Testing theories of labour market matching

Martyn Andrews  Steve Bradley
University of Manchester  Lancaster University

Dave Stott  Richard Upward
Lancaster University  University of Nottingham*

For submission to Labour Economics
January 2008

*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.
Abstract

We estimate models of two-sided search using micro-level data for a well-defined labour market, comparing the random matching model with the stock-flow matching model of Coles and collaborators. Using a dataset of complete labour-market histories for both sides of the market, we estimate hazard functions for both 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 equivalent robust result for vacancy hazards. Thus we find convincing evidence in favour of stock-flow matching. A random matching model with heterogeneity does not provide a competing explanation of the data. [111 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:

Dr. M.J. Andrews
School of Social Sciences
University of Manchester
Manchester, M13 9PL

Email: martyn.andrews@manchester.ac.uk
Phone: +44-(0)161-275-4874
1 Introduction

This paper is an empirical investigation into how job-seekers and employers meet and match 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. Pissarides’ (2000) text (first published in 1990) is the original two-sided search model applied to the labour market. This model, and others like them, (see, in particular Burdett & Wright (1998)), 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.

A recent alternative view is that matching occurs via a marketplace. In the marketplace, agents can search the other side of the market in a short period of time, particularly if there are employment agencies that facilitate speedy search. With increasing use of IT and Web resources, it is easy to see why this model might become more relevant. If an agent, say 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. Thus, most matches in this model occur between the stock on one side of the market and the inflow on the other, which is why this alternative model is known as the stock-flow matching model (and might also be thought of as a specific form of a non-random matching model). It is mainly associated with Melvyn Coles and collaborators—see Coles & Smith (1998) and Coles & Petrongolo (2008)—although there are other papers that do not use the stock-flow terminology (Taylor 1995, Lagos 2000).

As well as these two competing models giving quite different predictions, they also have quite different policy implications. The random matching model implies that an increase in search intensity reduces equilibrium unemployment, whereas the stock-flow matching model suggests that the unemployed who fail to find a match immediately must chase the flow of new vacancies when they come onto the market. In the stock-flow matching model, increasing search intensity has no effect on equilibrium unemployment, and a policy of reducing unemployment benefits to shorten unemployment durations does not necessarily have the desired effect. The stock-flow matching model is more consistent with frictions that arise from market failure in occupational or regional segments of markets, which suggests that regional policies that move employers closer to workers, or stimulate small-firm formation to absorb the pool of unemployed, might be appropriate. The stock-flow matching model also
has implications for firms who face skill shortages, a perennial problem in many economies, including the UK. Here, policies which lead to better or more suitable training for workers to reduce occupational mismatches would be appropriate.

The stock-flow matching model gives a plausible explanation as to why almost all estimates of hazards from unemployment slope downwards. Once a job-seeker has searched and failed to find a match amongst the stock of old vacancies, their hazard falls sharply because they are now only able to match the flow of new vacancies. This explanation has more in common with the unobserved heterogeneity class of models (where the heterogeneity comes from the ‘age’ of the stock of vacancies), rather than those models which describe the ‘scarring’ caused by the experience of unemployment. This is important, because what little evidence we have suggests that vacancy hazards also exhibit duration dependence, for which the scarring explanation is less plausible (Andrews, Bradley, Stott & Upward 2007).

There is no previous evidence on the stock-flow matching model using micro-level data; the only evidence 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; their findings are strongly supportive of the theory. Gregg & Petrongolo (2005) use similar data and come to similar conclusions. Coles & Petrongolo (2008) find evidence of one-sided stock-flow matching, whereby the stock of unemployed match with the inflow of vacancies, but not *vice versa*.

Our data are quite different, comprising detailed micro-level data from both sides of the same labour market. We observe matches between job-seekers and vacancies and we observe how long each agent has been in the market when they match. We also observe who matches with whom. These high frequency agent-level data are superior to those hitherto used for testing the stock-flow matching model against the random matching model, and allow us to conduct a formal test. This is because we are able to estimate the hazards of exit from the marketplace for both job-seekers and employers. Using agent-level data, we can control for observed and unobserved heterogeneity and we can control for aggregation bias, a potential problem with studies that use aggregate data (Burdett, Coles & van Ours 1994). With aggregate data, one cannot model the essential feature of this type of search model, that of individual agents changing behaviour in response to changing aggregate labour market conditions during the agents’ stay in the market.

There are testable predictions that separate the stock-flow and random matching models; implementing these empirically is the main aim of this paper. However, when one allows for heterogeneity in the stocks of job-seekers and vacancies, the distinction is less clear-cut. Suppose there are two types of job-seeker, labelled ‘good’
and ‘bad’, and two types of vacancy, also labelled good and bad. Assume that good job-seekers and good vacancies can match more quickly with either type on the other side of the market. Now consider job-seekers who do not match immediately. Their exit rate to jobs will be higher if there is a larger number of good vacancies on the other side of the market. But because good vacancies match more quickly, the flow of new vacancies will tend to be of higher quality than the stock of old vacancies. Thus, an increase in the number of new vacancies may appear to increase the exit rate of job-seekers, when, in fact, it is being driven by an increase in the number of good vacancies. Ignoring this heterogeneity in the stocks of job-seekers and vacancies can clearly bias the estimates, and give the stock-flow model a spurious credence unless one can model this heterogeneity. Because we have micro data with a rich set of covariates for both job-seekers and vacancies, we are able to estimate an amended random matching model, with heterogeneity.

The paper is organised as follows. In the next section, we present stylised versions of both the random matching model and the stock-flow matching model. This is developed into an estimable statistical model in Section 3. 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 and in Section 6 we discuss our results. Section 7 concludes.

2 Two Theories of Labour Market Matching

In this section, we first consider a stylised version of the random matching model, and then we compare it with a stylised version of the stock-flow matching model. There are stocks of vacancies $V$ and job-seekers $U$ (all of whom are assumed unemployed) attempting to meet and eventually form matched pairs. The rate at which they randomly contact each other per period is $\lambda(U, V)$, where $\lambda()$ has the same properties as a production function (concave and increasing in both arguments). If $\lambda(U, V)$ also exhibits constant returns to scale, the average number of contacts per vacancy is

$$\lambda^v(\theta) = \lambda/V = \lambda(U/V, 1)$$

and is decreasing in labour-market tightness $\theta \equiv V/U$. Similarly, the average number of contacts per job seeker is

$$\lambda^w(\theta) = \lambda/U = \lambda(1, V/U)$$
and is increasing in $\theta$. The corresponding hazards are:

$$
\begin{align*}
    h_e(\theta) &= \lambda_e(\theta)\mu(\theta), \\
    h_w(\theta) &= \lambda_w(\theta)\mu(\theta),
\end{align*}
$$

(1)

where $\mu$ is joint probability that a job-seeker finds an employer acceptable and an employer finds a job-seeker acceptable. There is little theory or evidence about the effect of $\theta$ on $\mu$.

