consider the probability of adjusting to n. Using Bayes' rule, equation (35), and a change of variable, we can write this as

$$\begin{aligned} 1 - \mathcal{H}(n+\lambda|n) + \mathcal{H}(n-\upsilon|n) &= 1 - \int_{n-\upsilon}^{n+\lambda} \hbar^\star(n|\nu) \frac{h_{-1}(\nu)}{h^\star(n)} d\nu \\ &= 1 - \int_{n-\upsilon}^{n+\lambda} \tilde{\hbar}^\star(n-\nu) \frac{h_{-1}(\nu)}{h^\star(n)} d\nu \\ &= 1 - \int_{-\lambda}^{\upsilon} \tilde{\hbar}^\star(z) \frac{h_{-1}(n-z)}{h^\star(n)} dz. \end{aligned} \tag{37}$$

Piecing this together, we have

$$\hat{h}(n) = \frac{h^\star(n) - \int_{-\lambda}^{\upsilon} \tilde{\hbar}^\star(z) h_{-1}(n-z)dz}{1 - \int_{-\lambda}^{\upsilon} \tilde{\hbar}^\star(z)dz}. \tag{38}$$

Multiplying both sides by n, using (36), and integrating yields

$$\begin{aligned} \widehat{\mathcal{N}} \equiv \int_{-\infty}^{\infty} n\hat{h}(n)dn &= \frac{\mathcal{N}^\star}{\phi} - \frac{1}{\phi}\int_{-\infty}^{\infty}\int_{-\lambda}^{\upsilon} n\tilde{\hbar}^\star(z)h_{-1}(n-z)dzdn \\ &= \frac{\mathcal{N}^\star}{\phi} - \frac{1-\phi}{\phi}\mathcal{N}_{-1} - \frac{1}{\phi}\int_{-\lambda}^{\upsilon} z\tilde{\hbar}^\star(z)dz. \end{aligned} \tag{39}$$

Since there is a constant-$\mathcal{N}^\star$ state prior to the aggregate shock, aggregate log employment is constant and equal to aggregate flow-balance employment, $\mathcal{N}_{-1} = \mathcal{N}_{-2} = \widehat{\mathcal{N}}_{-1}$. Imposing this and solving for $\widehat{\mathcal{N}}_{-1}$ yields

$$\widehat{\mathcal{N}}_{-1} = \mathcal{N}^{\star}_{-1} - \int_{-\lambda}^{\upsilon} z\tilde{\hbar}^{\star}_{-1}(z)dz. \tag{40}$$

Now consider a shock to aggregate log (quasi-) frictionless employment, $\Delta\mathcal{N}^{\star}$. On impact this will shift the mean of the distribution of desired employment adjustments, $\tilde{\hbar}^{\star}(\cdot)$, by $\Delta\mathcal{N}^{\star}$. Given the prior constant-$\mathcal{N}^{\star}$ state, substitution of (40) into (39) implies

$$\Delta\widehat{\mathcal{N}} = \frac{\Delta\mathcal{N}^{\star}}{\phi} - \frac{1}{\phi}\int_{-\lambda}^{\upsilon} z\Delta\tilde{\hbar}^{\star}(z)dz. \tag{41}$$

Noting that $\Delta\tilde{\hbar}^{\star}(z) = \tilde{\hbar}^{\star}_{-1}(z - \Delta\mathcal{N}^{\star}) - \tilde{\hbar}^{\star}_{-1}(z)$, a first-order approximation around $\Delta\mathcal{N}^{\star} = 0$ yields

$$\Delta\widehat{\mathcal{N}} \approx (1+\psi)\cdot\Delta\mathcal{N}^{\star}, \text{ where } \psi \equiv \frac{\upsilon\tilde{\hbar}^{\star}_{-1}(\upsilon) + \lambda\tilde{\hbar}^{\star}_{-1}(-\lambda)}{\phi_{-1}}. \tag{42}$$

The latter, combined with equation (28), implies the stated result.

Proof of Proposition 2. The proof mirrors the proof of Proposition 1 above. The adjustment rules again take the form $l(n) = n - \lambda$ and $u(n) = n + \upsilon$ for $\lambda > 0$ and $\upsilon > 0$. The density of log employment in flow balance is then defined by

$$\hat{h}(n) \equiv \frac{[1 - \mathcal{H}(n|n-\lambda)]h^{\star}(n-\lambda) + \mathcal{H}(n|n+\upsilon)h^{\star}(n+\upsilon)}{1 - \mathcal{H}^{\star}(n+\upsilon|n) + \mathcal{H}^{\star}(n-\lambda|n)}. \tag{43}$$

Since $\mathcal{H}^{\star}(\xi|\nu) = \Pr(n^{\star} - n_{-1} < \xi - \nu) \equiv \widetilde{\mathcal{H}}^{\star}(\xi - \nu)$, the probability of adjusting away from n is again independent of n,

$$1 - \mathcal{H}^{\star}(n+\upsilon|n) + \mathcal{H}^{\star}(n-\lambda|n) = 1 - \int_{-\lambda}^{\upsilon}\tilde{\hbar}^{\star}(z)dz \equiv \phi. \tag{44}$$

Now use Bayes' rule to write the probabilities of adjusting down and up to n as

$$\begin{aligned} 1 - \mathcal{H}(n|n-\lambda) &= \int_{n}^{\infty}\hbar^{\star}(n-\lambda|n)\frac{h_{-1}(\nu)}{h^{\star}(n-\lambda)}d\nu \\ &= \int_{n}^{\infty}\tilde{\hbar}^{\star}(n-\lambda-\nu)\frac{h_{-1}(\nu)}{h^{\star}(n-\lambda)}d\nu \\ &= \int_{-\infty}^{-\lambda}\tilde{\hbar}^{\star}(z)\frac{h_{-1}(n-\lambda-z)}{h^{\star}(n-\lambda)}dz, \end{aligned} \tag{45}$$

and, using an analogous method,

$$\mathcal{H}(n|n+\upsilon) = \int_{\upsilon}^{\infty}\tilde{\hbar}^{\star}(z)\frac{h_{-1}(n+\upsilon-z)}{h^{\star}(n+\upsilon)}dz. \tag{46}$$

Piecing this together, we have

$$\hat{h}(n) = \frac{\int_{-\infty}^{-\lambda}\tilde{\hbar}^{\star}(z)h_{-1}(n-\lambda-z)dz + \int_{\upsilon}^{\infty}\tilde{\hbar}^{\star}(z)h_{-1}(n+\upsilon-z)dz}{1 - \int_{-\lambda}^{\upsilon}\tilde{\hbar}^{\star}(z)dz}. \tag{47}$$

