The University of Manchester

Economics Discussion Paper

EDP-0630

Testing theories of labour market matching

By Martyn Andrews, Steve Bradley, Dave Stott
and Richard Upward

November 2006

Economics
School of Social Sciences
The University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom

Testing theories of labour market matching

Martyn Andrews
University of Manchester

Steve Bradley
Lancaster University

Dave Stott
Lancaster University

Richard Upward
University of Nottingham*

For submission to *Labour Economics*
November 2005

*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 for comments from various anonymous referees.

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. [96 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@man.ac.uk`
Phone: `+44-(0)161-275-4874`

1 Introduction

This paper is an empirical investigation into how workers 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 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 almost exclusively associated with Melvyn Coles and collaborators—see Coles & Smith (1998) and Coles & Petrongolo (2003).

These two competing models give quite different predictions and have 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 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 is not optimal. 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.

Finally, the stock-flow matching model gives a plausible explanation as to why unemployment hazards 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 whom the scarring explanation is less plausible.

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 (2003) 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 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 we would be unable to 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.

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

To set the scene, first consider a stylised version of the random 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

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

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 θ. The corresponding hazards are:

$$h^e(\theta) = \lambda^e(\theta)\mu(\theta) \qquad h^w(\theta) = \lambda^w(\theta)\mu(\theta), \tag{1}$$

where μ 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 θ on μ.

The aggregate matching (or hiring) function can be obtained by aggregating either hazard over the corresponding stock of market participants:

$$\delta(U,V) = Vh^e(\theta) = V\lambda^e(\theta)\mu(\theta) \tag{2}$$

$$= Uh^w(\theta) = U\lambda^w(\theta)\mu(\theta) = \lambda(U,V)\mu(\theta). \tag{3}$$

This shows how the matching function δ is decomposed into the contact function and the matching probability. It will exhibit constant returns to scale if $\lambda(\theta)$ does the same.

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. 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.

More formally, 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. The key assumption is that old job-seekers never 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.

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 time for many reasons, including duration dependence, unobserved heterogeneity and changing reservation utilities. A stronger test is therefore to estimate a model of exit 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 A Statistical Model of Non-random Matching

We now develop an estimable statistical model which has testable parametric restrictions that make the random matching model a special case of the non-random matching model. Our test formalises that proposed independently by Coles & Petrongolo (2003).

As above, the number of contacts per period are generated by

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

where, for estimation purposes, we will use the standard Cobb-Douglas specification $\lambda(U, V) = AU^\alpha V^\beta$. $\lambda(U, V)$ is the average number of contacts per period. In other words, 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.

Assume, for the moment, that it is the matching probabilities, conditional on contacting, that are different between types of pair. These are given by[1]

μ_{11} if new job-seeker, new vacancy
μ_{12} if new job-seeker, old vacancy
μ_{21} if old job-seeker, new vacancy
μ_{22} if old job-seeker, old vacancy.

In its basic form stock-flow matching implies that $\mu_{12} > 0$, $\mu_{21} > 0$, but $\mu_{22} = 0$. The pure theory rules out flow-flow matching by assumption and so μ_{11} is undefined. In real data, which is necessarily discrete, one might allow $\mu_{11} > 0$. We therefore redefine u and v as the stocks of 'new' job-seekers and vacancies, and redefine the 'stocks' of U and V, using the terminology of Coles and Smith, as 'old stocks'. The total stock of job-seekers and vacancies is thus decomposed as follows

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

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.

μ_{22} is zero because stock-stock matching cannot 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 occur, but with a low probability, especially if there are large numbers of each in the market. Then the prediction of the model is that $\mu_{12} > \mu_{22}$ and $\mu_{21} > \mu_{22}$.

