Marginal Jobs, Heterogeneous Firms, and Unemployment Flows*

Michael W. L. Elsby
University of Edinburgh and NBER

Ryan Michaels
University of Rochester

June 25, 2011

Abstract

This paper introduces a notion of firm size into a search and matching model with endogenous job destruction. The outcome is a rich, yet analytically tractable framework that can be used to analyze a broad set of features of both the cross section and the dynamics of the aggregate labor market. In a set of quantitative applications we show that the model can provide a coherent account of a) the salient features of the distributions of employer size, and employment growth across establishments; b) the amplitude and propagation of cyclical fluctuations in flows between employment and unemployment; c) the negative comovement of unemployment and vacancies in the form of the Beveridge curve; and d) the dynamics of the distribution of employer size over the business cycle.

*We are grateful to Gary Solon and Matthew Shapiro for their comments, support, and encouragement, to anonymous reviewers at the National Science Foundation for very constructive comments, and to John Haltiwanger, Ron Jarmin, and Javier Miranda for providing us with tabulations from the Longitudinal Business Database. We also thank Mark Bils, Bjorn Brugemann, Shigeru Fujita, William Hawkins, Bart Hobijn, Winfried Koeniger, Per Krusell, Toshi Mukoyama, Aysegul Sahin and Dmitriy Stolyarov for helpful discussions as well as seminar participants at the May 2007 UM/MSU/UWO Labor Day, the European Central Bank, the Kansas City Fed, UQAM, the NewYork/Philadelphia Fed Workshop on Quantitative Macroeconomics, Yale, the 2008 CEPR ESSLE conference, the Minneapolis Fed, Chicago Booth, Oslo and Oxford for helpful comments. In addition, we are grateful to William Hawkins for catching a coding error in an earlier draft of this paper. All remaining errors are our own. E-mail addresses for correspondence: mike.elsby@ed.ac.uk, and ryan.michaels@rochester.edu.

The study of the macroeconomics of labor markets has been dominated by two influential approaches in recent research: the development of search and matching models (Pissarides, 1985; Mortensen and Pissarides, 1994) and the empirical analysis of establishment dynamics (Davis and Haltiwanger, 1992). This paper provides an analytical framework that unifies these approaches by introducing a notion of firm size into a search and matching model with endogenous job destruction. The outcome is a rich, yet analytically tractable framework that can be used to analyze a broad set of features of both the cross section and the dynamics of the aggregate labor market. In a set of quantitative applications we show that the model can provide a coherent account of a) the salient features of the distributions of employer size, and employment growth across establishments; b) the amplitude and propagation of cyclical fluctuations in flows between employment and unemployment; c) the negative comovement of unemployment and vacancies in the form of the Beveridge curve; and d) the dynamics of the distribution of employer size over the business cycle.

A notion of firm size is introduced by relaxing the common assumption that firms face a linear production technology.[1] Though conceptually simple, incorporating this feature is not a trivial exercise. The existence of a non-linear production technology, and the associated presence of multi-worker firms, complicates wage setting because the surplus generated by each of the employment relationships within a firm is not the same—"the" marginal worker generates less surplus than infra-marginal workers. In section 1, we apply the bargaining solution of Stole and Zwiebel (1996) to derive a very intuitive wage bargaining solution for this environment, something that has been considered challenging in recent research (see Cooper, Haltiwanger and Willis, 2007; and Hobijn and Sahin, 2007). The solution is a very natural generalization of the wage bargaining solution in standard search and matching models. The simplicity of our solution is therefore a useful addition to the literature.[2]

The wage bargaining solution enables us to characterize the properties of the optimal labor demand policy of an individual firm in the presence of idiosyncratic firm heterogeneity. We demonstrate that the labor demand solution is analogous to that of a model of kinked hiring costs in the spirit of Bentolila and Bertola (1990), but where the hiring cost is en-

[1] In its simplest form, this manifests itself in a one firm, one job representation, as in Pissarides (1985) and Mortensen and Pissarides (1994).