Multiplying both sides by n and integrating yields

$$\begin{aligned}\widehat{\mathcal{N}} &= \frac{1}{\phi}\int_{-\infty}^{-\lambda}\tilde{\hbar}^{\star}(z)\int_{-\infty}^{\infty} n h_{-1}(n-\lambda-z)dn\,dz + \frac{1}{\phi}\int_{\upsilon}^{\infty}\tilde{\hbar}^{\star}(z)\int_{-\infty}^{\infty} n h_{-1}(n+\upsilon-z)dn\,dz \\ &= \frac{\mathcal{N}^{*}}{\phi} - \frac{1-\phi}{\phi}\mathcal{N}_{-1} + \frac{1}{\phi}\left[\lambda\int_{-\infty}^{-\lambda}\tilde{\hbar}^{\star}(z)dz - \upsilon\int_{\upsilon}^{\infty}\tilde{\hbar}^{\star}(z)dz\right] - \frac{1}{\phi}\int_{-\lambda}^{\upsilon} z\tilde{\hbar}^{\star}(z)dz,\end{aligned} \tag{48}$$

where we have used the fact that $\int_{-\infty}^{\infty} z\tilde{\hbar}^{\star}(z)dz = \mathcal{N}^{\star} - \mathcal{N}_{-1}$. Solving for $\widehat{\mathcal{N}}_{-1} = \mathcal{N}_{-1} = \mathcal{N}_{-2}$ in the prior constant-$\mathcal{N}^{\star}$ state yields

$$\widehat{\mathcal{N}}_{-1} = \mathcal{N}_{-1}^{\star} + \lambda\int_{-\infty}^{-\lambda}\tilde{\hbar}_{-1}^{\star}(z)dz - \upsilon\int_{\upsilon}^{\infty}\tilde{\hbar}_{-1}^{\star}(z)dz - \int_{-\lambda}^{\upsilon} z\tilde{\hbar}_{-1}^{\star}(z)dz. \tag{49}$$

Substitution of (49) into (48) implies that a shock to aggregate log (quasi-) frictionless employment that shifts the mean of $\tilde{\hbar}^{\star}(\cdot)$ by $\Delta\mathcal{N}^{\star}$ will induce a change in $\widehat{\mathcal{N}}$ relative to the prior constant-$\mathcal{N}^{\star}$ state equal to

$$\Delta\widehat{\mathcal{N}} = \frac{\Delta\mathcal{N}^{\star}}{\phi} + \frac{1}{\phi}\left[\lambda\int_{-\infty}^{-\lambda}\Delta\tilde{\hbar}^{\star}(z)dz - \upsilon\int_{\upsilon}^{\infty}\Delta\tilde{\hbar}^{\star}(z)dz\right] - \frac{1}{\phi}\int_{-\lambda}^{\upsilon} z\Delta\tilde{\hbar}^{\star}(z)dz. \tag{50}$$

Noting that $\Delta\tilde{\hbar}^{\star}(z) = \tilde{\hbar}_{-1}^{\star}(z-\Delta\mathcal{N}^{\star}) - \tilde{\hbar}_{-1}^{\star}(z)$, a first-order approximation around $\Delta\mathcal{N}^{\star} = 0$ yields

$$\Delta\widehat{\mathcal{N}} \approx \Delta\mathcal{N}^{\star}. \tag{51}$$

Combining with equation (28) yields the stated result.

C. Large-firm canonical search and matching model

In this appendix, we describe in more detail the theoretical results and the quantitative numerical model presented in section 1.5.

Theoretical results. The firm's problem for this model combines equations (16) and (17) in the main text to obtain:

$$\begin{aligned}&\Pi(n_{-1},x) \equiv \max_{n}\left\{Apxn^{\alpha} - (1-\eta)\omega n - \frac{c}{q(\theta)}\Delta n^{+} + \beta\mathbb{E}[\Pi(n,x')|x]\right\}, \\ &\text{where } A \equiv \frac{1-\eta}{1-\eta(1-\alpha)}.\end{aligned} \tag{52}$$

To establish Proposition 3 in the main text, we proceed as above.

Definition *(i) Quasi-frictionless employment $n^{\star}$ solves $Apx\alpha n^{\star\alpha-1} \equiv (1-\eta)\omega$, where ω is the worker's outside option; and (ii) frictionless employment n^{*} solves $px\alpha n^{*\alpha-1} \equiv w^{*}$, where w^{*} is the frictionless equilibrium wage.*

Remark *The change in aggregate log quasi-frictionless employment $\Delta\mathcal{N}^{\star}$ induced by a change in aggregate productivity $\Delta\ln p$ is related to the change in aggregate log frictionless employment $\Delta\mathcal{N}^{*}$ according to*

$$\Delta\mathcal{N}^\star = \frac{1-\epsilon_\omega}{1-\epsilon_{w^*}}\Delta\mathcal{N}^*, \tag{53}$$

where ϵ_ω and ϵ_{w^*}, respectively, denote the elasticities of the worker's outside option ω and the frictionless wage w^* to aggregate productivity p.

Lemma 1′ *If (i) $\ln x' = \ln x + \varepsilon'_x$ with ε'_x i.i.d., and (ii) $c = \gamma(1-\eta)\omega$, then (a) the adjustment triggers take the form in (10), are linear, $l(n) = l \cdot n$ and $u(n) = u \cdot n$ for time-varying $l < 1 < u$; and (b) desired (log) employment adjustments, $\ln(n^\star/n_{-1})$, are independent of initial firm size n_{-1}.*

Proof. Note that a conjecture that $\Pi(n_{-1}, x) = (1-\eta)\omega n^\star \widetilde{\Pi}(\zeta)$ yields

$$\widetilde{\Pi}(\zeta) \equiv \max_z \left\{ \frac{z^\alpha}{\alpha} - z - \frac{\gamma}{q(\theta)}(z-\zeta)^+ + \beta \mathbb{E}\left[e^{\varepsilon'_{n^\star}} \widetilde{\Pi}\left(e^{-\varepsilon'_{n^\star}} z \right) \right] \right\}. \tag{54}$$

Results (a) and (b) follow from the proof to Lemma 1 above.

