The Timing of Labor Market Expansions: New Facts and a New Hypothesis *

Giuseppe Moscarini†
Yale University
and
NBER

Fabien Postel-Vinay‡
University of Bristol
and
Université de Paris I
(Panthéon-Sorbonne)

September 2007 — IN PROGRESS

Abstract

We document three new facts about aggregate dynamics in US labor markets over the last 15 years, drawing in part from newly available datasets. These facts suggest a new view of how business cycles evolve and mature. We investigate whether this view is consistent with the transitional dynamics of the Burdett and Mortensen (1998) equilibrium search model, that we analyze in detail.

Keywords:
JEL codes:

*We would like to thank David Talan and Charles Carson from the Bureau of Labor Statistics for providing unpublished data on the firm size distribution in the BED data set. We are also grateful to seminar audiences at Bristol, Essex and the NBER Summer Institute for very useful feedback. The usual disclaimer applies.

†Address: Department of Economics, Yale University. E-mail `giuseppe.moscarini@yale.edu`.

‡Address: Department of Economics, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK. Tel: +44 117 928 8431. E-mail `Fabien.Postel-Vinay@bristol.ac.uk`. Internet `www.efm.bris.ac.uk/economics/staff/vinay/`. Postel-Vinay is also affiliated with CEPR (London), CMPO (Bristol), IZA (Bonn) and PSE (Paris).

1 Introduction

The cyclical behavior of (un)employment and wages still poses a formidable challenge to macroeconomists. No other aspect of the business cycle has been as widely studied and remains as poorly understood. While a consensus has slowly emerged on the importance of search frictions to explain equilibrium unemployment and residual wage inequality, how wages are set in practice is still followed by a big question mark.

The search-and-matching business cycle literature makes a common assumption of wage bargaining. Even abstracting from its poor quantitative performance, this solution remains a black box, silent about the source of workers' bargaining power. When labor demand rises, driven by higher productivity or real demand for goods, how is it exactly that workers obtain higher wages? And why do wages and labor productivity rise gradually and slowly over business cycle expansions?

The search literature on wage inequality, invariably cast in steady state and abstracting from aggregate fluctuations, typically assumes that firms have full monopsony power and make take-it-or-leave-it offers of employment contracts. To reconcile this intuitive and plausible assumption with the reality of worker rents and wage inequality, and to avoid the Diamond (1971) paradox, Burdett and Mortensen (1998) [BM] identify the source of workers' bargaining power in a form of moral hazard. Specifically, workers obtain more than their reservation wage in order to be induced to quit their previous employer and, once hired, to decline future outside offers. Poaching is the engine of wage growth and differentiation for individual workers.

In this paper we argue that the same poaching mechanism transmits aggregate shocks to wages and employment. Firms decide to pay more when they run out of unemployed job applicants and are forced to steal employees from their competitors. Our argument builds on three new facts about aggregate dynamics in US labor markets over the last 15 years, that we document by drawing in part from newly available datasets. These facts suggest a new view of how business cycles evolve and mature. We investigate whether this view is consistent with the transitional dynamics of the BM equilibrium search model, that we analyze in detail.

The facts are readily summarized. First, the Business Employment Dynamics (BED) data set provides annual firm-level data on job flows and firm sizes since 1992. The BED series show that small firms (in terms of employment) concentrated most of their job creation in the early part of the 1990's expansion, and promptly expanded their employment after 2001. Conversely, large firms concentrated most of their 1990's job creation after 1996, and again failed to create jobs in the first part of the 2000's expansion. This pattern is observed across nine firm size classes and is exemplified in Figure 1 which plots employment shares for four different classes.[1] The recoveries of the early 1990's and 2000's were "jobless" mainly at large firms, while the strong job creation of the late 1990's, in the mature phase of the expansion, was concentrated mainly in large firms.

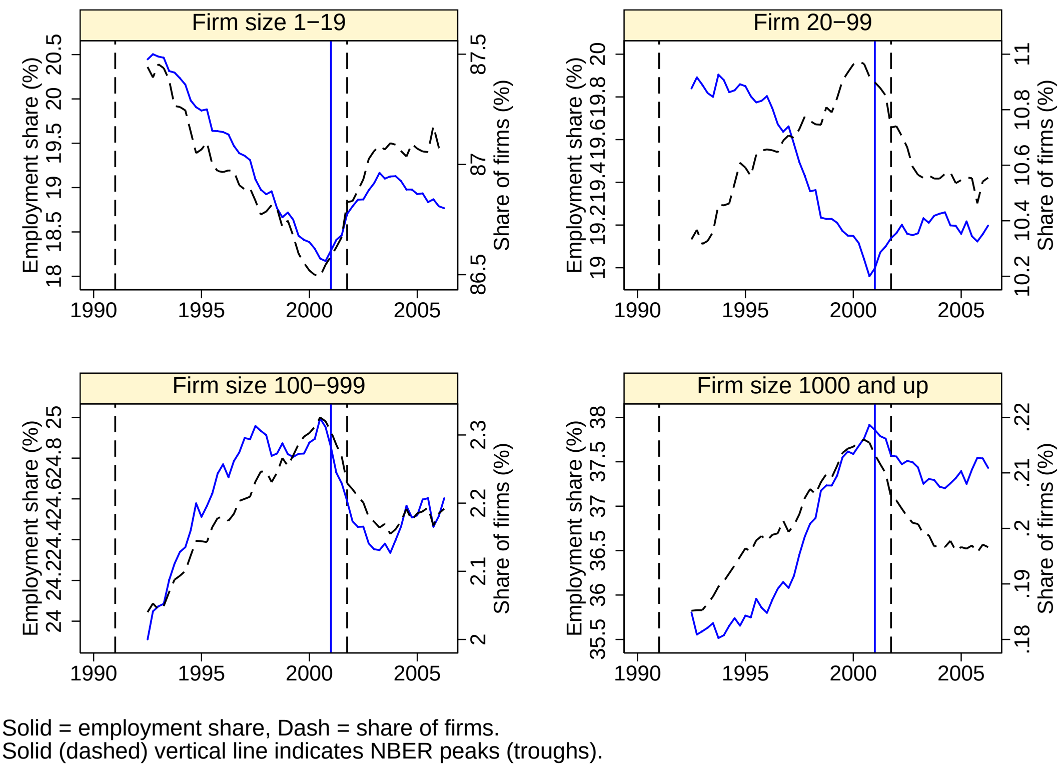
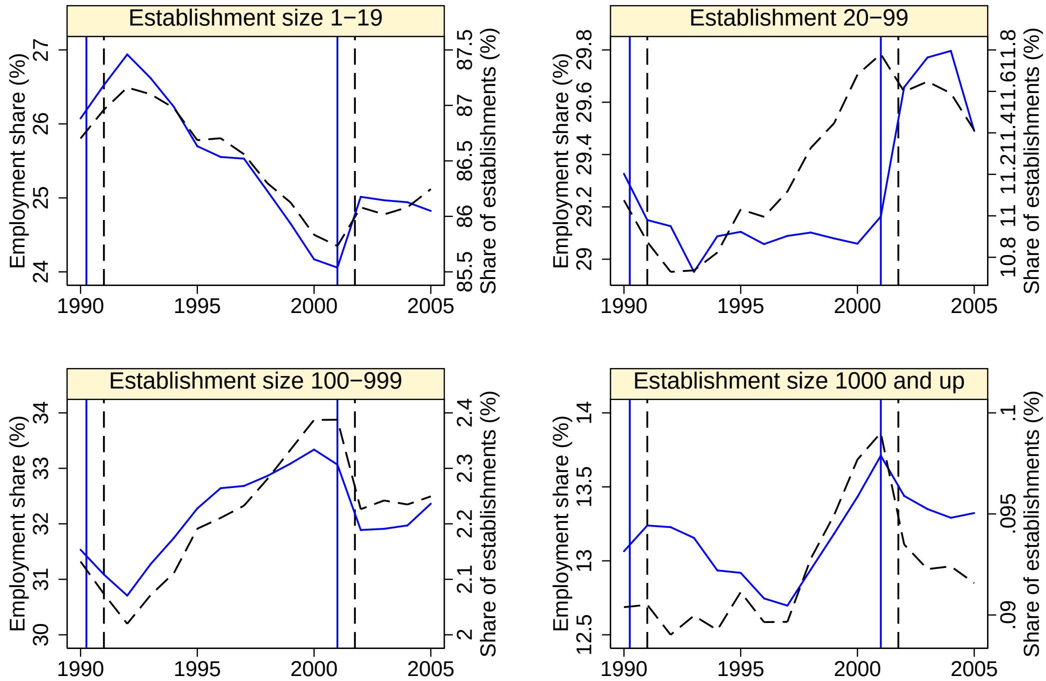


Fig. 1: Fraction of firms and employment shares — small *vs.* large firms

A similar graph can be constructed from establishment-level data using the County Business Pattern (CBP) data set (see Figure 2 — the CBP series starts in 1990). The pattern of establishment size dynamics over the last two business cycles closely resembles that of firm size dynamics. Part of this resemblance is due to the fact that most (small) firms are mono-establishment. More generally,

[1] On all figures, vertical lines are placed at NBER business cycle dates.



Solid = employment share, Dash = share of establishments.
Solid (dashed) vertical line indicates NBER peaks (troughs).

Fig. 2: Fraction of firms and employment shares — small *vs.* large establishments

however, large establishments tend to be part of large firms, as shown in Table 1.[2]

Second, monthly CPS data available since 1994 and compiled by Moscarini and Thomsson (2007) show that the employer-to-employer (EE) transition rate was actually falling in the first half of the 1990s, picked up late in that expansion, and again declined in the 2001 recession and thereafter, only recently showing signs of recovery (Figure 3).

Third, BLS public data on average real (weekly and hourly) earnings show a flat profile in the first part of both the 1990s and 2000s expansions, a sharp increase in 1997-1999 and (possibly) since the Fall of 2005 (Figure 4). The bigger picture exhibits similar patterns for the preceding five business cycles, with the notable exception of the 1980s expansion, over the second half of which wages dropped rather than peaked as they did in all other expansions over the past forty years.

[2]This pattern does not seem specific to US firms or establishments. Delli Gatti et al. (2004) find in a very large sample of Italian firms that the distribution of employment size becomes much less concentrated among large firms in the aftermath of the 1992 recession, and then regains concentration over the ensuing expansion. This fact is consistent with small firms accounting for a larger share of employment early in an expansion, as this corresponds to a drop in concentration.