A related point concerns whether stock-flow matching will be observed on both sides of the market. In a single market with *ex-ante* homogenous agents, stock-flow matching arises because of some form of market failure (e.g. efficiency wages). In this case, we have 'one-sided' stock-flow matching. For example, job-seekers chase vacancies as soon as they appear on the market, and so we do not have a model for old vacancies as above. One can model this by specifying $\mu_{12} = \mu_{22} < \mu_{21}$, or, if on the other side of the market, $\mu_{21} = \mu_{22} < \mu_{12}$. But in segmented labour markets differentiated by, for example, skill or location, 'two-sided' stock-flow matching may occur. In some sub-markets job-seekers have to chase new vacancies; in other sub-markets, the reverse is true.

[1]In this subscript notation the first subscript always refers to job-seekers, the second to vacancies; '1' always means new, '2' means old.

The aggregate matching function is defined for all four types of match:

$$\begin{aligned}
\delta_{11} &= \mu_{11}\frac{uv}{UV}\lambda(U,V) = A\mu_{11}uvU^{\alpha-1}V^{\beta-1}\\
\delta_{12} &= \mu_{12}\frac{u\bar{V}}{UV}\lambda(U,V) = A\mu_{12}u\bar{V}U^{\alpha-1}V^{\beta-1}\\
\delta_{21} &= \mu_{21}\frac{\bar{U}v}{UV}\lambda(U,V) = A\mu_{21}\bar{U}vU^{\alpha-1}V^{\beta-1}\\
\delta_{22} &= \mu_{22}\frac{\bar{U}\bar{V}}{UV}\lambda(U,V) = A\mu_{22}\bar{U}\bar{V}U^{\alpha-1}V^{\beta-1}.
\end{aligned}$$

where each of δ_{11}, δ_{12}, δ_{21}, and δ_{22} is the average number of matches of each type per period. Multiplying $\lambda(U,V)$ by $uv/UV,\ldots,\bar{U}\bar{V}/UV$ respectively splits the average number by type, which is then multiplied by the matching probability. 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 that makes old-old matches less frequent, and would be zero in the pure stock-flow matching model.

Note that one cannot separately identify A from four μs, and so we define $a_{ij} \equiv A\mu_{ij}$. It turns out that when we estimate these parameters, we can only identify their ratio. Providing one is prepared to assume that A does not vary by type, such a ratio can be interpreted as the corresponding ratio of μs.

Random matching is a special case when

$$H_0 : a_{11} = a_{12} = a_{21} = a_{22} \quad (= a, \text{ say}), \tag{4}$$

is true. Under H_0, summing the four δs, this aggregate matching function is given by

$$\delta = \mu\frac{[uv + u\bar{V} + \bar{U}v + \bar{U}\bar{V}]}{UV}\lambda(U,V) = \mu\lambda(U,V), \tag{5}$$

that is, generates Equation (3) above, except that here μ is no longer a function of labour-market tightness. This is because any effects of U and V via $\mu(U,V)$ cannot be identified separately from $\lambda(U,V)$; this is important when interpreting the estimates of α and β. It also follows that the way we have chosen to characterise stock-flow matching, namely random contacts with varying μs, is observationally equivalent to models that might have non-random contacts.

The corresponding hazard functions are given by:

$$h^w_{11} \equiv \delta_{11}/u = \mu_{11}\frac{uv}{UV}\lambda(U,V)/u = a_{11}vU^{\alpha-1}V^{\beta-1} \tag{6}$$

$$h^w_{12} \equiv \delta_{12}/u = \mu_{12}\frac{u\bar{V}}{UV}\lambda(U,V)/u = a_{12}\bar{V}U^{\alpha-1}V^{\beta-1} \tag{7}$$

$$h^w_{21} \equiv \delta_{21}/\bar{U} = \mu_{21}\frac{\bar{U}v}{UV}\lambda(U,V)/\bar{U} = a_{21}vU^{\alpha-1}V^{\beta-1} \tag{8}$$

$$h^w_{22} \equiv \delta_{22}/\bar{U} = \mu_{22}\frac{\bar{U}\bar{V}}{UV}\lambda(U,V)/\bar{U} = a_{22}\bar{V}U^{\alpha-1}V^{\beta-1} \tag{9}$$

For h^w_{11}, the $\lambda(U,V)/u$ term is the average number of contacts per job-seeker (and is directly analogous to λ^w in the random matching model); the $\mu_{11}uv/UV$ term is the matching probability (and is directly analogous to μ in the random matching model). There are another set of hazards for vacancies, labelled h^e_{11}, h^e_{12}, h^e_{21} and h^e_{22}.