The important feature of the random matching model is that it explicitly allows for search and congestion externalities, which cannot be eliminated by price adjustments. It is labour-market tightness that captures this feature. It has a negative effect on $h_w(\theta)$ because there are more job-seekers with which the unemployed have to compete, and there are fewer vacancies on the market to search over. It has a positive effect on $h_e(\theta)$ for analogous reasons. In the many empirical studies that estimate $h_w(\theta)$ using unemployment duration data and aggregate flows data, $\theta$ has the required negative effect. Moreover, if the contact function and matching function are specified as Cobb-Douglas, that is $\lambda(U, V) = U^\alpha V^{1-\alpha}$ and $\mu(\theta) = \theta^\gamma$, then the elasticity $\frac{\partial \log h_w}{\partial \log \theta}$ is given by $\gamma(1 - \alpha)$, and estimates of it are typically in the range $[-0.5, -0.3]$ (Petrongolo & Pissarides 2001). In other words, there is much empirical support for the random matching model because of the significant effect of the stocks of $U$ and $V$ on the exit probability from unemployment.

By contrast, there are no search frictions in the stock-flow matching model, because job-seekers and employers are able to search the whole market in a short period of time. Unemployment and vacancies persist because suitable partners were not available on this first search of the market, and so job-seekers and employers have to wait for new opportunities to flow into the market at a later date.

Suppose that agents arrive on the market at flow rates $u$ and $v$. A new job-seeker searches the stock of old vacancies $V$. If she matches, the stock of $V$ is reduced by one next period, but if she does not match, $U$ increases by one next period. One prediction of the model is that old job-seekers should not match with old vacancies, because, if there were gains to trade, they would have matched in an earlier period. The hazard for old job-seekers is therefore written $h_w(v, U)$, where $v$ has the positive effect just discussed and $U$ has a negative effect, because the stock of old job-seekers causes congestion from the same side of the market as the old unemployed job-seeker. Symmetrical arguments imply that the hazard for old vacancies is written $h_e(u, V)$, with positive and negative first derivatives respectively.

This comparison of the two models suggests that there are two testable implications of the stock-flow matching model. The first is that the exit rate of job-seekers [resp. vacancies] who match with old vacancies [resp. job-seekers] will fall sharply
once the job-seeker [resp. vacancy] has searched the market. Indeed, in the pure form of the theory the exit rate falls to zero. However, this is likely to be a very weak test of the stock-flow model, because (as is well known) hazard rates may fall with elapsed duration for many reasons, including duration dependence, unobserved heterogeneity and changing reservation utilities. A much stronger test is therefore to estimate the unemployment hazard $h^u(v, U)$ and the vacancy hazard $h^e(u, V)$, conditional on observed and unobserved heterogeneity. The hazard for an old job-seeker should depend positively on $v$ in addition to any effect from $V$, and the hazard for an old vacancy should depend positively on $u$ in addition to any effect from $U$.

3 Statistical Models

3.1 The Stock-flow Matching Model

We now develop an estimable statistical model which has testable parametric restrictions that make the random matching model a special case of the stock-flow matching model. Our test formalises that proposed independently by Coles & Petrongolo (2008). We also discuss how the random matching model with heterogeneity can be estimated.

Time is discrete. The number of contacts per period are generated by a Poisson process

$$C \sim \text{Poisson}[\lambda(U, V)]$$

where the Poisson parameter $\lambda$ is specified as a Cobb-Douglas: $\lambda(U, V) = AU^\alpha V^\beta$. $\lambda(U, V)$ is the average number of contacts per period. In this model, the contact function is ‘random’; pairs of agents of one type are no more/less likely to contact each other than pairs of another type.

In the stock-flow matching model, agents are different because they have been in the market for different lengths of time. We define $u$ and $v$ as the stocks of new job-seekers and vacancies. (In the terminology of Coles and Smith, $u$ and $v$ are the flows of job-seekers and vacancies). Hence, the total stock of job-seekers and vacancies decompose as follows

$$U = \bar{U} + u \quad V = \bar{V} + v,$$

where $\bar{U}$ and $\bar{V}$ are the stocks of old job-seekers and old vacancies respectively. Empirically, we have to determine how long it takes for an agent to become old.

In this model, it is the matching probabilities, conditional on contacting, that we
assume vary across pairs of agents. The probability that a new job-seeker matches with a new vacancy is given by $\mu_{11}$, with the remaining three types of match having probabilities:

- $\mu_{12}$ if a new job-seeker and an old vacancy contact;
- $\mu_{21}$ if an old job-seeker and a new vacancy contact;
- $\mu_{22}$ if an old job-seeker and an old vacancy contact.

In its basic form, the stock-flow matching model predicts that $\mu_{12} > 0$, $\mu_{21} > 0$, but $\mu_{22} = 0$. $\mu_{22}$ is zero because stock-stock matching is predicted not to occur. This might seem somewhat extreme. One can easily imagine that job-seekers and employers, having entered the old stocks themselves, might revise down their reservation utilities, and so re-examining the stock might then reveal potential matches. Stock-stock matches might then happen, but with a low probability, especially if there are large numbers of each in the market. Thus the key prediction of the model is amended to be $\mu_{12} > \mu_{22}$ and $\mu_{21} > \mu_{22}$.

In continuous time versions of the model, flow-flow matching cannot happen by assumption, and so $\mu_{11}$ is undefined. In real data, which is necessarily discrete, one might allow $\mu_{11} > 0$. However, it is not clear whether we would expect restrictions such as $\mu_{12} > \mu_{11}$ and $\mu_{21} > \mu_{11}$ to hold.

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

\[
\begin{align*}
    h_{11}^{w} &= \mu_{11} \frac{uv}{UV} \lambda(U, V) / u = A \mu_{11} v U^{\alpha-1} V^{\beta-1} \\
    h_{12}^{w} &= \mu_{12} \frac{u\bar{V}}{U \bar{V}} \lambda(U, V) / u = A \mu_{12} \bar{V} U^{\alpha-1} V^{\beta-1} \\
    h_{21}^{w} &= \mu_{21} \frac{\bar{U} v}{U \bar{V}} \lambda(U, V) / \bar{U} = A \mu_{21} v U^{\alpha-1} V^{\beta-1} \\
    h_{22}^{w} &= \mu_{22} \frac{\bar{U} \bar{V}}{U \bar{V}} \lambda(U, V) / \bar{U} = A \mu_{22} \bar{V} U^{\alpha-1} V^{\beta-1}
\end{align*}
\]

There are another set of hazards for vacancies, labelled $h_{11}^{e}$, $h_{12}^{e}$, $h_{21}^{e}$, and $h_{22}^{e}$.

Consider the expression for $h_{21}^{w}$, the exit rate for an old job-seeker who matches with a new vacancy. $\lambda(U, V)$ gives the mean total number of contacts per period. Multiplying $\lambda(U, V)$ by $\bar{U} v / U \bar{V}$ gives the mean total number of contacts per period between old job-seekers and new vacancies, because $\bar{U} v / U \bar{V}$ is the proportion of this type of contact out of all contacts. Multiplying by $\mu_{21}$ gives the mean total number

\footnote{In this subscript notation the first subscript always refers to job-seekers, the second to vacancies; ‘1’ always means new, ‘2’ means old.}
of matches per period of this type; dividing by $\bar{U}$ gives the hazard rate $h_{21}^w$.

Notice the role that $\mu_{22}$ has in the exit rate of old job-seekers with old vacancies. Old-old contacts could be relatively very frequent by the sheer numbers of old stocks $\bar{U}$ and $\bar{V}$. It is the matching probability $\mu_{22}$ that makes old-old matches less frequent, and would be zero in the pure stock-flow matching model.