Firm size category	Average number of establishments	Mean establishment size
all	1.255	15.577
0	1.002	0.000
1-4	1.002	2.101
5-9	1.012	6.490
10-19	1.054	12.751
20-99	1.316	29.801
100-499	3.819	50.712
500 and up	61.975	53.458

CBP data, 2004, and authors' calculations.

Table 1: Establishment and firm sizes

Finally, a breakdown of the above mean wage series by establishment size categories can be obtained from the CBP data. This is reported on Figure 5 (once again the CBP data only starts in 1990). This Figure is remarkable in at least three respects. First, mean wages are monotonically increasing in establishment size at all dates. This is another rendition of the well-documented "firm size-wage effect". Second, the pattern highlited at the aggregate level on Figure 4 holds roughly unchanged for all establishment size categories. Third, all wage profiles plotted on Figure 5 are nearly parallel. In other words, the distribution of mean wages by establishment size shows no clear sign of collapsing or fanning out over time.

All in all, EE rates, wages and employment in large firms appeared to comove: sluggish early in the last two expansions and brisk in the late stages of the 1990's expansion (and possibly of the current one, since late 2005).

Although they only pertain to the last two expansions, and as such are not enough to establish an empirical regularity, these facts suggest the following pattern. Early in an expansion, the large pool of unemployed workers sustains firms' monopsony power. Wages remain low, firms hire mostly from unemployment, relatively few workers quit from job to job. As the reservoir of unemployment dries out, more and more of the new hires arrive from other jobs. As poaching becomes the main source of hiring, average wages and earnings rise and the EE rate picks up. If workers quit mostly

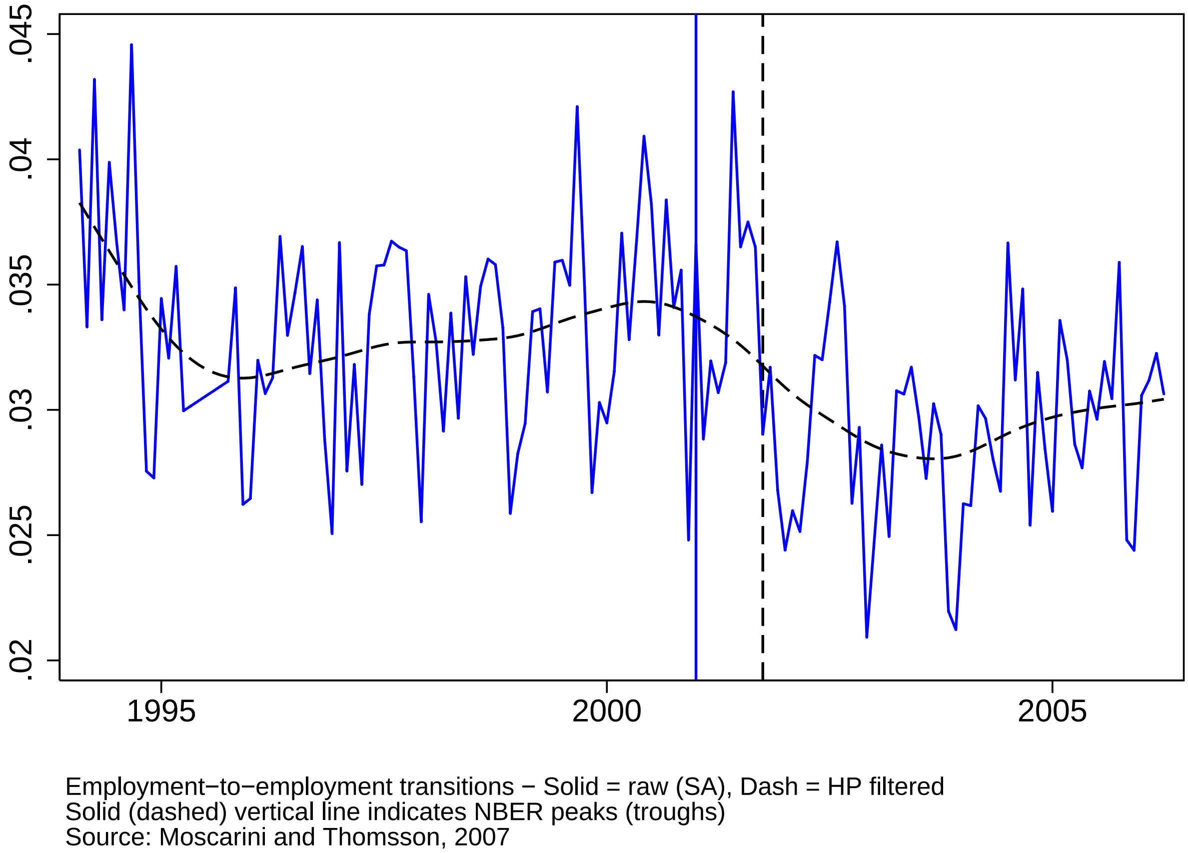


Fig. 3: Job-to-job quit rates

from small, low-paying firms to large, high-paying firms, the growth in the employment of large firms will be fuelled by the stock of employment at small firms, which takes some time to replenish after a recession. Hence, employment at small firms rises faster and peaks earlier than at large firms. The erosion in firms' monopsony power reduces average mark-ups several years into an expansion, potentially creating favorable conditions for a new recession.

While this description of labor market dynamics might appear plausible and intuitive, it remains to verify whether in fact it can be consistent with equilibrium behavior. To this purpose, we study the transitional dynamics of the BM wage posting model with heterogeneous firms. We assume that the economy is hit by an unanticipated, positive aggregate productivity shock, and we study the convergence to the new steady state. Firms post and commit to wage paths that depend only

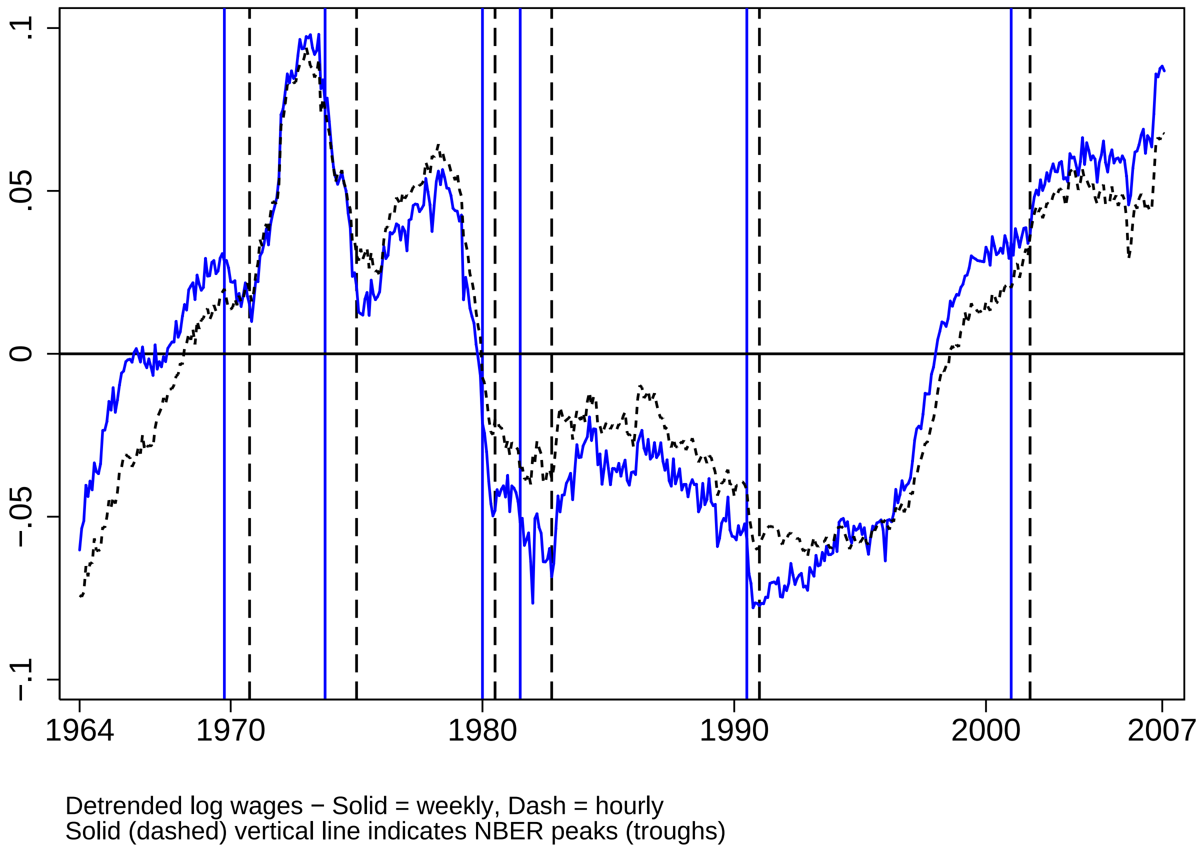


Fig. 4: Average wages

on calendar time (or, equivalently, on the unemployment rate).[3] We focus on Rank-Preserving Equilibria (RPE) of the wage-posting game, where at each point in time more productive firms always offer higher wages, so the rank of each firm in the wage offer distribution is invariant. Forward-looking workers make transitions based on the intertemporal value offered by each job. Forward-looking firms maximize the present discounted value of their profits. We obtain a system of partial differential equations in time and firm productivity that completely characterize RPE dynamics and that can be solved numerically.

We find that a baseline calibration of the model is qualitatively consistent with the main facts that we uncover. Firms backload wages both in order to piggyback on future poachers, who will

[3] To our knowledge, Shimer (2003) is the only attempt to analyze aggregate dynamics in a wage-posting search model. He considers a dynamic extension of the BM model where homogeneous firms can only commit to constant wage profiles also out of steady state. Because our analysis is motivated as much by the new evidence that we present as by the model per se, we assume that heterogeneous firms post time-dependent wage contracts.

Fig. 5: Wages by establishment size

sometimes deliver the promised raises, and to keep less productive competitors at bay by matching their own increasing (in calendar time) wage contracts. Small firms expand first and soon run out of job applicants. Large firms can keep hiring by poaching workers from small firms, although these are offered higher wages than the early hires, the majority of which arrive from unemployment. As a consequence, the employment share of small firms first rises and then declines, and the EE quit rate with it. The gradual selection of workers into larger, more productive and higher-paying firms contributes to generate a slow and gradual growth in productivity and wages.