Lemma 2 *If (i) the adjustment triggers are symmetric, $-\ln l = \ln u \equiv \mu$, and (ii) the distribution of innovations $\varepsilon_{n^\star}$ is symmetric, $\mathcal{E}(-\varepsilon_{n^\star}) = 1 - \mathcal{E}(\varepsilon_{n^\star})$, then the distribution of desired (log) employment adjustments $\ln(n^\star/n_{-1})$ is symmetric, $\widetilde{\mathcal{H}}^\star(-\varsigma) = 1 - \widetilde{\mathcal{H}}^\star(\varsigma)$.*

Proof. Note first that the distribution of the desired log change in employment, $n^\star - n_{-1}$, conditional on last period's log gap, $z_{-1} = n_{-1} - n^\star_{-1}$, takes the simple form $\Pr(n^\star - n_{-1} < \varsigma | z_{-1}) = \mathcal{E}(\varsigma - z_{-1})$, since $\varepsilon_{n^\star} \equiv n^\star - n^\star_{-1}$ is i.i.d. with distribution function $\mathcal{E}(\cdot)$. It follows that the unconditional distribution of $n^\star - n_{-1}$ is

$$\widetilde{\mathcal{H}}^\star(\varsigma) = \int_{-\mu}^{\mu} \Pr(n^\star - n_{-1} < \varsigma | z_{-1})\, g(z_{-1}) \mathrm{d}z_{-1} = \int_{-\mu}^{\mu} \mathcal{E}(\varsigma - z_{-1}) g(z_{-1}) \mathrm{d}z_{-1}, \tag{55}$$

where $g(z_{-1})$ is the ergodic density of z_{-1}. It is simple to verify that $\mathcal{E}(-\varepsilon_{n^\star}) = 1 - \mathcal{E}(\varepsilon_{n^\star})$ implies $\widetilde{\mathcal{H}}^\star(\varsigma) = 1 - \widetilde{\mathcal{H}}^\star(-\varsigma)$, provided $g(\cdot)$ also is symmetric, which we now establish.

Our strategy is to conjecture that $g(\cdot)$ is symmetric and verify that this is implied. Consider a firm with an initial $z_{-1} = z - \varepsilon$, such that $z \in (-\mu, \mu)$ lies strictly inside the inaction range. Clearly, this firm migrates to z if it draws ε. Thus, the mass of firms at z this period is given by

$$g(z) = \int_{z-\mu}^{z+\mu} g(z-\varepsilon)\, \mathrm{d}\mathcal{E}(\varepsilon) = \int_{-\mu}^{\mu} g(y)\, \mathrm{d}\mathcal{E}(z-y), \tag{56}$$

where we have used the change of variable $y = z - \varepsilon$. Under the conjecture that $g(y) = g(-y)$, one can confirm $g(z) = g(-z)$. To see this, evaluate $g(\cdot)$ at $-z$, use symmetry of $\mathcal{E}(\cdot)$, a change of variable $\tilde{y} = -y$, and standard rules of calculus to obtain

$$\begin{aligned} g(-z) &= \int_{-\mu}^{\mu} g(y)\,\mathrm{d}\mathcal{E}(-z-y) = \int_{-\mu}^{\mu} g(y)\,\mathrm{d}\mathcal{E}(z+y) = -\int_{\mu}^{-\mu} g(-\tilde{y})\,\mathrm{d}\mathcal{E}(z-\tilde{y}) \\ &= \int_{-\mu}^{\mu} g(-y)\,\mathrm{d}\mathcal{E}(z-y). \end{aligned} \tag{57}$$

Now consider the mass at the lower adjustment barrier, $z = -\mu$. This is composed of two parts: first, firms that begin at $-\mu$, draw a negative labor demand shock ($\varepsilon < 0$), and adjust to remain at $-\mu$; and second, firms that began away from $-\mu$ and then migrate there. Thus,

$$g(-\mu) = \mathcal{E}(0)g(-\mu) + \int_{0}^{2\mu} g(-\mu+\varepsilon)\mathrm{d}\mathcal{E}(-\varepsilon) = \frac{1}{\mathcal{E}(0)}\int_{0}^{2\mu} g(-\mu+\varepsilon)\mathrm{d}\mathcal{E}(\varepsilon), \tag{58}$$

where the second equality follows from symmetry of $\mathcal{E}(\cdot)$. A similar argument can be used to show that the mass at the upper adjustment barrier $z = \mu$ satisfies

$$g(\mu) = \frac{1}{\mathcal{E}(0)}\int_{0}^{2\mu} g(\mu-\varepsilon)\mathrm{d}\mathcal{E}(\varepsilon). \tag{59}$$

A conjecture of symmetry $g(-\mu+\varepsilon) = g(\mu-\varepsilon)$ is again confirmed, $g(-\mu) = g(\mu)$. It follows that $g(z) = g(-z)$ for all $z \in [-\mu, \mu]$, and symmetry of $\widetilde{\mathcal{H}}^{\star}(\cdot)$ obtains.

Proof of Proposition 3. The first-order conditions that define the triggers for optimal adjustment $z \in \{1/u\,, 1/l\}$ are given by

$$\begin{aligned} l^{1-\alpha} + \beta D(1/l\,;\theta) &\equiv 1, \\ u^{1-\alpha} + \beta D(1/u\,;\theta) &\equiv 1 + \frac{\gamma}{q(\theta)}, \end{aligned} \tag{60}$$

where $D(z;\theta) \equiv \mathbb{E}\left[\widetilde{\Pi}'\left(e^{-\varepsilon'_{n^\star}} z\right)\right]$. The latter satisfies the following recursion

$$\begin{aligned} D(z;\theta) = \int_{\ln(lz)}^{\ln(uz)} &\left[e^{(1-\alpha)\varepsilon'_{n^\star}} z^{\alpha-1} - 1 + \beta D\left(e^{-\varepsilon'_{n^\star}} z;\theta\right)\right] d\mathcal{E}(\varepsilon'_{n^\star}) \\ &+ \frac{\gamma}{q(\theta)}\left[1 - \mathcal{E}(\ln(uz))\right]. \end{aligned} \tag{61}$$

We first consider a first-order approximation to the firm's optimal policies around $\gamma = 0$.[27] To this end, note first that

[27] Equation (52) has the form $D(z) = \mathcal{C}(D,\gamma)(z)$, where $\mathcal{C}$ is a contraction map on the cross product of the space of bounded and continuous functions (where D "lives") and $[0,\Gamma]$, a closed subinterval of the nonnegative real line from which γ is drawn. By inspection, this map is continuously differentiable with respect to (w.r.t.) $\gamma \in [0,\Gamma]$. It then follows from Lemma 1 of Albrecht, Holmlund, and Lang (1991) that D is continuously differentiable w.r.t. γ and satisfies the recursion, $D_\gamma(z) = \mathcal{C}_\gamma(D,\gamma)(z) + \mathcal{C}_D\big(D_\gamma, D, \gamma\big)(z)$, where $\mathcal{C}_D$ is the Frechet derivative of $\mathcal{C}$. The right side of the latter expression defines a(nother) contraction map on a space of bounded and continuous functions. We have, then, that D_γ is bounded and continuous on $[0,\Gamma]$. Its calculation in (62) follows.