It is clear that we cannot identify all the parameters in this model. First, note that one cannot separately identify $A$ from four $\mu$s, and so we define $a_{ij} \equiv A\mu_{ij}$. Furthermore, below we show we can only identify ratios of $a_{ij}$. Providing one is prepared to assume that $A$ does not vary by type (as we have done here), such a ratio can be interpreted as the corresponding ratio of $\mu$s. Otherwise, one must interpret $a_{ij}$ as the autonomous matching propensity that does not vary with the stock of participants on both sides of the market.

Random matching is a special case when $H_0 : a_{11} = a_{12} = a_{21} = a_{22} = a$ (say), is true. Under $H_0$, summing the four $h_{ij}$s, the unemployment hazard becomes $\lambda(U, V)/U$, the standard hazard function for the random matching model.

The hazard for an old job-seeker $h_{2}^w$ is the sum of $h_{21}^w$ and $h_{22}^w$, so we have:

$$\log h_{2}^w(U, u, V, v) = \log(h_{21}^w + h_{22}^w) = \log[2a_{21}v + a_{22}(V - v)] + (\alpha - 1) \log U + (\beta - 1) \log V.$$ (7)

Similar expressions arise for the hazards for new job-seekers $h_{1}^w$, new vacancies $h_{1}^e$, and old vacancies $h_{2}^e$. Rather than estimate the non-linear model in Equation (7), it is much easier to estimate a model for $\log h_{2}^w$ which is linear in $\log U$, $\log u$, $\log V$, $\log v$, from which we can uniquely identify $\alpha$, $\beta$ and $a_{22}/a_{21}$:

$$\frac{\partial \log h_{2}^w}{\partial \log U} = \alpha - 1 \quad \frac{\partial \log h_{2}^w}{\partial \log V} = \frac{a_{22}V}{a_{21}v + a_{22}V} + \beta - 1 \equiv \pi_1$$

$$\frac{\partial \log h_{2}^w}{\partial \log u} = 0 \quad \frac{\partial \log h_{2}^w}{\partial \log v} = \frac{(a_{21} - a_{22})v}{a_{21}v + a_{22}V} \equiv \pi_2.$$ (8)

In interpreting the estimates, the following should be noted. 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 old or new 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 $\pi_1 + \pi_2 = \beta$). Effectively, the vacancy effect is split across both vacancy variables. Third, the coefficient on
$u$ should be zero if the stock-flow matching model is true, which itself nests the random matching model. If $H_0$ is true, then $a_{21} = a_{22}$, which implies $\pi_2 = 0$. In words, the effect of new vacancies onto the market has no effect on the hazard for old job-seekers. This is a one-sided test because, under the alternative, $\pi_2 > 0$. It is important to understand why this is so. 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 old and new 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

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

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

Analogous discussions apply to Equation (7) when considering $\log h_{11}^e(U, u, V, v)$, $\log h_1^g(U, u, V, v)$, and $\log h_2^b(U, u, V, v)$. To conclude, our test of stock-flow matching is not merely to see if hazards fall when agents have searched the market. Hazards can fall for other reasons. We also test whether the stock of new agents on the other side of the market has a significant impact on the hazard. However, data that record who matches with whom is required for the test to work; our data are unique in this respect.

### 3.2 A Random Matching Model with Heterogeneity

It is possible to develop a random matching model with heterogeneity that has similar predictions to the stock-flow matching model above. Instead of the stocks of job-seekers $U$ being made up of old and new job-seekers, suppose $U$ decomposes into ‘good’ and ‘bad’ job-seekers, labelled $U^g$ and $U^b$ respectively. Similarly, $V^g$ is the stock of good vacancies and $V^b$ is the stock of bad vacancies. All that is needed to generate the same statistical model is that the matching rate of good job-seekers with bad vacancies, $\mu_{gb}$, exceeds the matching rate of bad job-seekers with bad vacancies, $\mu_{bb}$, and that the matching rate of bad job-seekers with good vacancies, $\mu_{bg}$, also exceeds $\mu_{bb}$. Clearly, a good agent will match earlier than a bad agent. Intuitively, since all we have done is redefine ‘new’ as ‘good’ and ‘old’ as ‘bad’, it follows that Equation (7) is an appropriate model for estimating the job-seeker hazard for bad job-seekers when there is heterogeneity in the stocks of job-seekers and vacancies. This is written $\log h_{22}^e(U, U^g, V, V^g)$. Analogous considerations apply
to log $h_{1}^{e}(U, U^{g}, V, V^{g})$, log $h_{1}^{e}(U, U^{g}, V, V^{g})$, and log $h_{2}^{e}(U, U^{g}, V, V^{g})$.

4 The Data

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

The data comprise a longitudinal record of all youths in Lancashire aged 15–18, including those in education, employment, training and unemployment. For each individual we observe the start and end dates of every labour market spell over the sample period. The data also include a record of all vacancies notified to the Careers Service over the sample period. Approximately 20% of all job spells observed in the data resulted from a match with a vacancy posted with the Careers Service. Vacancies for which the Careers Service were not the method of search are not included in the data. However, in the youth labour market (in contrast to the adult labour market) vacancies posted with the Careers Service were generally representative of all vacancies available for this age group.\(^2\)

Job-seekers can come from 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 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. For example, a job-seeker finding a match from outside the Careers Service data, or a vacancy which matches a job-seeker who is still in education. The important point to note here is that we do not ignore these spells: they are included in the risk set up to the point where they are

\(^2\)See Upward (1998, ch. 4) for fuller details, especially Section 4.3 on the representativeness of our data.
censored, as is standard in duration analysis.

Also note that 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 first, narrow definition refers only to unemployed job-seekers. The second, 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. The narrow definition corresponds more closely to the existing literature.

Figure 1 illustrates the data in stylised form. Calendar time runs horizontally. For each job-seeker and each vacancy we observe the date at which they enter the market, denoted $E$. For a job-seeker this is the date on which they begin a spell of unemployment; for a vacancy this is the date on which the employer notifies the vacancy to the Careers Service. Each job-seeker [resp. vacancy] may then contact vacancies [resp. job-seekers], which may or may not result in a match. Contacts are not observed in this dataset. The pair illustrated in the figure match on date $M$. Finally, on date $X$ the pair exit the market, and the job spell begins.

The data are therefore also unusual in that we are able to distinguish between the search duration $M - E$ and the spell duration $X - E$ for each agent. Search duration is ended when a successful contact between a job-seeker and a vacancy is recorded in the data. Spell duration is ended when a job-seeker actually starts working in the 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 specification focuses on the duration of search, because this corresponds more closely to the theory. However, since almost all existing estimates of the matching function are forced to use spell durations, we also examine what happens when we use spell duration.

The point at which a job-seeker or a vacancy, when in the marketplace, changes from being new to old, is defined as $k^w$ for job-seekers and $k^e$ for vacancies. We refer to the first $k^w$ and $k^e$ weeks as the matching window. In Figure 1, the job-seeker was still searching after $k^w$ weeks have passed, and so enters the stock of old unemployed $\bar{U}$. On the other hand, using the search definition of duration, the employer entered the market later in calendar time, and had stopped searching before $k^e$ weeks had passed.
4.1 The Dependent Variable

We observe 2,761 matches in our data. They represent exits from both sides of the market, that is there are two hazards that can be estimated from this sample of matches, a job-seeker hazard $h^w$ and a vacancy hazard $h^e$. The data are organised into sequential binary response form (see, for example, Stewart 1996). In other words, for the $j$-th vacancy hazard, we define $y^e_{js} = 0$ for every week $s$ that the vacancy remains in the marketplace, except in the last week $t^e_j$, when $y^e_{js} = 1$, that is when the vacancy matches with job-seeker $i$. Figure 1 illustrates. If the vacancy is censored, $y^e_{js} = 0$ in the last period the vacancy is observed. Pooling over all the vacancies in the data, we generate an unbalanced panel of vacancies with $t^e_j$ observations for each vacancy $j$. Each row in this panel corresponds to a vacancy-week, of which there are 137,223. This defines the risk set for vacancies.