The rest of the paper discusses the theoretical model and its quantitative predictions. An outline follows. Section 2 describes the basic economic environment. Section 3 characterizes the dynamic labor market equilibrium and explains our solution strategy. Details and results of a simple calibration exercise are presented and discussed in Section 4. Finally, Section 5 concludes and gives some thoughts about possible further research avenues.

2 The Economy

Time is continuous. The labor market is populated by a unit-mass of workers who can be either employed or unemployed. It is affected by search frictions in that unemployed workers can only sample job offers sequentially at some finite Poisson rate $\lambda_0 > 0$. Employed workers are allowed to search on the job, and face a sampling rate of job offers of $\lambda_1 > 0$. Firm-worker matches are dissolved at rate $\delta > 0$. Upon match dissolution, the worker becomes unemployed. All workers are ex-ante identical: they are infinitely lived, risk-neutral, equally capable at any job, and they attach a common lifetime value of U_t to being unemployed at date t.

Workers face a measure N of active firms operating constant-return technologies with heterogeneous productivity levels $p \sim \Gamma(\cdot)$ among firms. For (quantitative) reasons that will become clear below, we assume that the sampling of firms by workers is not uniform in that a type-p firm has a sampling weight of $v(p) > 0$. Sampling weights are normalized in such a way that their cumulated sum $\Phi(p) = \int_{\underline{p}}^{p} v(x)\, d\Gamma(x)$ is a (sampling) cdf, i.e. $\Phi(\overline{p}) = 1$. The sampling density of a type-p firm is therefore $v(p)\gamma(p)$. This naturally encompasses the conventional case of uniform sampling which has $v(p) = 1$ for all p. As we shall see later in the analysis, however, a plausible calibration requires that $v(p)$ be increasing in p. This assumption can either be thought of as reflecting the greater visibility of large firms causing workers to apply unsolicited more often to large firms. Alternatively, it can be viewed as a shortcut for directed search: if search has any element of directness, people will apply more to high paying firms (which higher-p firms will turn out to be in equilibrium).

At some initial date which we normalize at $t_0 = 0$, each firm of a given type p commits to a wage profile $\{w_t(p)\}_{t\in[0,+\infty)}$ over the infinite future. We generalize the BM restrictions placed on the set of feasible wage contracts to a non-steady-state environment by preventing firms from making wages contingent on anything else than calendar time.[4] We thus rule out, among other things, wage-

[4]Or, less stringently, we allow firms to index wages to any aggregate variable that evolves monotonically over time (e.g. the unemployment rate).

tenure contracts (Stevens, 2004; Burdett and Coles, 2003), offer-matching or individual bargaining (Postel-Vinay and Robin, 2002; Dey and Flinn, 2005; Cahuc, Postel-Vinay and Robin, 2006), or contracts conditioned on employment status (Carrillo-Tudela, 2007).[5]

Any such profile $\{w_t(p)\}_{t\in[0,+\infty)}$ offered by any type-p firm yields a continuation value of $V_t(p)$ to any worker employed at that firm at any date t. The (time-varying) sampling distribution of job values is denoted as $F_t(\cdot)$, and its relationship to the sampling distribution of firm types $\Phi(\cdot)$ will be discussed momentarily. Because from the workers' viewpoint jobs are identical in all dimensions but the wage profile, employed jobseekers quit into higher-valued jobs only.[6] This gradual self-selection of workers into better jobs implies that the distribution of job values in a cross-section of workers—which will be denoted as $G_t(\cdot)$—differs from the sampling distribution $F_t(\cdot)$.

3 Equilibrium

3.1 The Contract Posting Problem

Firms post wage profiles over an infinite horizon that solve the following problem:

$$\Pi_0^\star(p) = \max_{\{w_t\}} \int_0^{+\infty} (p - w_t)\,\ell_t(p)\,e^{-rt}ds \tag{1}$$

subject to:

$$\dot{\ell}_t(p) = -\left(\delta + \lambda_1 \overline{F}_t(V_t(p))\right)\ell_t(p) + \frac{v(p)}{N}\left(\lambda_0 u_t + \lambda_1(1-u_t)\,G_t(V_t(p))\right) \tag{2}$$

$$\dot{V}_t(p) = \left(\rho + \delta + \lambda_1 \overline{F}_t(V_t(p))\right)V_t(p) - \lambda_1 \int_{V_t(p)}^{+\infty} x\,dF_t(x) - w_t - \delta U_t \tag{3}$$

$$w_t \geq \underline{w}, \tag{4}$$

where $\ell_t(p)$ denotes a type-p firm's workforce at date t,[7] $\underline{w}$ is the exogenous institutional minimum wage, r (ρ) is the firms' (workers') discount rate,[8] and $\overline{F}_t(\cdot) = 1 - F_t(\cdot)$ designates the survivor

[5]Note, however, that the model can be generalized to allow for time-varying individual heterogeneity under the assumption that firms offer the type of piece-rate contracts described in Barlevy (2005). In that sense experience and/or tenure effects can be introduced into the model.

[6]For simplicity, it is also assumed that any job offer posted in equilibrium is preferred to unemployment, i.e. $\inf_p V_t(p) \geq U_t$ at all t. Specifically, we assume the existence of a constant and exogenous institutional minimum wage $\underline{w}$ which is sufficiently high for unemployed workers to accept even the least valuable job offer.

[7]Incidentally, this implies that the *density* of firm types among workers at date t is given by $N\ell_t(p)\,\gamma(p)\,/\,(1-u_t)$.

[8]Although in most of what follows we will comply with standard practice and impose a common discount rate on firms and workers (i.e. assume $r = \rho$), this restriction is by no means essential. Indeed other cases, such as the case of myopic workers for example, are of potential interest (see below). We therefore begin by stating the general problem free of any assumption on relative discount rates.

function associated with $F_t(\cdot)$. When solving (1), the typical firm also is also constrained by its given initial size $\ell_0(p)$.

At the individual firm's level, the sampling and cross-sectional distributions of job values $F_t(\cdot)$ and $G_t(\cdot)$ are given macroeconomic quantities. So is the unemployment rate u_t which solves:

$$\dot{u}_t = \delta(1-u_t) - \lambda_0 u_t, \qquad u_0 \text{ given.} \tag{5}$$

Problem (1) is therefore a standard non-autonomous optimal control problem, the Lagrangian of which is defined by:

$$\begin{aligned}
\mathcal{L}_t(p) &= (p - w_t)\,\ell_t(p) + m_t(p)\,(w_t - \underline{w}) \\
&\quad + \pi_t(p)\left\{-\left(\delta + \lambda_1 \overline{F}_t(V_t(p))\right)\ell_t(p) + \frac{v(p)}{N}\left(\lambda_0 u_t + \lambda_1(1-u_t)\,G_t(V_t(p))\right)\right\} \\
&\quad + \nu_t(p)\left\{\left(\rho + \delta + \lambda_1 \overline{F}_t(V_t(p))\right)V_t(p) - \lambda_1 \int_{V_t(p)}^{+\infty} x dF_t(x) - w_t - \delta U_t\right\},
\end{aligned} \tag{6}$$

where $\nu_t(p)$ $[\pi_t(p)]$ is the costate associated with $V_t(p)$ $[\ell_t(p)]$ and $m_t(p) \geq 0$ is the Lagrange multiplier associated with the minimum wage constraint (4).

Optimality conditions are:

$$\nu_t(p) = -\ell_t(p) + m_t(p) \tag{7}$$

$$\begin{aligned}
\dot{\nu}_t(p) &= r\nu_t(p) - \left(\rho + \delta + \lambda_1 \overline{F}_t(V_t(p))\right)\nu_t(p) \\
&\quad - \lambda_1 f_t(V_t(p))\,\ell_t(p)\,\pi_t(p) - \frac{\lambda_1 v(p)}{N}(1-u_t)\,g_t(V_t(p))\,\pi_t(p)
\end{aligned} \tag{8}$$

$$\dot{\pi}_t(p) = \left(r + \delta + \lambda_1 \overline{F}_t(V_t(p))\right)\pi_t(p) - p + w_t(p) \tag{9}$$

$$m_t(p) \geq 0, \quad w_t(p) \geq \underline{w}, \quad m_t(p)\,(w_t(p) - \underline{w}) = 0 \tag{10}$$

$$\lim_{t\to+\infty} e^{-rt}\pi_t(p)\,\ell_t(p) = \lim_{t\to+\infty} e^{-rt}\nu_t(p)\,(V_t(p) - U_t) = 0. \tag{11}$$

Supplementing this latter set of conditions with the state equations (2), (3) and (5), we obtain a system of partial differential equations characterizing the solution to an individual firm's maximization problem *for a given path of sampling distributions* $\{F_t(\cdot)\}_{t\in[0,+\infty)}$. The main difficulty then lies in characterizing the *equilibrium* $\{F_t(\cdot)\}_{t\in[0,+\infty)}$, i.e. the path of sampling distributions which is consistent with the above dynamic system simultaneously for the whole population of firms. In the following subsection we introduce an equilibrium restriction which helps getting round this difficulty. Before we turn to that, however, it is worth spelling out some economic interpretation of the above optimality conditions.

As usual in economic applications of optimal control, the costate variables $\pi_t(p)$ and $\nu_t(p)$ are interpreted as the imputed unit value of the corresponding state variable at date t (i.e. $\ell_t(p)$ and $V_t(p)$, respectively). Note that ν_t is negative as it is costly for any firm to transfer a higher value to its employees.

Equation (9) describes the dynamics of the firm's shadow value of its marginal employee. Notice that the overall rate at which the firm will discount that value is the sum of sheer time discounting (at the interest rate r) plus a "depreciation rate" of $\delta + \lambda_1 \overline{F}_t(V_t(p))$ reflecting future match dissolution, either through job destruction or the worker quitting. With that in mind, equation (9) has a straightforward asset-pricing-type interpretation, whereby the firm's marginal employee is viewed as an asset priced at $\pi_t(p)$. The annuity value of the marginal employee, $\left(r + \delta + \lambda_1 \overline{F}_t(V_t(p))\right) \pi_t(p)$, must then equal the return on the corresponding asset which is the sum of a dividend term $p - w_t(p)$ plus a capital gain term $\dot{\pi}_t(p)$.

