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)

March 2008

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. The new facts concern a strong comovement between the employer-to-employer worker transition rate, various measures of wages, and the share of employment at large firms. All three remain below trend several years into the expansion. Then, simultaneously, large firms take over employment, workers start quitting more from job to job, and wages accelerate.

We investigate whether this new view of how business cycles evolve and mature is consistent with the transitional dynamics of the Burdett and Mortensen (1998) equilibrium search model, analyzed in a companion paper (Moscarini and Postel-Vinay, 2008). In our model, following a positive aggregate shock to labor demand, wages respond little on impact, and start rising only when firms run out of cheap unemployed hires and start competing to poach and to retain employed workers. Aggregate shocks are thus propagated by the hiring behavior of large firms. A calibrated example shows that the model qualitatively captures the essence of the three facts.

*We would like to thank David Talan and Charles Carson from the Bureau of Labor Statistics for providing unpublished data on the firm size distribution in the BED data set, Bige Kahraman and Rafael Melo for excellent research assistance, Daron Acemoglu, Pat Kline, Rob Shimer and Gianluca Violante for very helpful conversations on this project, and seminar audiences at Bristol, Carlos III, Edinburgh, Essex, Calgary, the Chicago FED, NYU, Southampton, UCL, Wurzburg, Yale, and conference participants from the 2007 NBER Summer Institute and the 2008 ADRES/EDHEC conference on 'Labor Market Outcomes: A Transatlantic Perspective' for very useful feedback. The usual disclaimer applies.

†Address: Department of Economics, Yale University, PO Box 208268, New Haven CT 06520-8268. Tel. +1-203-432-3596. E-mail giuseppe.moscarini@yale.edu. Web http://www.econ.yale.edu/faculty1/moscarini.htm

‡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), 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, as well as residual wage inequality unexplained by observable worker and firm characteristics, we still do not know how wages are set in practice. 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 rise gradually and slowly over business cycle expansions?

The search-and-matching business cycle literature commonly assumes that wages are settled through a bargaining process, but does not identify the source of parties' bargaining power. 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 line of research we argue that the same poaching mechanism also transmits *aggregate* productivity shocks to wages and employment. Firms offer higher wages only when they run out of cheap unemployed job applicants and find it profitable to steal employees from their competitors, who in turn fight back and start paying more to retain their workers. Our argument builds on a set of 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. In Moscarini and Postel-Vinay (2008) we develop the first theoretical analysis of the out-of-steady-state behavior of the BM model.[1] In the present paper, we build on those results and present a quantitative simulation exercise, to gage the extent to which the BM model's quantitative predictions are congruent with the facts that we uncover.

The three new facts concern three seemingly unrelated time series. First, the share of employment at small firms and establishments declines through the 1990s and 2000s expansions, and crosses below its overall average around the middle of each decade; in contrast, large firms are hit particularly hard by recessions and then slowly recover employment share. Alternatively, the difference between the growth rate of employment at large firms/establishments vis-à-vis small ones is strongly procyclical, its troughs and peaks roughly coinciding with NBER business cycle dates, and rising uniformly in between. Second, the monthly Employer-to-Employer (EE) worker transition rate declines through the first half of each decade and rebounds starting in 1996 and in 2004. Third, various measures of detrended worker compensation exhibit a behavior quite similar to the EE rate, both on economy-wide and for each establishment size class, with workers at larger establishments earning more at all point in time. All in all, the mid-point of the expansion, several years after the previous trough, marks the point when the unemployment rate crosses below trend, large firms take over employment, workers start quitting more from job to job, and wages accelerate, all at the same time.

These facts suggest the following view of aggregate expansions. Following a positive aggregate shock to labor demand, wages respond little on impact, and start rising when firms run out of cheap unemployed hires and start competing to poach and to retain employed workers. 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 productive firms find it profitable to start raising wages to raid workers from less productive competitors. These respond by paying their workers more to retain them. Wages rise both within firms and as workers upgrade by quitting to higher-paying

[1] To our knowledge, Shimer (2003) is the only prior attempt to analyze aggregate dynamics in a wage-posting search model. See below footnote 20 for a brief discussion.

employers. Workers quit mostly from small, low-paying firms to large, high-paying firms. The growth in the employment of large firms is fueled by the stock of employment at small firms, which takes some time to replenish after a recession. Hence, employment at small firms grows faster and peaks earlier than at large firms. As employment at small firms peaks and poaching becomes a more important source of hiring, the EE rate picks up.

The facts that we highlight and use to support the above view of labor market expansions are limited in time and scope by data availability. The bulk of our data only covers the last two expansions, which have been uniquely "jobless". Thus we do not claim to have identified new business cycle stylized facts. What we do claim is that the patterns we point out for the last two expansions shed some light on the workings of labor markets in general, mainly by identifying the hiring behavior of large firms as an important channel for the propagation of aggregate shocks.

We formalize those views within a wage-posting model à la BM. A calibrated example of the transitional dynamics of the wage-posting model, following an aggregate productivity shock, exhibits many of the qualitative features of our new facts. Larger firms pay more at any point in time. Employment growth is faster at small firms early on, and then large firms take over as they poach employees. Wages rise slowly at first, and then accelerate. The job-to-job quit rate rises for some time as the pool of workers at small firms, vulnerable to poaching, expands.

A detailed theoretical analysis of the transitional dynamics of the BM model, which we simulate in this paper, is available in Moscarini and Postel-Vinay (2008). Due to the known complexity of this problem, in the present paper we restrict attention to a certain class of equilibria (which we call *Rank-Preserving*), as defined and motivated in the main text. We further confine the analysis to the deterministic transition to a steady-state following an unanticipated, permanent aggregate productivity shock. Before the shock, firms pay constant wages in a stationary world, as in BM. After the shock, firms post and commit to new contracts that pay wages contingent only on either calendar time or the unemployment rate. All workers in a firm are paid the same wage, no matter when hired and from where. Workers receive offers both off and on the job, and decide which ones to accept based on their implied PDV of wages, taking future transitions into account.

In our computed example, firms backload wages to the late part of the expansion. In steady state analysis of wage-tenure contracts, wages are backloaded late in the employment relationship to piggy-back on future poachers. If workers are risk-neutral, backloading is extreme in the form of a step function of tenure (Stevens (2004)). Gradual wage growth requires risk aversion and extreme market incompleteness as a motive for consumption smoothing (Burdett and Coles, 2003). In our setup, wages grow slowly even if workers are risk neutral and capital markets are perfect; wage backloading occurs over calendar time, not over tenure. Beyond the piggybacking motive, the increasing scarcity of cheap hires from unemployment makes raising wages more attractive in order to poach employed workers, thus also to retain own employees.

While the process of upgrading from job to job is usually described as climbing a wage ladder, our non steady state analysis reveals that it is best described as jumping from a "wage escalator" up to a higher one. All workers benefit as wages rise within firms. In addition, job changers rise from one rising wage profile to a higher one at another firm. Therefore, aggregate wages rise for two reasons: on the intensive margin, all workers are paid progressively more and, on the extensive margin, workers move to higher-paying firms.

In our example, job changers exhibit faster (so, in this limited sense, more procyclical) wage growth than job stayers. This prediction is consistent with the main findings of the literature on real wage cyclicality (Bils, 1985; Beaudry and DiNardo, 1991; Solon, Barsky and Parker, 1994). This literature, however, has been hampered by the lack of high-frequency, reliable information on job-to-job transitions in a representative sample. We present such evidence from the monthly CPS starting in 1994, when it first became available, but we have not yet linked it to individual wage information. A similar view holds that workers flow from low-wage to high-wage industries in an expansion, a "Cyclical Upgrading of Labor" (Okun, 1973). This is based on the observation that high-wage industries have more cyclical employment. Barlevy (2001), however, shows that the inter-industry wage gains reflect to some extent the compensating differential associated to accepting riskier jobs. While the source of inter-industry wage differentials is still an open question, we show that the patterns of employment growth by firm and establishment size that we uncover hold within, and not across industries.

Our example also features a strong propagation in wages and labor productivity, which keep rising years after the initial shocks, although the unemployment rate's half life is just a few months. The main reason is that job-to-job transitions in the data are an order of magnitude slower than the reallocation from unemployment to employment. Thus, the upgrading process is slow, and so is the rise in labor productivity after an initial jump following the shock. The propagation is less pronounced for the EE rate. We are, however, ignoring further sources of propagation of the unemployment rate, such as endogenous labor force participation. This is likely to rise in the expansion, feeding the market with relatively cheap candidates for hiring from unemployment, and delaying the moment when large firms have to start raising wages aggressively to poach workers, the small firms have to respond to retain them, and the EE rate peaks.

The next section lays out the facts and offers an intuitive explanation of these based on the BM equilibrium search model. The rest of the paper discusses the theoretical model and its quantitative predictions. Section 3 describes the basic economic environment. Section 4 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 5. Finally, Section 6 concludes and gives some thoughts about possible further research avenues.

2 Aggregate Labor Market Fluctuations: New Evidence

2.1 Definitions and Overview

In order to organize the new evidence under a common heading, we open this section by providing our own definition of a tight labor market. We apply an HP filter to both the monthly civilian unemployment rate and the monthly civilian employment to working age population ratio (E/POP ratio) from the BLS in 1948:1-2008:1, exactly 60 years of data. We choose a smoothing parameter equal to $10^5 \times 3^{2.25}$, which corresponds to Shimer's (2005) choice of 10^5 at a quarterly frequency.[2] In

[2] While Shimer's value of 10^5 is unconventional and has been criticized as smoothing too much, we also tried a more conventional quarterly parameter of 1600; transformed into a monthly frequency by multiplying it by 3^n for $n = 2, 3$, or 4. The resulting trends cross well above and below the actual values even mid-way through NBER-dated expansions, suggesting that lower-frequency demographic changes are contaminating it. Our choice for monthly smoothing, equivalent to Shimer's quarterly value, is approximately equal to 1600×3^6.

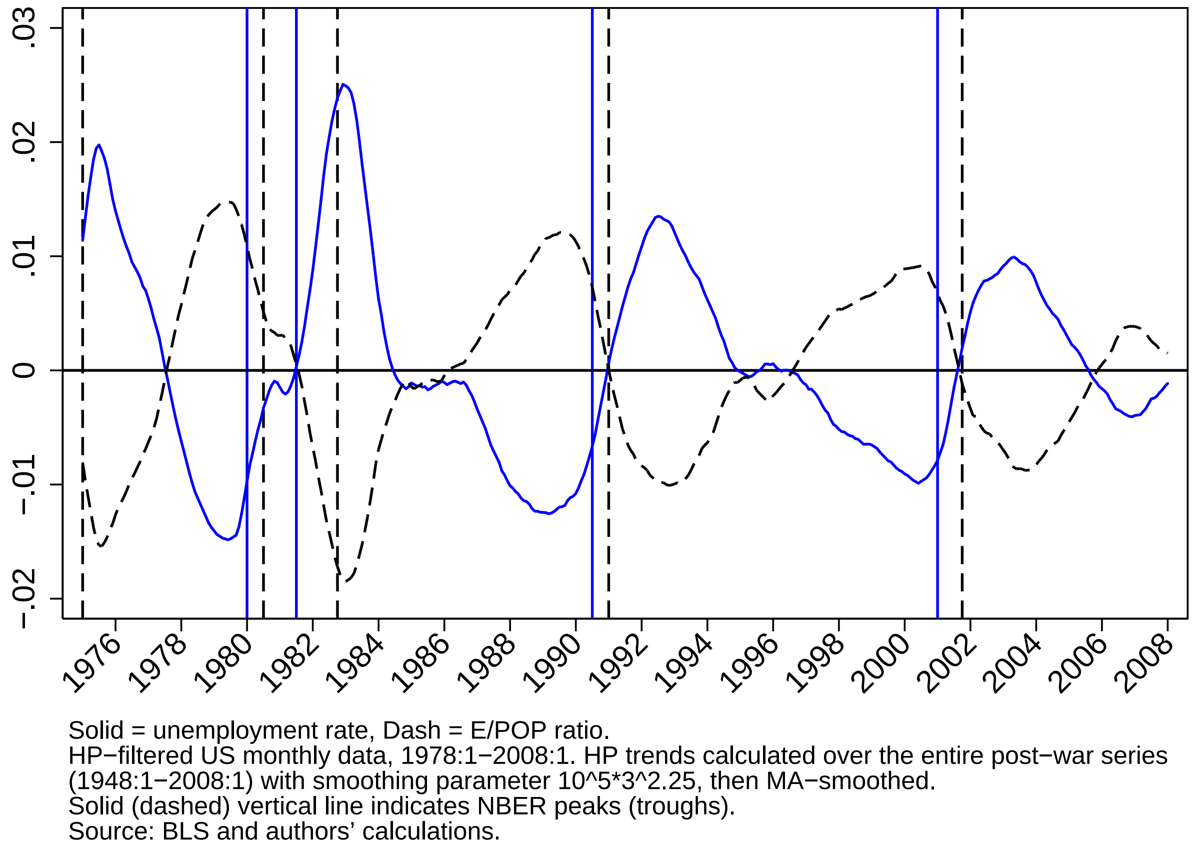


Fig. 1: Employment and unemployment rates

Figure 1 we plot the deviations of the unemployment rate and the E/POP ratio from their respective post-war trends.[3] We focus only on 1975-2008 (although the trends use data since 1948), because this is the largest span covered by our other data series. Clearly, the unemployment rate crosses its trend from above in 1977, 1984-1985, 1995-1996, and 2005. In the second and third episode, the unemployment rate hovers around trend for almost two years, but is higher before and lower after that point in time, within the cycle.[4] Armed with this graph, we define the labor market to be tight when the unemployment rate first falls below its trend and remains there for at least a year thereafter:

[3] An MA smoother was further applied to the HP-filtered series. All infra-yearly (i.e. monthly or quarterly) series plotted in this section are MA-smoothed for legibility.

[4] This "stickiness", reinforces the notion that this is a meaningful de-trending procedure: once it reaches its trend, the unemployment rate tends to stay there, until pressure on the labor market sets it into motion again, resuming the previous movement until the next recession that turns it around.

