Estimating the stock-flow matching model using micro data

Martyn Andrews  Steve Bradley
University of Manchester  Lancaster University

Dave Stott  Richard Upward
Lancaster University  University of Nottingham

Final draft for Journal of the European Economic Association
Draft based on JEEA Guidelines to Authors
October 2011

*The authors thank The Economic and Social Research Council (under grant R000239142) for financial assistance. The data were kindly supplied by Lancashire Careers Service and are also available from the UK Data Archive (http://www.data-archive.ac.uk). The comments from participants at various presentations are gratefully acknowledged. These include the Economics departments at Dundee, Erlangen, Leicester, Manchester, Nottingham, as well as EALE 2002 (Paris), ESEM 2003 (Stockholm) and RES 2003 (Warwick). We are grateful to Barbara Petrongolo for helpful clarifying conversations, and to three referees and the editor for comments that have improved the paper.
Abstract

We estimate the stock-flow matching model using micro-level data from a well-defined labour market. Using a dataset of complete labour-market histories for both sides of the market, we estimate hazard functions for job-seekers and vacancies. We find that the stock of new vacancies has a significant positive impact on the job-seeker hazard, over and above that of the total stock of vacancies. There is an even stronger robust result for vacancy hazards. Thus we find evidence in favour of stock-flow matching, even when controlling for unobserved search heterogeneity and stratifying into sub-markets defined by location and occupation. [97 words]

Keywords: two-sided search, matched job-seeker/vacancy data, stock-flow matching, hazards

New JEL Classification: C41, E24, J41, J63, J64

Address for Correspondence:

Professor S. Bradley
Management School
Lancaster University
Bailrigg
Lancaster
LA1 4YX, UK

Email: s.bradley@lancaster.ac.uk
1 Introduction

This paper is an empirical investigation into how job-seekers and employers meet and match with each other. The dominant model in the literature is one of friction and congestion: agents on both sides of the market take time to find a suitable partner. Burdett & Wright (1998) and Pissarides (2000) are the original two-sided search models applied to the labour market. Both models, and others like them, incorporate many of the same basic structures and assumptions, as surveyed by Burdett & Coles (1999). Because the process by which agents meet each other is random, these classical two-sided models of search are referred to as random matching models.\footnote{More generally, see the July 2011 issue of the American Economic Review for surveys from each of the three 2010 Novel Laureates “for their analysis of markets with search frictions” (Diamond 2011, Mortensen 2011, Pissarides 2011).}

More recently, alternative models of the matching process have stressed the role of sub-markets, separated by location or occupation. Models of directed search (e.g. Moen 1997, Acemoglu & Shimer 1999) allow for free entry by job-seekers and vacancies into these sub-markets, with firms posting wages and job-seekers choosing which sub-market to search in. In contrast, models of mismatch (e.g. Taylor 1995, Lagos 2000, Shimer 2007) posit that agents are fixed to particular locations or occupations, and cannot move without some cost.

Stock-flow matching offers another theory of mismatch. The key distinction here is that matching occurs via a marketplace. In a marketplace, agents can search the other side of the market in a short period of time. This will be relevant whenever there are technologies such as newspapers, employment agencies or the internet which allow simultaneous posting of vacancies or job-seekers. Unemployment and vacancies persist because suitable partners were not available on this first search of the market. For example, if a newly unemployed job-seeker searches the market and fails to find a match, she enters the stock of unemployed job-seekers and can then only match with the flow of new vacancies entering the marketplace. Symmetrically, employers enter the marketplace with vacancies, which they either fill, or the vacancy increases the stock. The lack of suitable
partners arises because of shortages in particular locations, or shortages of particular types. Thus, most matches occur between the stock on one side of the market and the inflow on the other. The stock-flow model was originally developed by Coles & Muthoo (1998) and Coles & Smith (1998); recent developments are Ebrahimy & Shimer (2010) and the closely related Shimer (2007).

Our contribution is to use microeconomic data to understand the underlying matching process. We specify and implement a test of random matching as a special case of stock-flow matching, which uses agent-level data on job-seekers and vacancies from the same market. The data we use are for a well-defined labour market in the UK between 1988 and 1992. We observe matches between job-seekers and vacancies and we observe how long each agent has been in the market when they match. Specifically, we focus on the job-seeker hazard when the job-seeker becomes old, whose covariates are the stock of market participants, namely the stock of unemployed job-seekers and the stock of vacancies. This describes a form of the classical random matching model estimated many times in the literature with aggregate data. We then add the stock of new vacancies, and see whether it has any impact on the hazard of getting a job over and above the effect of the stock of all vacancies. Exactly the reverse applies to the other side of the market, where the test examines the effect of the stock of new job-seekers on the vacancy hazard.

We find that the stock of new vacancies has a significant additional impact on the exit rate for old job-seekers. Our estimate implies that an old job-seeker is three times more likely to match with a new vacancy than an old vacancy. There is a larger effect on the other side of the market, with an old vacancy being nine times more likely to match with a new job-seeker than an old job-seeker. In short, with these data, we reject the random matching model.

The stock-flow matching model, along with other models of mismatch, have policy implications which are significantly different from those which arise from traditional frictional search models. There are two aspects to this. The first re-

\footnote{The data are from computerised records of the Lancashire Careers Service. The Careers Service was a Government-funded network which operated a free matching service for employers and young job-seekers.}
lates to the established literature on unemployment benefits, where, in a model with search frictions and unobserved search effort, an optimal unemployment insurance scheme is to reduce payments with unemployment duration to encourage greater search effort. Coles & Petrongolo (2008) note that, in the stock-flow model, greater search effort is not a productive investment because vacancies (assumed to be on the short side of the market) are always filled, regardless of search effort. Instead, as noted by Ebrahimy & Shimer (2010) it is the level of benefits rather than their duration which matters, because this changes the number of acceptable matches.

The second policy issue arises from the assumed heterogeneity across markets. For example, Lagos (2000) shows that, if the observed frictions in the aggregate matching function are the result of aggregating smaller markets each with stock-flow matching, then the effect of policies may be more fundamental than simply to shift a fixed matching function. Instead, policies will change the shape of the matching function itself. If stock-flow matching is the result of the uneven distribution of job-seekers and vacancies across locations (as in Lagos 2000) or across occupations (as in Shimer 2007), then regional policies that move firms closer to workers or workers closer to firms are needed. Coles, Jones & Smith (2010) also argue that, under stock-flow matching, the incidence and matching rate of the unemployed on the long side of the market are crucial measures of the benefits of a variety of other regional policies, such as propping up an industry or employer or encouraging the birth of new small firms. Training policies leading to more suitable training opportunities for workers and which offer them incentives to train, would also reduce occupational mismatches.

The paper is organised as follows. In the next section, we survey the existing empirical evidence. In Section 3, we write down a simple test of stock-flow matching in a reduced-form framework that parameterises variations in hazard rates across different types of match. In Section 4, we describe the data and show how they are used to construct the key variables in the stock-flow matching model. Section 5 sets out the econometric methodology. We discuss our results and report robustness checks of our preferred model in Section 6. We also examine
whether using aggregate data reveals any bias due to ignoring search intensity or whether we observe spurious stock-flow effects because our data aggregate over sub-markets separated by location or occupation. Section 7 concludes.

2 Literature

There are three types of evidence which assess whether non-random matching models better describe the data than traditional random models. First, there is a literature that uses macro aggregates to assess the predictions of theory-based versions of non-random matching models. On these, Shimer’s (2008) survey concludes that the stock-flow model provides a better description of the co-movements of aggregate labour-market variables—the flows in and out of employment, unemployment and vacancies, and associated stocks—than does the traditional random matching model. Specifically, Shimer (2007) and Ebrahimy & Shimer (2010) generate, when calibrated appropriately, two robust features of the U.S. labour market: namely the Beveridge curve and the reduced-form matching function. Further, in both models, the co-movement of aggregate labour-market variables appear to be more volatile than those predicted by Pissarides’s (1985) random matching model in response to aggregated productivity shocks. On the other hand, Pissarides (2008) argues strongly that aggregate matching function will continue to survive.

Second, there is indirect support of stock-flow matching from a large number of papers which estimate the hazard rate out of unemployment (starting with Lancaster 1979, Nickell 1979) and a smaller number which estimate the hazard rate for vacancies (e.g. Burdett & Cunningham 1998). The negative duration dependence in hazards is consistent with the stock-flow model because the subsequent inflow of new vacancies is smaller than the stock that the job-seeker examines when first entering the market.