Analogous considerations apply to the $i$-th job-seeker hazard. Pooling over all the job-seekers in the data, we generate an unbalanced panel of job-seekers with $t^w_i$ observations for each job-seeker $i$. The risk set for this side of the market is 477,868 unemployment weeks.$^3$

Summing over $y^e_{js}$ in the vacancy panel and over $y^w_{is}$ in the job-seeker panel gives the total number of matches in the data:

$$m = \sum_i \sum_s y^w_{is} = \sum_j \sum_s y^e_{js} = 2,761.$$  

Eventually, in Section 6 below, we need to decide how long a job-seeker or a vacancy is in the market before it changes from being new to old. Assume, for the moment, that $k^w = k^e = 4$ weeks. Then the job-seeker/vacancy match illustrated in Figure 1 would count as one match between an old job-seeker and a new vacancy. Define $m_{21}$ as the total number of matches between old job-seekers, ie those who have been unemployed for more than four weeks, and new vacancies, ie those that have been open for less than or equal to four weeks, then in our data we observe $m_{21} = 1,497$ stock-flow matches. There are also $m_{11} = 532$ flow-flow matches, another $m_{12} = 278$ stock-flow matches between new job-seekers and old vacancies, and $m_{22} = 454$ stock-stock matches. These four numbers total the 2,761 matches.

Table 1 summarises the raw data for both panels, using the ‘search’ definition of duration. The lower panel aggregates $y^e_{js}$ in various ways. There are 137,223 vacancy-weeks at risk, of which there are 2,761 matches ($y^e_{js} = 1$) and the rest where there are no matches ($y^e_{js} = 0$). The 2,761 matches are disaggregated by whether

\footnote{Strictly speaking, the unit of observation is a spell, not a job-seeker, as some job-seekers have multiple spells.}
the vacancy is old or new (recorded in the vacancy panel) and by whether or not the vacancy exits to an old or new job-seeker (recorded in the job-seeker panel). Thus the four types of match are recorded in the body of the table. In the upper panel, the 477,868 unemployment-weeks are disaggregated in the same way. There are more unemployment-weeks at risk than there are vacancy-weeks because the youth labour market in the UK in the early 1990s was particularly slack.

Table 1: Who matches whom? Search duration; \( k^w = k^e = 4 \) weeks

<table>
<thead>
<tr>
<th></th>
<th>new</th>
<th>old</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job-seekers ( y_{is}^w )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>zeros</td>
<td>123,338</td>
<td>351,769</td>
<td>475,107</td>
</tr>
<tr>
<td>exits to new vacancy</td>
<td>((m_{11})) 532</td>
<td>((m_{21})) 1,497</td>
<td>2,029</td>
</tr>
<tr>
<td>exits to old vacancy</td>
<td>((m_{12})) 278</td>
<td>((m_{22})) 454</td>
<td>732</td>
</tr>
<tr>
<td>Total</td>
<td>124,148</td>
<td>353,720</td>
<td>477,868</td>
</tr>
</tbody>
</table>

| Vacancies \( y_{js}^e \) |         |         |         |
| zeros                  | 38,547  | 95,915  | 134,462 |
| exits to new job-seeker| \((m_{11})\) 532 | \((m_{12})\) 278 | 810 |
| exits to old job-seeker| \((m_{21})\) 1,497 | \((m_{22})\) 454 | 1,951 |
| Total                  | 40,576  | 96,647  | 137,223 |

The cross-tabulations given in Table 1 reveal that the raw vacancy hazard to new unemployed job-seekers is given by:

\[
h_{11}^e = \frac{532}{40576} = 0.01311
\]

\[
h_{12}^e = \frac{278}{96647} = 0.00288
\]

and the raw vacancy hazard to the old unemployed job-seekers is given by:

\[
h_{21}^e = \frac{1497}{40576} = 0.03689
\]

\[
h_{22}^e = \frac{454}{96647} = 0.00470
\]

Notice that the drop in the hazard for vacancies matching with old unemployed job-seekers is \( h_{22}^e / h_{21}^e = a_{22} / a_{21} = 0.127 \), which is perfectly consistent with stock-flow matching. Without knowing who matches whom, the best we could do with vacancy data on their own is calculate \( h_{22}^e / h_{11}^e \).

A similar analysis applies to job-seekers. It turns out that the drop in the hazard for job-seekers matching with old vacancies is \( h_{22}^w / h_{12}^w = a_{22} / a_{12} = 0.573 \). This is also consistent with the theoretical characterisation of stock-flow matching, but
is less pronounced on this side of the market. To compute mean durations, one subtracts from the risk set all vacancies that do not match with job-seekers. This is \( \frac{15,895}{2761} = 5.76 \) weeks, which decomposes into 16.63 weeks for old vacancies and 1.83 weeks for new vacancies. Similarly, for all, new and old job-seekers, mean durations are 10.74, 2.57 and 14.13 weeks respectively.

Table 1 can be recomputed for any values of \( k^e \) and \( k^w \). In Figure 3 we plot the numbers of stock-stock, stock-flow, and flow-flow matches against window size, but keeping \( k^w = k^e \). It is obvious that the number of flow-flow matches must increase with \( k^w = k^e \) and that the number of stock-stock matches must decrease. But the number of stock-stock matches is never zero, and so a pure form of the theory does not occur in these data. The number of stock-flow matches \( m_{12} + m_{21} \) increases with window size, and then decreases. Notice that the number of stock-flow matches is largest when the window size is \( k^w = k^e = 4 \) weeks.

It is also possible that \( k^w \neq k^e \). For example, 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, \( k^w > k^e \). Our strategy when we come to estimation is therefore 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).

### 4.2 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 new job-seekers and vacancies, while the stocks are old job-seekers and vacancies. As for the dependent variable, the sizes of these new and old stocks will therefore depend on the values of \( k^w \) and \( k^e \).

During a given week \( t - 1 \), there is an inflow of ‘unemployed’ job-seekers \( u^+_{t-1} \) into the stock of ‘unemployed’ job-seekers \( U_{t-1} \), and an outflow \( u^-_{t-1} \), such that

\[
U_t = U_{t-1} + (u^+_{t-1} - u^-_{t-1}).
\]  (10)

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 \( u^+_t - u^-_t \) 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. These are from the National Online Management Information Service (NOMIS), but cannot be disaggregated into old and new stocks. Our data, being based on all job-seekers, also record a large inflow of school-leavers in April and June. 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 both $U_t$ and $V_t$ are disaggregated into old and new stocks as follows (with an analogous expression for $V_t$):

$$U_t = \left[ u^+_{t-1} - u^-_{t-1} | u^+_{t-1} \right] + \left[ U_{t-1} - u^-_{t-1} | U_{t-1} \right] \equiv u_t + \bar{U}_t. \quad (11)$$

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 $u^+_{t-1} - u^-_{t-1} | u^+_{t-1}$. Similarly, the old stock $\bar{U}_t$ is defined as the stock of unemployed at 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 $k$:

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

The data cover the whole of Lancashire, a county in the United Kingdom that comprises 14 towns/cities (in fact, local authority districts). When constructing the covariates $U$, $\bar{U}$, $u$, $V$, $\bar{V}$, and $v$, in fact we group Lancashire into just three labour markets (West, Central and East), recognising that job-seekers can travel between certain towns when looking for work. 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. See Figure 2. There is a similar annual variation in the data for new vacancy stocks, but less pronounced.