Notice two things. First, $h^w_{22}/h^w_{12} = a_{22}/a_{12}$ and $h^w_{21}/h^w_{11} = a_{21}/a_{11}$. The job-seeker's hazard to old vacancies will drop if $a_{12} > a_{22}$, as predicted by the theory. Similarly, the vacancy hazard to old job-seekers will fall if $a_{21} > a_{22}$. The size of the ratios a_{12}/a_{22} and a_{21}/a_{22} is therefore an important testable hypothesis.

The hazard for an old job-seeker $h^w_{2.}$ is just $h^w_{21} + h^w_{22}$, so we have:

$$\begin{aligned}\log h^w_{2.}(U,u,V,v) &= \log(h^w_{21} + h^w_{22}) \\ &= \log[a_{21}v + a_{22}(V-v)] + (\alpha-1)\log U + (\beta-1)\log V. \qquad (10)\end{aligned}$$

Similar expressions arise for the hazards for new job-seekers $h^w_{1.}$, new vacancies $h^e_{.1}$, and old vacancies $h^e_{.2}$. Rather than estimate the non-linear model in Equation (10), it is much easier to estimate a model for $\log h^w_{2.}$ which is linear in $\log U$, $\log u$, $\log V$, $\log v$, from which we can uniquely identify α, β and a_{22}/a_{21}. Differentiating:

$$\begin{aligned}\frac{\partial \log h^w_{2.}}{\partial \log U} &= \alpha - 1 \qquad \frac{\partial \log h^w_{2.}}{\partial \log V} = \frac{a_{22}V}{a_{21}v + a_{22}\bar{V}} + \beta - 1 \equiv \pi_1 \\ \frac{\partial \log h^w_{2.}}{\partial \log u} &= 0 \qquad \frac{\partial \log h^w_{2.}}{\partial \log v} = \frac{(a_{21}-a_{22})v}{a_{21}v + a_{22}\bar{V}} \equiv \pi_2. \qquad (11)\end{aligned}$$

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 β, 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 non-random 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 non-random matching, v going up leads to more stock-flow matches (δ_{21} increases) but $\bar{V}$ going down means fewer stock-stock matches (δ_{22} decreases). 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 π_2, which is given by

$$\frac{a_{22}}{a_{21}} = \frac{v}{V(1-\pi_2)^{-1} - \bar{V}}. \tag{12}$$

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

Analogous discussions apply to Equation (10) when considering $\log h^w_{1.}(U,u,V,v)$, $\log h^e_{.1}(U,u,V,v)$, and $\log h^e_{.2}(U,u,V,v)$. 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 records who matches with whom is required for the test to work; our data are unique in this respect.

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 30% 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.

Job-seekers can come from one of four labour market states: unemployment, employment, government-sponsored training or education. Each vacancy is filled by one of these types of job-seeker or it is withdrawn from the market, or it is censored. Job vacancies can either be filled via the Careers Service, or filled by some other means. Each job-seeker finds one of these types of vacancy, or she leaves the labour market and stops actively searching, or she is censored. A job-seeker who stops searching and leaves the labour market is the analogue of a vacancy which is withdrawn from the market.