Equation (8) next describes the dynamics of the firm's shadow value of a unit increase in the value it yields to its employees. It can also be viewed as an asset-pricing equation (even though in this case we are really talking about a cost as $\nu_t(p)$ is negative) whereby the annuity value $r\nu_t(p)$ is set equal to the capital gain $\dot{\nu}_t(p)$ plus a dividend term which represents the net benefit of increasing $V_t(p)$ by one unit through the effect of that increase on future profit streams (the effect of such an increase on current profits being nil). This latter term has two components, the first of which is $\pi_t(p) \cdot \frac{\partial \dot{\ell}_t(p)}{\partial V_t(p)} = \pi_t(p) \cdot \left[\lambda_1 f_t(V_t(p)) \ell_t(p) + \frac{\lambda_1 v(p)}{N}(1 - u_t) g_t(V_t(p))\right]$ and represents the future benefits of a larger workforce achieved through the higher retention and hiring rates resulting from the marginal increase in the value offered to workers. The second "dividend" component (in fact a cost as it is negative), $\nu_t(p) \cdot \frac{\partial \dot{V}_t(p)}{\partial V_t(p)} = \nu_t(p)\left[\rho + \delta + \lambda_1 \overline{F}_t(V_t(p))\right]$, has a somewhat less tangible interpretation. It measures the cost that the firm incurs through the change in the capital gain achieved by its workers caused by a marginal increase in the value currently transferred to them,

$V_t(p)$. This change in capital gain is proportional to the workers' overall discount rate which again results from a combination of pure time discounting (at rate ρ) and a risk of leaving the match (rate $\delta + \lambda_1 \overline{F}_t(V_t(p))$). A possible way to interpret this is to view an employer's commitment to transferring a certain value to its workers as that employer running up a debt to its employees. The consequence of a marginally higher current stock of debt is to increase the debt burden and speed up debt accumulation by an amount proportional to the interest paid on that debt, which here is indicated by the workers' discount rate.

Finally, equation (7) simply reflects the optimal balance between the instantaneous cost of increasing the current posted wage by \$1—it adds \$$\ell_t(p)$ to the current wage bill, plus possibly the instantaneous benefit of slackening the minimum wage constraint which is given by the Lagrange multiplier $m_t(p)$—and the future benefit of doing so, $-\nu_t(p)$. The debt analogy can be used for interpretation here as well: the future benefit of raising the wage at date t comes about through a reduced speed of debt accumulation (a smaller $\dot{V}_t(p)$) which follows from a higher installment (a higher wage) paid at date t.

3.2 Rank-Preserving Equilibria: Definition and Firm Size Evolution

We define a *Rank-Preserving Equilibrium* (RPE) as a dynamic equilibrium in which firms post values that are strictly increasing in p for all t. This has the consequence that workers rank firms according to productivity at all dates. Hence the following two properties hold true at all dates under the RP assumption:

$$F_t(V_t(p)) \equiv \Phi(p),$$
$$(1-u_t)\, G_t(V_t(p)) = N \int_{\underline{p}}^{p} \ell_t(x)\, d\Gamma(x).$$

The latter identity reflects a count of how many workers are employed at firms of type p or less.

We now consider the stock of workers employed at a firm of type-p or less, $N \int_{\underline{p}}^{p} \ell_t(x)\, d\Gamma(x)$. In a RPE (assuming one exists), those firms hire workers from unemployment and lose workers to their more productive competitors (firms of type higher than p). The stock of workers under consideration thus evolves according to:[9]

$$\int_{\underline{p}}^{p} \dot{\ell}_t(x)\, d\Gamma(x) = \frac{\lambda_0 u_t}{N}\Phi(p) - \left[\delta + \lambda_1 \overline{\Phi}(p)\right] \int_{\underline{p}}^{p} \ell_t(x)\, d\Gamma(x). \tag{12}$$

The latter equation now solves as:

$$\int_{\underline{p}}^{p} \ell_t(x)\, d\Gamma(x) = e^{-\left[\delta+\lambda_1\overline{\Phi}(p)\right]t} \left(\int_{\underline{p}}^{p} \ell_0(x)\, d\Gamma(x) + \frac{\lambda_0 \Phi(p)}{N} \int_0^t u_s e^{\left[\delta+\lambda_1\overline{\Phi}(p)\right]s} ds \right) \tag{13}$$

Now differentiating with respect to p, on obtains a closed-form expression for the workforce of any type-p firm:

$$\ell_t(p) = e^{-\left[\delta+\lambda_1\overline{\Phi}(p)\right]t} \left(\ell_0(p) + \lambda_1 t v(p) \int_{\underline{p}}^{p} \ell_0(x)\, d\Gamma(x) + \frac{\lambda_0 v(p)}{N} \int_0^t \left[1 + \lambda_1 \Phi(p)(t-s)\right] u_s e^{\left[\delta+\lambda_1\overline{\Phi}(p)\right]s} ds \right). \tag{14}$$

The steady-state versions of (13) and (14) are:

$$\ell_\infty(p) = \frac{\delta(1-u_\infty)(\delta+\lambda_1)}{N\left[\delta+\lambda_1\overline{\Phi}(p)\right]^2} v(p) \qquad \text{and} \qquad N\int_{\underline{p}}^{p} \ell_\infty(x)\, d\Gamma(x) = \frac{\delta(1-u_\infty)\Phi(p)}{\delta+\lambda_1\overline{\Phi}(p)}, \tag{15}$$

where $u_\infty = \frac{\delta}{\delta+\lambda_0}$ is the steady-state rate of unemployment.

This is the point at which the necessity for sampling weights appears. Note from equation (15) that the steady-state size ratio of the largest to the smallest firm in the market is $\left(1+\frac{\lambda_1}{\delta}\right)^2 \frac{v(\overline{p})}{v(\underline{p})}$. With uniform sampling ($v(p) \equiv 1$ throughout), this ratio would equal $\left(1+\frac{\lambda_1}{\delta}\right)^2$, which is in the order of 25-30 given standard estimates of λ_1 and δ. Now of course the data counterpart of that size ratio is virtually infinite. More generally, it appears that the BM model requires a sampling distribution that is very heavily skewed toward high-productivity firms in order to replicate the observed distribution of firm sizes.

Before going any further into characterizing Rank-Preserving Equilibria, we should notice that the analysis of firm size and employment dynamics carried out in this paragraph would apply to any job ladder model in which a similar concept of RPE can be defined. Indeed nothing in the dynamics of ℓ_t or u_t depends on the particulars of the wage setting mechanism, so long as this is

[9] Note that the following law of motion can also by obtained by integration of (2) w.r.t. p. Details available on request.

such that employed jobseekers move from lower-ranking into higher-ranking jobs in the sense of a time-invariant ranking. Therefore, this model's predictions about everything relating to firm sizes are in fact much more general than the wage- (or value-) posting assumption retained in the BM model.

3.3 Rank Preserving Equilibria: Characterization

We now go back to the dynamical system characterizing the behavior of the typical individual firm, and analyze it in a RPE. The system in question is comprised of the set of optimality conditions (7) - (11) plus the set of state equations (2), (3) and (5). We first focus on intervals of time when the solution is interior, i.e. such that $m_t(p) = 0$ and $w_t(p) > \underline{w}$. In this situation $\nu_t(p) = -\ell_t(p)$. For simplicity, we also assume equal discount rates for workers and employers from now on (i.e. $r = \rho$). Substitution of (7) into (8), and combination with (2) then yields:

$$\frac{v(p)}{N}\left(\lambda_0 u_t + \lambda_1(1-u_t)\,G_t(V_t(p))\right) = \lambda_1 \pi_t(p)\left(f_t(V_t(p))\,\ell_t(p) + \frac{v(p)}{N}(1-u_t)\,g_t(V_t(p))\right). \quad (16)$$

This latter equation reflects a balance between the firm's present-value cost and benefit of marginally changing its posted value at date t. The RHS of (16) equals $\pi_t(p) \cdot \frac{\partial \dot{\ell}_t(p)}{\partial V_t(p)}$ and clearly reflects the benefit of offering a marginally higher value stemming from the larger workforce achieved through the implied higher retention and hiring rates. To see how the LHS of (16) reflects the cost of a marginal increase in the value transferred to workers, it may help again to view $V_t(p)$ as an employer's debt to each of its employees. The (net) interest paid on that debt equals the workers' overall discount rate, $\rho + \delta + \lambda_1 \overline{F}_t(V_t(p))$, less the firm's discount (or interest) rate r. A unit increase in the value offered to all of the firm's employees then adds $\ell_t(p)$ to the firm's stock of debt. The marginal cost of such an addition to the stock of debt is an increase in the debt burden which in turn results from the net interest paid on that debt being raised by $\left[\rho - r + \delta + \lambda_1 \overline{F}_t(V_t(p))\right]\ell_t(p)$ plus an extrinsic expansion/contraction term $\dot{\ell}_t(p)$ reflecting the fact that the stock of debt is by nature indexed to workforce size. The sum of these latter two terms is equal to equation (16)'s LHS (under the assumption that $r = \rho$).

Next defining the shadow value to the *firm-worker match* (rather than to the firm) of the marginal unit of labor as $\mu_t(p) = \pi_t(p) + V_t(p)$, combination of (3) and (9) yields:

$$\dot{\mu}_t(p) = \left(r + \delta + \lambda_1 \overline{F}_t(V_t(p))\right) \mu_t(p) - \lambda_1 \int_{V_t(p)}^{+\infty} x dF_t(x) - \delta U_t - p, \tag{17}$$

which is supplemented by the transversality condition $\lim_{t \to +\infty} e^{-rt}\mu_t(p) = 0$. Interpretation of equation (17) is once again based on straightforward asset-pricing-type arguments and we shall therefore not dwell on it.