The third type of evidence directly tests the key prediction of the stock-flow model that the hazard for a job-seeker depends not just on overall labour market tightness, but on the numbers of new vacancies relative to the stock
of job-seekers. All previous evidence of this type comes from aggregate time-series data. Coles & Smith (1998) estimate job-seeker hazards using monthly aggregate time-series Job Centre data between 1987 and 1995 for the UK. They find that the outflows of unemployed workers with durations longer than one month are highly correlated with vacancy inflows, whilst those with durations shorter than one month are significantly correlated with vacancy stocks. Gregg & Petrongolo (2005) use similar data to construct a test, but their results suffer from aggregation bias, a well-known potential problem with studies that use aggregate data (Burdett, Coles & van Ours 1994). The problem is that the data record beginning-of-month stocks of unemployed and vacancies, rather than the stocks over the month in question, and this biases tests in favour of stock-flow matching. Coles & Petrongolo (2008) rework Gregg & Petrongolo (2005), but construct appropriate “at risk” measures of unemployed workers and vacancies within each month. They find evidence of one-sided stock-flow matching, whereby the stock of unemployed match with the inflow of vacancies, but not vice versa.

There are four features of our data that enhance an analysis of stock-flow matching. First, our data record who matches with whom. Coles & Petrongolo (2008) do not observe whether the match is between a new or old job-seeker and a new or old vacancy. Second, duration is recorded in days, which means we can model when agents go from being new to old. Aggregate data only record monthly flows. Third, our data record job-seeker and vacancy stocks at weekly intervals. Again, aggregate data only record monthly stocks. This allows us to address the problem of aggregation bias properly. Fourth, as with most micro data, we observe agent-level variables. This is important if they are correlated with the stocks of interest, but also allows us to look at distinct sub-markets defined by occupation and location to see whether stock-flow effects are, in fact, an artefact of aggregating over sub-markets.
3 A flexible structure for identifying hazard rates

In this section we describe a reduced form model which parameterises variations in job-seeker hazard rates (and vacancies), depending on how long they have been in the market. To motivate this, consider two key facts which emerge from our data. First, the matching rate for new job-seekers is 1.2 times the matching rate for old job-seekers. Second, new job-seekers are 1.5 times more likely to match with an old vacancy than are old job-seekers. On the other side of the market, these patterns are stronger. The matching rate for new vacancies is six times that for old vacancies, and new vacancies are more than six times more likely to match with new job-seekers than are old vacancies. This illustrates that it is too simplistic, as in random matching, to assume that job-seeker and vacancy hazard rates are driven by market tightness \( \theta = V/U \), where \( U \) is the stock of job seekers and \( V \) is the stock of vacancies. Our analysis attempts to understand the underlying matching process by re-estimating the above ratios, but controlling for observed covariates and unobserved heterogeneity.

We start with a standard random matching function of the form \( m = U^\alpha V^\beta \), with \( 0 < \alpha, \beta < 1 \). \( m \) is the average number of matches per week. In the stock-flow model, job-seekers are labelled as “old” and “new”—or equivalently labelled as “stock” and “flow”—depending on whether the job-seeker has already had an unsuccessful contact with a vacancy. We define an old job seeker as someone who has been on the market for more than \( k_w \) weeks, where \( k_w \) is a modelling choice. Given this, we decompose the total stock of job-seekers \( U \) into \( \bar{U} + u \), where \( \bar{U} \) is the stock of job seekers who have been unemployed for more than \( k_w \) weeks and \( u \) is the stock of those unemployed for \( k_w \) weeks or fewer. Similarly, the stock of vacancies \( V \) is decomposed into \( \bar{V} + v \), the stocks of vacancies whose duration is longer than, or shorter than, \( k_e \) weeks respectively. We refer to the first \( k_w \) and \( k_e \) weeks of a spell of search as the matching “window”, the period over which the current stock of all potential partners on the other side of the market can be searched.

---

\(^3\)The data are described fully in Section 3.2, and the reader can calculate these ratios directly from Table 1.
Random matching predicts that the number of matches between, say, new job seekers and old vacancies is given by

\[ m_{12} = \left[ \frac{u}{U + u} \right] \left[ \frac{\bar{V}}{V + v} \right] m \]

as each match is as likely as another. Instead, we consider a more flexible structure:

\[ m_{12} = a_{12} \left[ \frac{u}{U + u} \right] \left[ \frac{\bar{V}}{V + v} \right] m, \]

where now some types of matches may occur more often than others. By estimating these \( a_{ij} \) parameters on agent level data, we identify the micro-structure of the matching process. Substituting \( m = U^\alpha V^\beta \), the flow of matches between new job seekers and old vacancies is

\[ m_{12} = a_{12} u \bar{V} U^{\alpha - 1} V^{\beta - 1}. \]

There are similar expressions for \( m_{11}, m_{21} \) and \( m_{22} \).

We emphasise that we depart from random matching only by introducing the four parameters \( a_{11}, a_{12}, a_{21} \) and \( a_{22} \). Thus we cannot say anything about the precise mechanism by which the matching probability varies with duration in the marketplace. For example, we cannot tell if non-random matching occurs via the interview or contact process, or via the matching probability conditional on contacts. For this, one needs data on contacts as well as matches. A more general model allows \( \alpha \) and \( \beta \) to vary across match types, and in our empirical work we do not constrain these parameters to be equal.

Under random matching, \( a_{11} = a_{12} = a_{21} = a_{22} = a \). Under stock-flow matching, old job-seekers are more likely to match with new vacancies and so \( a_{21} > a_{22} \). The stock-flow matching model is silent about how \( a_{12} \) and \( a_{21} \) compare with \( a_{11} \), but we expect \( a_{11} \) to be positive with weekly data. We also note that matches between old agents do occur in the data, and so \( a_{22} > 0 \). This

\[ ^4 \text{This notation is used throughout the paper. The first subscript is 1 if the job-seeker is “new” and is 2 if job-seeker is “old”. The second subscript refers to vacancies.} \]
happens, for example, because job-seekers, having entered the old stock, might revise down their reservation utilities, and so re-examining the stock might then reveal potential matches.

Because there are four types of match, there are four different hazard rates out of unemployment. These are given by:

\[ h_{11}^w = \frac{m_{11}}{u} = a_{11} v U^{\alpha - 1} V^{\beta - 1} \]  
\[ h_{12}^w = \frac{m_{12}}{u} = a_{12} \bar{V} U^{\alpha - 1} V^{\beta - 1} \]  
\[ h_{21}^w = \frac{m_{21}}{\bar{U}} = a_{21} v U^{\alpha - 1} V^{\beta - 1} \]  
\[ h_{22}^w = \frac{m_{22}}{\bar{U}} = a_{22} \bar{V} U^{\alpha - 1} V^{\beta - 1}. \]

There are another set of hazards for vacancies, labelled \( h_{11}^e, h_{12}^e, h_{21}^e \) and \( h_{22}^e \). The log-hazard for a new job-seeker, \( h_{11}^w = h_{11}^v + h_{12}^w \), comes from summing Equations (1) and (2):

\[ \log h_{11}^w = \log[a_{11} v + a_{12} (V - v)] + (\alpha - 1) \log U + (\beta - 1) \log V + \epsilon^w, \]  

where \( \epsilon^w \) captures unobserved job-seeker heterogeneity. Similarly, using Equations (3) and (4), the hazard for an old job-seeker, \( h_{21}^w = h_{21}^v + h_{22}^w \), is:

\[ \log h_{21}^w = \log[a_{21} v + a_{22} (V - v)] + (\alpha - 1) \log U + (\beta - 1) \log V + \epsilon^w. \]

Rather than estimate the non-linear models in Equations (5) and (6), it is much easier to linearise the model:

\[ \log h_{21}^w = \pi_0 + \pi_1 \log V + \pi_2 \log v + \pi_3 \log U + \pi_4 \log u + \epsilon^w. \]