[2] Bertola and Caballero (1994) solve a related bargaining problem by taking a linear approximation to the marginal product function and specializing productivity to a two-state Markov process. The present paper relaxes these restrictions. More recent research that models endogenous separations has set worker bargaining power to zero in order to derive wages (Cooper et al., 2007; Hobijn and Sahin, 2007). In the presence of exogenous separations, Acemoglu and Hawkins (2006) characterize wages, but focus instead on a time to hire aspect to job creation, which leads to a more challenging bargaining problem. The wage bargaining solution for models with exogenous job destruction has been characterized by Smith (1999), Cahuc and Wasmer (2001), and Krause and Lubik (2007).

dogenously determined by frictions in the labor market. This yields an analytical solution for the optimal labor demand policy, summarizing microeconomic behavior in the model.

In section 2, we take on the task of aggregating this behavior to the macroeconomic level. This is a challenge because the presence of a non-linear production technology and idiosyncratic heterogeneity imply that a representative firm interpretation of the model doesn't exist. To address this, we develop a method that allows us to solve analytically for the equilibrium distribution of employment across firms (the firm size distribution). In turn, this allows us to determine the level of the aggregate (un)employment stock, which is implied by the mean of that distribution. We also provide a related method that allows us to solve for aggregate unemployment flows (hires and separations) implied by microeconomic behavior. Together, these characterize the aggregate steady state equilibrium of the model economy.

In section 3 we explore the dynamics of the model by introducing aggregate shocks. A difficulty that arises in the model is that, out of steady state, individual firms must forecast future wages, which involves forecasting the future path of the distribution of employment across firms, an infinite-order state variable. A useful feature of our analytical solution for optimal labor demand is that it allows us to simplify part of this problem. In particular, we are able to derive an analytical approximation to a firm's optimal labor demand policy in the presence of aggregate shocks, obviating the need for a numerical solution. Using this, we employ an approach that mirrors the method proposed by Krusell and Smith (1998) to solve for the transition paths for the unemployment stock and flows in the presence of aggregate shocks.

These results form the basis of a series of quantitative applications, which we turn to in section 4. An attractive feature of the model is that, by incorporating both a notion of firm size as well as idiosyncratic heterogeneity, it delivers important cross sectional implications. We show that the model can be used to match key features of the distribution of firm size, and of employment growth across establishments. This is achieved through two aspects of the model. First, due to the existence of kinked hiring costs, optimal labor demand features a region of inaction whereby firms choose neither to hire nor fire workers. This matches a key property of the distribution of employment growth—the existence of a mass point at zero establishment growth—noted at least since the work of Davis and Haltiwanger (1992).[3] Second, informed by the well-known shape of the distribution of firm size, we adopt a Pareto specification for idiosyncratic firm productivity. A surprising outcome of this approach is that the Pareto specification also provides a very accurate description of the tails

[3]Earlier work by Hamermesh (1989), which analyzed data from seven manufacturing plants, also drew attention to the "lumpy" nature of establishment-level employment adjustment.

of the distribution of employment growth, something that cannot be achieved using more conventional lognormal specifications of heterogeneity.

We then use these steady-state features of the model to provide a novel perspective on the cyclical dynamics of worker flows implied by the model. It is well-known that the cyclical amplitude of unemployment, and of the job-finding rate in particular, relies critically on the size of the surplus to employment relationships (Shimer, 2005; Hagedorn and Manovskii, 2007). Intuitively, small reductions in aggregate productivity can easily exhaust a small surplus, and lead firms to cut back substantially on hiring. The presence of large and heterogeneous firms in our model opens up a new approach to calibrating the payoff from unemployment, and thereby the match surplus. Because the model is capable of matching the observed cross-sectional distribution of employment growth, we obtain a sense of the plausible size of idiosyncratic shocks facing firms. Given this, a higher payoff from unemployment implies a smaller surplus, so that jobs will be destroyed more frequently, raising the rate of worker turnover. We discipline the model by choosing the payoff from unemployment that matches the empirical rate at which employed workers flow into unemployment.