The RP assumption finally changes the system (16) - (17) into:

$$\left(\frac{\lambda_0 u_t}{N} + \lambda_1 \int_{\underline{p}}^{p} \ell_t(x) d\Gamma(x)\right) V_t'(p) = 2\lambda_1 \gamma(p) \ell_t(p) \pi_t(p) \tag{18}$$

$$\dot{\mu}_t(p) = \left(r + \delta + \lambda_1 \overline{\Phi}(p)\right) \mu_t(p) - \lambda_1 \int_{p}^{+\infty} V_t(x) d\Phi(x) - \delta U_t - p \tag{19}$$

$$\lim_{t \to +\infty} e^{-rt}\mu_t(p) = 0. \tag{20}$$

Differentiation of (19) w.r.t. p yields (primes denote differentiation w.r.t. p while dots denote time differentiation):

$$\dot{\mu}_t'(p) = \left(r + \delta + \lambda_1 \overline{\Phi}(p)\right) \mu_t'(p) + \lambda_1 \gamma(p) v(p) \left(V_t(p) - \mu_t(p)\right) - 1. \tag{21}$$

This, together with (18), gives the following system of two PDEs in $(\mu_t'(p), \pi_t(p))$:

$$\begin{aligned} \dot{\mu}_t'(p) &= \left(r + \delta + \lambda_1 \overline{\Phi}(p)\right) \mu_t'(p) - \lambda_1 \gamma(p) v(p) \pi_t(p) - 1 \\ \mu_t'(p) &= \pi_t'(p) + \frac{2\lambda_1 \gamma(p) \ell_t(p)}{\frac{\lambda_0 u_t}{N} + \lambda_1 \int_{\underline{p}}^{p} \ell_t(x) d\Gamma(x)} \pi_t(p). \end{aligned} \tag{22}$$

This can be solved numerically, subject to some initial and boundary conditions. 'Initial' conditions are given by the steady-state solution to (22), which is characterized as:

$$\begin{aligned} \mu_\infty'(p) &= \frac{1 + \lambda_1 \gamma(p) v(p) \pi_\infty(p)}{r + \delta + \lambda_1 \overline{\Phi}(p)} \\ \pi_\infty(p) &= \frac{\left(\delta + \lambda_1 \overline{\Phi}(p)\right)^2}{r + \delta + \lambda_1 \overline{\Phi}(p)} \left(\int_{\underline{p}}^{p} \frac{dx}{\left(\delta + \lambda_1 \overline{\Phi}(x)\right)^2} + \frac{\pi_\infty(\underline{p})(r + \delta + \lambda_1)}{(\delta + \lambda_1)^2}\right). \end{aligned} \tag{23}$$

Now turning to boundary conditions, standard arguments prove that the lowest-type firms have no reason to pay more than the minimum wage: type $\underline{p}$ firms can only hire from unemployment and lose workers to poachers anyway, so trying to prevent poaching by raising wages is pointless for those firms in a RPE. While this implies that the minimum wage constraint (4) will bind at all dates for the lowest-type firm, it also implies that the following (time-invariant) boundary conditions are satisfied:

$$\begin{aligned}\pi_t\left(\underline{p}\right) &\equiv \frac{\underline{p}-\underline{w}}{r+\delta+\lambda_1} \\ \mu_t'\left(\underline{p}\right) &\equiv \frac{1+\lambda_1\gamma\left(\underline{p}\right)v\left(\underline{p}\right)\pi_t\left(\underline{p}\right)}{r+\delta+\lambda_1},\end{aligned} \tag{24}$$

where the second condition is obtained by combining the first one with the $\dot{\mu}_t'(p)$ equation in (22). These simple boundary conditions can be further simplified by imposing $\underline{p}=\underline{w}$, a kind of free-entry condition holding throughout the adjustment toward the new steady state, which implies $\pi_t\left(\underline{p}\right)\equiv 0$. The minimum productivity $\underline{p}$ that can survive in the market is $\underline{w}$, as any firm with $p>\underline{w}$ can make positive profits by offering $\underline{w}$, and possibly even more by offering a higher wage while no firm with $p<\underline{w}$ can ever make any profits.

Once (22) is solved for $\left(\mu_t'(p),\pi_t(p)\right)$, wages can be retrieved from (9) (written under the RP assumption):

$$w_t(p)=p-\left(r+\delta+\lambda_1\overline{\Phi}(p)\right)\pi_t(p)+\dot{\pi}_t(p), \tag{25}$$

which has the following familiar steady-state solution:

$$w_\infty(p)=p-\left(\delta+\lambda_1\overline{\Phi}(p)\right)^2\left(\int_{\underline{p}}^{p}\frac{dx}{\left(\delta+\lambda_1\overline{\Phi}(x)\right)^2}+\frac{\underline{p}-\underline{w}}{\left(\delta+\lambda_1\right)^2}\right). \tag{26}$$

This is exactly the BM solution for the heterogeneous firm case (see equation (47) in Burdett and Mortensen, 1998). This confirms that our contracts are consistent with the BM steady-state wage-posting equilibrium *if the labor market is at a steady state.* It is no longer the case off steady-state, however: posting a time-invariant wage is not, in general (although see Appendix A for a situation in which it is the case), a firm's best response to all other firms posting time-invariant wages.[10]

[10]To see this, notice that (9) and (18) yield two different growth rates for $\pi_t(p)$ if all wages are constant and the economy is off its steady state (so that firm sizes change over time). To see this, note that equation (9) gives a $\pi_t(p)$ which evolves as an exponential of time. But then with a constant wage and constant wages offered elsewhere, $V_t'(p)$ is constant over time, so dividing (18) by $\ell_t(p)$ tells us that $\pi_t(p)$ is proportional to the gross hiring rate, and so

We now look back to the minimum wage constraint. The only firm for which the minimum wage constraint (4) is binding at the steady state characterized above is the lowest-type firm, $\underline{p}$. It may be the case, however, that the constraint temporarily binds for some higher-type firms over the transition to that steady state, in which case the economy no longer behaves according to (22) as $m_t(p)$ becomes strictly positive for some p at some dates.

Appendix B describes an algorithm that constructs an equilibrium in which $\underline{w}$ is allowed to temporarily bind for some firms (at the lower end of the p-distribution) with the restriction that it only bind over some initial period. In other words, any firm can choose to post the minimum wage for a while right after the occurrence of the productivity shock, but onces it ceases to do so it is not allowed to return to the minimum wage. Simulations, however, will prove that the minimum wage is only offered by the lowest-p firms in equilibrium.

4 Quantitative Analysis of Rank Preserving Equilibria

4.1 Simulating an Expansion

In order to simulate the economy's response to a one-time, permanent and unanticipated aggregate productivity shock, we further specify the model as follows. We assume that any firm's productivity parameter p is the product of an aggregate productivity index y (common to all firms) and a firm-specific random effect θ. We further assume that there is an exogenous number N_0 of potential firms, each with a fixed value of θ drawn from some exogenous underlying distribution Γ_0. Because for any potential firm productivity is $p = y \times \theta$, the only profitable firms in the presence of a wage floor $\underline{w}$ are those with $\theta \geq \underline{w}/y$. The distribution of productivity levels among active firms will thus be given by:

$$\Gamma(p) = \frac{\Gamma_0(p/y) - \Gamma_0(\underline{w}/y)}{1 - \Gamma_0(\underline{w}/y)}, \tag{27}$$

and the number of active firms will be $N = N_0(1 - \Gamma_0(\underline{w}/y))$.

We model a 'boom' as a permanent 2 percent increase in y (from $y = 1$ to $y = 1.02$). We further

$\pi_t(p)$ cannot be exponential in time (because the hiring rate is not an exponential function of time in a RPE). All this implies that posting a constant wage in the face of competitors themselves posting constant wages violates the firm's set of necessary optimality conditions.

assume that this productivity increase causes the job finding rate λ_0 to increase by 8 percent,[11] and the arrival rate of offers to employed jobseekers, λ_1, to increase by 1.6 percent. If the wage floor $\underline{w}$ does not react, the shock causes entry of $\Delta N = N_0 \left(\Gamma_0 \left(\underline{w} \right) - \Gamma_0 \left(\underline{w}/1.02 \right) \right)$ firms at the bottom of the productivity distribution, all starting off with a size of zero. The distribution of productivity across active firms jumps instantly following (27).

4.2 Baseline Calibration

A sampling distribution of firm types is first calibrated following the Bontemps *et al.* (2000) procedure in such a way that the predicted steady-state wage distribution fits the business-sector wage distribution observed in the CPS. Specifically, equation (15) implies that the steady-state cross-section CDF of wages, $G_w^\star(\cdot)$ (say), is defined by

$$\Phi(p) = \frac{(\delta + \lambda_1) G_w^\star(w(p))}{\delta + \lambda_1 G_w^\star(w(p))} \quad \Rightarrow \quad \Phi'(p) = \frac{\delta(\delta + \lambda_1) g_w^\star(w(p)) w'(p)}{(\delta + \lambda_1 G_w^\star(w(p)))^2}. \tag{28}$$

Differentiation of (26) then yields:

$$w'(p) = 2\lambda_1 \Phi'(p) \frac{p - w(p)}{\delta + \lambda_1 \overline{\Phi}(p)} \quad \Rightarrow \quad p(w) = w + \frac{\delta + \lambda_1 G_w^\star(w)}{2\lambda_1 g_w^\star(w)}. \tag{29}$$

A lognormal distribution is fitted to a sample of wages from the 2006 CPS and then used to construct a sample of firm types using the above relationship. The sampling distribution $\Phi(\cdot)$ that rationalizes this sample in a steady state (and given values of δ and λ_1) is then retrieved using (28).

Once a sampling distribution has been obtained, the underlying distribution of firm types $\Gamma(p)$ and sampling weights $v(p)$ are calibrated based on the employment share-firm size relationship found in the BED data. Table 2 summarizes the information conveyed by the BED data about that relationship. The data in Table 2 is found to be well fitted by the following parametric relationship:

$$\Gamma(p) = \left(\frac{1 - e^{-\alpha_1 G_w^\star(w(p))}}{1 - e^{-\alpha_1}} \right)^{\alpha_2}, \tag{30}$$

with $\alpha_1 = 8.0661$ and $\alpha_2 = 0.5843$. Sampling weights are finally retrieved as $v(p) = \Phi'(p)/\gamma(p)$.

[11] This is based on an elasticity of labor market tightness with respect to productivity of 8 and an elasticity of the job finding rate w.r.t. labor market tightness of 0.5, both consensual numbers.

Firm size category	Cum. fraction of firms $[\Gamma(p)]$	Cum. emp. share $[G_w^\star(w(p))]$
1-4	0.535	0.052
5-9	0.742	0.114
10-19	0.868	0.192
20-49	0.949	0.303
50-99	0.976	0.387
100-249	0.991	0.493
250-499	0.996	0.565
500-999	0.998	0.633
1000 and up	1.00	1.00

BED data, all years pooled.

Table 2: Firm sizes and employment shares

Finally, we shift the support of the $\Gamma(\cdot)$ distribution thus obtained so that its infimum is at $p = 1$ and use it as our benchmark underlying distribution of firm types $\Gamma_0(\cdot)$ (given the normalization $y = 1$) as explained in the previous subsection.