Definition 1 (Tight/Slack Labor Market) *A labor market is tight in month t if the unemployment rate u_s is below its trend in months $s = t, \cdots, t+T$ where $T > 11$. The symmetric definitions holds for a slack labor market. The labor market is neutral at date t when it is neither tight nor slack at that date.*

The window of at least 12 months of consecutive observations above or below trend, before/after first crossing it, is necessary to eliminate neutral phases, when the unemployment rate hovers near trend. Notice that this definition does not coincide with that of an expansion or recession according to NBER dates (materialized on all figures by vertical lines, solid for peaks and dashed for troughs). Indeed, the phases of labor market neutrality occur typically roughly mid-way between troughs and peaks, although the unemployment rate itself does not have symmetric dynamics, typically rising faster than it declines.

Also note that, although we have chosen to couch this definition in terms of the unemployment rate, it is evident from Figure 1 that the E/POP ratio is an equally good indicator of labor market tightness as defined above, as the cyclical behavior of the E/POP ratio is essentially the mirror image of that of the unemployment rate (the correlation coefficient of the two series over the entire post-war period is -0.92). In other words, as far as labor market cycle dating is concerned, cyclical variations in labor force participation can be ignored to a first approximation. We will implicitly make this approximation throughout the paper, especially in the theory where the participation margin will be shut down altogether.

The three new facts that we uncover and document in this section concern three seemingly unrelated time series in the last few decades. Armed with our definition of tight and slack labor markets, they can be briefly characterized as follows. We first describe them verbally, and later graphically, rather than reporting correlations of detrended series, because many of the available time series cover only a couple of business cycles.

Fact #1. Small firms and establishments grow in size faster than large ones when the labor market is slack, and vice versa when the labor market turns tight. Therefore, the firm/establishment size distribution gets compressed in a loose labor market and becomes more unequal in a tight one.

Fact #2. The rate at which employed workers quit to other jobs is above trend in a tight labor market, and below trend in a slack labor market.

Fact #3. The annual growth rate of real wages or weekly earnings is above its trend in a tight labor market, and below in a slack labor market.

We now describe the three facts in detail.

2.2 The Cyclical Dynamics of the Firm/Establishment Size Distribution

Most of the new empirical evidence that we present in this paper pertains to Fact #1, the relative timing of net job creation within a business cycle, by size classes. Essentially, of the total job creation observed from cyclical trough to peak, more of it occurs early on in the expansion for small firms/establishments than for larger ones, and vice versa. In recessions, large firms/establishments are hit particularly hard. Notice that this fact does *not* say that most job creation occurs at small establishments, a well-know subject of political rhetoric and confusion. While the political debate is typically focused on growth rates, it is still the case that the *number* of jobs added by large establishments is higher than that added by small ones or by entrants at nearly all points in the aggregate expansion.[5]

Reflecting the dual structure of its statistical agencies, the US Federal Government has produced two comprehensive and independent sources of information on firms and establishments, specifically on their count, employment, payroll, industry and location. Business Employment Dynamics (BED) is a data set providing quarterly information on job flows and firm sizes since 1992. It is based on the Quarterly Census of Employment and Wages, which in turn collects information from the State and Federal Unemployment Insurance (UI) systems accruing to the Bureau of Labor Statistics for this purpose. The BED covers all employers subject to UI taxes, which account for 98% of

[5]Considering standard cutoff sizes to define "small" *vs.* "large" businesses, we can use data from the County Business Pattern files (see below for a presentation of this data set) to compare establishments with a workforce in excess of 500 employees to establishments in the category 1-19 employees. The latter create more jobs (or destroy fewer) around recession years, i.e. 1989-1990, 1990-1991, less so in 1991-1992, and again in 2001-2002. In all other years, 1992-2000 and 2002-2004, large establishments create more jobs in total. Details are available on request. On this subject see also Neumark, Wall and Zhang (2008).

employment. Whilst used primarily to measure job flows, information on employment stocks is publicly available when aggregated into nine classes of firm size.[6]

The second dataset is the County Business Patterns (CBP), which is maintained at the Bureau of the Census and derives from the Business Register and quinquennial Economic Census. The publicly available information from the CBP that we have been able to secure consists of annual information on counts, employment, industry, payroll, and location for all US firms (since 1988) and establishments (since 1990), as well on non-employers (self-employed) for 1992 and 1997-2005. This dataset forms the basis of a panel of all US establishments, the Longitudinal Business Database (LBD). Access to the microdata is restricted and we are currently in the process of submitting an application to gain it. Limited longitudinal information on employment growth by initial establishment size in CBP is publicly available, under the name of Business Information Tracking Series (BITS) at the Census Bureau and Dynamic Size Information at the Small Business Administration.[7]

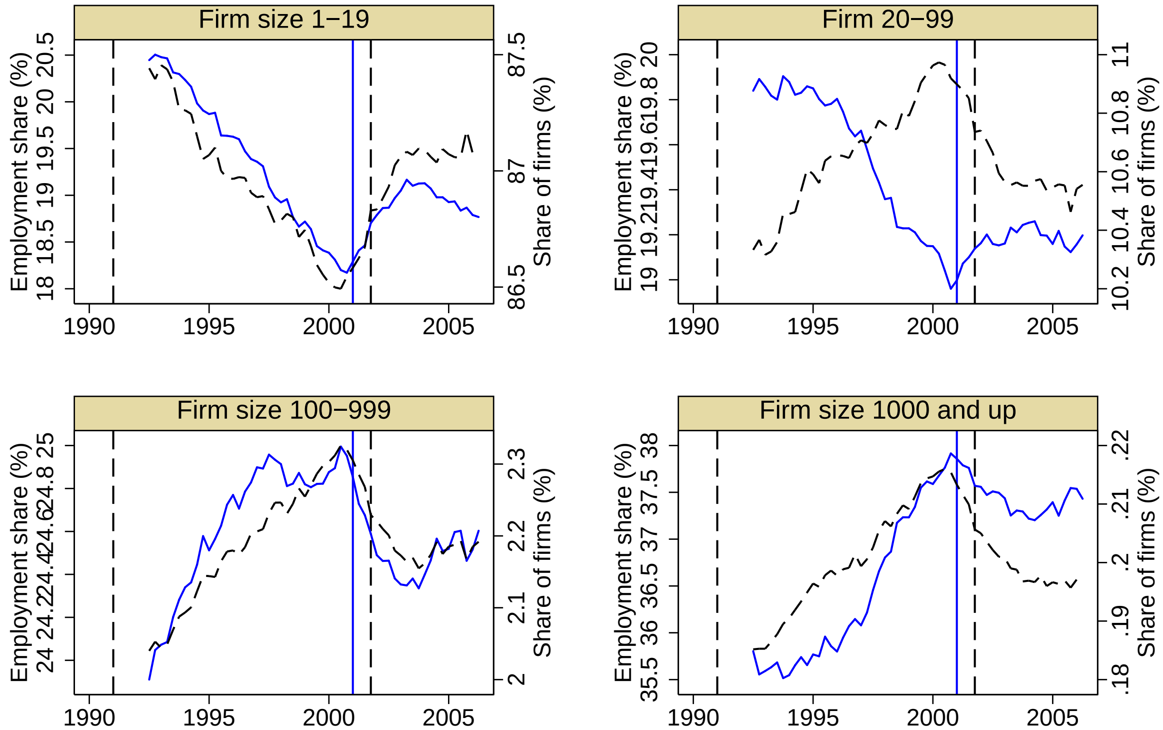
2.2.1 Employment Shares and Growth Rates by Size Classes

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 2 which plots employment shares for four different classes. 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.[8]

[6]For more information and to download the data: `www.bls.gov/bdm/`. David Talan and Charles Carson from the BLS kindly tabulated for us the firm size distribution for all available quarters.

[7]Visit `www.census.gov/csd/susb/susb.htm` and `www.sba.gov/advo/research/data.html` for more information and data files.

[8]This pattern is not 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 less concentrated on 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. Data from the UK Labor Force Survey that we have elaborated (not reported, available on request) convey information about the share of workers whose workplace has less than 25 employees. Also the UK Small Business Administration publishes data on the shares of UK employment in firms with less than 20 and more than 249 employees. Both sources suggest that small UK firms lost employment to large ones in expansions, with sudden reversals in the 1991



Solid = employment share, Dash = share of firms.
Solid (dashed) vertical line indicates NBER peaks (troughs).
Source: BED (quarterly) and authors' calculations.

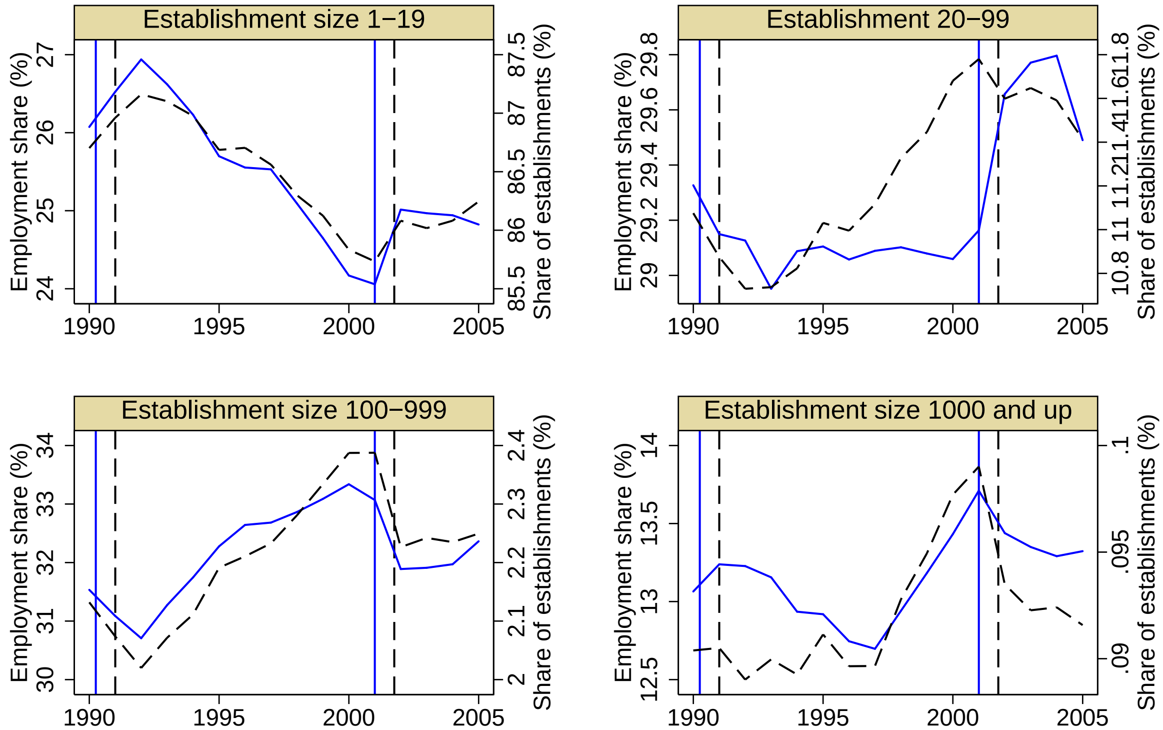
Fig. 2: Fraction of firms and employment shares — small *vs.* large firms

A similar graph can be constructed from *establishment*-level data using the CBP since 1990 (see Figure 3). 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, however, large establishments tend to be part of large firms, as shown in Table 1.[9]

The evidence for firms from BED can also be illustrated in terms of growth rates of employment rather than employment shares. The top two panels of Figure 4 plot average employment growth rates between large and small firms. The bottom two panels report the *difference* in average employment growth rates between large and small firms (the two series on the corresponding top

and 2002 aggregate slumps, just like in the US.

[9]Table 1 was constructed based on figures from the "Statistics about Business Size" page of the Census web site (`http://www.census.gov/epcd/www/smallbus.html`).



Solid = employment share, Dash = share of establishments.
Solid (dashed) vertical line indicates NBER peaks (troughs).
Source: CBP (yearly) and authors' calculations.

Fig. 3: Fraction of establishments and employment shares — small *vs.* large establishments

panel): when the series is positive, large firms grow faster, and vice versa.[10] The two columns of Figure 4 relate to two different definitions of the "small" and "large" classes, as indicated on each panel. This highlights more vividly the pattern of Figure 2: wherever one places the cutoff between small and large firms, small firms appear to have grown (relatively) faster than large firms at the beginning of the 1990's expansion, and slower later on in that same expansion. A similar pattern seems to appear again at the beginning of the 2000's expansion. Focusing on the top two panels of Figure 4, one can further decompose that pattern by noticing that average size growth of small firms actually trends down over the expansion, while that of large firms does not trend. This is why large firms take over on average. Hence these series of average growth rates by firm size class, albeit all clearly procyclical, diverge to some extent.

[10]On both bottom panels the straight line materializes a linear trend and the shaded area shows the 95% confidence band around that trend.

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.