The explanatory variables in the random matching model with heterogeneity are constructed as follows. We observe a number of characteristics for both job-seekers and vacancies. We define a good job-seeker as one whose exam grade at age 17 is 5+ GCSEs and who does not have a disadvantaged social background.\(^4\) We define a

\(^4\)GCSEs are national examinations taken in the school year a young person is aged 16. Career’s Officers assess whether someone has a disadvantaged social background when the young person
good vacancy as one that offers a skilled and non-manual job, according to standard definitions. Of the 26,113 job-seekers in the dataset, this means that 43.4% are good; of the 9,556 vacancy orders, 37.8% are good. From these, we can construct the following stocks:

\[ U_t = u_t + \bar{U}_t = U^g_t + U^b_t = u^g_t + u^b_t + \bar{U}^g_t + \bar{U}^b_t, \]

thereby generalising Equation (11). Figure 2 shows time-series plots of \( U_t, u_t, U^g_t \) and \( V_t, v_t, V^g_t \). It is clear that the good stocks move very closely with the total stocks through time; this is less true when comparing the new stock of vacancies with the total stock, and definitely less true when comparing the new stock of unemployed with the total stock.

5 Econometric Methodology

5.1 Specification of the Hazards

Having organised the data into sequential binary response form, the hazard for job-seeker \( i \) is modelled assuming proportional hazards and introducing a job-seeker specific random effect \( \varepsilon_i \), with density \( f_\varepsilon(\varepsilon) \). This is defined for a job-seeker, not a job-seeker spell. The likelihood \( L_i(\beta, \gamma) \) for each job-seeker with observed covariates \( \mathbf{x}'_{is} \) in this ‘mixed proportional hazards’ model is

\[
L_i(\beta, \gamma) = \int_{-\infty}^{\infty} \left[ \prod_{s=1}^{t_i} h_{is}(\cdot)^{y_{is}}[1 - h_{is}(\cdot)]^{1-y_{is}} \right] dF_\varepsilon(\varepsilon_i),
\]

with \( h_{is}(\cdot) = 1 - \exp[-\exp(\mathbf{x}'_{is}\beta + \gamma_s + \varepsilon_i)]. \)

Because of the proportional hazards assumption, the covariates \( \mathbf{x}'_{is} \) affect the hazard via the complementary log-log link. The \( \gamma_s \) terms are the log of a non-parametric piece-wise linear baseline hazard, and are collected into a vector \( \gamma \). Each interval corresponds to a week, but, because of data thinning, these are grouped into longer intervals at longer durations by constraining the appropriate \( \gamma_s \)s. We model the unobserved heterogeneity \( \varepsilon_i \) using Normal mixing, with variance \( \sigma^2 \). We also experiment with discrete mixing (Heckman & Singer 1984), to see whether the impact on the covariates and on the shape of the baseline hazard are the same. The same equation also applies to the vacancy hazard, replacing \( i \) by \( j \). The random effect \( \varepsilon_j \) is interviewed, on the basis of whether he or she comes from a poor home background, has been involved in crime, or has been in care. This applies to roughly 10% of the sample.
is defined for vacancy orders, not individual vacancies.

For the stock-flow model, the old job-seeker hazard, \( \log h_w(U, u, V, v) \), is log-linear in the four stocks (see Section 3). The specification for new job-seekers, \( \log h_w(U, u, V, v) \), has the same four stocks, and so we estimate a single model for all job-seekers by interacting each covariate with a dummy indicating whether \( s \) is less than \( k \). An agent is new if \( s \leq k \) and old if \( s > k \). So the specification for \( x'_{is} \) for our final model has eight covariates: \( \log U, \log u, \log V \) and \( \log v \) interacted with a 'new' and an 'old' dummy. For the random matching model with heterogeneity, the specification is \( \log U, \log U^g, \log V \) and \( \log V^g \) interacted with the new and old dummies. Throughout Huber/White standard errors correct for heteroskedasticity caused by unmodelled heterogeneity that is aggregated at the district level. All models include monthly dummy variables, to address the potential criticism that the model is identified mainly through the variation in the stocks of job-seekers over the calendar year.

It is worth emphasising that both stocks vary by duration \( s \) and job-seeker \( i/vacancy \ j \), because they vary through calendar time, across labour markets, and because each job-seeker/vacancy enters the marketplace at different calendar times. This is important for identification, and is an effect lost with aggregate data.

5.2 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 carefully that using daily data has very little impact on our results. 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. This one might label ‘pure’ aggregation bias. The alternative would be to collapse the flow data into months as well, thereby having both stocks and flows measured monthly. This is not ‘pure’
aggregation bias as there is additional measurement error in the durations.

6 Results

There are two testable implications of the stock-flow matching model. The first is that the exit rate of agents who match with old partners will fall sharply once the agent has searched the market. In the pure form of the theory the exit rate falls to zero, which is at odds with the data as there are always a substantial number of old-old matches (Figure 3). In Figure 4 we plot the raw hazards (together with two-piece hazards computed from Table 1). The raw hazards suggest that stock-flow matching is consistent with the raw data, especially on the employers’ side of the market. However, we have argued throughout that this is a very weak test of the stock-flow model, because hazard rates may fall with elapsed time for many reasons, including duration dependence, unobserved heterogeneity and declining reservation utilities. A second, stronger, testable implication is to estimate the parameters of models of exit, conditional on observed and unobserved heterogeneity, as outlined in Sections 3 and 5. Our strategy is to report a ‘Base Model’ which represents our preferred specification of the stock-flow matching model. We then assess whether a random matching model with heterogeneity provides an alternative explanation of the data. Finally, we look at departures from the Base Model to assess the robustness of the stock-flow matching model.

6.1 The Base Model

The Base Model is defined as follows. It is estimated using two binary response panels, one for job-seekers, the other for vacancies. In both, duration is measured in weeks. It specifies Normally distributed random effects for the unobserved heterogeneity. Its specification comprises log $U$, log $u$, log $V$ and log $v$ interacted with a new and an old dummy. Employer and job-seeker control variables are not included; this is because the essence of the model is to see whether individual behaviour responds to aggregate labour market conditions. Also, some of these controls are used to construct the good and bad stocks in the heterogeneity model. It specifies $k^w = k^e = 4$ weeks, because this is all the existing literature is able to do, and is also where the number of stock-flow matches is a maximum subject to $k^w \geq k^e$. Finally, we use weekly stocks to minimise the effects of aggregation bias. Departures from the Base Model are discussed more fully below. We interpret the results in the context of the statistical model developed in Section 3—see Equation (8) in particular. The
implied estimates of $\alpha$, $\beta$ and the $a$-ratios are also reported. Because we have two sets of stocks for job-seekers, narrow and wide definitions, we therefore have two variants of the Base Model.