From the estimates of \( \pi_1, \pi_2, \) and \( \pi_3 \), we can uniquely identify \( \alpha, \beta \) and \( a_{22}/a_{21} \):

\[ \frac{\partial \log h_{21}^w}{\partial \log V} = \frac{a_{22} V}{a_{21} v + a_{22} V} + \beta - 1 = \pi_1 \]  
\[ \frac{\partial \log h_{21}^w}{\partial \log U} = \frac{a_{22} V}{a_{21} v + a_{22} V} + \beta - 1 = \pi_3 \]  
\[ \frac{\partial \log h_{21}^w}{\partial \log v} = \frac{(a_{21} - a_{22}) v}{a_{21} v + a_{22} V} = \pi_2 \]  
\[ \frac{\partial \log h_{21}^w}{\partial \log u} = 0 = \pi_4. \]
These estimates should be interpreted as follows. First, an increase in the stock of unemployed job-seekers $U$ has the familiar effect of $\alpha - 1$, and it does not matter whether the congestion comes from new or old job-seekers, which is why the extra effect from new job-seekers $u$ is zero. Second, to obtain an estimate of $\beta$, one adds together the estimates on $\log V$ and $\log v$ (ie $\pi_1 + \pi_2 = \beta$). Third, the coefficient on $v$ should be zero if the random matching model is true, because $a_{21} = a_{22}$ implies $\pi_2 = 0$. This is a one-sided test because, under the alternative, $\pi_2 > 0$. It is important to understand what is happening in the stock-flow model when $\pi_2 > 0$. Suppose that the stock of new vacancies $v$ goes up whilst the stock of all vacancies $V$ remains fixed, which means that the stock of old vacancies $\bar{V}$ falls. Under random matching, this switch between new and old has no effect on the hazard. Under stock-flow matching, $v$ going up leads to more stock-flow matches but $\bar{V}$ going down means fewer stock-stock matches. The net effect is positive if $a_{22} < a_{21}$. The same can be seen from the estimate of $a_{22}/a_{21}$, obtained directly from the expression for $\pi_2$, which is given by

$$\frac{a_{22}}{a_{21}} = \left[ \frac{V}{v} (1 - \pi_2)^{-1} - \frac{\bar{V}}{v} \right]^{-1}. \quad (9)$$

If $\pi_2 > 0$, ie the effect of $v$ is significant and positive, then $a_{22}/a_{21} < 1$.

Under random matching, the log-hazard for all job-seekers, regardless of their "age", is simply

$$\log h^w = \log(a) + (\alpha - 1) \log U + \beta \log V + \epsilon^w, \quad (10)$$

This is the standard hazard function for the random matching model. It only depends on the total stock of job-seekers $U$ and vacancies $V$; it does not depend on the proportion of new and old in the stock.

Analogous discussions apply when we use data on vacancy spells. Here we estimate $\log h^e_2(U, u, V, v, \epsilon^e)$ and see whether $u$ is significant. This tests whether
\[ a_{12} = a_{22} \text{ and an estimate of } a_{22}/a_{12} \text{ is obtained from:} \]

\[
\frac{a_{22}}{a_{12}} = \left[ \frac{U}{u} (1 - \pi_2)^{-1} - \frac{U}{u} \right]^{-1},
\]

(11)

where \( \pi_2 \) now refers to the estimate on \( u \).

4 The Data

4.1 The Lancashire Careers Service data

The data we use are the computerised records of the Lancashire Careers Service over the period March 1988 to June 1992. The Careers Service was a Government-funded network which operated a free matching service for employers and youths.

Lancashire is a sub-region of the North West of England with a 25\% lower than average gross value added per head, which accounts for about one-fifth of employment in the North West Region. Over the last two decades the composition of employment has shifted from manufacturing to services, reflecting the long term structural changes which have occurred throughout the UK economy. The North West region has experienced persistently higher unemployment than the UK, but unemployment in the Lancashire sub-region is somewhat lower than for the region as a whole. The educational background of young workers in Lancashire does not differ substantially to that for the rest of the UK.

The matching technology used in the youth labour market was provided by the Careers Service rather than Job Centres; the latter provided the equivalent role in the adult labour market. Careers Services tended to offer more vocational guidance than Job Centres. However, the fundamental function of “matching” vacancies with job-seekers was essentially the same. Thus, our data do provide an insight into the public matching service more generally. The UK Office For National Statistics estimate that between a third and a half of all vacancies are advertised via public Job Centres (Machin 2003). In our data, about 20\% of job spells result from a match with a vacancy posted with the Careers Service.
The data comprise a longitudinal record of all youths in Lancashire aged 15–18. For each job-seeker, we observe the start date of every labour market spell over the sample period. Some spells are right-censored. The data also include a record of all vacancies notified to the Careers Service over the sample period: again, we observe completed and right-censored spells of vacancies placed on the market. Thus our job-seeker spells and vacancy spells comprise standard flow samples.

Job-seekers are observed in one of four labour market states: unemployment, employment, government-sponsored training or education. Vacancies are either posted with the Careers Service or not. Each job-seeker spell therefore has four possible outcomes: they can match with a vacancy posted with the Careers Service, they can match with a vacancy not recorded in our data, they can withdraw from the labour market, or their unemployment spell can be censored by the end of the sample period. A vacancy has six possible outcomes: it can match with a job-seeker from one of the four possible labour market states, it can be withdrawn from the market, or it can be censored.

We analyse matches between job vacancies and unemployed job-seekers. Matches involving school-leavers and those on training programmes are less relevant for the purpose of testing theories of labour market matching. A spell which ends in a different kind of match is treated as censored.

We need to consider other types of job-seeker when specifying the arguments of the matching function, because it might be the case that the stock of those engaged in on-the-job search affects the probability of a match between unemployed job-seekers and vacancies because they are competing for the same vacancies. We therefore use two definitions of job-seekers. The “narrow” definition refers only to unemployed job-seekers, whereas the “wide” definition includes those who are on training programmes and those who are in jobs, and who are registered as actively searching with the Careers Service.

We are also able to distinguish between the “search duration” and the “spell duration” for each agent. Search duration ends when a successful contact between a job-seeker and a vacancy is recorded in the data. Spell duration ends
when a job-seeker actually starts working in a new job, which is typically some
time after a successful contact. These durations form the dependent variables for
estimating job-seeker hazards $h^w$ and vacancy hazards $h^e$. Our preferred speci-
fication focuses on the duration of search, because this corresponds more closely
to the theory. However, as almost all existing estimates of the matching function
use spell durations, we also examine what happens when we use this measure.

### 4.2 The Dependent Variable

The data are organised into sequential binary response form. This means that
we pool over all the job-seekers in the data, and generate an unbalanced panel
of job-seeker spells with $t_i^w$ observations for each spell $i$. Each row in this panel
corresponds to a job-seeker week, of which there are 477,868. This defines the
risk set for job-seekers. The total number of job-seeker spells is 34,657. Some
job-seekers have multiple spells; the total number of job-seekers is 26,113.

| Table 1: Who matches whom? Search duration; $k^w = 4$, $k^e = 2$ weeks |
|---|---|---|
| **Job-seekers**$^a$ | New | Old | Total |
| Zeros | 123,338 | 351,769 | 475,107 |
| Exits to new vacancy $(m_{11})$ | 394 | (m$_{21}$) 1,133 | 1.527 |
| Exits to old vacancy $(m_{12})$ | 416 | (m$_{22}$) 818 | 1.234 |
| **Total** | 124,148 | 353,720 | 477,868 |
| **Vacancies**$^b$ | | | |
| Zeros | 23,334 | 111,128 | 134,462 |
| Exits to new job-seeker $(m_{11})$ | 394 | (m$_{12}$) 416 | 810 |
| Exits to old job-seeker $(m_{21})$ | 1,133 | (m$_{22}$) 818 | 1,951 |
| **Total** | 24,861 | 112,362 | 137,223 |

$^a$477,868 job-seeker weeks correspond to 34,657 job-seeker spells and
26,113 job-seekers.

$^b$137,223 vacancy weeks correspond to 14,154 vacancy spells, 9,556
vacancy orders, and 4,121 employers.

The dummy variable $y^w_{is}$ indicates whether the $i$-th job-seeker spell ends with
a match in week $s$. In other words, we have a sequence of observations $y^w_{is}, s =
1, \ldots, t_i^w$, all of which are zero except the last. For the last observation ($s = t_i^w$),
if the job-seeker matches with a vacancy, then $y^w_{is} = 1$; if the job-seeker spell is
censored, then $y^w_{is} = 0$. 

12
Analogous considerations apply to the $j$-th vacancy spell. Pooling over all the vacancies in the data, we generate an unbalanced panel of vacancies with $t_j^e$ observations for each vacancy spell $j$. The risk set for this side of the market is 137,223 vacancy weeks, corresponding to 14,154 vacancy spells. Some employers have multiple vacancy spells. This is because vacancies are often bundled together in so-called vacancy orders and some employers are observed more than once. The total number of vacancy orders is 9,556 and the total number of employers is 4,121.

Summing over $y_{jw}$ in the job-seeker panel or over $y_{jes}$ in the vacancy panel gives the total number of matches in the data, $m$, which is 2,761.

Table 1 summarises the raw data for both panels, using the “search” definition of duration, under the assumption that $k^w = 4$ weeks and $k^e = 2$ weeks. There are 477,868 job-seeker weeks at risk, of which 2,761 end as matches. These are disaggregated by whether the job-seeker spell is old (unemployed $> 4$ weeks) or new (unemployed $\leq 4$ weeks) and by whether or not the job-seeker exits to a new or old vacancy, where “old” means that the vacancy remains unfilled after 2 weeks. Thus, for example, there are $m_{21} = 1,133$ job-seekers who has been unemployed for more than 4 weeks and who match with a vacancy which has been open for 2 weeks or fewer.