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. However, we do 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. 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.[2]

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 $m_{21} = 1,496$ stock-flow matches. There are also $m_{11} = 533$ flow-flow matches, another $m_{12} = 277$ stock-flow matches between new job-seekers and old vacancies, and $m_{22} = 455$ 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 the vacancy is old or new (recorded in the vacancy panel) and by whether or not the

[2]Strictly speaking, the unit of observation is a spell, not a job-seeker, as some job-seekers have multiple spells. Similarly, some vacancies are posted in multiple vacancy orders.

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

	new	old	total
Job-seekers (y^w_{is})			
zeros	123,338	351,769	475,107
exits to new vacancy	(m_{11}) 533	(m_{21}) 1,496	2,029
exits to old vacancy	(m_{12}) 277	(m_{22}) 455	732
Total	124,148	353,720	477,868
Vacancies (y^e_{js})			
zeros	38,547	95,915	134,462
exits to new job-seeker	(m_{11}) 533	(m_{12}) 277	810
exits to old job-seeker	(m_{21}) 1,496	(m_{22}) 455	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:

$$h^e_{11} = 533/40576 = 0.01314$$
$$h^e_{12} = 277/96647 = 0.00287$$

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

$$h^e_{21} = 1496/40576 = 0.03687$$
$$h^e_{22} = 455/96647 = 0.00471.$$

Notice that the drop in the hazard for vacancies matching with old unemployed job-seekers is $h^e_{22}/h^e_{21} = a_{22}/a_{21} = 0.218$ 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 $h^e_{.2}/h^e_{.1}$.

A similar analysis applies to job-seekers. It turns out that the drop in the hazard for job-seekers matching with old vacancies is $h^w_{22}/h^w_{12} = a_{22}/a_{12} = 0.577$. 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 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 2 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.

4.2 The Explanatory Variables

The explanatory variables in our model are measures of the 'stock' and 'flow' of job-seekers and vacancies. In practice the flow are 'new' job-seekers and vacancies, while the stock 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

$$U_t = U_{t-1} + (u_{t-1}^+ - u_{t-1}^-). \qquad (13)$$

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. 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' as follows (with an analogous expression for V_t):

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

The 'new' stock u_t of unemployed are defined as the inflow of unemployed during the week less those who also exit during the week, namely $u^+_{t-1} - u^-_{t-1}|u^+_{t-1}$ and the 'old' stock $\bar{U}_t$ are defined as the stock of unemployed at the end of the previous week less those who also exit during the current 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. It follows that there is very little cross-section variation in the data. Identification is achieved through the so-called recruitment cycle, which generates a lot of time-series variation in the data. 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 unemployed stocks, but less pronounced.

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 as follows. We assume proportional hazards and introduce a positive-valued random variable (or mixture) ϵ:

$$h^w_s(\mathbf{x}'_{is}, \epsilon^w_i) = \bar{h}^w_s \epsilon^w_i \exp(\mathbf{x}'_{is}\boldsymbol{\beta}^w).$$

$\bar{h}^w_s$ is the baseline hazard, and does not vary by i. $\varepsilon^w_i \equiv \log \epsilon^w_i$ has density $f^w_\varepsilon(\varepsilon^w)$, and is a job-seeker specific random effect.

The likelihood $L_i(\boldsymbol{\beta}, \boldsymbol{\gamma})$ for each job-seeker with observed covariates $\mathbf{x}'_{is}$ in this 'mixed proportional hazards' model is

$$L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \int_{-\infty}^{\infty} \left[\prod_{s=1}^{t_i} h_s(\mathbf{x}'_{is}, \varepsilon_i)^{y^w_{is}} [1 - h_s(\mathbf{x}'_{is}, \varepsilon_i)]^{1-y^w_{is}} \right] dF_\varepsilon(\varepsilon_i),$$
$$\text{with} \quad h_s(\mathbf{x}'_{is}, \varepsilon_i) = 1 - \exp[-\exp(\mathbf{x}'_{is}\boldsymbol{\beta} + \gamma_s + \varepsilon_i)].$$