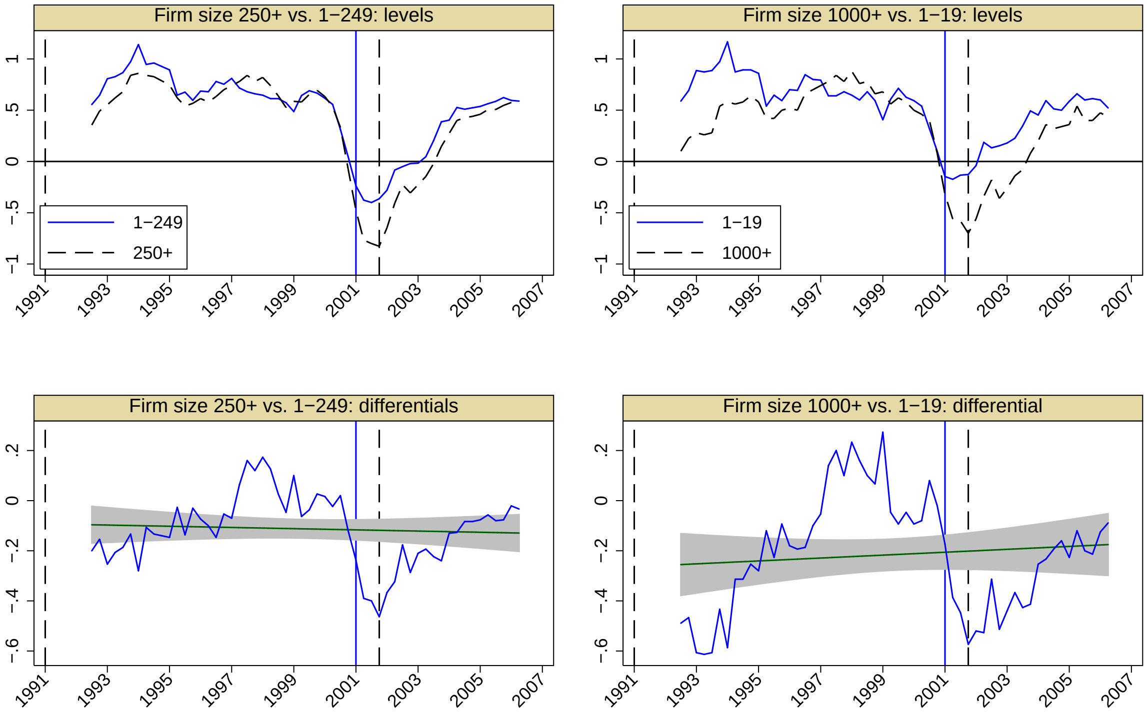
Table 1: Establishment and firm sizes

Similar firm-level evidence can be constructed at annual frequency, but beginning in 1988, from the CBP data which is available at the firm (as well as establishment) level. CBP only conveys information on employment shares, which is our main focus, and not on firm counts, as this particular version of the data aggregates non-employers with small firms in some years and not in others. This inconsistency makes cross-date comparisons of firm counts impossible, but does not affect employment as we have been excluding the self-employed throughout. Figure 5 repeats Figures 2 and 4 using CBP firm-level data and confirms the pattern already observed in BED, not only for the 1990's and early 2000's but also around the previous recession, as we now also encompass 1988-91.

Finally, and to further illustrate this pattern, Figure 6 shows the comovements between growth differentials across firm size classes (the series on the right panel of Figure 4) and the aggregate unemployment rate.[11] Although the two series fluctuate with a different amplitude, one sees that they exhibit very strong comovements: the correlation between growth rate differentials and the unemployment rate series is about -0.76 over the observation window.[12] Thus overall, recalling

[11] All in deviations from trend. The trend used for the unemployment rate is the HP-filtered series mentioned above, and the trend used for the (much shorter) series of growth rate differentials is just a linear trend. Note that a very similar-looking graph obtains when one uses the E/POP ratio instead of the unemployment rate as an indicator of labor market tightness.

[12] Interestingly it would seem that the growth rate difference is a leading indicator of the unemployment rate. The correlation between unemployment and the growth rate difference lagged one quarter is about -0.82.



Differential firm size growth, MA−smoothed.
Solid (dashed) vertical line indicates NBER peaks (troughs)
Source: BED and authors' calculations.

Fig. 4: Difference in average growth rates of employment across firm size classes.

our adopted definition of labor market tightness, we conclude that employment growth of small firms and establishments occurs relatively more when the labor market is slack, and vice versa for large firms.

2.2.2 The Reclassification Bias

While the statement concluding the last sub-section might appear plausible and intuitive, the following issue still needs to be resolved: the identity of firms in different size classes changes over time. Firms and establishments are reclassified at each observation date according to their new size — indeed the BED data set explicitly applies "dynamic allocation" of firms to size classes, i.e. it even changes class assignments at infra-quarterly frequency for firms crossing the line between two size classes. Thus, in a growing economy, where firms gain size on average, one would expect the distribution of firm size to rise in a stochastic dominance sense. We refer to this issue as the

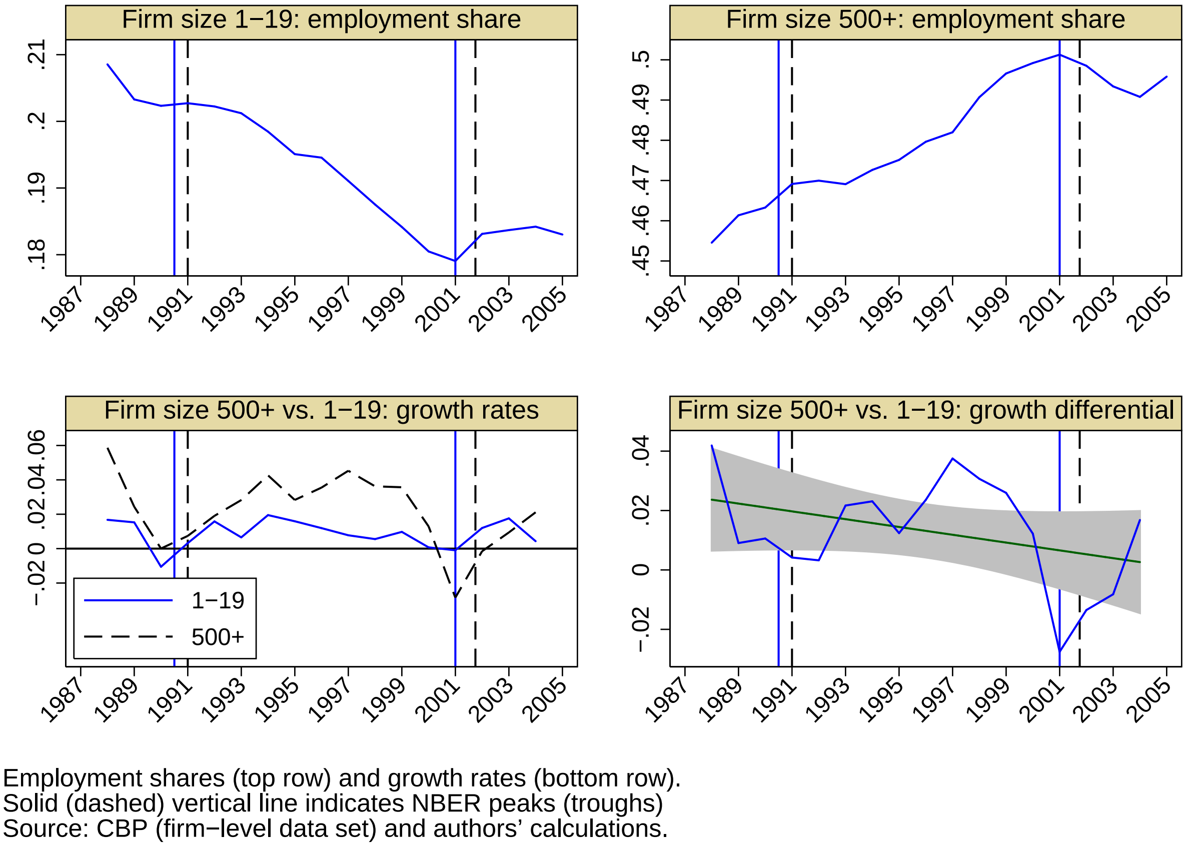


Fig. 5: Employment shares and employment growth by firm size classes, CBP firm-level data.

reclassification bias.

To tackle the reclassification issue, one can exploit data sources where longitudinal links on firms/establishments make it possible to avoid reclassification altogether. Ideally, we would like to fix the assignment of firms to size classes (or average wage classes, or value-added classes. . .) at a cyclical trough, then track over the expansion the shares of employment at these classes of firms without re-classifying them every period. This would allows us to verify, for example, whether firms/establishments that are initially small often leapfrog and overtake the initially larger ones, a phenomenon that would invalidate our interpretation of the facts.

To this purpose, we utilize two publicly available US longitudinal data sets on firms/establishments. The first is Compustat, which comprises only listed companies. We fix firm identities in 1975 and classify them in size bins (by employment). Then, for each year from 1976 to 2005, we calculate

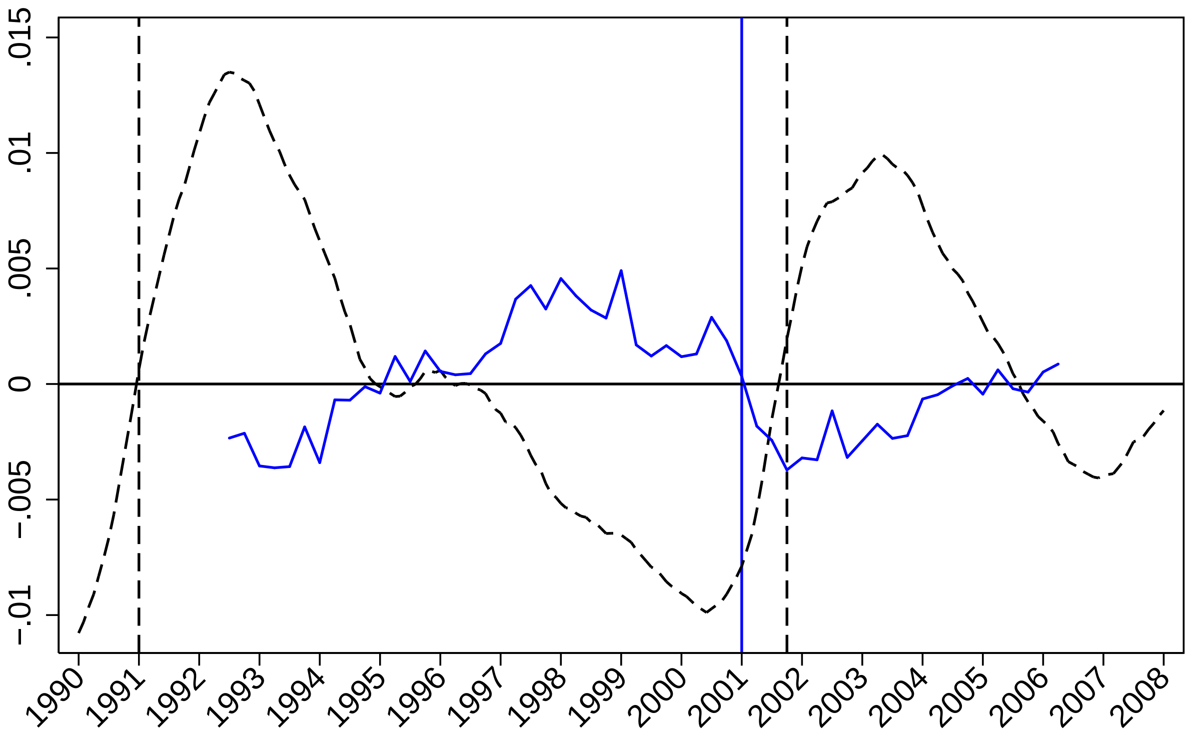


Fig. 6: Growth rate differentials, employment and unemployment rates.

the growth rate of employment over the past year at firms that had more than 5,000 employees in 1975 and subtract the growth rate of the other firms in the sample, which were smaller than 5,000 in 1975. The reason for the large size cutoff is that listed companies are very large. We plot this difference in growth rates in Figure 7, in a way that mimics Figure 4. Consistent with the pattern uncovered in the BED, over three full business cycles this difference in growth rates is procyclical, and crosses zero when the labor market turns tight.

Compustat is not a representative sample. The CBP is. Although the underlying source of micro data is not publicly available, the Census publishes employment growth by establishment size, where establishments are assigned to categories according to size in March each year and stay in the same class for one year.[13] This dataset, BITS, exploits the longitudinal links in CBP, for

[13]Visit `www.census.gov/csd/susb/susb.htm` for more information and data files.

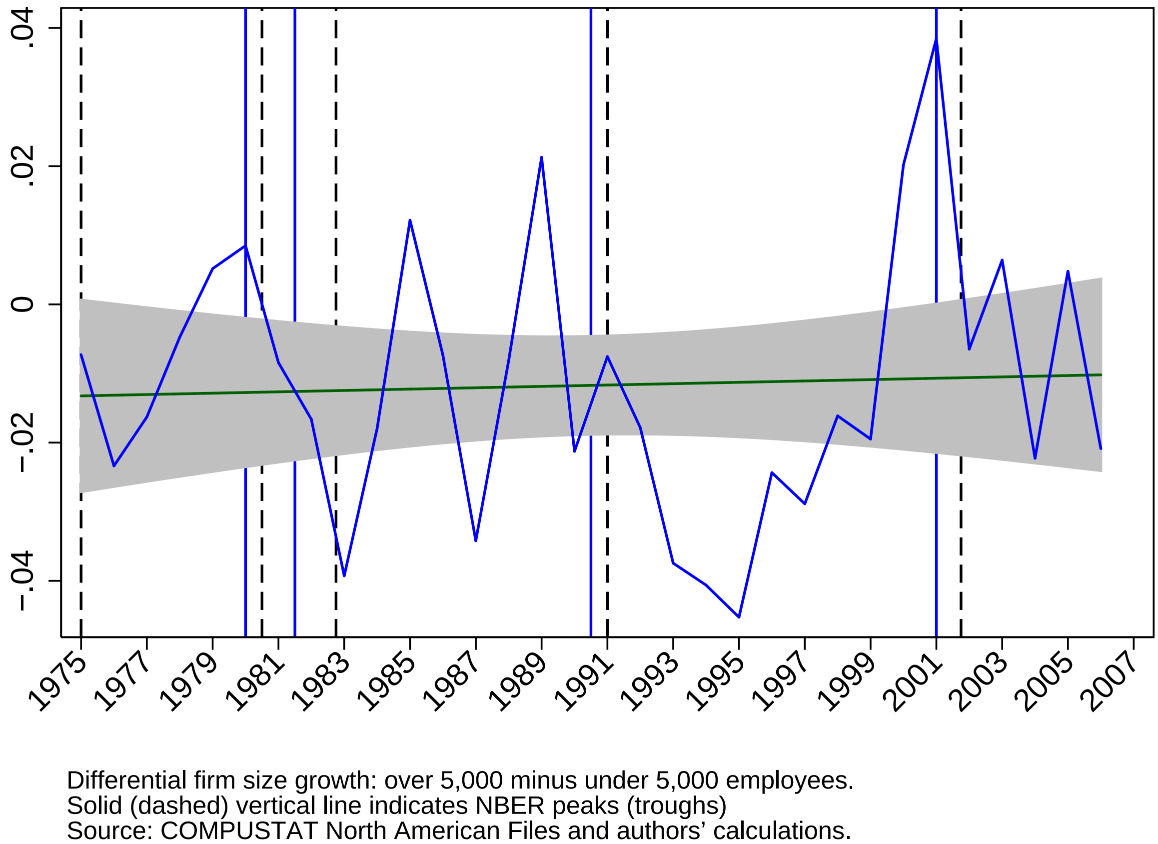


Fig. 7: Growth rate differentials across size classes for a fixed sample of publicly traded US companies classified by size in 1975.

just one year at a time. In Figure 8 we plot our findings from BITS, again in terms of difference in growth rates. Every year since 1989, that is since BITS data are publicly available, we compute the growth rate of employment over the subsequent year among establishments that started above 500 employees in March of year t and subtract the growth rate for the < 500 class. A positive number indicates that establishments that were larger in March of year t grew faster than the other ones until March of year $t+1$, when they were reclassified after computing the relevant growth rates. The evidence is again fairly clear-cut, whether one detrends or not: small establishments grow faster early in an expansion; the pattern then slowly reverses, to switch back abruptly during recessions. The differences in annual size growth rates are large.