In the lower panel, the 137,223 vacancy weeks are disaggregated in the same way. There are more unemployed job-seeker weeks at risk than there are vacancy weeks because the youth labour market in the UK in the early 1990s was particularly slack: compare the 34,657 job-seeker spells with the 14,154 vacancy spells.

Using Table 1, we compute the raw hazard for job-seeker spells which match with old vacancies. When the job-seeker is new, $h_{12}^w = 416/124,148 = 0.003351$, but when she becomes old, $h_{22}^w = 818/353,720 = 0.002313$. In other words, the drop in the raw hazard is $h_{22}^w/h_{12}^w = 0.690$, which is perfectly consistent with stock-flow matching, although, because 30% of our matches are between old job-seekers and old vacancies, a pure form of the theory does not occur in these data. For vacancies, the drop in the hazard for spells matching with old job-seekers is
\( h_{22} / h_{21} = 0.160 \). This is also consistent with stock-flow matching, and is more pronounced on this side of the market.

Table 1 can be recomputed for any values of \( k_e \) and \( k_w \). In deciding which values to choose, we note that employers, who have been in the market in previous years, know exactly what kind of job-seeker they are looking for. Young job-seekers, on the other hand, are relatively inexperienced and might only have a vague idea about what they want, and thus searching the market takes longer, on average. In other words, we constrain our choices to \( k_w \geq k_e \). Choosing \( k_e \) is relatively straightforward: after 5 weeks, many vacancies (62%) have been filled, which suggests that \( k_e = 5 \) is too long. Because it can take at least a week to organise interviews, \( k_e = 1 \) week is too short; hence we choose \( k_e = 2 \) weeks (after which 39% have been filled), but bear in mind that \( k_e = 3 \) or \( k_e = 4 \) are possibilities. Next, given \( k_e = 2 \), we compute the number of stock-flow matches for \( k_w = 2, 3, 4, \ldots \), and note that this is maximised when \( k_w = 4 \). Given this, our choice of \( k_w = 4 \), \( k_e = 2 \) weeks seems a sensible starting point, but our strategy when we come to estimation is to choose a small number of \( (k_w, k_e) \) pairs, such that \( k_w \geq k_e \), to see whether it makes any difference to the results.

Finally, we rule out the possibility that \( k_w \) varies over job-seekers and that \( k_e \) varies over vacancies; this is because job-seekers and vacancies are using the same matching technology (the Careers Service).

One important feature of these data is that, whatever the choice of \( k_w \) and \( k_e \), there are a substantial number of “old-old” matches. One possible explanation is that the time an agent has been on the market is not an accurate measure of actual search activity. For example, an old job-seeker might have changed their occupational preference and thus has not yet contacted vacancies in this new occupation. Alternatively, suppose some job-seekers only contact vacancies infrequently. If there is high vacancy turnover, the stock of sampled vacancies may in fact all be new to these job-seekers. In both cases, the possible mis-classification can only be corrected with data on contacts. A third related explanation is that the data reveal that such “old-old” matches tend to involve “skilled” job-seekers matching with “skilled” vacancies (precise definitions of “skilled” are discussed
later in Section 6.4). With such matches, interviewing/screening plays a more important role than with matches between “unskilled” agents.\(^5\)

### 4.3 The Explanatory Variables

The explanatory variables in the stock-flow model are measures of the “stocks” and “flows” of job-seekers and vacancies. In practice, the “flows” are stocks of new job-seekers and vacancies, while the “stocks” are stocks of old job-seekers and vacancies. As with the dependent variables, the sizes of these new and old stocks will depend on the values of \(k^w\) and \(k^e\).

During a given week \(t - 1\) in calendar time, there is an inflow of job-seekers \(u^+_{t-1}\) into the stock of job-seekers \(U_{t-1}\), and an outflow \(u^-_{t-1}\), such that

\[
U_t = U_{t-1} + (u^+_{t-1} - u^-_{t-1}).
\]

Unfortunately, these job-seeker data are a flow sample, which means that \(U_t\) is not observed. However, we observe data for thirty weeks before the sample period, and so \(U_t\) is built up recursively from the net inflow into unemployment \(u^+ - u^-\) each period. In other words, \(U_{-30}\) is set to zero. We have checked that our imputed measure essentially coincides with the equivalent measure of \(U_t\) from official unemployment stock data from October to April from 1989 onwards. The vacancy stock data are from a stock sample and so we observe \(V_t\) for all \(t\).

If \(k^w = k^e = 1\) week, then \(U_t\) is disaggregated into new and old stocks as follows (with an analogous expression for \(V_t\)):

\[
U_t = [u^+_{t-1} - u^-_{t-1}] + [U_{t-1} - u^-_{t-1}] = u_t + \bar{U}_t.
\]

The new stock \(u_t\) of unemployed is defined as the inflow of unemployed during the week less those who also exit during the week from the inflow, namely \(u^+_{t-1} - u^-_{t-1}\). Similarly, the old stock \(\bar{U}_t\) is defined as the stock of unemployed at

\(^5\)Of matches involving skilled vacancies, 39.3% are old-old, whereas only 24.0% of unskilled vacancies are old-old. Similarly, of matches involving skilled job-seekers, 34.0% are old-old, but of matches involving unskilled job-seekers, only 26.0% are old-old.
the end of the previous week less those who also exit during the current week from the stock at the beginning of the week, namely $U_{t-1} - u_{t-1}^- | U_{t-1}$. The above expression generalises for any window size $k$:

$$U_t = \left[ \sum_{i=1}^{k} u_{t-i}^+ - \sum_{i=1}^{k} u_{t-i}^- \right] + \left[ U_{t-k} - \sum_{i=1}^{k} u_{t-i}^- | U_{t-k} \right] \equiv u_t^k + \bar{U}_t^k.$$ 

When constructing the covariates $u, \bar{U}, U, v, \bar{V},$ and $V$, we group Lancashire into three labour markets ("districts": West, Central and East). 96% of all matches take place between a job-seeker and vacancy from the same labour market. This number drops to 75% when Lancashire is treated as 14 towns/cities. There are very large peaks in both new and old unemployed stocks, arising from young people leaving school between May and August each year, which, of course is when employers post their vacancies. There is a similar annual variation in the data for new vacancy stocks, but less pronounced.

### 4.4 Temporal Aggregation Bias

Temporal aggregation bias is an important issue in this literature, and is discussed at length by Burdett et al. (1994), Gregg & Petrongolo (2005) and Coles & Petrongolo (2008). In the context of monthly data, the problem arises in not observing the instantaneous hiring rate, but rather flows over a discrete period (a month). The assumptions one needs to adjust the stock measures depend on how quickly agents are matching, which itself is being modelled, and so there is a simultaneity bias. Coles & Petrongolo (2008) estimate matching functions using a maximum likelihood technique to deal with this problem. In our data this will not be a problem as we observe weekly flows together with stocks that also vary weekly. (Had we used daily stocks, the issue would completely disappear; we have checked that using daily data has very little impact on our results below.)

What we are able to do, specifically, is assess the extent to which using monthly stocks data biases the estimates. Using the *same* flows data, we use two sets of the stocks data: (a) stocks measured weekly, ie the value observed on the Monday of each week and (b) stocks measured monthly, ie the value observed on
the first week of the month.

5 Econometric Methodology

In this section, we describe how we estimate the hazard to matching on both sides of the market. The econometric framework we use is a reduced-form mixed proportional hazards (MPH) model, a framework widely used in this literature. We estimate a discrete-time version of this model, using weekly data, because it allows us to estimate the baseline hazard non-parametrically.