where, for notational clarity, we have suppressed the superscript w. The same equation also applies to the vacancy hazard, replacing i by j. Because of the proportional hazards assumption, the covariates affect the hazard via the complementary log-log link. The γ_s terms are interpreted as the log of a non-parametric piece-wise linear baseline hazard, as $\gamma_s \approx \log \bar{h}_s$ when $\mathbf{x}'_{is}\boldsymbol{\beta} = 0$. The γ_s terms are collected into a vector $\boldsymbol{\gamma}$. Each interval corresponds to a week, but, because of data thinning, these are grouped into longer intervals at longer durations by constraining the appropriate γ_ss. We model the unobserved heterogeneity ε_i using Normal mixing, with variance σ^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. In both cases, details on how the likelihood functions are amended are given in Stewart (1996).

The appropriate specification for $\mathbf{x}'_{is}$ was derived in Section 3, namely the stocks of old and new job-seekers and vacancies. Instead of (10), we estimate a log-linear model with marginal effects given in (11). Because the hazard is predicted to change when the job-seeker or vacancy changes from new to old, we must interact 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 our final model has eight covariates: $\log U$, $\log u$, $\log V$ and $\log v$ interacted with a 'new' and an 'old' dummy.

It is worth emphasising that both stocks vary by duration s and job-seeker i/vacancy j, because they vary through calendar time and because each job-seeker/ vacancy enters the market place at different calendar times. This is important for identification, and is an effect lost with aggregate data. As just noted, instead of having just two dummies for the baseline hazard, for new and old, we estimate a non-parametric piece-wise linear version for reasons discussed in the next section.

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 (2003). 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 (2003) estimate matching functions using a quite sophisticated technique that deals 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 preceding Monday and (b) stocks measured monthly, ie the value observed on the first day of the preceding 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 2). In Figure 3 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.

6.1 The Base Model

Our strategy is to report a 'Base Model' which represents our preferred specification. We then re-estimate this model by making one-move departures in various dimensions to see whether our assumptions are important or innocuous. 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*

	Job-seekers, h^w		*Vacancies, h^e*	
	narrow	*wide*	*narrow*	*wide*
(a) New				
$\log u$	-0.008 (0.053)	-0.049 (0.038)	-0.161 (0.057)	0.137 (0.044)
$\log U$	-0.175 (0.082)	-0.143 (0.059)	0.902 (0.075)	0.661 (0.071)
$\log v$	0.242 (0.096)	0.227 (0.097)	-0.223 (0.078)	-0.231 (0.078)
$\log V$	0.234 (0.086)	0.265 (0.089)	-0.077 (0.072)	-0.127 (0.074)
α, β	0.817, 0.476	0.808, 0.492	0.741, 0.700	0.798, 0.642
$\alpha + \beta$	1.292 (0.071)	1.300 (0.068)	1.441 (0.068)	1.440 (0.076)
a-ratio[a]	0.462 [0.006]	0.482 [0.009]	2.218 [0.998]	0.396 [0.001]
(b) Old				
$\log u$	-0.134 (0.051)	-0.211 (0.041)	0.121 (0.063)	0.522 (0.050)
$\log U$	-0.296 (0.060)	-0.263 (0.055)	1.102 (0.111)	0.607 (0.121)
$\log v$	0.288 (0.058)	0.280 (0.058)	-0.003 (0.109)	0.030 (0.108)
$\log V$	0.258 (0.054)	0.285 (0.055)	-0.276 (0.116)	-0.319 (0.118)
α, β	0.570, 0.546	0.526, 0.565	1.223, 0.721	1.129, 0.711
$\alpha + \beta$	1.116 (0.051)	1.091 (0.054)	1.943 (0.103)	1.840 (0.117)
a-ratio[b]	0.404 [0.000]	0.413 [0.000]	0.649 [0.027]	0.087 [0.000]
SE[c]	0.650 (0.075)	0.651 (0.075)	1.865 (0.070)	1.868 (0.073)
$\log L$	-16625.6	-16631.5	-11349.9	-11393.4
Obs	477868	477868	137223	137223

*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, 58 and 212 respectively for the narrow definition of the stocks. The corresponding numbers for the wider definition are 207, 1987, 58 and 212.