Applying this approach to an otherwise standard calibration reveals that our generalized model can replicate both the observed procyclicality of the job finding rate, as well as the countercyclicality of the employment to unemployment transition rate in the U.S.[4] We show that this is a substantial improvement over standard search and matching models. As shown by Shimer (2005), these are unable to generate enough cyclicality in job creation. To overcome this, the standard model must reduce the size of the surplus, which in turn yields excessive employment to unemployment transitions.[5] The generalized model does not face this tension between reproducing the cyclicality of job creation and the rate of worker turnover. Due to the diminishing marginal product of labor, the model generates simultaneously a large average surplus and a small marginal surplus to employment relationships. The former allows the model to match the rate at which workers flow into unemployment, the latter the volatility of the job-finding rate over the cycle.[6]

[4] For evidence on the countercyclicality of employment to unemployment flows, see Perry (1972); Marston (1976); Blanchard and Diamond (1990); Elsby, Michaels, and Solon (2009); Fujita and Ramey (2009); Pissarides (2007); Shimer (2007); and Yashiv (2006).

[5] This formalizes the intuition of recent research that has argued that the average surplus required for the standard model to match the observed cyclicality of the job finding rate is implausibly small (Mortensen and Nagypal, 2007). A small average surplus also jars with widespread evidence for the prevalence of long term employment relationships in the US economy, which researchers have taken to imply substantial rents to ongoing matches (Hall, 1982; Stevens, 2005).

[6] One might imagine that a symmetric logic holds on the supply side of the labor market if there is heterogeneity in workers' valuations of leisure so that "the" marginal worker obtains a low surplus from employment. Interestingly, Mortensen and Nagypal (2007) argue that this is not the case. They show that

A potential concern in models that incorporate countercyclical job destruction, such as the model in this paper, has been that they often cannot generate the observed procyclicality of vacancies (Shimer, 2005; Mortensen and Nagypal, 2008). Importantly, we find that our model makes considerable progress in this regard: Our calibration of the model generates most of the observed comovement between vacancies and output per worker. As a result, it reproduces a key stylized fact of the U.S. labor market: the negative comovement between unemployment and vacancies in the form of the Beveridge curve. The model therefore provides a coherent and quantitatively accurate picture of the joint cyclical properties of both flows of workers in and out of unemployment, as well as the behavior of unemployment and vacancies.

A less well-documented limitation of the standard search and matching model relates to the propagation of the response of the job finding rate to aggregate shocks to labor productivity. The job finding rate is a jump variable in the standard model, responding instantaneously to aggregate shocks, while it exhibits a sluggish response in U.S. data. An appealing feature of the generalized model is that it delivers a natural propagation mechanism: The job finding rate is a function of the distribution of employment across firms, which we show is a slow-moving state variable in our model. Simulations reveal that this aspect of the model can help account for the persistence of the decline in job creation following an adverse shock.

Our efforts to understand the amplification and propagation of labor market flows over the business cycle are related to the important early work of Cooper, Haltiwanger, and Willis (CHW, 2007). They also study a random matching model with decreasing returns. This paper's model differs from theirs in at least two regards. First, CHW suppress worker bargaining power. This simplifies wage bargaining, but at the cost of severing the link between the bargained wage, on the one hand, and market tightness on the other. This paper investigates a more general bargaining rule that allows workers to extract some of the rents. Second, CHW include additional labor market frictions, such as fixed costs to hire and fire. We agree that the interaction of these adjustment costs with matching frictions deserves further attention. However, we find it helpful to abstract from these for a number of reasons. In particular, we are able to make substantial headway in characterizing the analytics of optimal establishment-level labor demand and aggregate job creation. This in turn allows us to highlight the effects of decreasing returns and to contrast our model more

if firms cannot differentiate workers' types when making hiring decisions, they will base their decision on the average, rather than the marginal, valuation of leisure among the unemployed. The same is unlikely to be true of the model studied here, since firms presumably know their production technology when making hiring decisions.

tightly with that in Mortensen and Pissarides. Thus, we see our paper as complementary to CHW, who study a richer model but rely principally on numerical solution methods.