Consider the hazard for a new job-seeker spell $i$, where the job-seeker matches with a new vacancy. Equation (5) describes how the observed covariates and unobserved heterogeneity affect the log–hazard. We write the hazard as $h^w_i(x_{is}, \epsilon^w_i)$, defined as the probability that this job-seeker matches at some point between elapsed duration $s - 1$ and $s$, conditional on having survived to $s - 1$:

$$h^w_i(x_{is}, \epsilon_i) = \Pr\{T_i \in [s - 1, s) | T_i \geq s - 1, x_{is}, \epsilon_i\} \quad s = 1, 2, \ldots, t_i \leq k^w. \quad (12)$$

Here $T_i$ is the latent duration of spell $i$, $t_i$ is the completed duration of spell $i$, $x_{is}$ is a vector of observed covariates, and $\epsilon_i$ is the spell-invariant unobserved heterogeneity term.$^6$ $x_{is}$ is specified as $[\log u_{is}, \log v_{is}, \log U_{is}, \log V_{is}]$, together with the employer and job-seeker controls. For this job-seeker, completed duration does not exceed $k^w$. If a job-seeker becomes old, the hazard is written

$$h^w_{2i}(x_{is}, \epsilon_i) = \Pr\{T_i \in [s - 1, s) | T_i \geq s - 1, x_{is}, \epsilon_i\} \quad s = k^w + 1, \ldots, t_i. \quad (13)$$

To model the effect of covariates on the hazard rate, we make the proportional hazards assumption. Then the precise form of the discrete hazard is given by the complementary log-log link function:

$$h^w(x_{is}, \epsilon_i) = 1 - \exp\{-\exp[d_i x_{is} \beta_1 + (1 - d_i) x_{is} \beta_2 + \gamma_s + \epsilon_i]\} \quad s = 1, \ldots, t_i. \quad (14)$$

$^6$For notational clarity, where possible we drop the superscript $w$. The subscript $i$ tells the reader that this is a job-seeker hazard; the subscript $j$ is used for a vacancy hazard.
We write the model like this because these two hazards are actually estimated as one regression model by pooling the data in and interacting each covariate with a dummy variable indicating whether or not the job-seeker is new, \( d \equiv 1(s \leq k^w) \). The \( \gamma_s \) terms are interpreted as the log of a non-parametric piecewise linear baseline hazard. Each interval corresponds to a week, but, because of data thinning, these are grouped into longer intervals at longer durations by constraining corresponding duration dummies. Note that the random effect refers to a job-seeker, not a job-seeker spell.

The log-linear specification for the hazard, Equation (7), means that the parameters on the four covariates are easily interpreted. Write either of the two models in Equation (14) as:

\[
\begin{align*}
    h^w = 1 - \exp[-\exp(x\beta + \gamma + \epsilon^w)]
\end{align*}
\]

Then

\[
\log[-\log(1 - h^w)] \approx \log h^w = x\beta + \gamma + \epsilon^w.
\]

Because the model is linear in the logs of the covariates, we can interpret the parameters as elasticities in the usual way.

Although identification in duration models is an important issue, van den Berg (2001, Section 5) shows that identification of the MPH model requires that the hazard is multiplicative in elapsed duration \( t \), the covariates \( x \) and the heterogeneity term \( \epsilon \). However, allowing the effects of the covariates to vary over elapsed duration is a departure from these standard requirements. More recently, Brinch (2007) shows that, because the stocks of job-seekers and vacancies vary through calendar time and vary across labour markets, our model is identified.

As explained in Section 4.2, a standard approach for estimating this model is to expand the data so that each job-seeker spell contributes \( t_i \) rows/weeks. As also explained, the dependent variable \( y_{its} \) is binary. Estimation of binary choice models with a complementary log-log link and random effects is standard; here we model the unobserved heterogeneity \( \epsilon_i \) either using Gaussian mixing or discrete mixing. The same applies to the exit hazard for a vacancy spell, replacing \( i \) by \( j \) and \( w \) by \( e \). The random effect \( \epsilon_j \) is defined for a vacancy order, not an employer.
6 Results

6.1 Base model

Table 2: Base Model*

<table>
<thead>
<tr>
<th></th>
<th>Search duration</th>
<th>Spell duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Job-seekers, h^w</td>
<td>Vacancies, h^w</td>
</tr>
<tr>
<td>(a) New</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log u</td>
<td>-0.185 (0.065)</td>
<td>0.204 (0.075)</td>
</tr>
<tr>
<td>log U</td>
<td>0.363 (0.130)</td>
<td>1.207 (0.208)</td>
</tr>
<tr>
<td>log v</td>
<td>0.319 (0.084)</td>
<td>-0.065 (0.078)</td>
</tr>
<tr>
<td>log V</td>
<td>0.020 (0.121)</td>
<td>-0.651 (0.142)</td>
</tr>
<tr>
<td>(b) Old</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log u</td>
<td>-0.215 (0.062)</td>
<td>0.440 (0.072)</td>
</tr>
<tr>
<td>log U</td>
<td>0.360 (0.147)</td>
<td>1.282 (0.220)</td>
</tr>
<tr>
<td>log v</td>
<td>0.273 (0.056)</td>
<td>0.097 (0.071)</td>
</tr>
<tr>
<td>log V</td>
<td>-0.077 (0.100)</td>
<td>-0.633 (0.144)</td>
</tr>
<tr>
<td>Gaussian st. err.</td>
<td>0.659 (0.073)</td>
<td>1.700 (0.083)</td>
</tr>
<tr>
<td>log L</td>
<td>-16,382.0</td>
<td>-11,026.9</td>
</tr>
<tr>
<td>Obs</td>
<td>477,868</td>
<td>137,223</td>
</tr>
</tbody>
</table>

Derived parameter estimates

(a) New

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>1.179 (0.118)</td>
<td>1.411 (0.188)</td>
</tr>
<tr>
<td>β</td>
<td>0.339 (0.110)</td>
<td>0.284 (0.133)</td>
</tr>
<tr>
<td>α + β</td>
<td>1.518 (0.143)</td>
<td>1.694 (0.197)</td>
</tr>
<tr>
<td>α-ratio^b</td>
<td>0.258 [0.000]</td>
<td>0.289 [0.006]</td>
</tr>
</tbody>
</table>

(b) Old

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>1.145 (0.134)</td>
<td>1.722 (0.197)</td>
</tr>
<tr>
<td>β</td>
<td>0.195 (0.095)</td>
<td>0.463 (0.136)</td>
</tr>
<tr>
<td>α + β</td>
<td>1.340 (0.152)</td>
<td>2.186 (0.209)</td>
</tr>
<tr>
<td>α-ratio^c</td>
<td>0.302 [0.000]</td>
<td>0.117 [0.000]</td>
</tr>
</tbody>
</table>

*Wide definition of job-seeker stocks. For the search definition, the weighted averages across the three 'districts' for u, U, v and V are 207, 1991, 33 and 207 respectively. For the spell definition, these are 208, 1991, 33 and 207. The matching window is k^w = 4 and k^e = 2.

Estimates based on 2761 matches between 34,657 unemployed job-seeker spells (26,113 job-seekers) and 14,154 LCS job vacancies (4,121 employers and 9,556 orders). Standard errors in parentheses. All regressions contain dummy variables for month (11), year (4), and 'district' (2), and also contain job-seeker and vacancy covariates. For job-seeker regressions, these are gender (1 dummy), grades at age 16–17 (so-called GCSEs) (3), ethnicity (1), disadvantaged social background (1); for vacancy regressions, these are whether the vacancy requires a skilled employee (1), a non-manual employee (1), a written method of application (1), firm size (3) and wage (4).

*a Standard error for Gaussian heterogeneity. The number of quadrature points is 12 for both job-seeker regressions and 24 for both vacancy regressions.

^b a_{12}/a_{11} for job-seekers, a_{21}/a_{11} for vacancies. The a-ratios calculated from Equations (9) and (11). p-values are for one-sided alternative hypothesis.

^c a_{22}/a_{21} for job-seekers, a_{22}/a_{12} for vacancies.

The model described thus far is referred to as the Base Model, and is reported in Table 2. As discussed in Section 4.1, there are two variants, because we have a definition of duration based on the time searching for a new job, as well as a definition based on the duration whilst unemployed. The latter is the more
common definition, whereas the former corresponds more closely to the theory. Our test of stock-flow matching amounts to whether an increase in the number of new vacancies on the market significantly increases the exit probability for old job-seekers. In the old job-seeker hazard, using search duration (first column), this effect is estimated as \( \partial \log h^w_2 / \partial \log v = 0.273 \), and is significant with a standard error of 0.056. This converts to a point estimate for \( a_{22}/a_{21} = 0.302 \). In words, an old job-seeker is three times more likely to match with the new vacancy than an old vacancy. A weaker, but still significant, estimate occurs with the spell duration definition (third column), that is 0.194, converting to \( a_{22}/a_{21} = 0.398 \).

On the other side of the market, much stronger effects are estimated, and are the same for both search and spell regressions: the effect of \( \log u \) in the old vacancy hazard, search definition, \( \partial \log h^v_2 / \partial \log u \) is 0.440 which converts to \( a_{22}/a_{12} = 0.117 \) (second column). When using the spell definition, the effect of \( \log u \) is \( \partial \log h^v_2 / \partial \log u = 0.452 \) and \( a_{22}/a_{12} = 0.112 \). Here, an old vacancy is now nine times more likely to match with the new job-seeker than an old job-seeker, rather than just three times. Given there are many more job-seekers than vacancies in this particular labour market (Table 2), it makes good sense that there are stronger stock-flow effects in the vacancy regression. Remember, also, that a vacancy becomes old after just two weeks, not four weeks, as for job-seekers. It also makes good sense that there are stronger effects for the “better” search definition for job-seekers, but that it makes no difference for the vacancy regressions. Our analysis focusses on \( a_{22}/a_{21} \) for job-seekers, \( a_{22}/a_{12} \) for vacancies. This is because stock-flow matching is silent about how \( a_{12} \) and \( a_{21} \) compare with \( a_{11} \). Nonetheless, it seems intuitively sensible that a new agent is more likely to match with the new rather than old stock, which is what we find.