Table 2: Base Model

<table>
<thead>
<tr>
<th></th>
<th>Job-seekers, $h^c$</th>
<th>Vacancies, $h^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>narrow</td>
<td>wide</td>
</tr>
<tr>
<td>(a) New</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\log u$</td>
<td>-0.162 (0.080)</td>
<td>-0.187 (0.064)</td>
</tr>
<tr>
<td>$\log U$</td>
<td>-0.075 (0.096)</td>
<td>-0.045 (0.068)</td>
</tr>
<tr>
<td>$\log v$</td>
<td>0.410 (0.104)</td>
<td>0.409 (0.104)</td>
</tr>
<tr>
<td>$\log V$</td>
<td>0.205 (0.083)</td>
<td>0.213 (0.082)</td>
</tr>
<tr>
<td>$\alpha, \beta$</td>
<td>0.764, 0.615</td>
<td>0.768, 0.621</td>
</tr>
<tr>
<td>$\alpha + \beta$</td>
<td>1.379 (0.092)</td>
<td>1.389 (0.089)</td>
</tr>
<tr>
<td>$a$-ratio$^a$</td>
<td>0.281 [0.000]</td>
<td>0.281 [0.000]</td>
</tr>
<tr>
<td>(b) Old</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\log u$</td>
<td>-0.193 (0.068)</td>
<td>-0.323 (0.060)</td>
</tr>
<tr>
<td>$\log U$</td>
<td>-0.263 (0.072)</td>
<td>-0.074 (0.062)</td>
</tr>
<tr>
<td>$\log v$</td>
<td>0.480 (0.066)</td>
<td>0.483 (0.066)</td>
</tr>
<tr>
<td>$\log V$</td>
<td>0.020 (0.054)</td>
<td>0.046 (0.054)</td>
</tr>
<tr>
<td>$\alpha, \beta$</td>
<td>0.545, 0.500</td>
<td>0.604, 0.529</td>
</tr>
<tr>
<td>$\alpha + \beta$</td>
<td>1.044 (0.074)</td>
<td>1.133 (0.077)</td>
</tr>
<tr>
<td>$a$-ratio$^b$</td>
<td>0.227 [0.000]</td>
<td>0.225 [0.000]</td>
</tr>
<tr>
<td>SE$^c$</td>
<td>0.688 (0.072)</td>
<td>0.670 (0.073)</td>
</tr>
<tr>
<td>$\log L$</td>
<td>-16559.4</td>
<td>-16568.5</td>
</tr>
<tr>
<td>Obs</td>
<td>477868</td>
<td>477868</td>
</tr>
</tbody>
</table>

*Estimated hazards for job-seekers and vacancies, sequential binary response panel, weekly stocks, random effects stock-flow matching models, 4-week window. Estimates based on 2761 matches between 34657 unemployed job-seeker spells (26113 job-seekers) and 14154 LCS job vacancies (9556 orders). The weighted averages across the 3 LADs for $u, U, v, V$ are 191, 755, 55 and 203 respectively for the narrow definition of the stocks. The corresponding numbers for the wider definition are 207, 1987, 55 and 203. All regressions contain monthly dummy variables.

$^a$For job-seekers, $a_{21}/a_{11}$ for vacancies. The $a$-ratios calculated from Equation (9) and analogous expressions. We do not report standard errors, as the $a$-ratios are not normally distributed. By definition, $p$-values are the same as for underlying parameter estimates (alternative hypothesis is one-sided).

$^b$For job-seekers, $a_{22}/a_{12}$ for vacancies.

$^c$Standard error ($\sigma$) for Normally distributed random effects. We use 12 quadrature points in the job-seeker regressions and 24 quadrature points in the vacancy regressions.

The primary focus of this subsection is our test of stock-flow matching. Described simply, we need to see whether an increase in the number of new vacancies on the market [resp job-seekers] significantly increases the exit probability for old job-
seekers [resp vacancies]. For job-seekers, under stock-flow matching, an increase in \( v \) and a fall in \( \dot{V} \) (such that \( V \) is constant) leads to more matches because the increase in old-new matches outweighs the fall in old-old matches because \( a_{22} < a_{21} \). In the old job-seeker hazard, using wide stocks (second column), this effect is estimated as \( \frac{\partial \log h_w^{e}}{\partial \log v} = 0.483 \), and is significant. This converts to a point estimate for \( a_{22}/a_{21} = 0.225 \). An old job-seeker is four times more likely to match with the new vacancy than an old vacancy. A very similar estimate occurs with narrow stocks.

On the other side of the market, for the narrow definition, almost the same effect is detected: the effect of \( \log u \) in the old vacancy hazard, \( \frac{\partial \log h_e^{w}}{\partial \log u} \) is 0.401, which converts to \( a_{22}/a_{12} = 0.274 \). However, when using the wider definition, the effect of \( \log u \) is a lot stronger, with \( \frac{\partial \log h_e^{w}}{\partial \log u} = 0.616 \) and \( a_{22}/a_{12} = 0.061 \). Here, an old vacancy is now fifteen times more likely to match with the new job-seeker than an old job-seeker, rather than just four times, which makes obvious sense given there are more old job-seekers in the wider definition.

This strong evidence of stock-flow matching is the key result of the paper. However, before estimating a competing random matching model with heterogeneity, and assessing the robustness of the stock-flow model, we briefly examine three other features of the Base Model.

First, in Figure 5, we plot and compare the hazards from the Base Model with the raw hazards. We also plot the hazards controlling for the eight stock variables only (ie without controlling for unobserved heterogeneity). This is for wide stocks, but identical figures are obtained for narrow. Looking at the three vacancy hazards first, Figure 5(b) shows that the severe fall in the raw hazard over the first 8 weeks almost completely disappears in the Base Model, and that this is primarily due to controlling for unobserved heterogeneity (adding the eight covariates to the raw hazard model makes little difference). This is why looking at hazards on their own tells us nothing about stock-flow matching. On the other side of the market, the shape of the job-seeker hazard is unaffected by either adding covariates or controlling for unobserved heterogeneity. This is different to the vacancy hazards, where the standard error of the heterogeneity is about three times bigger (1.83 compared with 0.67). But again, looking at these job-seeker hazards to examine stock-flow matching would give the wrong impression (that agents, when old, are more likely to exit), whereas the regression-based estimate, using vacancy data, suggests that \( a_{22} \leq a_{12} \). As already noted, our view is that the hazard increases initially because job-seekers are learning how to search.

Second, the stylised theory outlined earlier is silent about flow-flow (that is, new-new) matches. In the data, for \( k^w = k^e = 4 \), there are 532 new-new matches (out of
Looking at the results for new agents, the effect of log $v$ in the new job-seeker regressions (0.410, 0.409 for narrow, wide respectively) means that a new job-seeker is more likely to match with a new vacancy than an old vacancy ($a_{12}/a_{11} = 0.281$, twice). On the other side of the market, $a_{21}/a_{11}$ is estimated as 1.374 for narrow stocks (but is insignificantly different from one), whereas $a_{21}/a_{11} = 0.265$ for wide stocks.