$$
\begin{aligned}
D_\gamma(z;\theta) &\approx \frac{1}{q(\theta)}[1-\mathcal{E}(\ln(uz))] + \beta\int_{\ln(lz)}^{\ln(uz)} D_\gamma\left(e^{-\varepsilon'_{n^\star}}z;\theta\right)d\mathcal{E}(\varepsilon'_{n^\star}) \\
&= \frac{1}{q(\theta)}[1-\mathcal{E}(\ln z)] \text{ when } \gamma = 0.
\end{aligned} \tag{62}
$$

Thus, we can write $D(z;\theta) \approx \gamma[1-\mathcal{E}(\ln z)]/q(\theta)$. Substituting into the first-order conditions and noting that $l = e^{-\lambda}$ and $u = e^{\upsilon}$ yields

$$
\begin{aligned}
e^{-(1-\alpha)\lambda} + \beta\frac{\gamma}{q(\theta)}[1-\mathcal{E}(\lambda)] &\approx 1, \\
e^{(1-\alpha)\upsilon} + \beta\frac{\gamma}{q(\theta)}[1-\mathcal{E}(-\upsilon)] &\approx 1 + \frac{\gamma}{q(\theta)}.
\end{aligned} \tag{63}
$$

Next, linearizing the leading terms around $\lambda = 0$ and $\upsilon = 0$, respectively, leads to

$$
\begin{aligned}
-(1-\alpha)\lambda + \beta\frac{\gamma}{q(\theta)}[1-\mathcal{E}(\lambda)] &\approx 0, \\
(1-\alpha)\upsilon + \beta\frac{\gamma}{q(\theta)}[1-\mathcal{E}(-\upsilon)] &\approx \frac{\gamma}{q(\theta)}.
\end{aligned} \tag{64}
$$

Imposing $\beta \approx 1$, and $\mathcal{E}(-\varepsilon) = 1-\mathcal{E}(\varepsilon)$ yields $\lambda \approx \upsilon$.

Now return to the relationship between $\Delta\widehat{\mathcal{N}}$ and $\Delta\mathcal{N}^\star$ in equation (50). The time-variation in the adjustment triggers alters the approximations around small aggregate shocks. Specifically, with $\lambda \approx \upsilon \approx \mu$, equation (50) becomes

$$
\Delta\widehat{\mathcal{N}} \approx \frac{\Delta\mathcal{N}^\star}{\phi} + \frac{1}{\phi}\left[\Delta\left(\mu\int_{-\infty}^{-\mu}\tilde{\hbar}^\star(z)dz\right) - \Delta\left(\mu\int_{\mu}^{\infty}\tilde{\hbar}^\star(z)dz\right)\right] - \frac{1}{\phi}\Delta\left(\int_{-\mu}^{\mu} z\tilde{\hbar}^\star(z)dz\right). \tag{65}
$$

In order to take a first-order approximation around $\Delta\mathcal{N}^\star = 0$, note that

$$
\begin{aligned}
\left.\frac{\partial\Delta\left(\mu\int_{-\infty}^{-\mu}\tilde{\hbar}^\star(z)dz\right)}{\partial\Delta\mathcal{N}^\star}\right|_{\Delta\mathcal{N}^\star=0} &= -\mu_{-1}\tilde{\hbar}^\star_{-1}(-\mu_{-1}) \\
&\quad + \left\{\tilde{\mathcal{H}}^\star_{-1}(-\mu_{-1}) - \mu_{-1}\tilde{\hbar}^\star_{-1}(-\mu_{-1})\right\}\left.\frac{\partial\mu}{\partial\Delta\mathcal{N}^\star}\right|_{\Delta\mathcal{N}^\star=0}.
\end{aligned} \tag{66}
$$

Similarly,

$$
\begin{aligned}
\left.\frac{\partial\Delta\left(\mu\int_{\mu}^{\infty}\tilde{\hbar}^\star(z)dz\right)}{\partial\Delta\mathcal{N}^\star}\right|_{\Delta\mathcal{N}^\star=0} &= \mu_{-1}\tilde{\hbar}^\star_{-1}(\mu_{-1}) \\
&\quad + \left\{1 - \tilde{\mathcal{H}}^\star_{-1}(\mu_{-1}) - \mu_{-1}\tilde{\hbar}^\star_{-1}(\mu_{-1})\right\}\left.\frac{\partial\mu}{\partial\Delta\mathcal{N}^\star}\right|_{\Delta\mathcal{N}^\star=0}
\end{aligned} \tag{67}
$$

and

$$
\begin{aligned}
\left.\frac{\partial\Delta\left(\int_{-\mu}^{\mu} z\tilde{\hbar}^\star(z)dz\right)}{\partial\Delta\mathcal{N}^\star}\right|_{\Delta\mathcal{N}^\star=0} &= 1 - \phi_{-1} - \mu_{-1}\left[\tilde{\hbar}^\star_{-1}(\mu_{-1}) + \tilde{\hbar}^\star_{-1}(-\mu_{-1})\right] \\
&\quad + \mu_{-1}\left[\tilde{\hbar}^\star_{-1}(\mu_{-1}) - \tilde{\hbar}^\star_{-1}(-\mu_{-1})\right]\left.\frac{\partial\mu}{\partial\Delta\mathcal{N}^\star}\right|_{\Delta\mathcal{N}^\star=0}.
\end{aligned} \tag{68}
$$

Using these and (65) it follows that, to a first-order approximation around $\Delta\mathcal{N}^\star = 0$,

$$\Delta\widehat{\mathcal{N}} \approx \Delta\mathcal{N}^{\star} + \frac{1}{\phi_{-1}}\left\{\widetilde{\mathcal{H}}^{\star}_{-1}(-\mu_{-1}) - \left[1 - \widetilde{\mathcal{H}}^{\star}_{-1}(\mu_{-1})\right]\right\}\left.\frac{\partial\mu}{\partial\Delta\mathcal{N}^{\star}}\right|_{\Delta\mathcal{N}^{\star}=0} \cdot \Delta\mathcal{N}^{\star}. \tag{69}$$

To complete the proof, note from Lemma 2 that symmetry of the adjustment barriers, and of $\mathcal{E}(\cdot)$, implies that $\widetilde{\mathcal{H}}^{\star}(\cdot)$ is also symmetric. It follows that $\widetilde{\mathcal{H}}^{\star}_{-1}(-\mu_{-1}) - \left[1 - \widetilde{\mathcal{H}}^{\star}_{-1}(\mu_{-1})\right] \approx 0$, and (69) collapses to $\Delta\widehat{\mathcal{N}} \approx \Delta\mathcal{N}^{\star}$. The result then follows from equation (53).