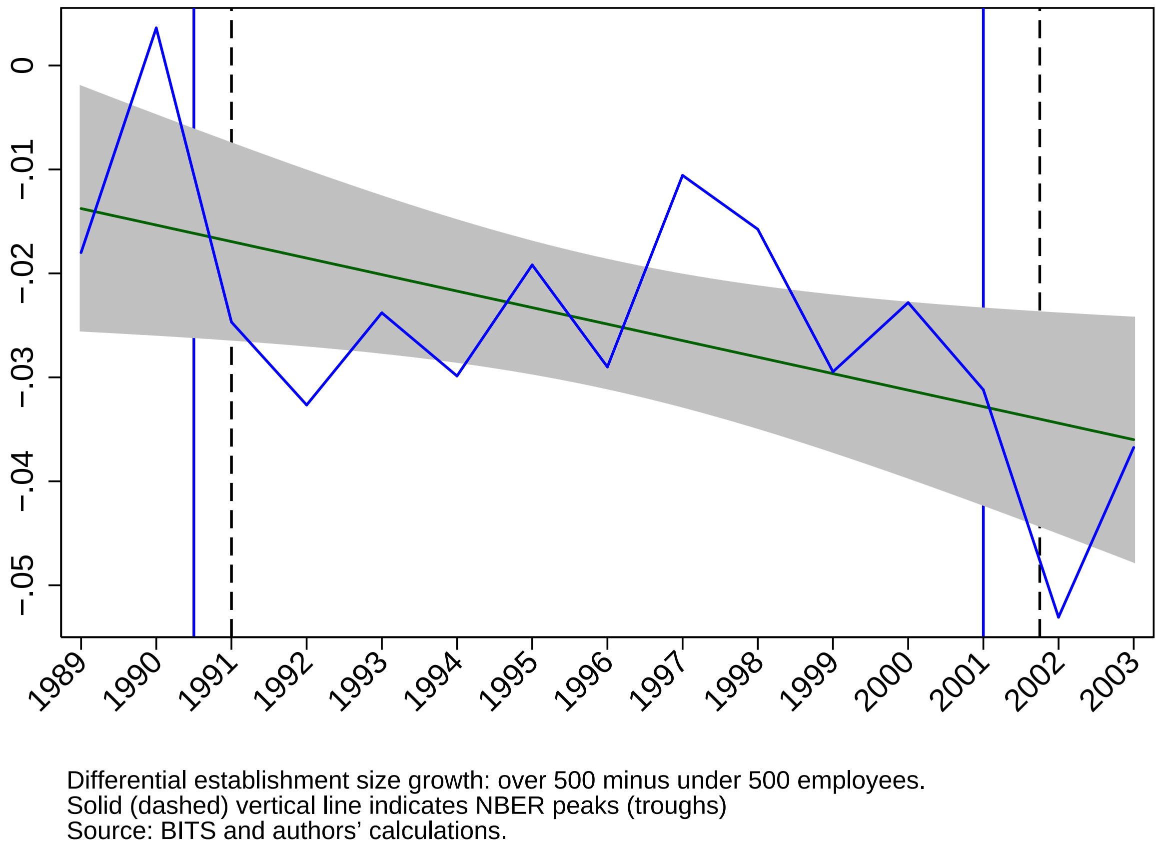
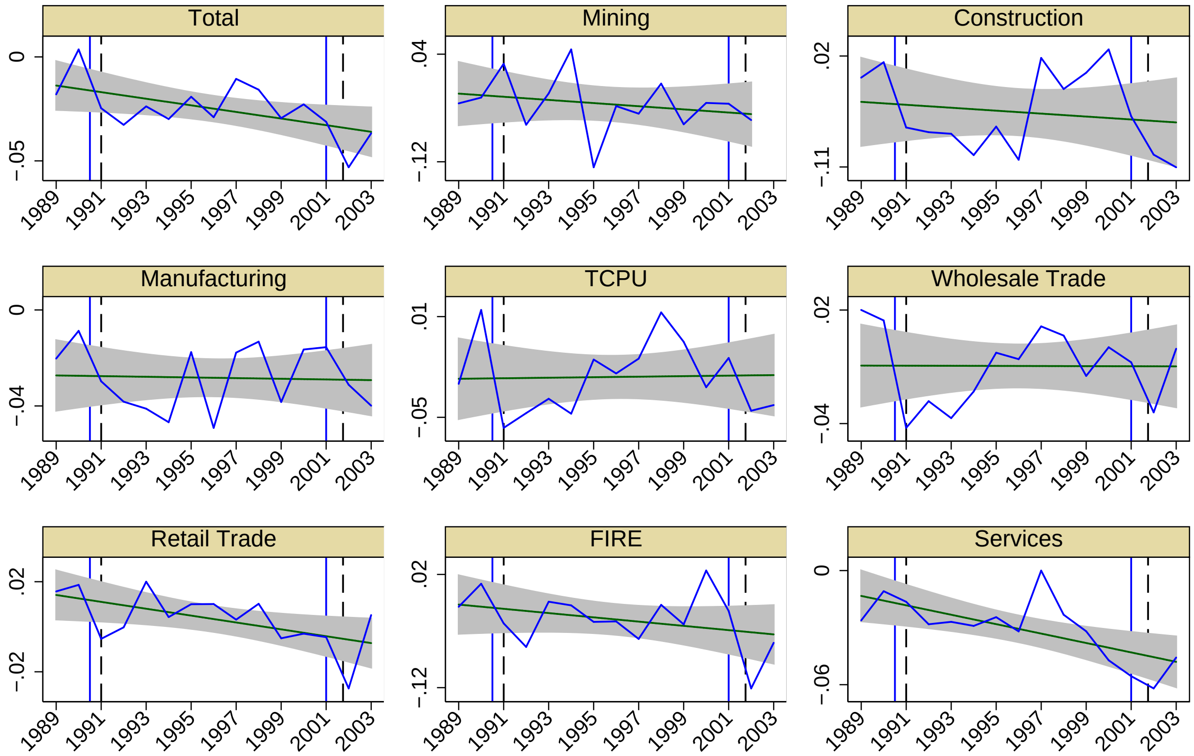


Fig. 8: Growth rate differentials across size classes for all US establishments, reclassified every year in March *after* computing the growth rate over the past year.

2.2.3 Industry-level Evidence

An additional appeal of the CBP/BITS data set is that it provides industry-level information. This enables us to gage whether the pattern identified in this subsection is also observed within industries or whether it mostly reflects a reallocation of labor across industries. More specifically, is it the case that small firms exhibit faster growth at early stages of an expansion while large firms take over later on *uniformly across industries*, or is it that job creation early on is concentrated at industries that tend to be populated by smaller firms, while industries with larger firms have a "slower start"?

Figure 9 provides a crude answer to these questions by plotting equivalent series to the one dis-



Differential establishment size growth: over 500 minus under 500 employees.
Solid (dashed) vertical line indicates NBER peaks (troughs)
Source: BITS and authors' calculations.

Fig. 9: Growth rate differentials across size classes for all US establishments, size classes fixed over a year, by industry.

played on Figure 8 for eight broad industries.[14] Although industry-level data are inevitably more noisy, the aggregate pattern shown on the upper-left panel (a repetition of Figure 8) seems by and large to apply across industries. In all cases, without exceptions, the disaggregated industry series drops on or around recessions. They cross the trend from below only around 1996 in five of the eight industries: Construction, Transportation Communication and Public Utilities (TCPU), Wholesale Trade, Finance Insurance and Real Estate (FIRE), Services. The pattern is similar in Manufacturing, except for the sharp drop in 1996. Manufacturing is the industry where the firm/employment size distribution is most skewed towards large values, an outlier in the economy at large, so small establishments play a lesser role. The pattern is less clean in Retail Trade, but even there small

[14]Details of how we have mapped SIC and (subsequently) NAICS codes into these eight industries are available on request.

establishments grow faster early on after the recessions. Mining is the only (small) industry that does not fit the general picture at all. The effect on the other (between-industry) margin can also be assessed using the BITS data, for instance by looking at the difference in employment growth between industries populated by establishments whose size is above the economy-wide average and industries with comparatively smaller establishments. A plot of this difference (not reported here) suggests that this between-industry difference has no particular cyclical pattern, contrary to the within-industry series plotted in Figure 9. This finding will make an interesting contribution to our discussion of how to interpret the facts below.

2.3 The Cyclical Dynamics of the Employer-to-Employer (EE) Transition Rate

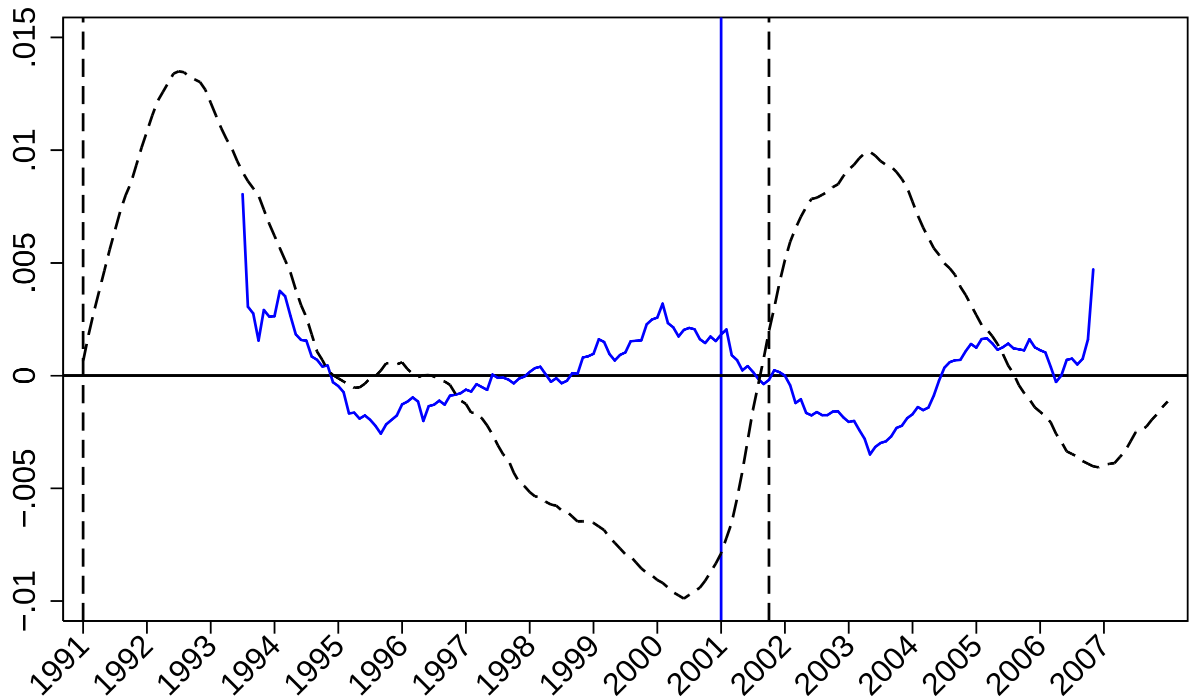
Monthly CPS data available since 1994 allow the construction of monthly EE transition rate series.[15] Such a series, compiled by Moscarini and Thomsson (2008), is plotted in Figure 10 (in deviation from HP trend), together with the unemployment rate series from Figure 1 to highlight the EE rate's cyclical behavior in terms of our adopted definition of labor market tightness.

Figure 10 shows that the EE transition rate was actually falling below trend 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. Therefore, the job-to-job transition rate is nearly perfectly procyclical if we take our measure of labor market tightness (based on the the unemployment rate) as the cyclical indicator, while it lags more conventional indicators such as GDP- or NBER-dated troughs by many years.

Finally, publicly available data from the Census Survey of Income and Program Participation (SIPP)[16] conveys information about the workforce size of an individual's current employer (employing establishment), as well as individual job histories. This allows a crude analysis of the poaching

[15]The critical change in CPS interviewing procedures, which made measurement of the EE transition rate possible, is the introduction of Dependent Coding of industries and occupations. Before asking specific questions about those, starting in 1994 CPS interviewers first read to the survey respondent his/her answers from previous month, and asked whether anything, including the worker's employer, had changed. The resulting EE transition rate cannot detect unemployment spells that last less than three weeks. So, strictly speaking, it is not a job-to-job quit rate. The definition of such a quit rate is however, to some extent subjective. The worker may take a voluntary break before starting a previously lined up new job. Furthermore, Moscarini and Thomsson (2008) also control for job search activity in the intervening month, and find little difference.

[16]Information about the SIPP, as well as data files are available at `www.bls.census.gov/sipp/`.