In the closing sections of the paper, we move beyond characterizing the business cycle properties of the model and turn to evaluating its implications for a number of additional cross-sectional outcomes. First, recent literature has emphasized empirical regularities in the cyclical behavior of the cross-sectional distribution of establishment size: While the share of small establishments with fewer than 20 workers rises during recessions, the shares of larger firms decline (Moscarini and Postel-Vinay, 2009). The model replicates this observation: For each establishment size class considered, it broadly matches the comovement with unemployment over the business cycle observed in U.S. data. Given that these implications of the model are venturing farther afield from the moments it was calibrated to match, we view these results as an important achievement.

In our final quantitative application, we evaluate the model's ability to account for the observation that workers employed in larger firms are often paid higher wages—the employer size-wage effect (Brown and Medoff, 1989). A distinctive attribute of the model is that, by incorporating large firms with heterogeneous productivities, it can speak to this empirical regularity. The magnitude of the size-wage effect implied by the model is mediated by two competing forces, as noted by Bertola and Garibaldi (2001). On the one hand, the existence of diminishing returns in production might lead one to anticipate a negative relation between employer size and wages. On the other, larger firms also tend to be more productive. Quantitatively, the latter dominates, generating one quarter of the empirical size-wage effect.

The remainder of the paper is organized as follows. Section 1 describes the set-up of the model, and characterizes the wage bargaining solution together with the associated optimal labor demand policy of an individual firm. Given this, section 2 develops a method for aggregating this microeconomic behavior up to the macroeconomic level, and uses it to characterize the steady state equilibrium of the model. Section 3 introduces aggregate shocks to the analysis. It presents an approach to computing the out of steady state dynamics of the model through the use of analytical approximations. We then use the model in section 4 to address a wide range of quantitative applications. Finally, section 5 summarizes our results, and draws lessons for future research.

4.6 Extensions

The previous sections have shown that the model derived in sections 2 and 3 can account for many of the cross-sectional and cyclical features of the U.S. labor market. In this section, we push the model harder. We consider two additional outcomes which the model can speak to, but was not designed to account for: the cyclical dynamics of the cross-sectional distribution of establishment size, and the employer size-wage effect.

Cyclical Dynamics of the Employer Size Distribution Until now, we have focused separately on the implications of the model for the cross section of employers and for the aggregate dynamics of labor stocks and flows. Recent literature, however, has sought to understand the joint dynamics of the cross section. Moscarini and Postel-Vinay (2009) in particular emphasize empirical regularities in the cyclical behavior of the cross-sectional distribution of establishment size: The share of small establishments rises during recessions, while the shares of larger firms decline. Figure 6 reiterates this finding. It uses annual data from County Business Patterns for the years 1986 to 2007 on the number of establishments by employer size, and plots the log deviations from trend of the establishment size shares against the unemployment rate.[42] The dots plot the data, and the dot-dashed lines the corresponding least squares regression lines. The share of establishments with 1 to 19 workers rises with unemployment, while the shares of establishments with more than 20 employees decline as unemployment rises.

The blue dashed lines in Figure 6 plot the analogous relationships implied by simulations of the model of sections 2 and 3. The results are encouraging: The model replicates the observation that the share of establishments with fewer than 20 employees increases in times of recession, while the shares of the larger size classes decline. In addition, the model also provides a reasonable account of the magnitude of the cyclicality of the establishment shares. It comes very close to replicating the cyclical sensitivities for the 20 to 99 employee and the 1000+ employee groups, and implies around one half of the cyclical sensitivity of the 1 to 19 and 100 to 999 employer size classes.

The observation that the share of smaller establishments rises in a recession in both the model and the data is not in itself a surprising fact, since aggregate employment, the mean of the distribution of establishment size, falls during recessions. What is noteworthy about the model is that it replicates the position in the distribution—at around 20 employees—at

[42] Here, we use County Business Patterns data rather than data from the Small Business Administration (see Section 4.2) because the former affords us a longer consistent time series for the size classes that we are interested in for this exercise.

10 percent. As shown in Figure 7, the counterpart implied by the model is a wage premium closer to 2.3 percent.