Numerical model. Our numerical results are derived from a model that augments the firm's problem (16) with a time-invariant per-worker hiring cost k as follows:

$$\begin{aligned} &\Pi(n_{-1}, x) \equiv \max_{n}\left\{Apxn^{\alpha} - (1-\eta)\omega n - \left(\frac{c}{q(\theta)} + k\right)\Delta n^{+} + \beta\mathbb{E}[\Pi(n, x')|x]\right\}, \\ &\text{where } A \equiv \frac{1-\eta}{1-\eta(1-\alpha)}, \end{aligned} \tag{70}$$

and the workers' threat point ω is given by (19). The model is solved at a biweekly frequency as a means to approximate the continuous time nature of labor market flows.

We now describe how we set values of the structural parameters not described in section 1.5. Three of the remaining parameters are the size of the labor force L, the standard deviation of idiosyncratic productivity draws σ_x, and the flow payoff from unemployment b. In any steady state, total hires equal the outflows from unemployment. Hence, the labor force, for a given level of hires, determines the outflow rate f, which we target to equal its empirical counterpart of 0.232 at a biweekly frequency. (This is calculated in the data based on the method of Shimer (2005) for the period 1951 to 2015.) To replicate an average unemployment rate u of 6 percent, we set the inflow rate into unemployment at 0.0145 per fortnight by adjusting σ_x. Finally, conditional on an unemployment rate, we can set b to ensure an average establishment size in the range of 17-21 employees, consistent with Census Bureau data on average establishment and firm size.

To map the job finding rate f to labor market tightness θ, we assume a conventional Cobb-Douglas matching function that implies the job-finding rate $f = m\theta^{\varphi}$ where m denotes matching efficiency. Using data for the period 1951 to 2015, and the methods of Shimer (2005), we estimate a matching elasticity $\varphi = 1/3$. We then choose matching efficiency $m = 0.29$ so that the observed mean of tightness is consistent with a job-finding rate of $f = 0.232$ (as in Pissarides, 2009). This yields $\theta = 0.53$ and a vacancy rate of 3.2 percent. The values of these additional parameters underlying the resulting calibrations are summarized in Table 1.

Table 1. Parameters by inaction rate underlying numerical search model

Parameter	Meaning	Values by inaction rate			Reason
		52.5%	67%	80%	
η	Worker bargain power	0.063	0.026	0.009	Real wage as procyclical as employment
b	Unemployment payoff	0.22	0.21	0.19	Average employer size
L	Labor force	19.15	19.83	19.57	Mean job-finding rate
σ_x	Std. of x innovations	0.28	0.47	0.83	Mean unemployment inflow rate
k	Hiring cost	0.72	2.42	7.20	Inaction rate

D. Sensitivity analyses

This appendix describes our parameterization of the process for idiosyncratic shocks, contrasts it with alternatives in the literature, and performs sensitivity analyses on our main results to reasonable changes in the parameters of the shock process.

Idiosyncratic shock process. Recall that we specify the process of idiosyncratic productivity as a geometric AR(1),

$$\ln x_{t+1} = \rho_x \ln x_t + \varepsilon_{xt+1}, \tag{71}$$

where $\varepsilon_{xt+1} \sim N(0, \sigma_x^2)$ and t denotes *quarters*. Note that $\ln x$ is implicitly de-meaned here. In what follows, we infer ρ_x and σ_x from estimates of the persistence and volatility of idiosyncratic productivity in data of *different* frequencies.

Low(er) frequency data. Our baseline parameterization is informed in large part by Abraham and White (2006), who measure plant-level total factor productivity (TFP) using annual microdata for the U.S. manufacturing sector. We interpret their measurement of annual log TFP as being $X_s \equiv \sum_{t=1}^{4} \ln x_{st}$, where $\ln x_{st}$ is the realization in quarter t of year s. In other words, the measurement of annual log TFP reflects the accrual of log points (in excess of mean) throughout the four quarters of the year.

Abraham and White run least squares regressions of X_s on $X_{s-1} \equiv \sum_{t=1}^{4} \ln x_{s-1t}$. An average of estimated slope coefficients from their weighted and unweighted regressions is

$$\frac{cov(X_s, X_{s-1})}{var(X_{s-1})} = 0.39. \tag{72}$$

Assuming that the annual data are generated from the quarterly process (71), one can relate the covariance and variance terms in (72) to ρ_x and σ_x as follows:

$$cov(X_s, X_{s-1}) = [\rho_x + 2\rho_x^2 + 3\rho_x^3 + 4\rho_x^4 + 3\rho_x^5 + 2\rho_x^6 + \rho_x^7] \frac{\sigma_x^2}{1 - \rho_x^2}, \tag{73}$$

and,

$$var(X_{s-1}) = [4 + 6\rho_x + 4\rho_x^2 + 2\rho_x^3]\frac{\sigma_x^2}{1-\rho_x^2}. \tag{74}$$

Solving for the implied quarterly persistence yields $\rho_x = 0.68$. To infer σ_x, we use Abraham and White's estimate of $var(X_{s-1}) = 0.21$, which in turn implies that $\sigma_x = 0.1034$.

Foster et al. (2008) also provide estimates of the productivity process based on plant-level measurements of TFP. Unlike Abraham and White, they restrict their sample to several industries for which real output can be most credibly measured. This mitigates measurement error at the expense of a more representative panel. In addition, Foster et al.'s data are not annual but are drawn instead from the quinquennial Census of Manufacturers. Nonetheless, we can convert their estimates of the persistence and volatility of productivity to a quarterly frequency by following the same steps above. We find that $\rho_x = 0.943$ and $\sigma_x = 0.022$.[28]

Quarterly frequency. Cooper, Haltiwanger, and Willis (2015) estimate a dynamic labor demand model using quarterly data. They specify a geometric AR(1) for idiosyncratic productivity, as in (71), and parameterize the process in order to replicate certain moments concerning labor inputs. They estimate a persistence parameter of $\rho_x = 0.4$ and an innovation standard deviation of $\sigma_x = 0.5$.

Monthly frequency. Cooper, Haltiwanger, and Willis (2007) estimate a search and matching model using monthly data. Accordingly, they specify an idiosyncratic productivity process that follows a geometric AR(1) at a *monthly* frequency. Their estimates of the monthly analogues to ρ_x and σ_x are $\rho_x^m = 0.395$ and $\sigma_x^m = 0.0449$.[29]

Now taking their monthly AR(1) as the underlying driving force, we can calculate the implied persistence and volatility of quarterly data. Specifically, letting x_{mt} denote the realization in month $m \in \{1,2,3\}$ of quarter t, we compute the persistence and volatility of $\sum_{m=1}^{3} \ln x_{mt}$. This leads to

$$\rho_x = \frac{\rho_x^m + 2(\rho_x^m)^2 + 3(\rho_x^m)^3 + 2(\rho_x^m)^4 + (\rho_x^m)^5}{3 + 4\rho_x^m + 2(\rho_x^m)^2} = 0.194, \tag{75}$$

and

$$\sigma_x^2 = (1-\rho_x^2)\frac{3 + 4\rho_x^m + 2(\rho_x^m)^2}{1-(\rho_x^m)^2}(\sigma_x^m)^2 = 0.251. \tag{76}$$

This implies a standard deviation $\sigma_x = 0.501$.