Together with the estimated \( \alpha \)-ratios, estimates of \( \alpha \) and \( \beta \) are reported below each regression. These vary by \( i, j \), because we are effectively running four separate regressions, for old/new job-seekers and old/new vacancies. The four pairs of estimates vary from specification to specification. In particular, the estimate of \( \alpha \) is stronger in the old vacancy regression than the old job-seeker
regression. The estimates of $\alpha$ are bigger than unity, whereas the estimates of $\beta$ range between 0.3 and 0.6, implying significant increasing returns to scale ($\alpha + \beta > 1$). This is not to do with stock-flow matching *per se*, as it turns out that scale effects persist when we estimate the random matching version in the next sub-section.

6.2 Departures from the Base Model

In this subsection, we look at various departures from the Base Model to assess the robustness of the stock-flow matching model, to assess whether the assumptions we have made are important or innocuous. The first two rows of Table 3 summarise the Base Model.

The first departure, Row (1), shows that the results are robust to the way the unobserved heterogeneity is modelled, because here we used discrete (Heckman-Singer) mixing. In the job-seeker regression the log-likelihood is unaffected, and the number of parameters being estimated for discrete and Gaussian mixing is the same. In the vacancy regression, the log-likelihood is 4.5 log-points higher, but there are 11 more parameters to estimate. The Akaike Information criterion suggests that Gaussian mixing is the better way of modelling heterogeneity.

Row (2) investigates whether the heterogeneity should be defined for a job-seeker or a job-seeker spell. This would make sense for job-seekers who experience repeated unemployment events and then experience scarring effects. As 25.2% of job-seekers experience multiple spells in the dataset, this is potentially an issue. However, Row (2) shows that the results do not change; moreover, the log-likelihood falls by 9.3 log-points. For the vacancy regressions, it does not make sense to define the heterogeneity for a vacancy rather than a vacancy order, because all the vacancies in an order are identical. About one half of employers have more than one spell; again, when we define the heterogeneity for an employer, the likelihood falls by 115.2 log-points.
Table 3: Summary of departures from Base Model∗

| Departures | Search duration (Table 2) | Spell duration (Table 2) | (1) Heckman-Singerb | (2) Job-seeker spell random effectsc | (2) Employer random effectsd | (3) Without heterogeneity | (4) Without covariates | (5) Monthly stocks | (6) Classical random matchinge | (7) Ditto, monthly stocksf | (8) Narrow stocks | (9) k∗ = 5, kε = 5 | (10) k∗ = 4, kε = 4 | (11) k∗ = 4, kε = 1 | (12) k∗ = 3, kε = 3 | (13) k∗ = 3, kε = 2 | (14) k∗ = 2, kε = 2 | (15) k∗ = 2, kε = 1 |
|------------|--------------------------|--------------------------|---------------------|----------------------------------|----------------------------|--------------------------|---------------------|------------------|---------------------------|--------------------------|------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|            | 207 33 0.273 (0.056) 0.302 1.145 0.195 -16382.0 | 208 33 0.194 (0.054) 0.398 1.101 0.177 -16334.7 | 207 33 0.275 (0.057) 0.300 1.152 0.219 -16391.3 | 207 33 0.275 (0.057) 0.300 1.152 0.219 -16391.3 | 207 33 0.484 (0.062) 0.100 1.339 0.623 -11142.1 | 207 33 0.271 (0.056) 0.304 1.152 0.184 -16393.8 | 207 33 0.270 (0.056) 0.305 1.202 0.18 1 -16504.8 | 221 56 0.163 (0.074) 0.589 1.340 -0.006 -16392.0 | 207 33 0f 1f 202 0.167 -16410.4 | 221 56 0f 1f 202 0.167 -16410.4 | 207 33 0.266 (0.056) 0.310 0.755 0.214 -16383.0 | 207 33 0.266 (0.056) 0.310 0.755 0.214 -16383.0 | 207 18 0.257 (0.045) 0.204 1.103 0.208 -16369.8 | 158 45 0.344 (0.062) 0.297 1.203 0.217 -16382.1 | 158 33 0.266 (0.054) 0.309 1.189 0.208 -16380.9 | 107 33 0.265 (0.052) 0.311 1.119 0.218 -16388.4 | 107 18 0.263 (0.041) 0.199 1.085 0.238 -16376.5 | 0.440 (0.072) 0.117 1.722 0.463 -11026.9 |

<table>
<thead>
<tr>
<th></th>
<th>log u</th>
<th>a22/a21</th>
<th>α</th>
<th>β</th>
<th>log L</th>
<th>log u</th>
<th>a22/a21</th>
<th>α</th>
<th>β</th>
<th>log L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.440 (0.072)</td>
<td>0.117</td>
<td>1.722</td>
<td>0.463</td>
<td>-11026.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.452 (0.072)</td>
<td>0.112</td>
<td>1.678</td>
<td>0.626</td>
<td>-11257.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

∗In each row, the Base Model is re-estimated with one dimension altered (a single departure).

aAverage U is 1987, average V is 203 except for narrow stocks (average U is 755) and monthly stocks (1965 and 201 respectively).
bFor job-seeker regressions, M = 2 mass points were used; for vacancy regressions, M = 5 mass points were used.
cRandom effects defined for 34,657 job-seeker spells rather than 26,113 job-seekers.
dRandom effects defined for 4,121 employers rather than 9,556 vacancy orders.
eEstimates of Equation (10) for job-seekers and log hε = log(a) + α log U + (β - 1) log V + ε for vacancies, pooled across old and new agents. f Imposed.
Row (3) shows the result of estimating the Base Model without unobserved heterogeneity. Apart from a moderate fall in $\alpha$ in the vacancy regression, again very little changes, even though the likelihood is a lot lower. Row (4) reports what happens when the individual-level control variables are dropped from the Base Model. There is very little change in any of the estimates, which implies that observable characteristics of job-seekers and vacancies are not correlated with the aggregate numbers of job-seekers and vacancies in a particular market.

In Row (5) we examine the effects of aggregation bias by replacing stocks observed at weekly intervals with those observed at monthly intervals. Now, for every week in a given month, the value of the stock is the same and equal to that of the first week of the month. The results show that aggregation bias might be a problem for investigators with monthly data. First, the estimate of $\alpha$ is bigger in the job-seeker regression (moving from 1.145 to 1.340) and is smaller for $\beta$ (moving from 0.195 to −0.006). The effect in the vacancy hazards is the other way round, with $\alpha$ falling from 1.722 to 1.572 and $\beta$ increasing from 0.463 to 0.699. Thus aggregation bias really does bias the estimates. More importantly, aggregation bias affects our estimates of the coefficient on log $v$ and the $a$-ratios in the job-seeker regression. Now $a_{22}/a_{21}$ is estimated as 0.589 rather than 0.302. In the vacancy regression, $a_{22}/a_{12}$ is estimated as 0.028 rather than 0.117; the stock-flow effect becomes almost insignificant in the job-seeker regression, but is unrealistically pronounced (1/30) in the vacancy regression.

Row (6) reports estimates of the classical random matching model. By construction, the $a$-ratios are unity. Moreover, the estimates of $\alpha$ and $\beta$ are unaffected, as already noted, which means that the “non-standard” estimate of $\alpha > 1$ is nothing to do with stock-flow matching. Row (7) shows that aggregation bias is also a problem with the random matching model. In row (8) we use the “narrow” definition of job-seeker stocks which includes only unemployed job-seekers. The estimate of $a_{22}/a_{21}$ is unaffected, but the estimate of $a_{22}/a_{12}$ increases substantially to 0.682. Since we have an a priori reason to prefer the wide definition (unemployed job-seekers are likely to face competition from employed job-seekers), and since the narrow estimates are less convincing because
they imply weaker stock-flow matching on the employers’ side of the market, we regard the estimates based on the “wide” definition the more preferable.