Third, in terms of classical matching elasticities $\alpha$ and $\beta$, the estimates are generally sensible, but always show a significant degree of increasing returns to scale, or scale effects. Finding scale effects is contrary to what is usually found in the literature, which of course comes from mainly aggregate data. Most studies find constant returns, see, for example, Broersma & van Ours (1999, Table 1) and more comprehensively, Petrongolo & Pissarides (2001). (Their survey also suggests that $\alpha > \beta$ for unemployment-to-job transitions, as we find with our job-seeker data.) For the narrow definition, the scale effects are much stronger when estimating vacancy hazards; estimates are much closer to unity using unemployment data, which is what most of the literature uses. $\alpha + \beta$ is estimated as 1.38 for new job-seekers and 1.04 for old, whereas for new vacancies it is 1.44 and even higher for old vacancies at 2.05 (although this has a larger standard error). Using wide stocks makes little difference. There are good reasons why we might expect scale effects. Petrongolo & Pissarides (2006) develop and estimate a model that has increasing returns to quality of matches, with better matches occurring in larger markets. If agents respond by increasing their reservation utilities in proportion to the match quality, the hazard function should be independent of scale. Our results therefore imply that employers do not adjust their reservation utility when facing an increase in the quality of job-seekers, whereas job-seekers do when better quality vacancies arrive onto the market. This might be because employers have more market power, but this remains conjecture unless one can disentangle arrival rate effects from matching probability effects.

### 6.2 A Random Matching Model with Heterogeneity

We now examine whether a model based on observable differences between job-seekers and observable differences between vacancies provides an alternative explanation of the data. We focus on the wide-stock variant of the Base Model because it has the stronger stock-flow effects.

As noted above, the specification for this model is $\log U$, $\log U^g$, $\log V$ and $\log V^g$ interacted with a new and an old dummy. Effectively, ‘good’ variables replace ‘new’ variables in the pure stock-flow models discussed immediately above. Our approach
is to make single departures away from the stock-flow model, examining the competing roles of \( \log v \) and \( \log V^g \) in the old job-seeker regressions and the competing roles of \( \log u \) and \( \log U^g \) in the old vacancy regressions.

In the first model we estimate, \( \log V^g \) replaces \( \log v \) in the old job-seeker regression in the Base Model. Compare this model with the Base Model below it:

\[
\log h^w_2 = -0.306 \log u - 0.054 \log U - 0.569 \log V^g + 0.496 \log V + \ldots \\
\quad (0.060) \quad (0.063) \quad (0.323) \quad (0.102)
\]

\[
\log h^w_2 = -0.323 \log u - 0.074 \log U + 0.483 \log v + 0.046 \log V + \ldots \\
\quad (0.060) \quad (0.062) \quad (0.060) \quad (0.054)
\]

Clearly, the estimated effect on \( \log V^g \) is \(-0.569\) and is wrongly signed and badly determined (the standard error being five times bigger). Also, the maximised log-likelihood falls by 26.5 log-points. Similarly, when \( \log V^g \) is added to the stock-flow model, all the estimates remain virtually unchanged:

\[
\log h^w_2 = -0.326 \log u - 0.076 \log U + 0.477 \log v - 0.377 \log V^g + 0.147 \log V + \ldots \\
\quad (0.060) \quad (0.063) \quad (0.066) \quad (0.324) \quad (0.111)
\]

but the estimated effect on \( \log V^g \) changes to \(-0.377\) and is insignificant. Also, the gap between this model—one that nests the two competing models—and the stock-flow model (Equation 13) is only 0.5 log-points, but is 25.0 log-points for the random-matching model with heterogeneity (Equation 12).

This clear rejection of the random matching model with heterogeneity in favour of the stock-flow matching model is because \( \log v \) is highly correlated with \( \log V^g \), being 0.93 in both datasets. Because the stock of good vacancies moves exactly in line with total vacancies over the recruitment cycle (see Figure 2), this means that the random matching model with heterogeneity does not capture the essential feature of the stock-flow model. If the pure heterogeneity model was to provide a competing explanation of the data, good job-seekers would exit at different rates to bad job-seekers over the cycle, and the stocks would not move in line. Note that the correlations between \( \log V \) and \( \log v \) are 0.74 and 0.62 for the job-seeker and vacancy datasets respectively; these are much lower than above, which is why the new stocks have considerable explanatory power in the stock-flow model.

On the other side of the market, \( \log U \) is highly correlated with \( \log U^g \), being 0.94 and 0.97 in the two datasets, whereas the correlations between \( \log U \) and \( \log u \)
are 0.40 and 0.61. Repeating the same as above for the old vacancy regressions, when adding \( \log U^g \) to the stock-flow model and when replacing \( \log u \) by adding \( \log U^g \), the models do not converge to sensible parameter values and have much lower log-likelihoods. However, when one replaces \( \log u \) in the stock-flow model

\[
\log h'_{2} = 0.616 \log u + 0.311 \log U + 0.042 \log v - 0.099 \log V + \ldots \\
(0.073) \quad (0.112) \quad (0.112) \quad (0.103)
\]

by a decomposition of new stocks into new good stocks and new bad stocks \( u = u^g + u^b \), then the estimates are as follows:

\[
\log h'_{2} = 0.860 \log u^g - 0.325 \log u^b + 0.129 \log U + 0.082 \log v - 0.297 \log V + \ldots \\
(0.133) \quad (0.154) \quad (0.117) \quad (0.113) \quad (0.106)
\]

Instead of \( \log u \) having an estimate of 0.616, the effect of \( \log u^g \) is estimated as 0.869 and the effect of \( \log u^b \) is estimated as –0.325. These estimates make perfect sense, and the log-likelihood increases by 11.9 log-points. Hence, on this side of the market, the stock-flow matching model becomes even richer when augmented by these good/bad variables.

To conclude, the stock-flow matching model provides a much better explanation of the data than the random matching model with heterogeneity. This is because good stocks of data cannot be seen as proxies for the new stocks because ‘good’ is highly correlated with ‘total’, not highly correlated with ‘new’. Even though some elements of the good/bad model enhance the stock-flow model, in what follows, we focus exclusively on the pure stock-flow model.

### 6.3 Departures from the Base Model

To summarise, we find strong evidence of stock-flow matching on both sides of the market. For old job-seekers, \( \frac{\partial \log h'_w}{\partial \log v} = 0.483 \), giving \( a_{22}/a_{21} = 0.225 \). For old vacancies, \( \frac{\partial \log h'_e}{\partial \log u} = 0.616 \) is bigger, and converts to a lower \( a \)-ratio of \( a_{22}/a_{12} = 0.061 \). See top row of Table 3. We now report estimates of various departures from the Base Model, to assess whether our assumptions are important or innocuous.

Row (1) of the table shows that the results are robust to the way the heterogeneity is modelled, because here we use discrete (Heckman-Singer) mixing. Row (2) shows the result of estimating the Base Model without unobserved heterogeneity. Again, very little changes, even though the likelihood is a lot lower. (There is a moderate fall in \( \alpha \) in the vacancy regression.)
Table 3: Summary of departures from Base Model, wide stocks