Robustness of results. We now probe the robustness of our results to variation in the parameterization of the productivity process. To conserve space, we look at two models:

[28] These estimates correspond to their measure of "physical TFP".

[29] These are the average of Cooper et al.'s results for two specifications of adjustment costs that seem most germane to our application: (i) a fixed cost to adjust and linear cost to hire, and (ii) a fixed cost to adjust and linear cost to fire.

One includes a fixed cost of adjusting, and the other includes symmetric per-worker costs to hire and fire.

We conduct four simulations of each model. In one, $\rho_x = 0.9$, which places it in the neighborhood of the estimate in Foster et al. In another, $\rho_x = 0.3$, which is the midpoint between the two estimates in Cooper et al. For both of these simulations, σ_x is set to its baseline value of 0.15. In the third and fourth simulations, ρ_x is returned to its baseline value (of 0.7), and σ_x is varied. In the third, the baseline value of σ_x is halved, which reduces it in the direction of Foster et al. In the fourth, σ_x is doubled, which raises it in the direction of Cooper et al. (For results for still larger σ_xs, see the simulations of the search model in Appendix C.) In *every* parameterized version of each model, the cost of adjusting is set to induce an inaction rate of 67 percent, which is in the middle of the range we consider in the main text.

Figures B and C in this Appendix present the results. Figure B illustrates the impulse responses from models with a fixed cost of adjusting. The top panel varies ρ_x, whereas the bottom panel varies σ_x. These results are difficult to distinguish from the baseline case presented in the main text (Figure 2.B), with one exception, namely, the case where $\rho_x = 0.3$. The rise in flow-balance employment on impact is smaller when ρ_x is smaller, though still substantially larger than that of its frictionless counterpart.

Figure C presents results from models with a (symmetric) linear cost of adjusting. Again, flow-balance employment rises slightly less in the case where ρ_x is small. But, the impulse responses are very similar to what we present in the main text (Figure 3.B).

E. Details of numerical methods

We detail how we solve the labor market equilibrium models in the main text. As in many other applications of heterogeneous agent models, our state space is, in principle, infinite dimensional. The reason for this is that firms forecast the future wage to make their labor demand decision. The future wage is, in turn, jointly determined with aggregate employment, which is a function of the full distribution of firm size. We follow Krusell and Smith's (1998) bounded rationality approach to prune the state space. Specifically, we find that firms are able to make a very accurate projection of aggregate employment one period ahead, N' , based on only knowledge of the current mean of the firm-size distribution, N, as well as current productivity p,

$$\ln N' = \nu_0 + \nu_N \ln N + \nu_p \ln p. \tag{77}$$

This is the forecast rule for aggregate employment that firms use in all of our models in section 1.

In the search model, firms have to forecast one more price, namely, market tightness, θ. This, again, is a function of the distribution of firm size as well as productivity. Fortunately, we find that variation in tightness can be almost entirely accounted for (in

a statistical sense) by variation in mean firm size, N, and p. Therefore, we assume that firms anticipate that tightness obeys the relation

$$\ln \theta = \Theta_0 + \Theta_N \ln N + \Theta_p \ln p. \tag{78}$$

The coefficients in (77) and (78) solve a fixed-point problem. Our general approach is to, first, conjecture the coefficients and solve the firms' optimal labor demand policies. We then simulate the decision rules for 250,000 firms over 200 quarters (the latter excludes the "burn-in" time).[30] This yields a time series for N, and we run the regression (77). If the implied coefficients differ from our initial conjecture, we update and repeat. We implement this routine for the models in sections 1.3 and 1.4.

Our approach to the search model (section 1.5) differs in two respects. First, time aggregation is a more acute problem in the search model because of the worker flows. As noted in the main text, we solve and simulate the model at a biweekly frequency. Second, the size of the idiosyncratic shocks (σ_x) is much larger. As a result, the law of large numbers does not "kick in" even if we use a very large number of firms, which means the simulated paths of N and θ are noisy. We thus switch to using Young's (2010) nonstochastic simulator, which constructs the joint distribution of n and x in order to aggregate up to N. This yields smoother time series. For this simulator, we need at least 500 grid points in the productivity (x) dimension.

Table 2 summarizes the estimates of (77) and (78). We make two observations. First, the pattern in the coefficient estimates is reasonably intuitive. For instance, when adjustment frictions are higher, the coefficient ν_p on $\ln p$ in (77) is typically lower and the coefficient ν_N on lagged employment is typically higher. Second, the goodness of fit, as summarized by the R^2, is especially excellent in the models of fixed and time-invariant linear costs. As den Haan (2010) has noted, though, R^2 measures only the quality of the one-step-ahead forecast. But, one may use the estimates in the table to verify, as we have, that the entire impulse response of aggregate employment implied by (77) also closely matches the impulse response simulated from the heterogeneous-agent models and shown in Figures 2 through 4. The fit of the regressions in the search model is slightly worse, and as a result, there is a slight bit of "daylight" between the impulse responses simulated from the model and those implied directly from the forecast equations. For instance, for the baseline parameterization (of 52.5% inaction), the model-generated IRF of aggregate employment peaks after 15 fortnights at 0.446%, whereas the IRF implied by (77) peaks after 18 fortnights at 0.441%.

The last item to address is the simulation of the impulse response of flow-balance employment, $\widehat{N}$. As noted in the main text, we track the migration of firms into and out of each integer-valued employment level, n, to calculate the two ingredients of the flow-balance density, namely, the share of firms that flow into each n and the probability of

[30] Note that (77) and (78) are not perfect forecasts, and thus we are required to solve for a fixed point in N (and θ, where applicable) at every period of the simulation.

outflow from each n. This mimics what we do in the data, with one slight exception. When we work with the data, we noted that, to reduce noise in our estimates of the flows at high levels of employment, we compute the flows within bins that pool together a range of employer sizes (i.e., 501-510). We do not have to implement these bins in our model-generated data. Bins are necessary only if we aim to replicate the share of very large establishments (i.e., with over 500 workers), which can be done if we introduce a distribution of time-invariant productivity into the model. This is computationally cumbersome, but we have done it for the baseline parameterization, in which inaction is 52.5 percent per quarter. The impulse responses are virtually identical to what we presented in the main text, where we omit a fixed productivity distribution.