[a] a_{12}/a_{11} for job-seekers, a_{21}/a_{11} for vacancies. The a-ratios calculated from Equation (12) 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] a_{22}/a_{21} for job-seekers, a_{22}/a_{12} for vacancies.

[c] Standard error (σ) for Normally distributed random effects.

The Base Model is defined as follows. It is estimated using two binary response panels, where the unit of observation is a week, one for job-seekers, the other for vacancies. 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. 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, but include changing the econometric specification, altering k^w and k^e, adding covariates, and using monthly stocks. We interpret the results in the context of the statistical model developed in Section 3—see Equation (11) in particular. The implied estimates of α, β and the a-ratios are also reported.

Table 2 shows that the estimates for the job-seeker hazards h^w are unaffected by whether we use the narrow or wide definitions of U and u. The biggest differences occur on the estimates for $\log u$ and $\log U$ for old job-seekers; as the differences partly offset each other, they give a similar estimate of α. On the other hand, the estimates for the vacancy hazard very much depend on the narrow/wide specifications.

First, we examine our test of stock-flow matching. Recall that this is 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 $\bar{V}$ (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 $\frac{\partial \log h^w_{2.}}{\partial \log v} = 0.280$, and is significant. This converts to a point estimate for $a_{22}/a_{21} = 0.413$. A very similar estimate occurs with narrow stocks. By contrast, on the other side of the market, the estimate very much depends on whether narrow or wide stocks are used. For the narrow definition, the effect of $\log u$ in the old vacancy hazard, $\frac{\partial \log h^e_{.2}}{\partial \log u}$ is 0.121 and is only just significant (p-value =0.027). Said differently, the implied point estimate of $a_{22}/a_{12} = 0.649$ has a one-sided confidence interval that just excludes unity. However, when using the wider definition, the effect of $\log u$ is a lot stronger, with $\frac{\partial \log h^e_{.2}}{\partial \log u} = 0.522$ and $a_{22}/a_{12} = 0.087$.

It is important to emphasise that in our test, job-seeker data is being used to identify an effect about employer behaviour. This might seem very counter-intuitive. An estimate of $a_{22}/a_{21} = 0.413$ can be compared with that found in the raw vacancy data, that is $a_{22}/a_{21} = 0.138$. Similarly, vacancy data is being used to identify an effect about job-seeker behaviour; we estimate a_{22}/a_{12} as either 0.649 or 0.087, whereas the raw job-seeker data gives $a_{22}/a_{12} = 0.577$. In other words, data from different sides of the same market can be used to estimate the same ratio. We have argued throughout this paper that looking at the shape of agents' baseline hazards is not necessarily a test of stock-flow matching, simply because hazards can fall for other reasons. Our estimates control for the effects of adding aggregate stocks, and also control for unobserved heterogeneity. These conditional hazards in the Base Model are plotted in Figure 4. This is for wide stocks, but identical figures are obtained for narrow. Also plotted are the raw hazards and what happens when we re-estimate the Base Model without controlling for unobserved heterogeneity.

Looking at the three vacancy hazards first, Figure 4(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 8 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 (1.87 compared with 0.65). 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.

In terms of classical matching elasticities α and β, 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 1.29 for new job-seekers and 1.12 for old, whereas for new vacancies it is 1.44 and even higher for old vacancies at 1.94 (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 (2005) 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.

Finally, recall that it should not matter whether the congestion comes from old or new job-seekers. In other words, $\log u$ should be insignificant in the job-seeker hazard, twice, and, similarly, $\log v$ should be insignificant in the vacancy hazard, twice. This only happens half the time. In the job-seeker hazards, $\log u$ is insignificant for new job-seekers but has a negative impact for old job-seekers. It is the other way round for the vacancy hazards. This suggests that the appropriate stock-flow matching model is *not* one that drops all four of these variables.