Rows (9)–(15) report what happens when we alter the window sizes away from $k^w = 4$, $k^e = 2$ weeks. The most important departure is Row (10), where $k^w = k^e = 4$, because this is what Coles & Petrongolo (2008) have to assume (having monthly data). Even though the effects of $\log v$ and $\log u$ are stronger, the $a$-ratios are unaffected. This is because $a_{22}/a_{21}$ depends on the size of the average stocks (see Equation 9), which change in size as $k^w$ and $k^e$ vary. Looking at the other ($k^w, k^e$) pairs, the estimate of $\log v$ in the job-seeker hazard tends to fall with smaller windows, ranging from 0.45 to 0.26. The net effect is that $a_{22}/a_{21}$ is robustly estimated between 0.20 and 0.30. The same happens for the vacancy hazards: the effect of $\log u$ falls with $k^w$ and $k^e$, but leaves $a_{22}/a_{12}$ robustly estimated in the range 0.05 to 0.12.

To summarise: the stock of new vacancies $\log v$ is robustly significant in the old job-seeker regression, and the stock of new job-seekers $\log u$ is robustly significant in the old vacancy regression. This implies that $a_{22}/a_{21} \approx 1/3$ and $a_{22}/a_{12} \approx 1/9$ for all these departures from the Base Model. In particular, the result appears robust to the choice of window size. The only assumptions that really matter in the Base Model are using weekly rather than monthly stocks and using the wide rather than narrow definition of job-seeker stocks.

### 6.3 Unobserved Search Heterogeneity

The models we have estimated assume that the heterogeneity is uncorrelated with the covariates; in particular, the stocks of unemployment and vacancies. However, search intensity, which is unobserved, may vary systematically with labour market conditions; job-seekers might search harder when the market is slack, and employers might search harder when the market is tight. Our estimates may therefore be biased, although the direction of the bias is ambiguous.

Coles & Petrongolo (2008) argue that aggregated data can be used to address this problem. Because of congestion between job-seekers, if one job-seeker searches harder, another job-seeker must have a lower matching probability. Ag-
ggregating over job-seekers nets out these congestion effects, and so the heterogeneity bias is smaller, and might be zero. Even if it is non-zero, Coles & Petrongolo argue that, providing aggregate search intensity changes slowly and systematically over time, its correlation with the stocks of interest can be controlled for using month and year dummies when using aggregated data. To see whether using aggregate data does ameliorate the effects of unobserved search intensity, we aggregate our data to three weekly time-series, one for each local labour market.

Consistent estimates of the parameters of interest can be obtained by running Poisson regressions of the number of matches per district/week on the weekly stocks of new and total agents in each district. As noted, these regressions include month, year and district dummies. To control for duration dependence in the counts, we specify Gamma mixing with an over-dispersion parameter “α”, where $\alpha = 0$ corresponds to no mixing, and a higher $\alpha$ means more duration dependence.

<table>
<thead>
<tr>
<th></th>
<th>Job-seeker matches</th>
<th>Vacancy matches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New</td>
<td>Old</td>
</tr>
<tr>
<td>log $u$</td>
<td>0.749 (0.155)</td>
<td>0.036 (0.092)</td>
</tr>
<tr>
<td>log $U$</td>
<td>0.877 (0.276)</td>
<td>1.008 (0.188)</td>
</tr>
<tr>
<td>log $v$</td>
<td>0.276 (0.129)</td>
<td>0.240 (0.076)</td>
</tr>
<tr>
<td>log $V$</td>
<td>-0.061 (0.244)</td>
<td>-0.051 (0.143)</td>
</tr>
<tr>
<td>Over-disp ‘α’</td>
<td>0.719 (0.116)</td>
<td>0.197 (0.032)</td>
</tr>
<tr>
<td>Log pseudo-likelihood</td>
<td>-846.7</td>
<td>-1,274.4</td>
</tr>
<tr>
<td>Obs</td>
<td>663</td>
<td>663</td>
</tr>
<tr>
<td>Matches</td>
<td>810</td>
<td>1,951</td>
</tr>
</tbody>
</table>

**Derived parameter estimates**

- $\alpha = 1.626 (0.237)$
- $\beta = 0.215 (0.229)$
- $\alpha + \beta = 1.841 (0.280)$
- $\alpha$-ratio = 0.299 [0.033] $\times$ 0.339 [0.001] = 0.177 [0.001] $\times$ 0.166 [0.002]

*Poisson regressions using aggregated data, ie long-thin panel comprising 221 weeks for three ‘districts’. The matching window is $k^w = 4$ and $k^e = 2$. The dependent variable is the number of matches inside the window (‘new’) or the number of matches outside the window (‘old’), ie based on a total of 2,761 matches. Robust standard errors in parentheses. All regressions contain dummy variables for month (11), year (4), and ‘district’ (2), but do not contain job-seeker and vacancy covariates. We use the search definition of duration and the wide definition of job-seeker stocks.

*See Table notes for Table 2.
We report our estimates in Table 4. The dependent variable in the column labelled “Job-seeker matches (New)” is the number of matches per district/week involving job-seekers who have been in the market for less than four weeks. In contrast, the dependent variable in the column labelled “Vacancy matches (New)” is the number of matches involving vacancies which have been in the market for less than two weeks. These estimates can be compared directly with those in Table 2. The estimates are similar and not significantly different, given the larger standard errors in Table 4. Crucially, the estimate of $a_{22}/a_{21}$ increases only slightly from 0.302 to 0.339, while $a_{22}/a_{12}$ has increased from 0.117 to 0.166. Thus, it would appear that there is some support for the argument that there might be some bias in the stock-flow effects reported in this paper, in that they weaken a little, but the point estimates of the $a$-ratios are still very strong. This is nothing to do with the absence of control variables in the Poisson regressions, because these had no effect on the parameters of interest in the Base model.

6.4 Horizontal Heterogeneity

A second possible criticism is that the random matching model may be rejected because we observe matches across sub-markets separated by occupation or location. To explain, suppose there are just two sub-markets. In the first, there are more job-seekers than vacancies, whereas in the second, there are more vacancies than job-seekers. A randomly chosen old job-seeker is more likely to come from the first sub-market, while a randomly chosen old vacancy is more likely to come from the second. Observing relatively few old-old matches can occur because of this type of segmentation, rather than because of stock-flow considerations. We refer to this as “horizontal heterogeneity” because it is not about search effort, but just about different types. In this sub-section, we report what happens when we re-estimate our Base Model for separate sub-markets defined by occupation and location, to see whether the stock-flow effects remain intact. In principle, as we define narrower and narrower sub-markets we would expect two-sided stock-flow effects to weaken, because, absent heterogeneity, the agent on the short side of the market (in this case vacancies) will always be able to find a match in the
initial period.\textsuperscript{7} This would not be possible with the kind of aggregated data used hitherto in this literature.

Although our data can be sub-divided into three geographically distinct districts, there are not enough matches in the data to allow us to stratify the data into sub-markets further defined by occupation (in practice, skilled and unskilled matches). Instead, we test for stock-flow matching effects within each district and, separately, within each occupation.

When we re-estimate the Base Model separately for each district, the effect of $u$ in the vacancy hazard remains significant for all three districts, and the implied $a$-ratios are all significantly less than one (0.174, 0.157 and 0.088). Nothing changes on this side of the market. For job-seekers, the effect of $v$ is only significant in the third district, with $a$-ratios of 0.538, 0.891 and 0.110. This is clearly an example of two-sided stock-flow matching becoming one-sided as the market is divided into sub-markets.

As noted, our data only allow us to consider two sub-markets for occupation. We define a skilled job-seeker as one whose exam performance at age 16–17 is 5+ high grade (A*-C) GCSEs, and we define a skilled vacancy as one that offers a skilled \textit{and} non-manual job, according to standard definitions. Of 26,113 job-seekers, 43.4\% are labelled skilled, and of the 9,556 vacancy orders, 37.8\% are labelled skilled. We expect that skilled job-seekers will match with the skilled jobs and that unskilled job-seekers will match with the remaining vacancies, but we recognise that there will be a certain degree of non-assortative matching in the data, especially as educational attainment is the only observable proxy for skill for young job-seekers.\textsuperscript{8}

Using these definitions, we construct the following stocks:

\[ U = u + \bar{U} = U_s + U_u = u_s + u_u + \bar{U}_s + \bar{U}_u, \]

\textsuperscript{7}In Shimer (2007), there are a large number of markets sub-divided by occupation and location, with wages set competitively so that unemployed workers and vacancies cannot coexist in the same market.

\textsuperscript{8}There is a substantial degree of non-assortative matching in these data: of 1,013 unskilled vacancies, 26.1\% end up with skilled job-seekers; of 1,748 skilled vacancies, 29.3\% end up with unskilled job-seekers.
where “s” refers to skilled and “u” to unskilled. We split the sample of old job-seekers into skilled and unskilled sub-samples, and run separate regressions for both. Because there is non-assortative matching, we must also allow for the relevant stocks from the other sub-market. In other words, we split each stock into its skilled and unskilled components (e.g., \(v\) into \(v_s\) and \(v_u\)), giving

\[
\log h_{w2}^s(U_s, u_s, V_s, v_s, U_u, u_u, V_u, v_u, \epsilon_w^w)\]

as the specification for the old skilled regression. For parsimony, we constrain the effects of \(U_s\) and \(U_u\) to be equal, and the same for \(u_s\) and \(u_u\). (The estimates are not of interest and nothing changes by doing this.) Similar models are estimated for \(\log h_{w2}^u\), \(\log h_{s2}^s\), \(\log h_{s2}^u\), so that the four regressions each have six covariates.