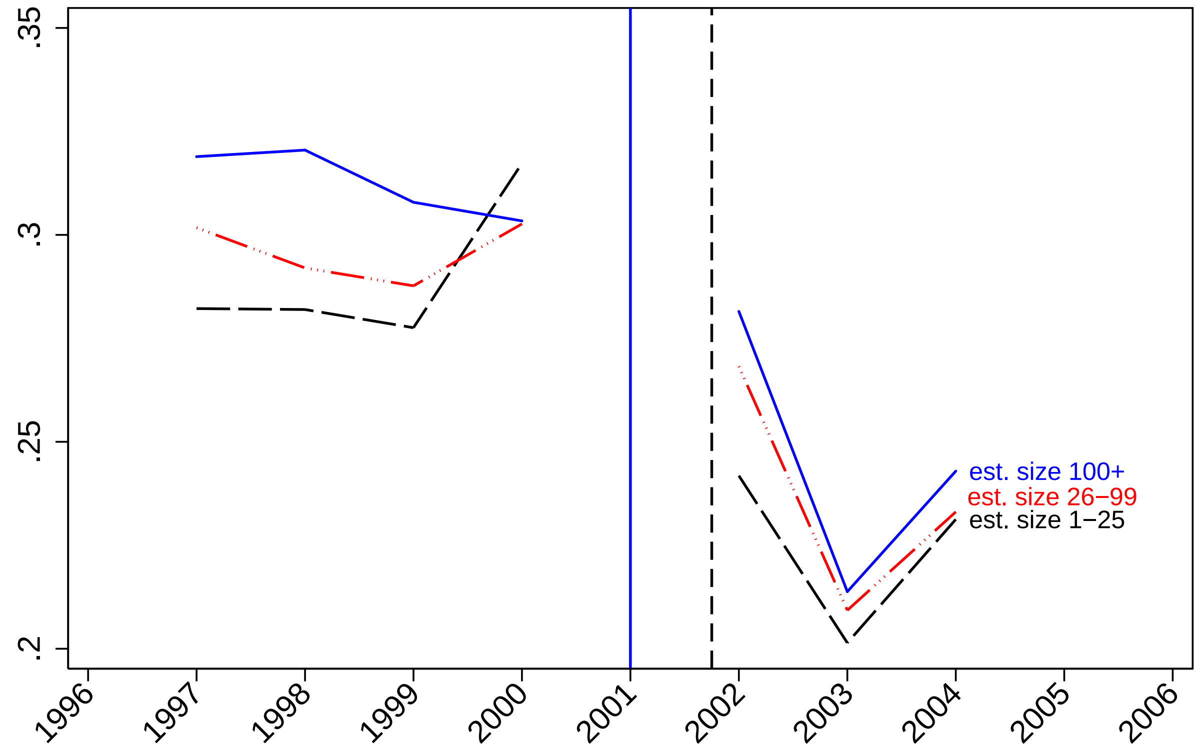


Solid = EE rate, Dash = unemployment rate.
EE rate: US monthly data, deviation from linear trend, MA−smoothed.
Solid (dashed) vertical line indicates NBER peaks (troughs).
Source: CPS, BLS and authors' calculations.

Fig. 10: Monthly transition rate of employed workers to other employers

activity of establishments as a function of their size. Figure 11 plots a measure of the fraction of new hires coming from another employer (i.e. following an EE transition) for three categories of establishment size.[17] In other words, it plots a measure of the importance of poaching in the recruitment activity of establishments, by establishment size category. Changes in the design of the SIPP and other data limitations restrict the period over which that indicator can be constructed to the years shown on Figure 11. While this admittedly constitutes very limited evidence, we still notice the following two points. First, and consistently with the evidence presented in 10, poaching was more intense in the latter half of the 1990s expansion than in the immediate aftermath of the 2001 recession. This is true for all three categories of establishment size. Second, larger estab-

[17]Specifucally, it is constructed as the fraction of workers who have changed employers in the previous year and are now employed at an establishment in size category X *without work interruption* among all workers having changed employers in the previous year and now employed at an establishment in size category X.



Fraction of new hires coming from other employers, by establishment size category.
Solid (dashed) vertical line indicates NBER peaks (troughs)
Source: SIPP and authors' calculations.

Fig. 11: Poaching and establishment size

lishments almost always poach more than smaller ones. This difference in "poaching intensity", however, is more pronounced in 1997-1999, when the labor market turns tight, than in 2002-2004, when it is slack.

2.4 The Cyclical Dynamics of Wages and Earnings

Third, publicly available data from the BLS Current Employment Statistics on average monthly 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 12). The bigger picture exhibits similar patterns for the preceding five business cycles—with the notable exception of the 1980s expansion, when a sharp decline in real wages of unskilled workers gave rise to the well-known increase in wage inequality.

In order to control for composition effects in employment, we gather information on earnings,

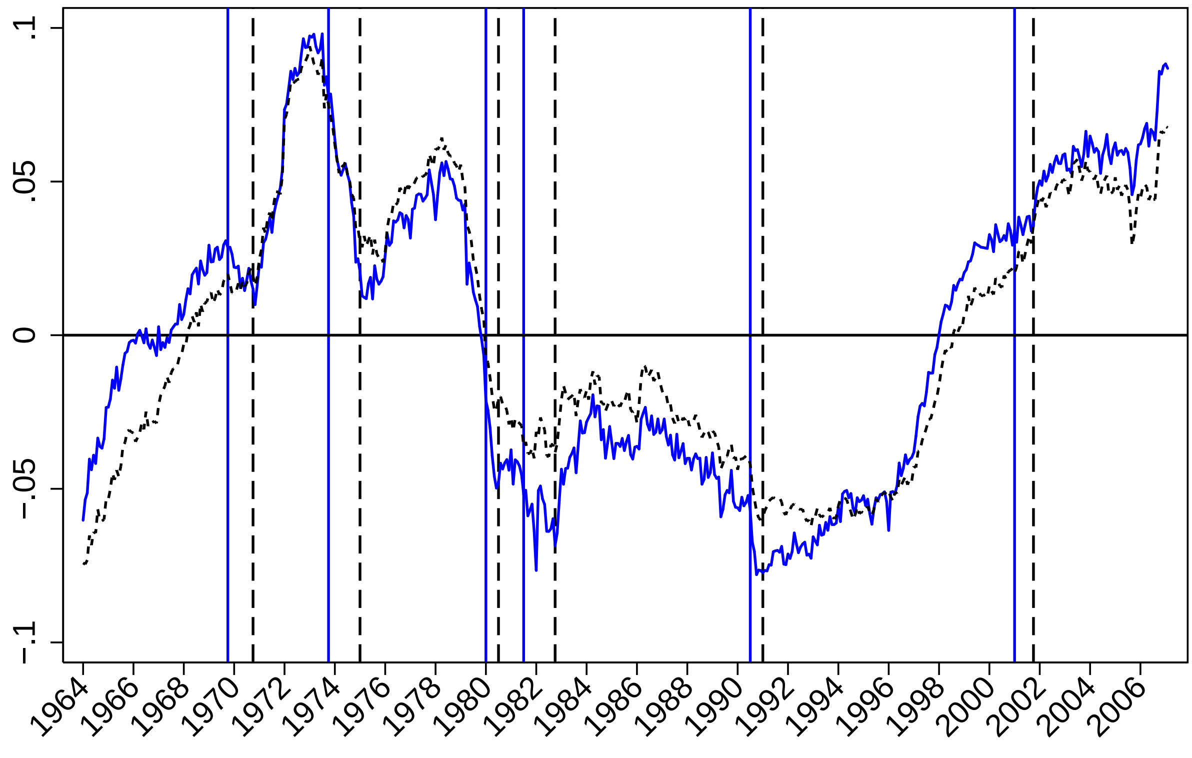


Fig. 12: Average earnings and wages

hours and demographics from the Merged Outgoing Rotations of the monthly CPS, starting in 1982 when they became available. We take percentage wage growth over subsequent 12 months for the same individual as the object to be explained, in order to eliminate fixed effects in wage levels. We regress wage growth on many demographics, to capture also composition effects in wage changes. Then, we plot the median residual from this regression month by month in Figure 13. We also plot the 52nd percentile of the residual distribution, in order to take into account the proportion of topcoded earnings (between 0 and 2% per month) which create a spurious mass of zero or very low wage growth. The unemployment rate (again in deviation from HP trend) is also superimposed to gage cyclicality. The overall picture is qualitatively unchanged. Relative to the first month, January 1982, rescaled to zero, unexplained wage growth is positive in the late 1980s and especially late 1990s when the labor market was tight according to our definition, and negative in early 1990s

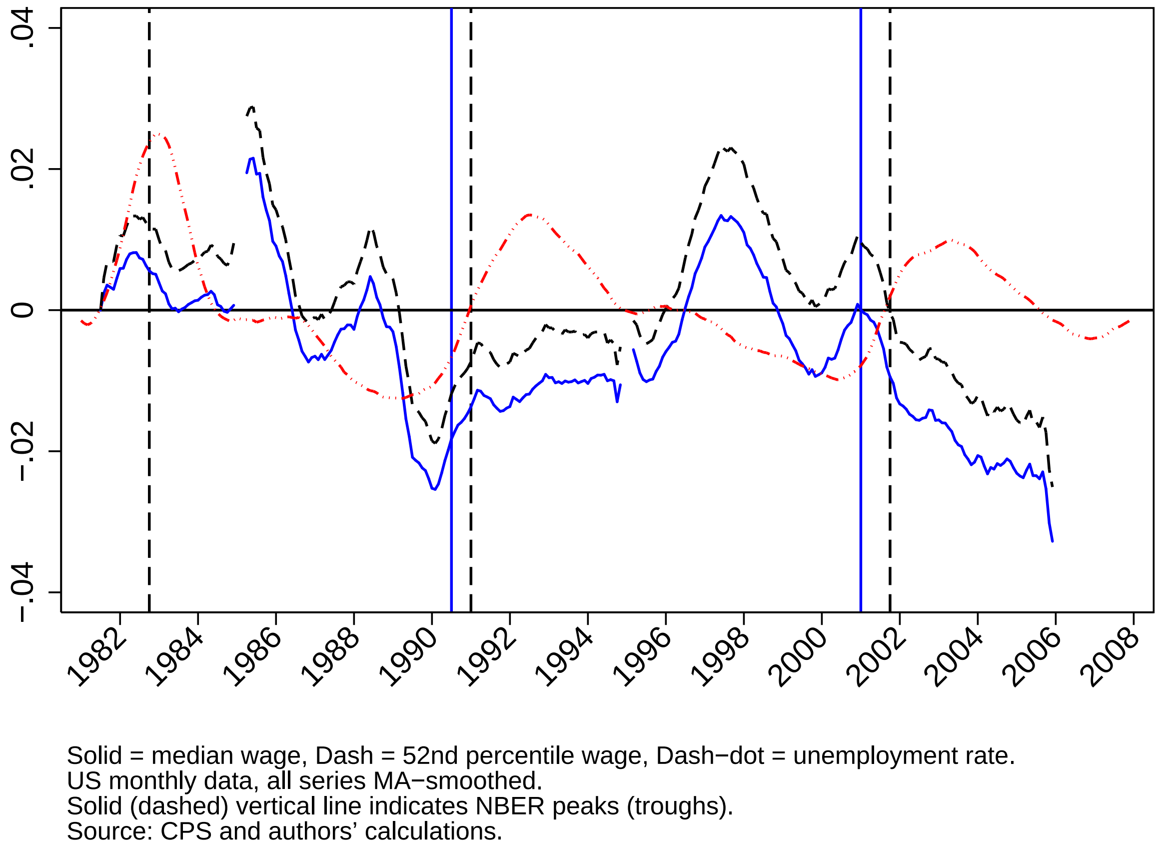


Fig. 13: Median and 52nd percentile of residual growth rate of individual hourly wages unexplained by employment composition effects

and 2000s when it was slack.

A breakdown of the mean wage series by establishment size categories can be obtained from the CBP data. This is reported on Figure 14 (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 gap (see Oi and Idson, 1999, for an overview).[18] Second, the pattern highlighted at the aggregate level on Figure 12 holds roughly unchanged for all establishment size categories. Third, all wage profiles plotted on Figure 14 are nearly parallel. In other words, the distribution of mean wages by

[18] Although some of the size wage gap is explained by workforce composition (which is obviously ignored in Figure 14), the voluminous literature on this matter establishes that significant firm size-wage effects remain after controlling for various worker and job characteristics. Those results are discussed in great detail in Oi's and Idson's (1999) Handbook chapter.

establishment size shows no clear sign of collapsing or fanning out over time.

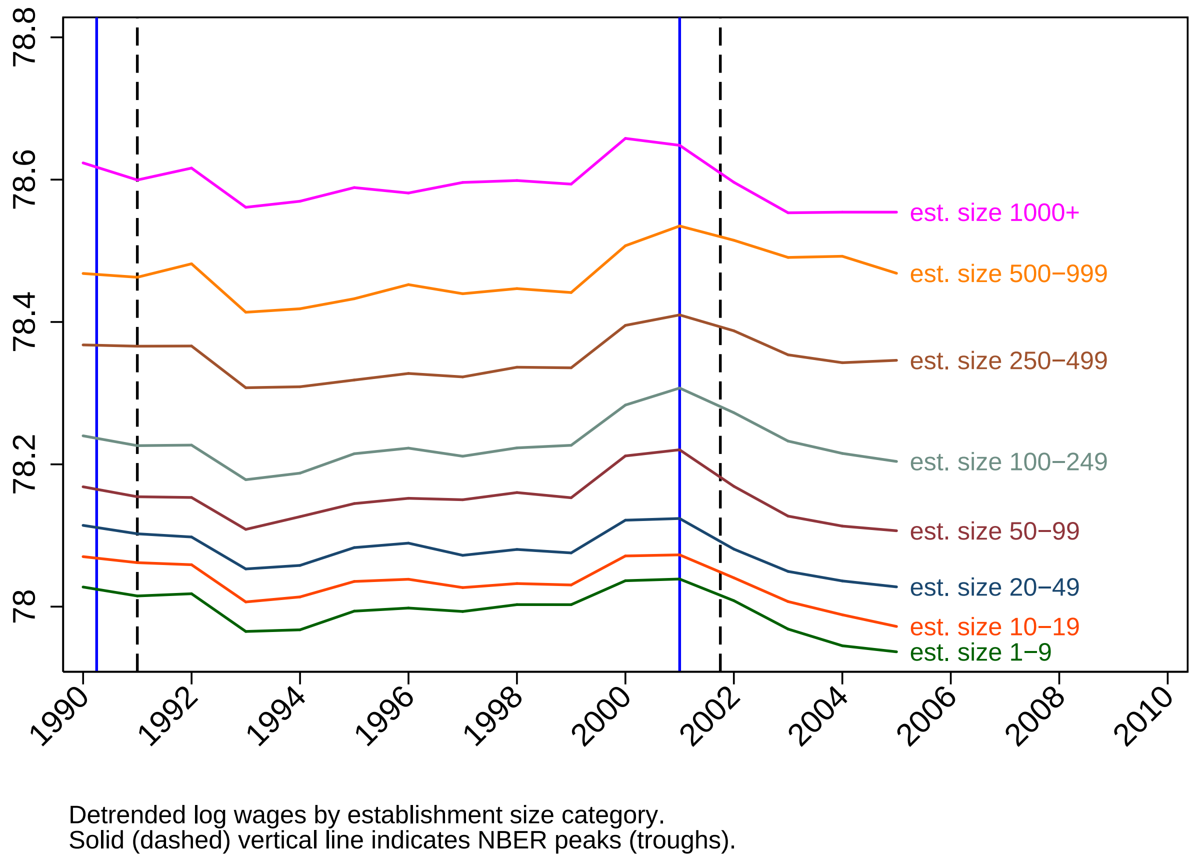


Fig. 14: Wages by establishment size

2.5 Taking Stock of the Facts: A Hypothesis on Aggregate Labor Market Fluctuations

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).