6.2 Departures from the Base Model

We now report estimates of various departures from the Base Model. Because we have two variants of the Base Model (wide and narrow definition of job-seekers), this exercise is done twice. However, we only report departures from the wide variant, see Table 3. We choose to focus on the wide-stock variant for the following reason. Define W and w as the old and new stocks of job-seekers for the wide definition. When we add $\log(W-U)$ and $\log(w-u)$ to the narrow vacancy regression, interacted with the old and new dummies, this gives three significant effects out of four, which indicates that job-seekers in employment or training programmes do compete with the unemployed when vacancies arrive on the market.

Row (1) of Table 3 shows the result of estimating the Base Model without unobserved heterogeneity. We already know that there is more heterogeneity on the vacancy side of the market; it is therefore not surprising that this has an impact on the vacancy estimates, whereas the job-seeker estimates are unaffected. Now $\partial \log h^e_{\cdot 2} / \partial \log u = 0.641$, rather than 0.522, implying that a_{22}/a_{12} falls from 0.087 to 0.055. Row (2) shows that the results are robust to the way the heterogeneity is modelled by also using discrete (Heckman-Singer) mixing.

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 unobervables as well. This justifies modelling the heterogeneity using random effects techniques.[3]

[3] This assumption is absolutely standard—see Wooldridge (2002, Chapter 20) for a clear discussion of these identification issues.

Table 3: Summary of departures from Base Model, wide stocks*

			Job-seekers, h^w					*Vacancies, h^e*				
	avge u	avge v[a]	$\log v$	a_{22}/a_{21}	α	β	$\log L$	$\log u$	a_{22}/a_{12}	α	β	$\log L$
Base Model	207	58	0.280 (0.058)	0.413	0.526	0.565	-16631.5	0.522 (0.050)	0.087	1.129	0.711	-11393.4
Departures												
(1) Without heterogeneity	207	58	0.276 (0.058)	0.418	0.546	0.547	-16642.7	0.641 (0.042)	0.055	0.718	0.791	-11827.1
(2) Heckman-Singer[b]	207	58	0.280 (0.058)	0.413	0.526	0.565	-16631.4	0.522 (0.049)	0.087	1.243	0.647	-11361.6
(3) With covariates[c]	207	58	0.300 (0.059)	0.390	0.493	0.533	-16498.9	0.550 (0.048)	0.079	1.059	0.824	-11206.3
(4) Classical random matching[d]	207	58	0[e]	1[e]	0.696	0.475	-16663.3	0[e]	1[e]	0.939	0.703	-11461.2
(5) Monthly stocks	221	60	-0.070 (0.057)	1.297	0.816	0.405	-16685.6	0.619 (0.055)	0.065	0.113	1.173	-11479.6
(6) Spell duration[f]	213	58	0.153 (0.056)	0.602	0.510	0.548	-16589.3	0.305 (0.047)	0.169	1.169	0.594	-11712.9
(7) $k^w = 5$, $k^e = 5$	254	67	0.238 (0.066)	0.503	0.608	0.561	-16638.8	0.657 (0.054)	0.063	0.898	0.751	-11384.6
(8) $k^w = 4$, $k^e = 2$	207	35	0.266 (0.048)	0.313	0.526	0.581	-16624.6	0.409 (0.041)	0.131	1.175	0.728	-11399.8
(9) $k^w = 4$, $k^e = 1$	207	19	0.248 (0.040)	0.214	0.527	0.587	-16614.1	0.368 (0.037)	0.152	1.011	0.707	-11411.4
(10) $k^w = 3$, $k^e = 3$	158	47	0.308 (0.051)	0.332	0.595	0.587	-16627.6	0.431 (0.044)	0.095	0.967	0.698	-11406.1
(11) $k^w = 3$, $k^e = 2$	158	35	0.281 (0.046)	0.297	0.597	0.590	-16623.8	0.373 (0.039)	0.118	1.192	0.725	-11404.6
(12) $k^w = 2$, $k^e = 2$	107	35	0.270 (0.043)	0.309	0.584	0.588	-16631.9	0.337 (0.037)	0.096	1.212	0.725	-11409.0
(13) $k^w = 2$, $k^e = 1$	107	19	0.261 (0.036)	0.202	0.582	0.600	-16621.0	0.295 (0.034)	0.114	1.030	0.706	-11421.3