Thus, while the model yields a positive size-wage effect, it generates only around one quarter of the magnitude of the effect observed in the data. We do not view this necessarily as a problem, however. The mechanism that accounts for the size-wage effect in the model—the interaction of surplus sharing with heterogeneity in employer productivity—is only one of a large number of proposed channels. Oi and Idson (1999) present a summary of these, including efficiency wages, market power, specific human capital, among others. In a model of wage posting, Burdett and Mortensen (1998) demonstrate that on-the-job search in the presence of labor market frictions also can generate a positive employer size-wage effect, as higher-paying firms recruit and retain more workers. The results of Figure 7 suggest that the model presented in this paper leaves room for these additional explanations.

5 Summary and Discussion

In this paper, we have introduced a notion of firm size into a search and matching model with endogenous job destruction. This yields a rich, yet analytically tractable framework. In a series of quantitative applications, we show that the model provides a useful laboratory for analyzing the salient features of both the dynamics and the cross section of the aggregate labor market. Specifically, a calibrated version of the model provides a coherent account of the distributions of establishment size and employment growth; the amplitude and propagation of the cyclical dynamics of worker flows; the Beveridge curve relation between unemployment and vacancies; and the dynamics of the distribution of firm size over the business cycle.

We conclude by placing this paper within the context of more recent research on labor market dynamics and then discuss avenues for future work. Since this paper's first writing, there have been a number of contributions to the search-and-matching literature that study the cross section of employment growth and the cyclical dynamics of labor market flows. The model in Fujita and Nakajima (2009) is closest to our own. They introduce exogenous separations to drive a wedge between gross worker flows and net job changes, which enables them to potentially speak to facts on both of these flows. Hawkins (2011) incorporates a costly entry decision into a framework that is otherwise similar to that introduced in this paper. A few other papers consider decreasing returns within search-and-matching frameworks that depart more noticeably from that studied here. For instance, Kaas and Kircher (2010) study a model of directed search under decreasing returns. The application of directed search enables them to replicate the finding in Davis, Faberman, and Haltiwanger

(2010) that fast-growing firms fill a greater share of their posted vacancies.[44] Schaal (2010) generalizes the on-the-job search model of Menzio and Shi (2009) to include decreasing returns at the firm level, which enables him to engage a rich set of cross-sectional facts (see footnote 10).

Going forward, there are a number of additional avenues for further related work. First, since the model has a well-defined concept of a firm, it lends itself to estimation using establishment level data. As a result, the analytical framework developed here will complement recent research efforts that have sought to solve and estimate search models using numerical methods (e.g. Cooper, Haltiwanger and Willis, 2007).

Second, our interpretation of the standard search and matching model as a model of kinked adjustment costs raises the question of the aggregate implications of other forms of adjustment costs in the labor market. Recent research has emphasized the importance of fixed adjustment costs in explaining the empirical properties of labor demand at the micro level (see for example Caballero, Engel, and Haltiwanger, 1997, and Cooper, Haltiwanger, and Willis, 2004). Incorporating these adjustment costs into the model will provide a unification of the joint insights of the two dominant approaches to the modelling of aggregate labor markets—the search and matching framework, and models of adjustment costs.

A final extension relates to the nature of wage setting. An attractive feature of incorporating firm size into models of the labor market is that an assessment of the multilateral dimension to wage bargaining between a firm and its many workers becomes feasible. This has been of particular interest in recent literature that has emphasized the importance of rigidities in the structure of wages within a firm, as well as of individual wages over time, for determining the volatility of unemployment (Bewley, 1999; Hall, 2005). While the wage bargaining solution derived in the present paper seeks to improve upon approaches in previous work, it is in many ways an idealized environment in which the wages of all workers can be renegotiated costlessly. This idealized setting, however, provides a fruitful benchmark for analyzing the implications of rigidities in renegotiation of wages within a firm, as well as across time.

6 References

Acemoglu, Daron and William Hawkins, "Wages and Employment Persistence with Multi-worker Firms," mimeo, University of Rochester, 2010.

[44] Firms in their model accomplish this by posting higher wages. In our random matching framework, firms do not post wages in order to affect the rate at which they fill jobs.