Surely, the facts that we have highlighted only pertain to the last two expansions, and as such are not enough to establish an empirical regularity. Yet from following the current guessing game on a possible 2007-2008 recession in the US economy, we have learned that practitioners of macroeconomic forecasting had already identified, at least qualitatively, some cyclical pattern of job creation by employer size.[19] Commercial forecasters rely not only on some of the statistics produced by the Federal Government that we exploit here, but also on occasional private surveys of businesses, of inferior statistical quality but often going back in time much longer. Although we do not put much weight on an established conventional wisdom of commercial forecasters, at least it does not contradict the possibility that we may, indeed, have identified new business cycle stylized facts. We stress, however, that our search for facts has been guided mostly by our theoretical hypothesis spelled out below. Our conceptual framework will provide both an interpretation of the facts and a motive to look further for additional facts that might appear prima facie unrelated, and that were previously unknown to academic or commercial economists. Examples are the evolution of the firm size distribution in BED at quarterly frequency and the composition of new hires by size of new employers from SIPP.

Whatever their actual degree of regularity, those 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 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

[19] The following excerpts are from CNNMoney.com reports and relate to the perceived 2007-08 turning point in the US labor market.

> " 'Job seekers now are in the driver's seat,' the Chicago-based outplacement firm Challenger Gray & Christmas observes. 'Low unemployment throughout last year forced companies to increase wages and offer new perks in order to attract and retain the most talented people.' " *March 28, 2007.*

And a year later, precursory signs of a looming recession are posted by commercial forecasters:

> "Another sign from workers that the labor market is getting difficult: There has been a sharp decline in the number of workers willing to quit their jobs. [...] 'When the quit rate is low, it's a very bad sign', said economist Robert Bruca of FAO Economics." *March 6, 2008.*

> "Joel Prakken, chairman of Macroeconomic Advisors, which processes the ADP payroll services data to produce the report, says what is notable is the decline in employment among medium size employers, those with 50 to 499 employees." *March 6, 2008.*

several years into an expansion, potentially creating favorable conditions for a new recession.

This pattern is reminiscent of Okun's (1973) idea of *Cyclical Upgrading* (see Bils and McLaughlin, 2001 for a recent look on cyclical upgrading). Cyclical Upgrading, however, is generally thought of as a cross-industry pattern whereby labor reallocates itself from low- to high-paying industries. Instead the phenomena that we emphasize in this section holds within industries (Figure 9) and not across. This is surely worth noticing, although it does not pose a particular problem for our proposed interpretation, which seems to apply equally well to many industries. It is in fact natural to expect that, if workers have any significant attachment to an industry, then they should upgrade within industries more than across.

While our proposed 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, in the following sections we study the transitional dynamics of the BM wage posting model with heterogeneous firms.[20] The BM model is the canonical framework for the analysis of frictional labor markets which explicitly addresses firm size, job-to-job quits, wage dispersion, and unemployment, the four key ingredients of our facts. We are not aware of any other model which can account for all four.

3 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.

[20] As mentioned in the Introduction, Shimer (2003) is the only prior attempt that we know of to analyze the out-of-steady-state behavior of a search/wage-posting 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.

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 $q(p) > 0$. Sampling weights are normalized in such a way that their cumulated sum $\Phi(p) := \int_{\underline{p}}^{p} q(x) \, d\Gamma(x)$ is a (sampling) cdf, i.e. $\Phi(\overline{p}) = 1$. The sampling density of a type-p firm is therefore $\varphi(p) := q(p)\gamma(p)$. This naturally encompasses the conventional case of uniform sampling which has $q(p) = 1$ for all p. As we shall see later in the analysis, however, a plausible calibration requires that $q(p)$ be increasing in p.[21]

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.[22]

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. 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 $F_t(\cdot)$.

[21] 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). Finally, if each firm faces a convex cost of posting vacancies, in equilibrium more productive firms will post more vacancies, as they are more profitable, so they will be sampled more often. This last assumption allows to reconcile increasing sampling weights with free entry in vacancy creation and to endogenize the contact rates.

[22] Or, less stringently, we allow firms to index wages to any aggregate variable that evolves monotonically over time (e.g. the unemployment rate). We thus rule out, among other things, wage-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). 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.

4 Equilibrium

4.1 The Contract Posting Problem

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

$$\Pi_0\left(L_0\left(p\right);p\right)=\max_{\{w_t\}}\int_0^{+\infty}\left(p-w_t\right)L_t\left(p\right)e^{-rt}dt \tag{1}$$

subject to:

$$\rho V_t\left(p\right)=\dot{V}_t\left(p\right)+w_t-\delta\left[V_t\left(p\right)-U_t\right]+\lambda_1\int_{V_t(p)}^{+\infty}\left[x-V_t(p)\right]dF_t\left(x\right) \tag{2}$$

$$\dot{L}_t\left(p\right)=-\left[\delta+\lambda_1\overline{F}_t\left(V_t\left(p\right)\right)\right]L_t\left(p\right)+\frac{q\left(p\right)}{N}\left[\lambda_0 u_t+\lambda_1\left(1-u_t\right)G_t\left(V_t\left(p\right)\right)\right] \tag{3}$$

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

where (with a slight notational abuse) $L_t\left(p\right)$ denotes a type-p firm's workforce at date t,[23] $\underline{w}$ is the exogenous institutional minimum wage, U_t is the workers' lifetime value of unemployment, r (ρ) is the firms' (workers') discount rate,[24] and $\overline{F}_t\left(\cdot\right)=1-F_t\left(\cdot\right)$ designates the survivor function associated with $F_t\left(\cdot\right)$. When solving (1), the typical firm of productivity p also is also constrained by its given initial size $L_0\left(p\right)$.

The firm's problem has two state variables that the firm controls through the wage. First, the chosen path of wages translates through the Hamilton-Jacobi-Bellman equation (2) into a value $V_t\left(p\right)$ for the worker of employment at that type-p firm. The worker's opportunity cost $\rho V_t\left(p\right)$ equals the capital gain plus the flow wage minus the capital loss when the match is destroyed exogenously at rate δ, plus the capital gain that occurs at rate $\lambda_1\overline{F}_t\left(V_t\left(p\right)\right)$ when the worker receives and offer which also turns out to provide him with a higher value. This offer is drawn from the endogenous offer distribution $F_t\left(\cdot\right)$, which is the cross-section distribution at time t of all such values offered by other firms.

The value $V_t\left(p\right)$ offered by a type-p firm translates into inflows and outflows of workers. The only friction in the model is search, so the boundaries of the firm are defined by attrition, retention and hiring. Equation (3), describes the evolution of the firm's employment. Following standard

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

[24] 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.

practice, we impose a law of large numbers at the individual firm's level and we treat the evolution of firm size as deterministic, although it is the result of various random events. These include separations—both exogenous at rate δ and endogenous at rate $\lambda_1 \overline{F}_t(V_t(p))$ when a worker receives a better offer—which reduce employment, and accessions from both unemployment (at rate λ_0) and from other firms that are paying their workers less than $V_t(p)$.

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 that no individual firm can affect with its choice. Given all firm's choices of wages, and the implied worker values $V_t(p)$ and firm sizes $L_t(p)$, they are defined by

$$F_t(W) = \int_{\underline{p}}^{\overline{p}} \mathbb{I}\{V_t(x) \leq W\}\, q(x)\, d\Gamma(x) \tag{5}$$

$$G_t(W) = \frac{\int_{\underline{p}}^{\overline{p}} L_t(x)\, \mathbb{I}\{V_t(x) \leq W\}\, d\Gamma(x)}{\int_{\underline{p}}^{\overline{p}} L_t(x)\, d\Gamma(x)} \tag{6}$$

where $\mathbb{I}\{\cdot\}$ is an indicator function. Notice that both are normalized to be proper c.d.f.'s. Also notice an important restriction that was kept implicit so far: the definitions in (5) and (6) are only valid in symmetric equilibria where there is no dispersion in firm size conditional on p (i.e. $p \mapsto V_t(p)$ and $p \mapsto L_t(p)$ are well-defined mappings for all t). Although this restriction will receive some further discussion below, we will essentially limit our attention to such equilibria in the rest of the paper.

Similarly, a single firm cannot affect the value of unemployment, which solves the HJB equation:[25]

$$\rho U_t = \dot{U}_t + b + \lambda_0 \int_{U_t}^{+\infty} (x - U_t)\, dF_t(x) \tag{7}$$

with b denoting the income flow in unemployment, or the unemployment rate u_t, which solves

$$\dot{u}_t = \delta(1 - u_t) - \lambda_0 u_t, \qquad \text{with} \quad u_0 = 1 - N \int_{\underline{p}}^{\overline{p}} L_0(x)\, d\Gamma(x) \ \text{ given.} \tag{8}$$

Before we move on to solving (1), we should clarify that our formulation of the contract-posting game and the firm's best-response problem contains the assumption that a firm must pay all of

[25] In formulating (1), we assume for simplicity that any job offer posted in equilibrium is preferred to unemployment, i.e. $\inf_p V_t(p) \geq U_t$ at all t. This is achieved by assuming that the minimum wage $\underline{w}$ is sufficiently higher than b for unemployed workers to find even the least valuable job offer worth accepting.

its workers the same wage, irrespective of when they were hired, from where, and of the outside offers that some of them may have received. In particular, the firm does renege on its promised wage, cannot condition the wage on tenure or received outside offers, and more generally does not respond to outside offers to its employees, but lets them go if they are offered more.[26] Furthermore, the solution is time-inconsistent, so the equilibrium we are studying is not sequential.

4.2 Optimality Conditions

The current value Hamiltonian of problem (1) is defined by:

$$\begin{aligned}\mathcal{H}_t(p) &= (p - w_t) L_t(p) + m_t(p)(w_t - \underline{w}) \\ &\quad + \pi_t(p)\left\{-\left(\delta + \lambda_1 \overline{F}_t(V_t(p))\right) L_t(p) + \frac{q(p)}{N}\left(\lambda_0 u_t + \lambda_1 (1 - u_t) G_t(V_t(p))\right)\right\} \\ &\quad + \nu_t(p)\left\{\left(\rho + \delta + \lambda_1 \overline{F}_t(V_t(p))\right) V_t(p) - \lambda_1 \int_{V_t(p)}^{+\infty} x dF_t(x) - w_t - \delta U_t\right\}, \quad (9)\end{aligned}$$

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

Optimality conditions are:

$$\nu_t(p) = -L_t(p) + m_t(p) \quad (10)$$

$$\begin{aligned}\dot{\nu}_t(p) &= r\nu_t(p) - \left(\rho + \delta + \lambda_1 \overline{F}_t(V_t(p))\right)\nu_t(p) \\ &\quad - \lambda_1 f_t(V_t(p)) L_t(p) \pi_t(p) - \frac{\lambda_1 q(p)}{N}(1 - u_t) g_t(V_t(p)) \pi_t(p) \quad (11)\end{aligned}$$

$$\dot{\pi}_t(p) = \left(r + \delta + \lambda_1 \overline{F}_t(V_t(p))\right)\pi_t(p) - p + w_t(p) \quad (12)$$

$$m_t(p) \geq 0, \quad w_t(p) \geq \underline{w}, \quad m_t(p)(w_t(p) - \underline{w}) = 0 \quad (13)$$

$$\lim_{t\to+\infty} e^{-rt}\pi_t(p) L_t(p) = \lim_{t\to+\infty} e^{-rt}\nu_t(p)(V_t(p) - U_t) = 0. \quad (14)$$

Supplementing this latter set of conditions with the state equations (3), (7), and (8) we obtain a system of partial differential equations characterizing the solution to an individual firm's max-

[26] As argued in Moscarini (2005), not responding to outside offers is a sequential equilibrium of an ascending (English) auction between the incumbent and the poacher, and the unique equilibrium which survives natural refinements. The more productive of the two firms wins without offering more than it does to its other workers, because it can always respond to any attempt by the competitor to outbid it, even if the competitor trembles. In this case, our assumption of no ex-post competition is not particularly restrictive. If the auction is instead simultaneous with either one bid or a sealed bid, as in Bertrand (Postel-Vinay and Robin, 2002), then firms would bid their maximum valuation and our assumption has bite.

imization problem *for a given path of sampling distributions* $\{F_t(\cdot)\}_{t\in[0,+\infty)}$. Given a solution to that system, the optimal wage path can be retrieved using (2). The main difficulty, however, 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. This task will be carried out in the following subsections. 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. $L_t(p)$ 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 (12) 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 (12) 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 (11) 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 $\pi_t(p) \cdot \frac{\partial \dot{L}_t(p)}{\partial V_t(p)} = \pi_t(p) \cdot \left[\lambda_1 f_t(V_t(p)) L_t(p) + \frac{\lambda_1 q(p)}{N}(1 - u_t) g_t(V_t(p))\right]$ 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 (10) simply reflects the optimal balance between the instantaneous cost of increasing the current posted wage by \$1—it adds \$$L_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.