<table>
<thead>
<tr>
<th>Departures</th>
<th>Old job-seekers, $h^w$</th>
<th>Old vacancies, $h^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\text{avge } u$</td>
<td>$\text{avge } v^a$</td>
</tr>
<tr>
<td>Base Model</td>
<td>207</td>
<td>55</td>
</tr>
<tr>
<td>(1) Heckman-Singer(^b)</td>
<td>207</td>
<td>55</td>
</tr>
<tr>
<td>(2) Without heterogeneity</td>
<td>207</td>
<td>55</td>
</tr>
<tr>
<td>(3) With covariates(^c)</td>
<td>207</td>
<td>55</td>
</tr>
<tr>
<td>(4) Classical random matching(^d)</td>
<td>207</td>
<td>55</td>
</tr>
<tr>
<td>(5) Monthly stocks</td>
<td>221</td>
<td>55</td>
</tr>
<tr>
<td>(6) Spell duration(^f)</td>
<td>208</td>
<td>56</td>
</tr>
<tr>
<td>(7) $k^w = 5$, $k^e = 5$</td>
<td>254</td>
<td>63</td>
</tr>
<tr>
<td>(8) $k^w = 4$, $k^e = 2$</td>
<td>191</td>
<td>33</td>
</tr>
<tr>
<td>(9) $k^w = 4$, $k^e = 1$</td>
<td>191</td>
<td>18</td>
</tr>
<tr>
<td>(10) $k^w = 3$, $k^e = 3$</td>
<td>158</td>
<td>45</td>
</tr>
<tr>
<td>(11) $k^w = 3$, $k^e = 2$</td>
<td>158</td>
<td>33</td>
</tr>
<tr>
<td>(12) $k^w = 2$, $k^e = 2$</td>
<td>107</td>
<td>33</td>
</tr>
<tr>
<td>(13) $k^w = 2$, $k^e = 1$</td>
<td>107</td>
<td>18</td>
</tr>
</tbody>
</table>

\(^a\) In each row, the Base Model is re-estimated with one dimension altered (a single departure). Information refers to old agents only, except for “Classical random matching”.

\(^b\) For job-seekers regressions, 2 mass points were used; for vacancy regressions, 7 mass points were used.

\(^c\) For job-seekers regressions, these are gender (1 dummy), grades at age 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). In all regressions, they are interacted with old and new.

\(^d\) Estimates of $\log h^w = \log(A_{\mu}) + (\alpha - 1) \log U + \beta \log V$ and $\log h^e = \log(A_{\mu}) + \alpha \log U + (\beta - 1) \log V$.

\(^e\) Imposed.

\(^f\) The number of observations in the spell duration datasets is 480,423 for job-seekers and 139,505 for vacancies.
Row (3) reports what happens when various observed covariates are added to 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. This is not surprising, and applies to unobservables as well. This justifies modelling the heterogeneity using random effects techniques.

Row (4) reports estimates of the classical random matching model. This involves constraining $\alpha$ and $\beta$ across old and new, a constraint which is rejected by the data. We find a slight degree of increasing returns in job-seeker regressions $\alpha + \beta = 1.237$. Our estimate of $\alpha + \beta = 1.447$ for vacancies is a new finding, mainly because very few studies use vacancy data. Notice that we also have increasing returns in the Base Model, which implies that increasing returns arises because we are using micro-level data.

In Row (5) we examine the effects of aggregation bias by replacing stocks observed at weekly intervals with those observed at monthly intervals. For every week in a given month, the value 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 0.604 to 0.759) and is smaller for $\beta$ (moving from 0.528 to 0.443). The effect in the vacancy hazards is the other way round, with $\alpha$ falling from 0.927 to 0.579—a very large change—and $\beta$ decreasing from 0.943 to 0.841, so that $\alpha + \beta$ falls from 1.870 to 1.420. In the first three cases, it is possible to show that these movements in $\alpha$ and $\beta$ go the right way, given standard results when signing omitted variables biases. 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 job-seeker regressions. Now $a_{22}/a_{21}$ is estimated as 0.394 rather than 0.225.

In Row (6) we replace our preferred measure of search duration with that used hitherto in the literature, which we label spell duration (see Section 4). This has a small effect on the estimates in the vacancy regressions, with $a_{22}/a_{12}$ increasing from 0.061 to 0.079, but a much bigger effect in the job-seeker regressions, with $a_{22}/a_{21}$ increasing from 0.225 to 0.309. In other words, there are stronger stock-flow effects when we use search duration data, which might be expected.

Rows (7)–(13) report what happens when we alter the window sizes away from $k^w = k^e = 4$ weeks. We choose the following $(k^w, k^e)$ pairs: (5,5), (4,2), (4,1), (3,3), (3,2), (2,2), and (2,1). The estimate of $\log v$ in the job-seeker hazard tends to fall with smaller windows, ranging from 0.52 to 0.32. In addition, $a_{22}/a_{21}$ depends on the size of the average stocks (see Equation 9), which change in size as $k^w$ and $k^e$.
vary. The net effect is that $a_{22}/a_{21}$ is robustly estimated between 0.15 and 0.23. 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 0.04 to 0.11.

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 $a_{22}/a_{21} < 1$ and $a_{22}/a_{12} < 1$ for all these departures from the Base Model when stocks are measured using the wide definition. 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 search duration rather than spell duration.

7 Conclusion

In this paper we report estimates of job-seeker and vacancy hazards using micro-level data from both sides of a single market. In particular, we examine whether there is any evidence in favour of Coles & Smith’s stock-flow matching model, or whether, alternatively, the random matching model adequately describes the data. Our test is a simple one, but needs data that record which job-seeker matches with which vacancy. 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. If the effect is positive and significant, this implies that employers find it harder to match to old job-seekers once their vacancies become old. Exactly the reverse applies to the other side of the market, where the test examines the effect of the stock of new job-seekers. The test does not examine whether vacancy hazards or job-seeker hazards fall at certain durations, because this can happen for other reasons.

Our results for the stock-flow model are summarised as follows. The stock of new vacancies has a significant additional impact on the exit rate for old job-seekers, as is predicted by stock-flow matching theory, and is robust across choice of window and whether or not we use a narrow or wide definition of the stock. For the wide definition (which additionally includes those searching whilst not unemployed), this implies that the hazard rate for vacancies falls by about three-quarters when vacancies become old ($a_{22}/a_{21} \approx 0.23$). There is an equivalent robust effect for the exit rate of old job-seekers on the other side of the market for the wide definition, with the
hazard falling drastically when job-seekers become old, to less than one-tenth of its value \((a_{22}/a_{12} \approx 0.06)\).

We also check whether a random matching model with heterogeneity in the stocks of job-seekers and vacancies provides a competing explanation of the data. Because the stocks of good job-seekers move in line with total stocks over the recruitment cycle—and the same for the stocks of vacancies—the stock-flow matching model fits the data far better than its competitor (although the stock-flow model can be enhanced if one splits the stock of new job-seekers into good and bad job-seekers in the vacancy regressions).

We have convincing evidence that aggregation bias is a problem with data that do not record information more frequently than at monthly intervals. In particular, \(a_{22}/a_{21}\) is estimated as 0.394 rather than 0.225. Throughout, we find evidence of increasing returns to scale for all the regressions we run. This has more to do with the fact we are using micro-level data rather than estimating stock-flow matching models, because increasing returns occurs in the random matching version. We cannot say why this is so without estimating the underlying contact function; for this, we need data on contacts as well as matches.

To conclude, we find convincing evidence in favour of stock-flow matching, using a unique dataset with high quality agent-level information from both sides of the same market. We are thus able to observe both stocks and flows over intervals shorter than one month, which is the best one can do using aggregate data. All of the analysis in this paper is conducted at the level of individual matches. Most importantly, with aggregate data we would be unable to model the essential feature of search models, that of individual agents changing behaviour in response to changing aggregate labour market conditions during the agents’ stay in the market.

References


Stewart, M. (1996), Heterogeneity specification in unemployment duration models, mimeo, University of Warwick.


Figures

Figure 1: Stock-flow matching and sequential binary response form
Figure 2: Unemployment and vacancy stocks: total, new and good
Figure 3: Stock-flow counts by window size, $k^w = k^e$
Figure 4: Raw job-seeker and vacancy hazards split by old and new; $k^w = k^e = 4$
Figure 5: Conditional and raw baseline hazards