Table 2. Estimated parameters of Krusell-Smith forecast equations

Adjustment cost	Inaction rate	Employment regression				Tightness regression			
		ν_0	ν_N	ν_p	R^2	Θ_0	Θ_N	Θ_p	R^2
Fixed	52.5%	2.951	0.013	0.417	0.9999	—	—	—	—
	67%	2.875	0.036	0.406	0.9999	—	—	—	—
	80%	2.711	0.087	0.383	0.9998	—	—	—	—
Linear	52.5%	2.411	0.174	0.350	0.9999	—	—	—	—
	67%	2.007	0.308	0.289	0.9999	—	—	—	—
	80%	1.431	0.505	0.200	0.9998	—	—	—	—
Search	52.5%	0.441	0.847	0.085	0.9990	-48.31	16.480	12.181	0.9991
	67%	0.396	0.864	0.071	0.9993	-65.78	22.258	13.056	0.9996
	80%	0.306	0.895	0.056	0.9996	-92.15	31.403	12.404	0.9994

References

Abraham, Arpad, and Kirk White. 2006. "The Dynamics of Plant-level Productivity in U.S. Manufacturing." Center for Economic Studies WP 06-20.

Albrecht, James W., Bertil Holmlund, and Harald Lang. 1991. "Comparative Statics in Dynamic Programming Models with an Application to Job Search." *Journal of Economic Dynamics and Control* 15(4): 755-769.

Caballero, Ricardo, and Eduardo M. R. A. Engel. 1999. "Explaining Investment Dynamics in U.S. Manufacturing: A Generalized *(S,s)* Approach." *Econometrica* 67(4): 783-826.

Cooper, Russell, John Haltiwanger, and Jonathan L. Willis. 2007. "Search Frictions: Matching Aggregate and Establishment Observations." *Journal of Monetary Economics* 54(1): 56-78.

Cooper, Russell, John Haltiwanger, and Jonathan L. Willis. 2015. "Dynamics of Labor Demand: Evidence from Plant-level Observations and Aggregate Implications." *Research in Economics* 6(1): 37-50.

Den Haan, Wouter J. 2010. "Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents." *Journal of Economic Dynamics and Control* 34(1): 79-99.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review* 98(1): 394-425.

Krusell, Per, and Anthony Smith. 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106(5): 867-896.

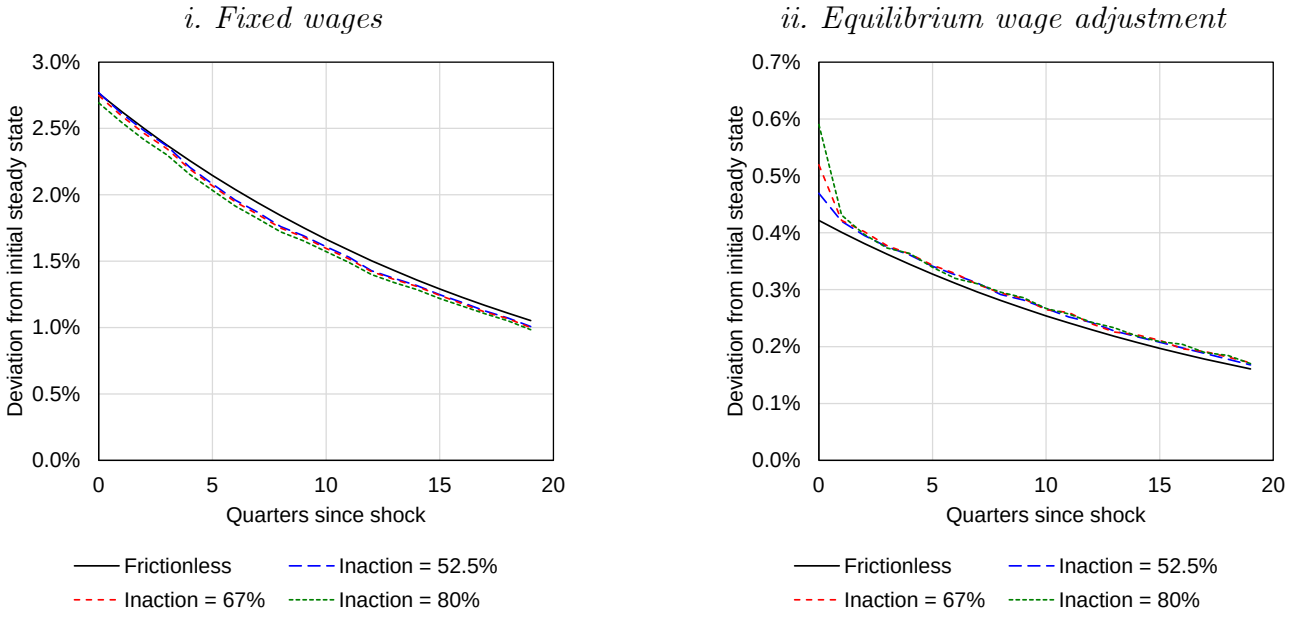
Pissarides, Christopher A. 2009. "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica* 77: 1339-1369.

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

Young, Eric R. 2010. "Solving the Incomplete Markets Model with Aggregate Uncertainty Using the Krusell–Smith Algorithm and Non-stochastic Simulations." *Journal of Economic Dynamics and Control* 34(1): 36-41.

Additional figures

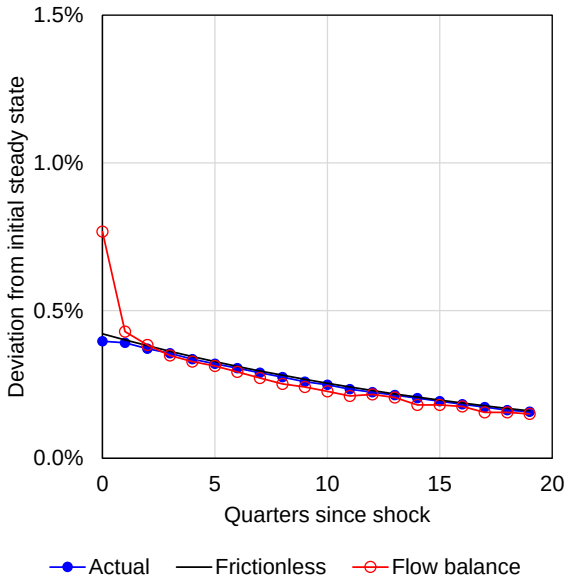
Figure A. Impulse responses of mandated and frictionless employment: Fixed costs



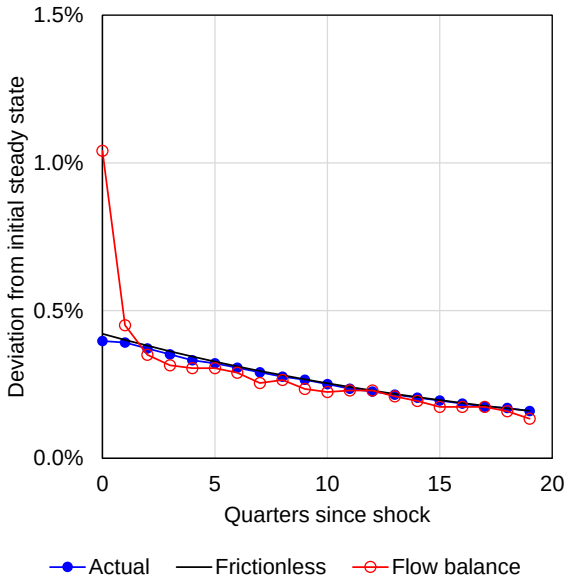
Notes: Each panel plots the impulse response of aggregate frictionless employment in contrast to the impulse responses of mandated employment for each stated average quarterly inaction rate. These results can be compared to the impulse responses in Figure 2.