4.3 Rank-Preserving Equilibria

All further formal analysis of the model will build on the following definition:

Definition 2 (Rank-Preserving Equilibrium) *A Rank-Preserving Equilibrium [RPE] is a dynamic equilibrium in which firms post values that are strictly increasing in p for all t.*

A direct consequence of the above definition is that in a RPE workers rank firms according to productivity at all dates. The following two properties hold true at all dates under the RP assumption:

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

In addition to considerably simplifying equilibrium determination (see below), the RP assumption is theoretically appealing for at least two reasons. First, it parallels a well-known property of the static equilibrium characterized by BM, which is to have a unique equilibrium where workers rank firms according to productivity. Second, RPE feature constrained-efficient labor reallocation at all dates: if workers consistently rank more productive firms higher than less productive ones, then job-to-job moves will always be up the productivity ladder.[27] Exactly how generic these RPE are is a theoretical question that we address in detail in Moscarini and Postel-Vinay (2008), to which we refer the interested reader.

Let us consider the stock of workers employed at a firm of type-p or less, $\int_{\underline{p}}^{p} L_t\left(x\right) d\Phi\left(x\right)$. 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:[28]

$$\int_{\underline{p}}^{p} \dot{L}_t\left(x\right) d\Phi\left(x\right) = \frac{\lambda_0 u_t}{N} \Phi\left(p\right) - \left[\delta + \lambda_1 \overline{\Phi}\left(p\right)\right] \int_{\underline{p}}^{p} L_t\left(x\right) d\Phi\left(x\right).$$

The latter equation now solves as:

$$\int_{\underline{p}}^{p} L_t\left(x\right) d\Phi\left(x\right) = e^{-\left[\delta+\lambda_1\overline{\Phi}(p)\right]t} \left(\int_{\underline{p}}^{p} L_0\left(x\right) d\Phi\left(x\right) + \frac{\lambda_0 \Phi\left(p\right)}{N} \int_0^t u_s e^{\left[\delta+\lambda_1\overline{\Phi}(p)\right]s} ds \right) \tag{15}$$

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

$$L_t\left(p\right) = e^{-\left[\delta+\lambda_1\overline{\Phi}(p)\right]t} \Bigg[L_0\left(p\right) + \lambda_1 t q\left(p\right) \int_{\underline{p}}^{p} L_0\left(x\right) d\Phi\left(x\right) + \frac{\lambda_0 q\left(p\right)}{N} \int_0^t \left[1 + \lambda_1\left(t-s\right)\Phi\left(p\right)\right] u_s e^{\left[\delta+\lambda_1\overline{\Phi}(p)\right]s} ds \Bigg] \tag{16}$$

[27]We thank Pat Kline for pointing this out to us.

[28]Note that the following law of motion can also by obtained by integration of (3) w.r.t. p. Details available on request.

The steady-state versions of (15) and (16) are:

$$L_\infty(p) = \frac{\delta\lambda_0(\delta+\lambda_1)\,q(p)}{N(\delta+\lambda_0)\left[\delta+\lambda_1\overline{\Phi}(p)\right]^2} \quad \text{and} \quad N\int_{\underline{p}}^{p} L_\infty(x)\,d\Phi(x) = \frac{\delta\lambda_0\Phi(p)}{(\delta+\lambda_0)\left[\delta+\lambda_1\overline{\Phi}(p)\right]}. \tag{17}$$

This is the point at which the necessity for sampling weights appears. Note from equation (17) that the steady-state size ratio of the largest to the smallest firm in the market in units of (non-normalized) employment is

$$\frac{L_\infty(\overline{p})}{L_\infty(\underline{p})} = \left(1+\frac{\lambda_1}{\delta}\right)^2 \frac{q(\overline{p})}{q(\underline{p})}.$$

With uniform sampling ($q(p) \equiv 1$ 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. We discussed several possible interpretations/formalizations of these sampling weights. This skew may be interpreted as a measure of the inability of search frictions alone to generate a plausible firm size distribution. Alternative forces that may contribute to shaping the size distribution include credit constraints (Cooley and Quadrini, 2001) and industry-specific human capital accumulation (Rossi-Hansberg and Wright, 2007). We note, however, that our frictions-based approach has the distinct advantage of generating observable implications in terms of job-to-job quit rates by workers, wages within and across size classes, which add considerable empirical discipline to the theoretical exercise.

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 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-posting assumption retained in the BM model.

4.4 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 (10) - (14) plus the set of state equations (3), (2) and (8). 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 $\nu_t(p) = -L_t(p)$. For simplicity, we also assume equal discount rates for workers and employers from now on (i.e. $r = \rho$). Substitution of (10) into (11), and combination with (3) then yields:

$$\frac{q(p)}{N}\left(\lambda_0 u_t + \lambda_1(1-u_t)G_t(V_t(p))\right) = \lambda_1 \pi_t(p)\left(f_t(V_t(p))L_t(p) + \frac{q(p)}{N}(1-u_t)g_t(V_t(p))\right). \quad (18)$$

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 (18) equals $\pi_t(p)\cdot\frac{\partial \dot{L}_t(p)}{\partial V_t(p)}$ 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 (18) 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 $L_t(p)$ 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 $\left[\rho - r + \delta + \lambda_1\overline{F}_t(V_t(p))\right]L_t(p)$ plus an extrinsic expansion/contraction term $\dot{L}_t(p)$ 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 (18)'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 (2) and (12) yields:

$$\dot{\mu}_t(p) = \left(r + \delta + \lambda_1\overline{F}_t(V_t(p))\right)\mu_t(p) - \lambda_1\int_{V_t(p)}^{+\infty} x dF_t(x) - \delta U_t - p, \quad (19)$$

which is supplemented by the transversality condition $\lim_{t\to+\infty} e^{-rt}L_t(p)\left[\mu_t(p) - U_t\right] = 0$, obtained from adding the two conditions in (14) together and substituting the first-order condition (10).

Interpretation of equation (19) 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 (18) - (19) into:

$$\left(\frac{\lambda_0 u_t}{N} + \lambda_1 \int_{\underline{p}}^{p} L_t(x)\, d\Gamma(x)\right) V_t'(p) = 2\lambda_1 \gamma(p) L_t(p) \pi_t(p) \tag{20}$$

$$\dot{\mu}_t(p) = \left(r + \delta + \lambda_1 \overline{\Phi}(p)\right) \mu_t(p) - \lambda_1 \int_p^{+\infty} V_t(x)\, d\Phi(x) - \delta U_t - p \tag{21}$$

$$\lim_{t \to +\infty} e^{-rt} L_t(p) \left[\mu_t(p) - U_t\right] = 0. \tag{22}$$

Differentiation of (21) w.r.t. p yields (primes denote differentiation w.r.t. p while dots denote time differentiation):

$$\dot{\mu}_t'(p) = \left(r + \delta + \lambda_1 \overline{\Phi}(p)\right) \mu_t'(p) + \lambda_1 \gamma(p) q(p) \left(V_t(p) - \mu_t(p)\right) - 1. \tag{23}$$

This, together with (20), gives the following system of two PDEs in $(\mu_t'(p), \pi_t(p))$:

$$\begin{aligned} \dot{\mu}_t'(p) &= \left(r + \delta + \lambda_1 \overline{\Phi}(p)\right) \mu_t'(p) - \lambda_1 \gamma(p) q(p) \pi_t(p) - 1 \\ \mu_t'(p) &= \pi_t'(p) + \frac{2\lambda_1 \gamma(p) L_t(p)}{\frac{\lambda_0 u_t}{N} + \lambda_1 \int_{\underline{p}}^{p} L_t(x)\, d\Gamma(x)} \pi_t(p). \end{aligned} \tag{24}$$

This can be solved numerically, subject to some initial and boundary conditions. 'Initial' conditions are given by the steady-state solution to (24), which is characterized as:

$$\begin{aligned} \mu_\infty'(p) &= \frac{1 + \lambda_1 \gamma(p) q(p) \pi_\infty(p)}{r + \delta + \lambda_1 \overline{\Phi}(p)} \\ \pi_\infty(p) &= \frac{\left(\delta + \lambda_1 \overline{\Phi}(p)\right)^2}{r + \delta + \lambda_1 \overline{\Phi}(p)} \left(\int_{\underline{p}}^{p} \frac{dx}{\left(\delta + \lambda_1 \overline{\Phi}(x)\right)^2} + \frac{\pi_\infty(\underline{p})(r + \delta + \lambda_1)}{(\delta + \lambda_1)^2}\right). \end{aligned} \tag{25}$$

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:

$$\pi_t\left(\underline{p}\right) \equiv \frac{\underline{p}-\underline{w}}{r+\delta+\lambda_1}$$

$$\mu_t'\left(\underline{p}\right) \equiv \frac{1+\lambda_1\gamma\left(\underline{p}\right)q\left(\underline{p}\right)\pi_t\left(\underline{p}\right)}{r+\delta+\lambda_1}, \tag{26}$$

where the second condition is obtained by combining the first one with the $\dot{\mu}_t'(p)$ equation in (24). 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 (24) is solved for $(\mu_t'(p), \pi_t(p))$, wages can be retrieved from (12) (written under the RP assumption):

$$w_t(p) = p - \left(r+\delta+\lambda_1\overline{\Phi}(p)\right)\pi_t(p) + \dot{\pi}_t(p), \tag{27}$$

which has the following familiar steady-state solution:

$$w_\infty(p) = p - \left(\delta+\lambda_1\overline{\Phi}(p)\right)^2\left(\int_{\underline{p}}^{p}\frac{dx}{\left(\delta+\lambda_1\overline{\Phi}(x)\right)^2} + \frac{\underline{p}-\underline{w}}{\left(\delta+\lambda_1\right)^2}\right). \tag{28}$$

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.[29]

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

[29]To see this, notice that (12) and (20) 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). Equation (12) 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 (20) by $L_t(p)$ tells us that $\pi_t(p)$ is proportional to the gross hiring rate, and so $\pi_t(p)$ 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.

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 (24) 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.

5 Quantitative Analysis of Rank Preserving Equilibria

5.1 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 (17) implies that the steady-state cross-section CDF of wages, $G_w(\cdot)$ (say), is defined by

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

Differentiation of (28) then yields:

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

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 (29).

Once a sampling distribution has been obtained, the underlying distribution of firm types $\Gamma(p)$ and sampling weights $q(p)$ 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

Firm size category	Cum. fraction of firms $[\Gamma(p)]$	Cum. emp. share $[G_w(w(p))]$
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

relationship:

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

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

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 (see the next subsection).

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

5.2 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 $\Gamma_0(\cdot)$. Because for any potential firm productivity is given by $p = y \times \theta$, 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:

$$\Gamma(p) = \frac{\Gamma_0(p/y) - \Gamma_0(\underline{w}/y)}{1 - \Gamma_0(\underline{w}/y)}, \tag{30}$$

and the number of active firms will be $N = N_0(1 - \Gamma_0(\underline{w}/y))$. The distribution of potential firm types $\Gamma_0(\cdot)$ is then calibrated by shifting the support of the $\Gamma(\cdot)$ distribution obtained as explained in the previous subsection so that its infimum is at $p = 1$, and use that as our benchmark $\Gamma_0(\cdot)$ (given the normalization $y = 1$). Finally, the number chosen for N_0 (see Table 3) reflects an average firm size of 20.

We finally model a 'boom' as a permanent 2 percent increase in y (from $y = 1$ to $y = 1.02$). We further assume that this productivity increase causes the job finding rate λ_0 to increase by 8 percent,[30] 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 $\Delta N = N_0(\Gamma_0(\underline{w}) - \Gamma_0(\underline{w}/1.02))$ 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 (30).

5.3 Results

As can very easily be inferred from Equation (8), 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

[30]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.

δ and λ_0, 90% of the distance between the initial and the final steady state is covered in less than six months.

Figure 15 then shows how unemployment adjusts at single firms: it shows a plot of $L_t(p)/L_0(p)$ for four different values of p corresponding to the 50th, 90th, 95th and 99.9th percentiles of the (post-shock) distribution of firm types, $\Gamma(\cdot)$.[31] 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 15 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.[32]

For comparison with the descriptive evidence shown in Section 2, 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 16 to 20 which parallel Figures 2 to 5 from Section 2.

The response of the average job-to-job quit rate, $\frac{\lambda_1 N}{1-u_t}\int_{\underline{p}}^{\overline{p}} \overline{\Phi}(x) L_t(x) d\Gamma(x)$, is plotted on Figure

[31] 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.

[32] It actually equals $N\int_{\underline{p}}^{\overline{p}} \overline{\Phi}(x) L_t(x) d\Gamma(x) / (1-u_t)$. This becomes $\frac{\delta}{\lambda_1}\left\{\left(1+\frac{\delta}{\lambda_1}\right)\ln\left(1+\frac{\lambda_1}{\delta}\right)-1\right\}$ at a steady state, i.e. about 0.23 with our parameterization.