Apart from productivity dispersion, our baseline parameterization is explicated in Table 3. The time unit is one month. The value of r reflects an annual discount rate of five percent. The minimum wage is binding (in the sense that $\underline{p} = \underline{w}$) since, being equal to 5, it exceeds the lower support of the distribution of potential firm productivity levels which was normalized at 1. Finally, the number chosen for N_0 reflects an average firm size of 20.

Parameters (post-shock monthly values)						
r	δ	λ_0	λ_1	$\underline{w}$	y	N_0
0.0043	0.025	0.40	0.12	5	1.02	0.0509

Table 3: Baseline parameterization

4.3 Results

As can very easily be inferred from equation (5), the response of the unemployment rate to the positive shock hitting the economy is a simple monotonic adjustment toward the new (lower) steady-state value. The interesting feature of that adjustment is its speed: given our calibrated values of δ and λ_0, 90% of the distance between the initial and the final steady state is covered in less than six months.

Figure 6 then shows how unemployment adjusts at single firms: it shows a plot of $\ell_t(p)/\ell_0(p)$ for four different values of p corresponding to the 50th, 90th, 95th and 99.9th percentiles of the (post-shock) distribution of firm types, $\Gamma(\cdot)$.[12] Patterns of employment adjustment differ markedly across firm types—which translates into differences across firm size categories as low-p firms are also smaller firms in the initial state of the labor market. One sees on Figure 6 that "large" firms tend to increase in size monotonically and gradually (the higher the firm in terms of p, the more gradual the adjustment). Conversely, "smaller" firms experience a short episode of rapid growth soon after the shock and then start shrinking back toward their final steady-state size, which they overshoot in the adjustment process. Firms at the 50th percentile of the $\Gamma(\cdot)$ distribution (which places them at the 21st percentile of the sampling distribution $\Phi(\cdot)$ and at the 4.5th percentile in terms of steady-state cumulated employment shares) even end up being smaller after the increase in productivity than in the initial steady state.

This pattern conforms with intuition: in the few months following the shock, most of the new hires are workers coming from unemployment and get disproportionately allocated to small (low-p) firms. After six months or so (given the magnitude of λ_0), the unemployment pool dries out and poaching becomes the main channel of hiring. Poaching benefits larger, higher-p, better-paying firms at the expense of smaller ones. It occurs later on in the expansion and is a much slower process than the initial siphoning of the unemployment pool as λ_1 is about a third of λ_0 in magnitude and the average offer acceptance rate of an employed jobseeker is less than one.[13]

[12] The normalization by $1/\ell_0(p)$ is just there to rescale the paths and keep the picture legible. Moreover, on all Figures, circles on the axes indicate initial (steady-state) values of the various indicators plotted.

[13] It actually equals $N\int_{\underline{p}}^{\overline{p}} \overline{\Phi}(x)\,\ell_t(x)\,d\Gamma(x)\,/\,(1-u_t)$. This becomes $(1+\delta/\lambda_1)\{1-\delta\ln(1+\lambda_1/\delta)/\lambda_1\}$ at a steady state, i.e. about 0.76 with our parameterization.

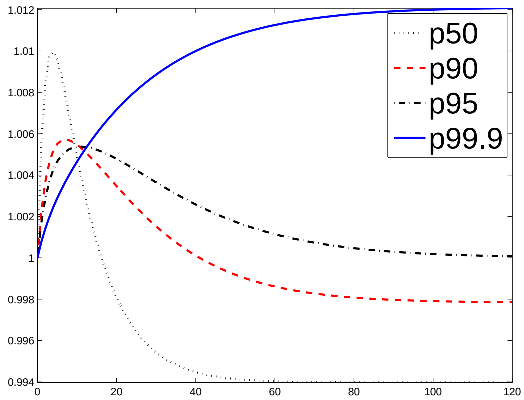


Fig. 6: Firm size dynamics

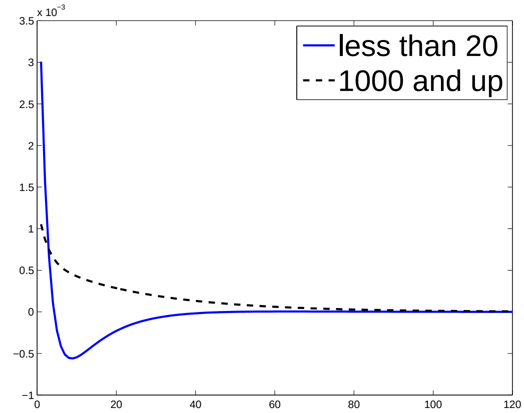


Fig. 7: Firm growth—small *vs.* large firms

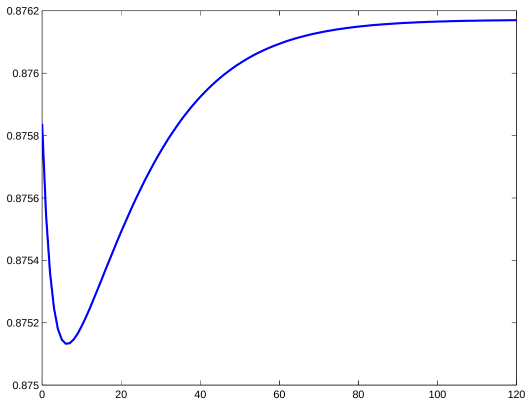


Fig. 8: Firm shares (firms < 20)

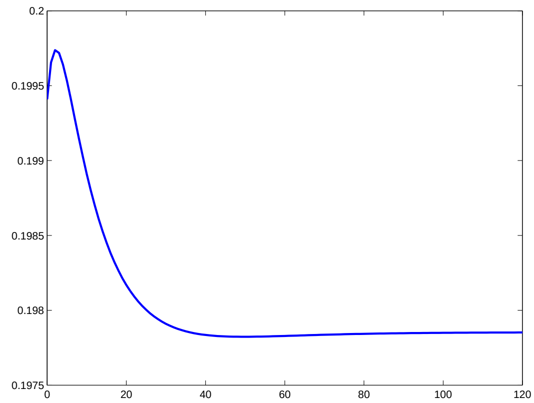


Fig. 9: Emp. shares (firms < 20)

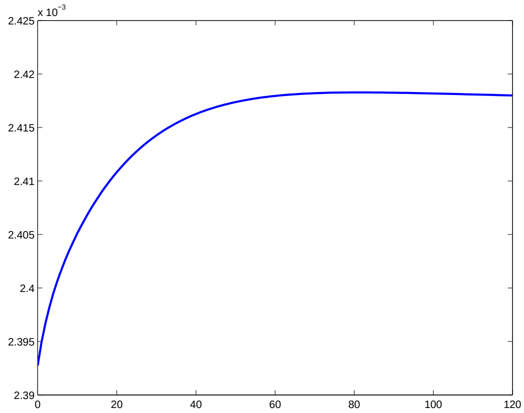


Fig. 10: Firm shares (firms ≥ 1000)

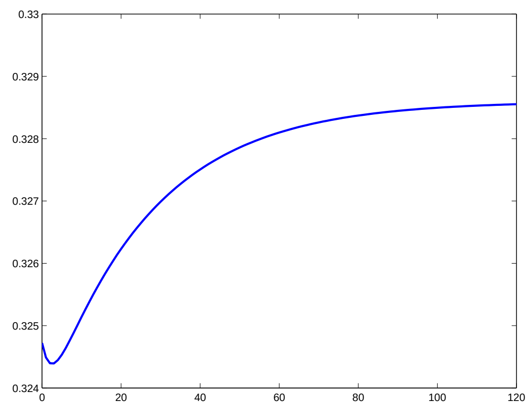


Fig. 11: Emp. shares (firms ≥ 1000)

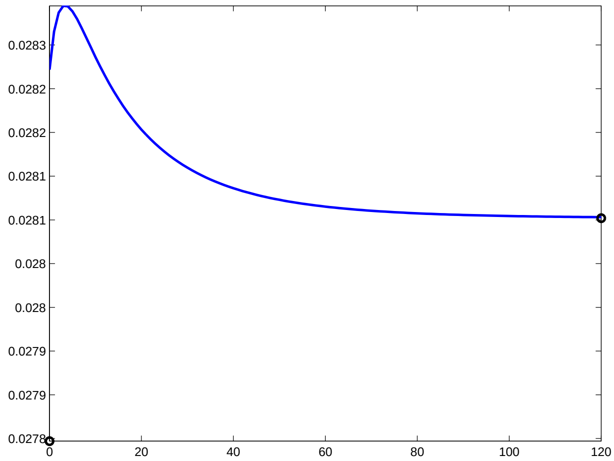


Fig. 12: Average job-to-job quit rate

For comparison with the descriptive evidence shown in the Introduction, the mechanism just described can be depicted in terms of employment shares and average growth rates by firm size category. This is done in Figures 7 to 11 which parallel Figures 1 and 2 from the Introduction.

The response of the average job-to-job quit rate, $\frac{\lambda_1 N}{1-u_t}\int_{\underline{p}}^{\overline{p}} \overline{\Phi}(x)\,\ell_t(x)\,d\Gamma(x)$, is plotted on Figure 12. Apart from the initial jump caused by the assumed instant response of λ_1 to the productivity shock, the average quit rate has an initial increasing phase which reflects the initial disproportionate inflow of new hires into small, low-productivity firms. These workers start getting poached away by larger firms relatively easily, while at the same time the unemployment pool quickly gets depleted and the excess inflow of workers into easy-to-poach positions slows down. As workers gradually get reallocated toward more productive, better-paying firms, poaching becomes more difficult (the acceptance rate of outside offers falls) and the quit rate falls.

Finally, Figures 13 and 14 plot the dynamic responses of mean wages and mean output per worker.

The path followed by mean output per worker results from a pure composition effect. After the initial upward jump caused by the sudden 2 percent increase in the productivity levels of all established firms, mean output per worker adjusts quasi-monotonically to its higher final steady-state value following the gradual reallocation of newly hired workers into more productive firms.