The results are reported in Table 5. Our Base Model tells us that an increase in the stock of new vacancies in the market increases the exit rate for old job-seekers: recall that in the Base Model \(\partial \log h_{w2}^w / \partial \log v\) is estimated as 0.270 (0.056). The issue is whether this is because of the newness of the stock, or because of the type (skilled or unskilled), or both. In the skilled regression, splitting the effect of \(v\) into \(v_s\) and \(v_u\), gives estimates of 0.156 (0.063) and 0.118 (0.078) respectively, and in the unskilled regression, we get 0.148 (0.058) and 0.136 (0.069). Three things are transparent. First, 3 of the 4 stock-flow effects are significant, and the fourth is marginally rejected (\(\log v_u\) in the skilled regression, with a p-value of 0.12). Second, the effect of increasing the stock of new skilled vacancies has a positive impact significant on unskilled job-seekers, but, as just noted, the other effect is weaker. This is consistent with the non-assortative matching in the data. Third, it does not matter whether the stock of new vacancies is skilled or unskilled. Testing that the effects of \(v_s\) and \(v_u\) are the same has a p-value of 0.66 in the skilled regression and 0.48 in the unskilled regression (in which case, the variable becomes \(\log(v_s + v_u)\)). In short, for job-seekers, it does not matter whether the job-seeker is skilled or unskilled, or whether the new vacancy is skilled or unskilled, the stock-flow model remains intact.
Table 5: Occupational sub-markets and stock-flow effects

<table>
<thead>
<tr>
<th></th>
<th>Job-seeker matches</th>
<th>Vacancy matches</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Skilled</td>
<td>Unskilled</td>
</tr>
<tr>
<td>log $u$</td>
<td>-0.256 (0.094)</td>
<td>-0.199 (0.086)</td>
</tr>
<tr>
<td>log $U$</td>
<td>0.392 (0.256)</td>
<td>0.375 (0.185)</td>
</tr>
<tr>
<td>log $v_s$</td>
<td>0.156 (0.063)</td>
<td>0.148 (0.058)</td>
</tr>
<tr>
<td>log $v_u$</td>
<td>0.118 (0.078)</td>
<td>0.136 (0.069)</td>
</tr>
<tr>
<td>log $V_s$</td>
<td>-0.106 (0.103)</td>
<td>-0.080 (0.095)</td>
</tr>
<tr>
<td>log $V_u$</td>
<td>-0.185 (0.114)</td>
<td>0.194 (0.104)</td>
</tr>
<tr>
<td>SE</td>
<td>0.520 (0.161)</td>
<td>0.758 (0.082)</td>
</tr>
<tr>
<td>log $L$</td>
<td>-7.173.7</td>
<td>-9.254.5</td>
</tr>
<tr>
<td>Obs</td>
<td>163,678</td>
<td>314,190</td>
</tr>
<tr>
<td>Job-seekers</td>
<td>11,331</td>
<td>14,782</td>
</tr>
</tbody>
</table>

Derived parameter estimates

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>1.136 (0.233)</td>
<td>1.176 (0.169)</td>
<td>1.605 (0.248)</td>
<td>1.194 (0.116)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.492 (0.071)</td>
<td>0.699 (0.063)</td>
<td>0.137 (0.249)</td>
<td>0.543 (0.171)</td>
</tr>
<tr>
<td>$\alpha + \beta$</td>
<td>1.627 (0.230)</td>
<td>1.875 (0.170)</td>
<td>1.742 (0.296)</td>
<td>1.737 (0.188)</td>
</tr>
<tr>
<td>$a$-ratios</td>
<td>0.282 [0.014]</td>
<td>0.185 [0.011]</td>
<td>0.049 [0.013]</td>
<td>-0.010 [0.000]</td>
</tr>
<tr>
<td></td>
<td>0.424 [0.128]</td>
<td>0.494 [0.048]</td>
<td>-0.219 [0.348]</td>
<td>-0.199 [0.039]</td>
</tr>
</tbody>
</table>

*Estimates of Base Model, without covariates, split into skilled and unskilled sub-samples. Sample means for $u = u_s + u_u$ are 202 = 77+124, for $U = U_s + U_u$ are 1925= 712+1213, for $v = v_s + v_u$ are 32= 14+18, and for $V = V_s + V_u$ are 191= 106+85.

For vacancies, the results are rather different. Recall that the Base Model effect $\partial \log h^e_2 / \partial \log u$ is estimated as 0.405 (0.074), giving an $a$-ratio of 0.133. For the old skilled vacancy regression, there is a very strong stock-flow effect from the stock of new skilled job-seekers, but the stock of new unskilled job-seekers has no effect. The stock-flow effect does not become weaker when we split $u$ into $u_s$ and $u_u$. But in the unskilled vacancy regression the stock-flow effect also comes from the new stock of skilled job-seekers, which is a non-assortative effect in the sense that it is the increase in the stock of new skilled vacancies which increases the unskilled hazard. In a very slack labour market such as this one, it is plausible that employers with unskilled vacancies can still choose from the pool of skilled job-seekers at the expense of unskilled job-seekers.

To summarise, for old job-seekers, stock-flow effects remain intact. For old vacancies, the stock-flow effects are tied up with which type of new job-seeker comes into the market: irrespective of whether the old vacancy is skilled or unskilled, the stock-flow effect occurs through the inflow of skilled, not unskilled, job-seekers. Ultimately, one could always argue that stock-flow effects are the result of aggregating across heterogeneous sub-markets with different degrees of
labour market tightness. Nevertheless, it seems to be the case in our data that stock-flow effects are still observed even in quite small districts and in occupational sub-markets.

7 Conclusion

The stock-flow matching model has recently become a serious alternative to the random matching model in describing how job-seekers and employers interact in the labour market. However, empirical evidence is limited and comes from only aggregate time-series data. In this paper, we use agent-level data to examine the matching process by modelling job-seeker and vacancy hazards data from both sides of a single market. Our data are very different to those used previously.

Our main findings are as follows. We find that the stock of new vacancies has a significant additional impact on the exit rate for old job-seekers, implying that an old job-seeker is three times more likely to match with a new vacancy than an old vacancy. There is a larger effect on the other side of the market, with an old vacancy being nine times more likely to match with a new job-seeker than an old job-seeker.

To deal with the possibility that unobserved search intensity is correlated with stocks of job-seekers and vacancies, we aggregate to the level of the local labour market. Our findings remain largely unaltered, suggesting that this type of heterogeneity bias is not a problem. There is also the possibility that our results are driven by the existence of sub-markets with differing degrees of labour-market tightness. We also find that stock-flow effects remain even within local labour markets and within occupations, although it is also clear that these occupational sub-groups are not well-defined because there is a significant degree of cross-matching. Moreover, any remaining unobserved differences across sub-markets might also generate the observed relationship, and in this sense our results may be indicative of a more general class of search models with heterogeneity.

As emphasised in the Introduction, the policy implications for labour markets characterised by non-random matching are very different to the traditional ran-
dom matching model. In the UK in 1988, benefits for 16-17 year old unemployed job-seekers were removed to encourage young people to take available vacancies. This was just as our sample was being collected, and, given our results, this was exactly the wrong policy, in that any resulting increase in search intensity by the unemployed would have very little impact on the rate at which vacancies are filled, which are always on the short side of the market. Furthermore, receipt of unemployment benefits leads to a greater information flows to the unemployed about vacancies and stronger search incentives simply because of contact with the benefit system. Our findings also lead to the policy implication that the kind of regional and training policies discussed in the Introduction are needed to complement a benefits’ policy.

Our results relate to a specific region and specific labour market in the UK. However, the key feature of the matching process in our data is a centralised database of vacancies and job-seekers which is coordinated by a matching agency (the Careers Service). This is undoubtedly a common feature of modern labour markets. Future research would benefit from analysing similar agent-level data for different (adult) markets, or in different countries. Finally, we emphasise that this type of research would benefit if data on contacts as well as matches were available, as such data would also allow the investigator to actually observe when an agent becomes “old”, that is, after the agent experiences an unsuccessful contact. Because this varies by agent, this would represent an improvement on our key assumption, that all agents become old after some fixed period of time.

References