Figure B. Impulse responses of aggregate employment under alternative calibrations: Fixed costs

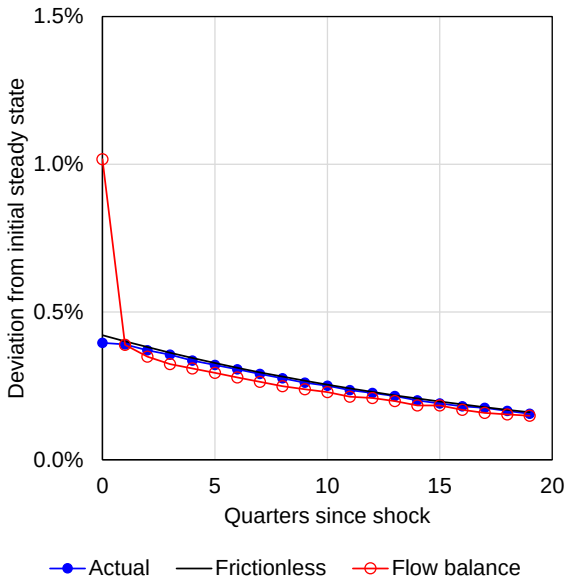
i. $\rho_x = 0.3$



ii. $\rho_x = 0.9$



iii. $\sigma_x = 0.075$



iv. $\sigma_x = 0.3$

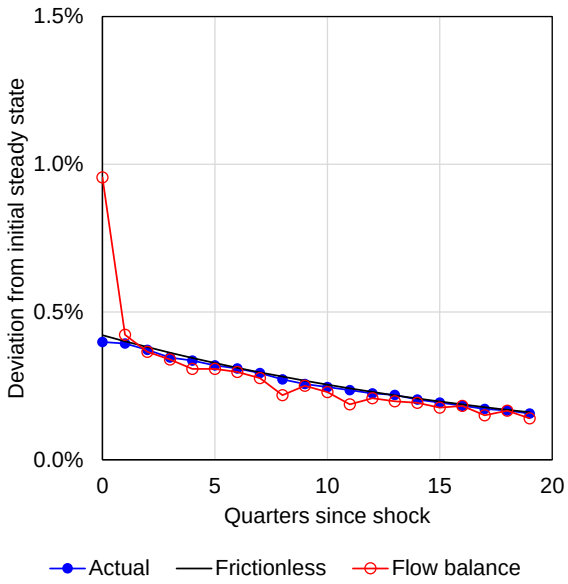
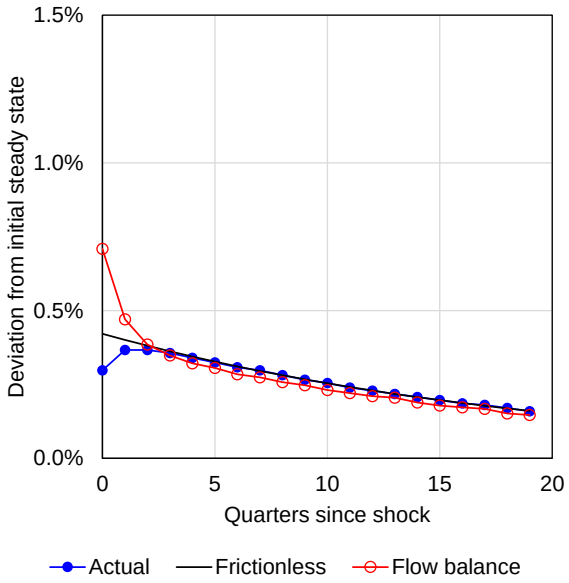
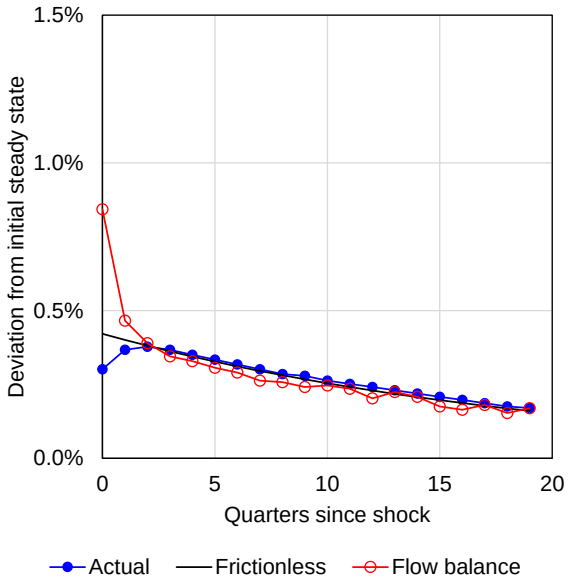


Figure C. Impulse responses of aggregate employment under alternative calibrations: Linear costs

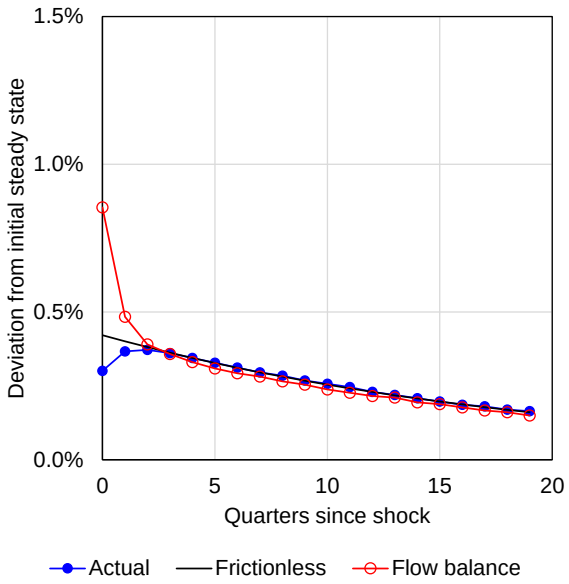
i. $\rho_x = 0.3$



ii. $\rho_x = 0.9$



iii. $\sigma_x = 0.075$



iv. $\sigma_x = 0.3$