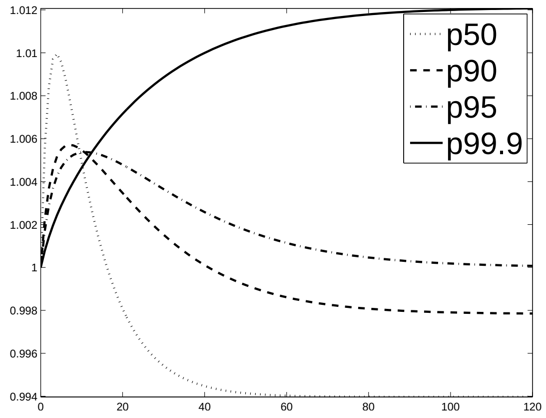


Fig. 15: Firm size dynamics

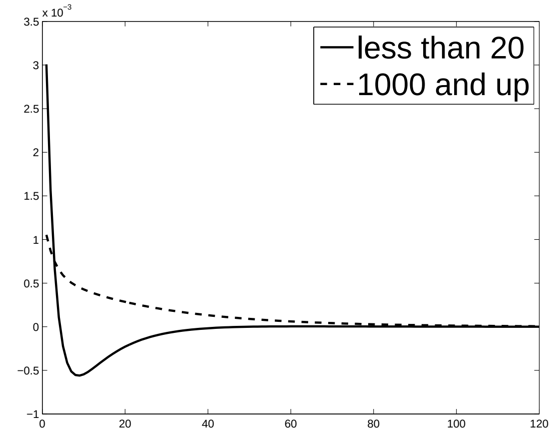


Fig. 16: Firm growth—small *vs.* large firms

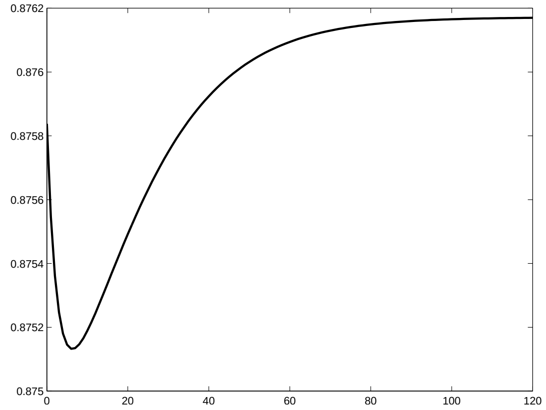


Fig. 17: Firm shares (firms < 20)

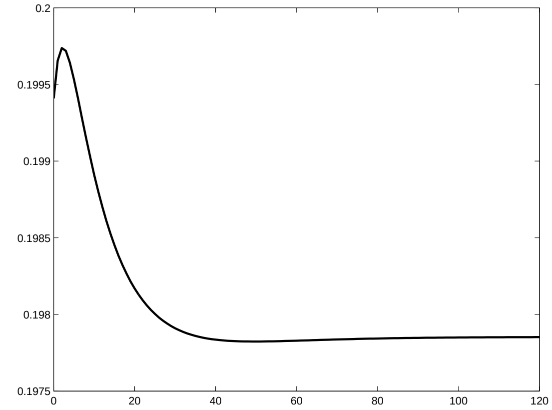


Fig. 18: Emp. shares (firms < 20)

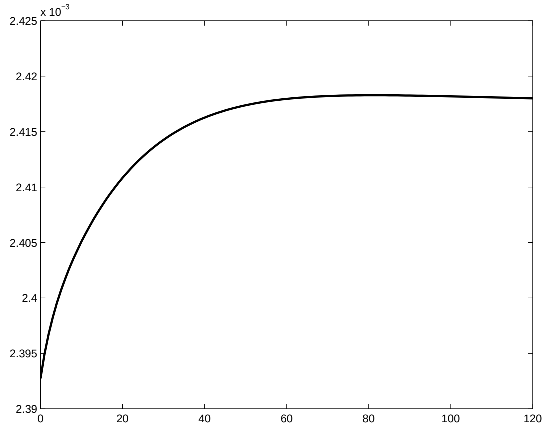


Fig. 19: Firm shares (firms $\geq$ 1000)

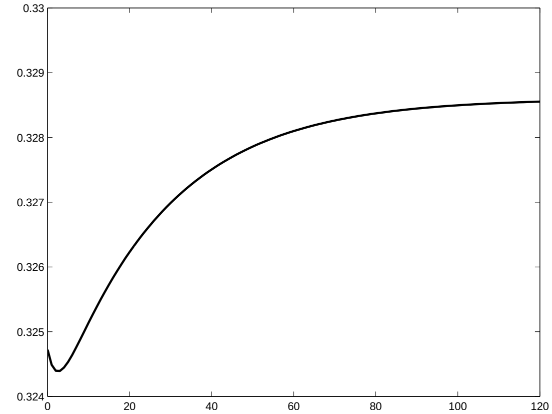


Fig. 20: Emp. shares (firms $\geq$ 1000)

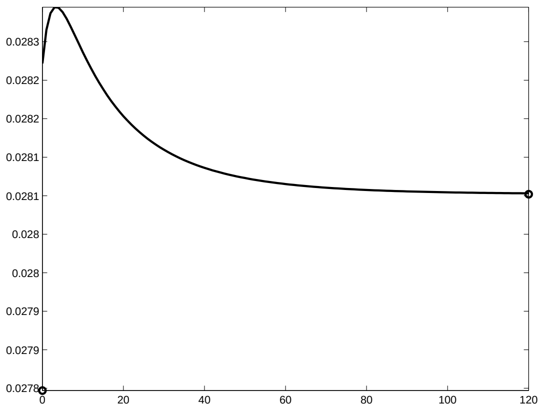


Fig. 21: Average job-to-job quit rate

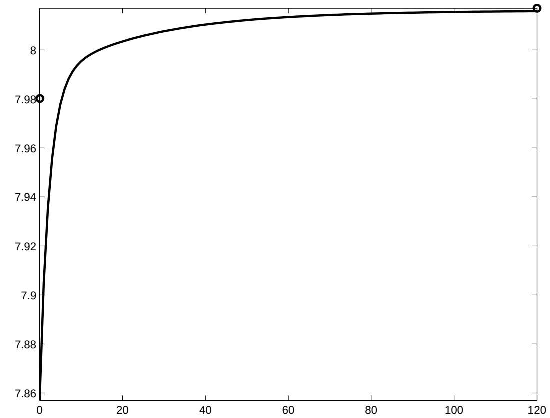


Fig. 22: Mean wage

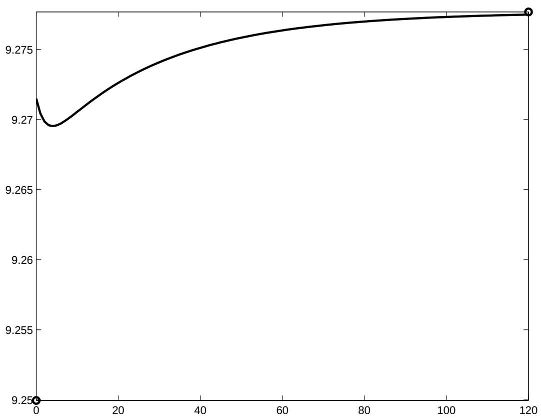


Fig. 23: Mean output per worker

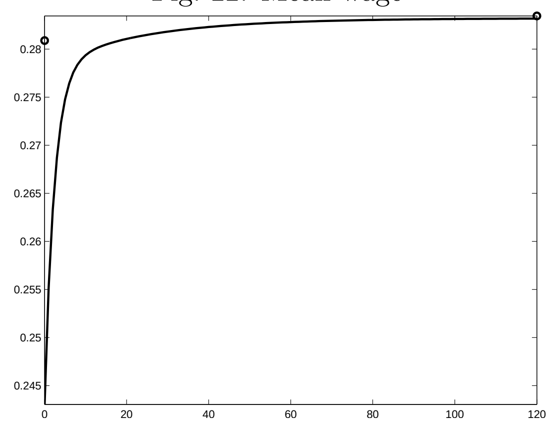


Fig. 24: Mean labor share

21. 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 22, 23 and 24 plot the dynamic responses of mean wages, mean output per worker, and the mean labor share, respectively.

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. 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.[33] Second, each firm-level wage follows a dynamic path of its own. Wages are backloaded to the late part of the expansion. The composition of these individual dynamic paths causes the initial downward jump in the mean wage: the effect on the intensive margin, the within-firm backloading, dominates the aggregate wage at first.

Combining the output and wage series one can visualize the dynamic response of the labor share (Figure 24). This is an interesting plot to look at in the light of a recent paper by Choi and Ríos-Rull (2008), who document a number of facts about the cyclical behavior of the labor share. Most notably, they show that the labor share is countercyclical and persistent. Our model replicates these facts, in that Figure 24 resembles the impulse response function produced by Choi's and Ríos-Rull's VAR analysis: the labor share decreases on impact of a positive aggregate shock, and then gradually increases back toward its new steady-state value. This, however, is only a half-success, as our model also fails on two important points. First, Choi and Ríos-Rull (2008) also identify an overshooting property of the labor share: the labor share IRF peaks after five years at

[33] 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.

a higher level than the initial and final steady-states.[34] Although the initial dip in average output per worker (see Figure 23) would in principle help generating this overshooting of the labor share in our model, it doesn't seem to be enough.[35] Second, our model predicts a labor share of 25-30%, which is about half the number observed in the data. This is related to a fundamental problem of the BM model, which can only replicate an empirically sensible wage density given a distribution of firm productivity with an implausibly long right tail, resulting in profit rates close to 100% for a substantial fraction of highly productive firms (see e.g. Postel-Vinay and Robin, 2006, for a discussion of this problem).

Our example features a strong propagation in wages, labor productivity and the labor share, which keep rising years after the initial shocks, although the unemployment rate's half life is just a few months. The main reason is that job-to-job transitions in the data are an order of magnitude slower than the reallocation from unemployment to employment. Thus, the upgrading process is slow, and so is the rise in labor productivity after an initial jump following the shock. The propagation is less pronounced for the EE rate. We are, however, ignoring further sources of propagation of the unemployment rate, such as endogenous labor force participation. This is likely to rise in the expansion, feeding the market with relatively cheap candidates for hiring from unemployment, and delaying the moment when large firms have to start raising wages aggressively to poach workers, the small firms have to respond to retain them, and the EE rate peaks. On the other hand, our Figure 1 shows that participation seems to play a very minor role at business cycle frequencies. Finally, our quantitative exercise does not feature a strong amplification of aggregate productivity shocks on unemployment, a subject of much debate in recent years (Shimer, 2005). This is not our main focus, and at any rate the present exercise is just one initial attempt to gain traction on the new evidence that we present.

[34]Incidentally, the final steady-state labor share is slightly higher than the initial one in our calibration exercise. This results from the comparative static properties of the BM model.

[35]This could potentially be fixed by 'prolonging' the dip in output per worker, which could be achieved by assuming that the initial entry of low-productivity firms is only gradual.

5.4 Discussion: The Sources of Wage Backloading

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 at the individual worker 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.

Aggregate dynamics introduce an independent motive for backloading wages. When unemployment declines and the resulting inflow of cheap workers into firms with it, firm growth slows down. At that point, more productive firms, which have more to lose from not employing workers, find it profitable to raise wages to raid smaller, less productive competitors. The latter, in turn, respond by raising wages to retain their workers. This direct impact of unemployment on the level of wages gives rise to a backloading in calendar time that originates purely from the time-varying scarcity of cheap job applicants.

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.

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 4, 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.

6 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, including downturns that we have so far ignored in the analysis. 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.

Our example can claim very limited success from a quantitative viewpoint. This is due in part on our exclusive focus on one friction, job search, as the common source of the firm's boundaries and size, wage dynamics, and worker flows. It is natural to think about additional frictions that affects firm size and that operate differentially at different levels of size, such as borrowing constraints.

References

[1] Barlevy, G. (2001), "Why are the Wages of Job Changers so Procyclical?", *Journal of Labor Economics*, 19(4), 837-878.

[2] 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).

[3] Beaudry, P. and DiNardo, J. (1991), "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Micro Data,", *Journal of Political Economy*, 99(4), 665-88.

[4] Bils, M. (1985), "Real Wages Over the Business Cycle: Evidence From Panel Data,", *Journal of Political Economy*, August.

[5] Bils, M. and K. McLaughlin (2001), "Inter-Industry Mobility and the Cyclical Upgrading of Labor", *Journal of Labor Economics*, 19(1), 94-135.

[6] 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.

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

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

[9] 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.

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

[11] Choi, S.-K. and J.-V. Ríos-Rull (2008) "The Labor Share: Disciplining Device for RBC Models?", University of Minnesota *mimeo*.

[12] Cooley, T. F. and V. Quadrini (2001), "Financial Markets and Firm Dynamics", *American Economic Review*, 91(5), 1286-310.

[13] 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), 1–18.

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

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

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

[17] Moscarini, G. (2005), "Job Matching and the Wage Distribution", *Econometrica*, 73(2), 481-516.

[18] Moscarini, G. and F. Postel-Vinay (2008), "Non-stationary Search Equilibrium", Yale University and University of Bristol *mimeo*.

[19] Moscarini, G. and K. Thomsson (2007), "Occupational and Job Mobility in the US", *Scandinavian Journal of Economics*, 109(4), 807836.

[20] Neumark, D., B. Wall and J. Zhang (2008), "Do Small Businesses Create More Jobs? New Evidence from the National Establishment Time Series", NBER Working Paper No. 13818.

[21] Oi, W. and T. L. Idson (1999), "Firm Size and Wages", in O. Ashenfeller O. and D. Card (eds.), *Handbook of Labor Economics*, Amsterdam: Elsevier Science B.V.

[22] Okun, A. (1973), "Upward Mobility in a High Pressure Economy", *Brooking Papers on Economic Activity*, 207-52.

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

[24] Postel-Vinay, F. and J.-M. Robin (2006), “Microeconometric Search-Matching Models and Matched Employer-Employee Data”, *in* R. Blundell, W. Newey and T. Persson, editors, *Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress*, Cambridge: Cambridge University Press.

[25] Rossi-Hansberg, E. and M. L. J. Wright (2007), “Establishment Size Dynamics in the Aggregate Economy”, *American Economic Review*, 97(5), 1639-1666.

[26] Solon, G., R. Barsky and J. Parker (1994), “Measuring the Cyclicality of Real Wages: How Important Is Composition Bias?”, *Quarterly Journal of Economics*, 109(1), pp. 1-25.

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

[28] Shimer, R. (2005), “Reassessing the Ins and Outs of Unemployment”, University of Chicago *mimeo.*

[29] 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:

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

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

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

$$1 = \pi_t(p) \times \left[\lambda_1 f_t(w_t(p)) + \lambda_1 \frac{v(p)}{N L_t(p)} (1 - u_t) g_t(w_t(p))\right] \tag{33}$$

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

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

Now focusing on RPE's, where $(1 - u_t) g_t(w_t(p)) = N L_t(p) f_t(w_t(p)) / v(p) = N L_t(p) \gamma(p) / w_t'(p)$, the first order condition (33) becomes:

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

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

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

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):

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

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 (16) 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, $w_t\left(\underline{p}\right) = \underline{p} = \underline{w}$ for all t. To show that the invariant solution is unique, integrate equation (37) between $\underline{p}$ and p:

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

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

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

and

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

We now establish 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

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

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 (26).) Pick a step size h.

Step 2. Construct a candidate $\pi_t(p_0+h)$ using the second (static) differential equation in (24), such as:[36]

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

Step 3. Construct a candidate wage path for type-(p_0+h) firms from $\widetilde{\pi}_t(p_0+h)$ and equation (12):

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

[36]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 (24):

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

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