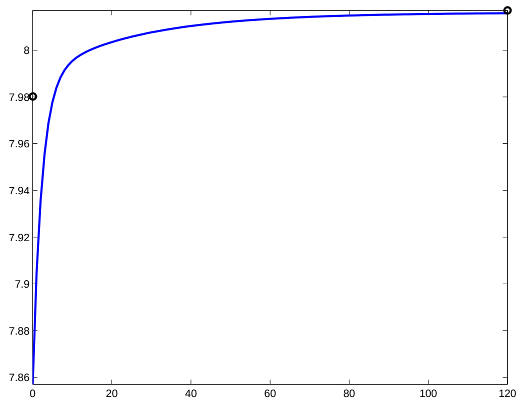


Fig. 13: Mean wage dynamics

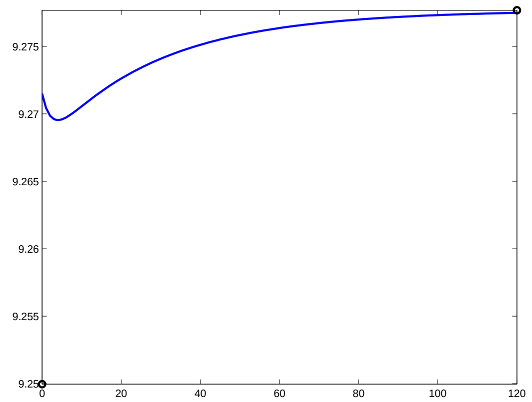


Fig. 14: Mean output per worker dynamics

The slight dip observed in the initial phase of that adjustment is due to the mass of low-productivity firms suddenly becoming viable as a result of the positive aggregate shock on y and thus entering the market with an initial size of zero. These entrant firms drag average output per worker down in the early phase of the expansion as they hire some workers into low-productivity jobs.

The mechanisms generating the path followed by the mean wage are more intricate. First, the same composition effect as for mean output per worker operates for wages: there is an initial excess inflow of workers into low-paying firms and those workers gradually reallocate themselves into better-paying firms, thereby causing a sluggish positive response of the mean wage to the aggregate productivity shock. Note that, because of this composition effect, the aggregate mean wage would exhibit this sluggish adjustment pattern even if all firm-level wages would jump right onto their new steady-state values upon impact of the productivity shock.[14] Second, each firm-level wage follows a dynamic path of its own. The composition of these individual dynamic paths causes the initial downward jump in the mean wage.

4.4 Discussion

Intuitively, it is in the firms' interest to backload wage payments. In this version of the BM model, because firms are not allowed to index wages to individual tenure, they cannot backload

[14]This is precisely the situation that would arise under the special assumption of infinitely impatient workers (worker with an infinite rate of future discount). The full details of that special case are in Appendix A.

at the individual level (as they would do in the wage-tenure models of Stevens, 2004 and Burdett and Coles, 2003). However they can index contracts to calendar time and benefit from future competition from higher-paying firms. Specifically, the prospect of receiving an offer from a better-paying firm later on makes up for the low wage that a single firm offers today. In other words, superior firms impose a "top-down" externality on inferior firms through future poaching which encourages the latter to backload wages. Furthermore, this effect is reinforced by a "bottom-up" strategic complementarity whereby a superior firm's response to an inferior firm's backloading is to backload itself by offering slightly more at all dates—just enough to maintain its rank and poach workers away from the lower-p firm. Note that, unlike the former top-down externality, this latter bottom-up mechanism has no time dimension per se. Another difference from the wage-tenure models cited above is that our contract-posting model delivers smooth backloading despite risk neutrality, whereas Stevens (2004) shows that the optimal (backloaded) wage-tenure contract offered to risk-neutral workers is a step contract (while Burdett and Coles, 2003 show how worker risk aversion entails gradual backloading). In our case, the gradual nature of backloading is purely driven by strategic considerations.

This discussion highlights a twofold motive for firms to backload wages: piggybacking on future poaching by superior firms and just beating inferior firms that do backload. Of course the extent to which firms can piggyback on their future competitors depends on the workers' horizon relative to the firms' own horizon. Most of the analysis so far was based on the assumption that workers and firms were equally patient in that they shared the same finite discount rate $\rho = r$. As argued in section 3, however, the model is perfectly well defined with different discount rates for firms and workers. With different discount rates, an additional motive for backloading comes into play: intertemporal trading.

To illustrate that, consider the following two polar cases. First, suppose workers were finitely patient and firms were myopic, i.e. suppose $\rho < +\infty = r$. In this simple case firms do not care about the future, although workers do, so all firms would promise the minimum wage today to myopically save on wage bills, and very high wages in the future. Patient workers will like this strategy. So wage paths tend to become vertical and their growth rate becomes infinite. Now suppose on the contrary that *firms* were finitely patient and *workers* were myopic, i.e. suppose $r < +\infty = \rho$. This case is formally analyzed in Appendix A where we show that the unique RPE features all firms offering constant wages—i.e., no backloading. This is intuitive: if workers cease to care about the future, firms can no longer play the piggybacking game and backloading motives become ineffective.

Intuitively, relative firm and worker discount rates determine the intertemporal trading. Indeed this discussion suggests the possibility, to be explored, that the slope of the wage profile during an expansion decreases in the patience of firms relative to workers.

5 Conclusion

We identify and illustrate several new facts about the pattern of aggregate employment and wage movements in the US economy over the last 15 years. In particular, we find that three apparently unrelated labor market series — average (weekly or hourly) real earnings, the rate at which workers quit from job to job, and the employment share of large firms — either stagnate or decline for several years after each of the last two recessions, and then rise as the ensuing expansion enters it mature phase. Both in 1994-1996 and in 2001-2004 earnings are flat or declining, the worker job-to-job quit rate falls, and small firms account for most of job creation. In 1997-2000 and in 2005-2006 earnings rise sharply, as does the worker job-to-job quit rate, and employment shifts towards large firms. The 2001 recession causes a downturn in earnings and quit rate and an increase in the employment share of small firms. In addition, we document that wages are monotonically increasing in firm size at all points in time, and rise late in an expansion, with no discernible pattern of either convergence or fanning out across all firm size classes.

While the period under consideration is, due to data availability, too short to establish any new stylized facts about business cycles, this evidence suggests a new view of how labor markets function, or at least functioned in the last two cycles. More productive firms pay higher wages, thus hire, employ and retain more workers. Workers quit from low-wage, small firms to high-wage, large ones. Early in an expansion, when unemployed job applicants are plentiful, all firms exploit their monopsony power and pay low wages. As few workers are employed, in particular at small firms, the aggregate job-to-job quit rate is small. As the pool of unemployment dries out, small firms have a harder time hiring workers, while large firms can now poach from small firms their larger employment pool. So the aggregate quit rate rises and the share of employment at large firms increases. Aggregate wages rise for two reasons, once the quit-poaching machine gets going. First, workers climb to higher-paying firm, so there is a composition effect. Second, firms offer wage profiles that increase over time. The increased competition for employed workers erodes firms' monopsony power and leads to a redistribution of rents from profits to salaries late in an expansion.

We propose and analyze a model of the labor market which captures these features. We study convergence to steady state equilibrium in the Burdett and Mortensen (1998) wage-posting model with firms of heterogeneous labor productivity. We allow firms to commit to wage contracts that depend on calendar time (or on the unemployment rate.) We restrict attention to equilibria of the contract-posting game where workers always quit from less to more productive firms. We find that firms post wage profiles that increase smoothly over time. Since workers are risk-neutral and have no motive for wage smoothing, this gradual increase is due entirely to strategic considerations across firms. Specifically, firms backload wages for two reasons. First, in order to let future poachers sometimes deliver the promised higher future wages to its current workers. Second, because less productive firms do, so offering low and then increasing wages is sufficient to poach workers from them. A calibrated version of the model delivers aggregate and disaggregated dynamics that are qualitatively consistent with all the facts presented in this paper. Following an unanticipated aggregate productivity shock, the economy starting from an initial steady state converges to a new one. In the transitional dynamics, quits, productivity and wages rise slowly due to the composition effect. Wages also rise due to backloading. The pattern of employment growth across firm size classes is accurately replicated.

Our analysis presents several limitations. On the empirical side, as firm productivity is not easily observable, we proxy it by firm size, as suggested by the model. Firm size is, however, endogenous and evolving over time. Thus, to firmly establish that the employment share of small firms peaks right after the end of a recession we need a panel of firms, to identify those that are small at the end of the recession. So far we have exploited repeated cross-sections of firms, to obtain a meaningfully long time series. We are currently working and plan to work on a variety of firm panels from different countries, both to extract direct information on firm productivity and to fix the identity of small firms after a recession. On the theoretical side, in order to focus on the role of aggregate dynamics in the contract-posting game, we abstract from the possibility that such contracts may be conditioned on worker tenure, employment status of the applicant, or other features. Also we adopted a rather minimal description of the search technology (exogenous and constant worker-firm contact rates), mainly in order to maintain tractability. Next, the thought experiment is the adjustment to a one-shot aggregate shock, but ideally we would like to characterize dynamics in an explicitly stochastic model, with aggregate uncertainty recognized by all agents. Finally, we are aiming to obtain a full analytical characterization of the dynamic equilibrium. On the quantitative side, our results still present a large margin of improvement. The half-life of the main time series of interest produced by the simulation is an order of magnitude shorter than in the data. We expect that introducing partially persistent aggregate shocks and/or relaxing some of the theoretical restrictions listed above will fill much of this gap.

References

[1] Barlevy, G. (2005), "Estimating Models of On-the-Job Search using Record Statistics", Federal Reserve Bank of Chicago Working Paper (http://www.chicagofed.org/publications/workingpapers/papers/wp2003-18.pdf).

[2] Bontemps, C., J.-M. Robin, and G. J. Van den Berg (2000), "Equilibrium Search with Continuous Productivity Dispersion: Theory and Non-Parametric Estimation", *International Economic Review*, 41 (2), 305-58.

[3] Burdett, K. and M. Coles (2003), "Equilibrium Wage-Tenure Contracts", *Econometrica*, 71(5), 1377-1404.

[4] Burdett, K. and D. T. Mortensen (1998), "Wage Differentials, Employer Size and Unemployment", *International Economic Review*, 39, 257-73.

[5] Cahuc, P., F. Postel-Vinay and J.-M. Robin (2006), "Wage Bargaining with On-the-job Search: Theory and Evidence", *Econometrica*, 74(2), 323-64.

[6] Carrillo-Tudela, C. (2007), "Offers Conditioned on Employment Status: An Equilibrium Search Approach", University of Leicester mimeo.

[7] Delli Gatti, D., C. Di Guilmi, E. Gaffeo, G. Giulioni, M. Gallegati and A. Palestrini (2004), "Business Cycle Fluctuations and Firms' Size Distribution Dynamics", *Advances in Complex Systems*, 7(2), 118.

[8] Dey, M. S. and C. J. Flinn (2005), "An Equilibrium Model of Health Insurance Provision and Wage Determination", *Econometrica*, 73(2), 571-627.

[9] Diamond, P. (1971), "A Model of Price Adjustment", *Journal of Economic Theory*, 3, 156-168.

[10] Mortensen, D. (2003), *Wage Dispersion: Why Are Similar Workers Paid Differently?*, MIT Press.

[11] Moscarini, G. and K. Thomsson (2007), "Occupational and Job Mobility in the US", forthcoming *Scandinavian Journal of Economics*.

[12] Postel-Vinay, F. and J.-M. Robin (2002), "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity", *Econometrica*, 70 (6), 2295-350.

[13] Shimer, R. (2003), "Out-of-Steady-State Behavior of the Burdett-Mortensen Model", University of Chicago *mimeo*.

[14] Stevens, M. (2004), "Wage-Tenure Contracts in a Frictional Labour Market: Firms' Strategies for Recruitment and Retention", *The Review of Economic Studies*, 71, 535-551.

Appendix

A The Case of Myopic Workers: $\rho = +\infty$

If workers are (infinitely) impatient, they only care about current wages and the firm's problem simplifies to:

$$\Pi_0^\star(p) = \max_{\{w_t\}} \int_0^{+\infty} (p - w_t(p))\,\ell_t(p)\,e^{-rt}dt \tag{31}$$

$$\text{subject to:} \quad \dot{\ell}_t(p) = -\left(\delta + \lambda_1 \overline{F}_t(w_t)\right)\ell_t(p) + \frac{v(p)}{N}\left(\lambda_0 u_t + \lambda_1(1-u_t)\,G_t(w_t)\right), \tag{32}$$

which has one less state variable ($V_t(p)$) than the original problem (1). (Readers will pardon the notational abuse whereby $F(\cdot)$ and $G(\cdot)$ now take w_t, rather than V_t, as an argument.)

Denoting the costate associated with $\ell_t(p)$ (i.e. the firm's shadow value of the marginal worker) as $\pi_t(p)$, the optimality conditions for (31) write down as:

$$1 = \pi_t(p) \times 2\lambda_1 f_t(w_t(p)) \tag{33}$$

$$\dot{\pi}_t(p) = \left(r + \delta + \lambda_1 \overline{F}_t(w_t(p))\right)\pi_t(p) + w_t(p) - p, \tag{34}$$

$$\lim_{t\to+\infty} e^{-rt}\pi_t(p) = 0, \tag{35}$$

where the first order condition (33) uses the fact that $(1-u_t)\,g_t(w_t(p)) = N\ell_t(p)\,f_t(w_t(p))/v(p)$.

Now focusing on RPE's, the first order condition becomes:

$$w_t'(p) = 2\lambda_1\gamma(p)\,\pi_t(p). \tag{36}$$

Substitution into (34) delivers the following PDE in $w_t(p)$:

$$\dot{w}_t'(p) = \left(r + \delta + \lambda_1\overline{\Gamma}(p)\right)w_t'(p) + 2\lambda_1\gamma(p)\,(w_t(p) - p). \tag{37}$$

This has a simple time-invariant solution, which is rank preserving (it is the customary steady-state wage equation in the BM model with heterogeneous firms):

$$w_\infty(p) = p - \left(r + \delta + \lambda_1\overline{\Gamma}(p)\right)^2 \int_{\underline{p}}^{p} \frac{dx}{\left(r + \delta + \lambda_1\overline{\Gamma}(x)\right)^2} - \left(\underline{p} - w_\infty(\underline{p})\right). \tag{38}$$

This time-invariant solution satisfies the RP property and the optimality conditions (33) - (35). It is therefore an RPE, in which all firms jump right on to the new steady-state wage policy after a shock. Firm sizes then evolve according to (14) and the cross-section distribution of wages also gradually shifts toward its new steady-state shape as labor gets reallocated between firms.

The model can be closed by assuming the free-entry condition $\underline{p} = \underline{w}$. Under this assumption, $w_t(\underline{p}) = \underline{p} = \underline{w}$ for all t. We renormalize Γ to have mass one. To show that the invariant solution is unique, integrate equation (37) between $\underline{p}$ and p:

$$\int_{\underline{p}}^{p} \dot{w}_t'(x)\,dx = (r+\delta)\left[w_t(p) - w_t(\underline{p})\right] + \lambda_1\int_{\underline{p}}^{p}\overline{\Gamma}(x)\,w_t'(x)\,dx + 2\lambda_1\int_{\underline{p}}^{p}\gamma(x)\,(w_t(x) - x)\,dx.$$

Integrating by parts the middle term on the RHS yields (using $d\underline{w}/dt = 0$):

$$\dot{w}_t(p) = \left(r+\delta+\lambda_1\overline{\Gamma}(p)\right) w_t(p) - \left(r+\delta+\lambda_1\overline{\Gamma}(\underline{p})\right)\underline{p} + 3\lambda_1 \int_{\underline{p}}^{p} w_t(x)\,\gamma(x)dx - \int_{\underline{p}}^{p} x\gamma(x)\,dx,$$

and

$$\ddot{w}_t(p) = \left(r+\delta+\lambda_1\overline{\Gamma}(p)\right)\dot{w}_t(p) + 3\lambda_1 \int_{\underline{p}}^{p} \dot{w}_t(x)\,\gamma(x)dx.$$

We can show that the invariant distribution is the unique solution, so the equilibrium jumps right away to the new steady state. Since $\dot{w}_t(p)$ is differentiable in p, there exists $\hat{p} > \underline{p}$ such that $\dot{w}_t(p)$ preserves the sign for $p \in [\underline{p}, \hat{p}]$. If this sign is zero, $\dot{w}_t(p) = 0$ for all p: we have the stationary solution. If it is weakly positive with strict inequality on a set of positive measure, then from the above equation $\ddot{w}_t(p) > 0$. But then $\dot{w}_t(p)$ rises and becomes even more positive on some set of p's. By induction, $\dot{w}_t(p)$ and thus $w_t(p)$ grow unbounded, ultimately make profits negative, and cannot converge to the new steady state. By the same reasoning, if $\dot{w}_t(p) \leq 0$ for all $p \in [\underline{p}, \hat{p}]$ with strict inequality on a set of positive measure, then $w_t(p)$ grows unbounded below on some set of productivities, violating the minimum wage requirement and any reservation wage. Using the entry condition, wages are

$$w_t(p) = p - \left(r+\delta+\lambda_1\overline{\Gamma}(p)\right)^2 \int_{\underline{p}}^{p} \frac{dx}{\left(r+\delta+\lambda_1\overline{\Gamma}(x)\right)^2}.$$

B Numerical Equilibrium Determination

The algorithm we use to numerically characterize the dynamic equilibrium is based on the restriction that, if the minimum wage constraint binds for some firms, it will do so at early stages of the expansion only. In other words, any firm can choose to post the minimum wage for a while right after the productivity shock, but onces it ceases to do so it is not allowed to return to the minimum wage. Simulations will prove that an equilibrium with exactly this pattern exists.

In order to construct that equilibrium, we proceed through the following steps.

Step 1. Consider some productivity level p_0 such that the functions $\pi_t(p_0)$ and $\mu_t'(p_0)$ are known. (In effect the algorithm is started at $p_0 = \underline{p}$ for which those functions are known from (24).) Pick a step size h.

Step 2. Construct a candidate $\pi_t(p_0+h)$ using the second (static) differential equation in (22), such as:[15]

$$\widetilde{\pi}_t(p_0+h) = \pi_t(p_0) + h \times \left(\mu_t'(p_0) - \frac{2\lambda_1\gamma(p_0)\,\ell_t(p_0)}{\frac{\lambda_0 u_t}{N} + \lambda_1\int_{\underline{p}}^{p_0}\ell_t(x)\,d\Gamma(x)}\pi_t(p_0)\right). \tag{39}$$

Step 3. Construct a candidate wage path for type-(p_0+h) firms from $\widetilde{\pi}_t(p_0+h)$ and equation (9):

$$\widetilde{w}_t(p_0+h) = p_0 + h - \left(r+\delta+\lambda_1\overline{\Gamma}(p_0+h)\right)\widetilde{\pi}_t(p_0+h) + \dot{\widetilde{\pi}}_t(p_0+h). \tag{40}$$

[15]The following uses a simple Euler approximation. In practice we use a 2-step Runge-Kutta approximation for numerical accuracy.

Step 4. Construct $w_t(p_0+h)$ and $\pi_t(p_0+h)$ as follows:

- If $\widetilde{w}_t(p_0+h) \geq \underline{w}$ at all dates, set $w_t(p_0+h) = \widetilde{w}_t(p_0+h)$ and $\pi_t(p_0+h) = \widetilde{\pi}_t(p_0+h)$ for all t.
- If $\widetilde{w}_t(p_0+h) < \underline{w}$ for $t \in [0, t^*]$, set $w_t(p_0+h) = \widetilde{w}_t(p_0+h)$ and $\pi_t(p_0+h) = \widetilde{\pi}_t(p_0+h)$ for all $t > t^*$ and set $w_t(p_0+h) = \underline{w}$ and:

  $$\pi_t(p_0+h) = \widetilde{\pi}_{t^*}(p_0+h)\, e^{-\left(r+\delta+\lambda_1\overline{\Gamma}(p)\right)(t^*-t)} + \frac{p_0+h-\underline{w}}{r+\delta+\lambda_1\overline{\Gamma}(p_0+h)}\left(1-e^{-\left(r+\delta+\lambda_1\overline{\Gamma}(p_0+h)\right)(t^*-t)}\right)$$

  for $t \in [0, t^*]$. (Note that t^* may depend on p_0.)

Step 5. Use $w_t(p_0+h)$ and $\pi_t(p_0+h)$ constructed at step 4 to solve for $\mu_t'(p_0+h)$ in the first equation of (22):

$$\mu_t'(p_0+h) = \int_t^{+\infty} \left[1+\lambda_1\gamma(p_0+h)\, v(p_0+h)\, \pi_t(p_0+h)\right] e^{-\left[r+\delta+\lambda_1\overline{\Phi}(p_0+h)\right](s-t)} ds.$$

Step 6. Start over at step 1 substituting p_0+h for p_0.