*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".

[a] Average U is 1987, average V is 212 except for spell duration (2386 and 216 respectively) and monthly stocks (1965 and 210 respectively).

[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 (4) reports estimates of the classic random matching model. This involves constraining α and β across old and new, a constraint which is rejected by the data. We find a slight degree of increasing returns in job-seeker regressions. Our estimate of $\alpha + \beta = 1.642$ for vacancies is a new finding, mainly because very few studies use vacancy data.

In Row (5) we examine the effects of aggregation bias by replacing stocks observed at weekly intervals with those observed at monthly intervals. The results show that aggregation bias is a very serious problem. First, the estimate of α is much bigger in the job-seeker regression (moving from 0.526 to 0.816) and is smaller for β (moving from 0.565 to 0.405), and so $\alpha + \beta$ increases from 1.091 to 1.221. The effect in the vacancy hazards is the other way round, with α falling from 1.129 to 0.113—a very large change—and β increasing from 0.711 to 1.173, so that $\alpha + \beta$ falls from 1.840 to 1.286. It is possible to show that these movements in α and β go the right way, given standard results is 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 1.297 (and insignificantly different to 1) rather than 0.413, meaning that one would come to quite different conclusions about stock-flow matching when using monthly stocks, completely reversing those made so far.

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 little effect on the estimates.

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 is very robust, ranging from 0.24 to 0.31, which means that a_{22}/a_{21} falls with k^w and k^e as the average stocks change in size (see Equation 12). For the vacancy hazards, the effect of $\log u$ does fall with k^w and k^e, but leaving a_{22}/a_{12} robustly estimated in the range 0.06 to 0.15.

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 only one case (and that where aggregation bias is deliberately imposed) is $\log v$ insignificant, and hence a_{22}/a_{21} estimated to be equal to one. In particular, the result appears robust to the choice of window size.

When we repeat this exercise using the narrow definition of the stocks of job-seekers (not reported), estimates of $\log v$ in the job-seeker regression are unaffected, and our conclusion is equally robust. However, estimates of $\log u$ in the vacancy regression are less robust. The exclusion of employed job-seekers reduces the estimated coefficient on $\log u$ in every case, and in some cases renders it insignificant. This in turn implies that $a_{22}/a_{12} \approx 1$ in these cases. In one case, when using the spell duration definition, the estimate becomes negative at –0.220, making the a-ratio a very implausible 3.482. The reason for this general movement in the estimates away from two-sided stock-flow matching might be that those searching whilst employed or on training schemes compete successfully with some of the unemployed because they are seen as being better quality applicants by employers.

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. 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 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 about two-thirds when vacancies become old. 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 by about ninetenths when job-seekers become old. This result, that there is two-sided stock flow matching, is very plausible, and is consistent with the raw data from the other side of the market. The raw hazards have similar shapes—sharp declines for vacancies, much flatter for job-seekers—but when stocks are added to these regressions and unobservables are controlled for, we conclude that the sharp decline in the vacancy hazard at very short durations is also driven by unobserved heterogeneity rather than stock-flow matching on its own. Our own view is that the results using the wider definition are more convincing.

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. We have convincing evidence that aggregation bias is a serious problem with data that do not record information more frequently than at monthly intervals. 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

Broersma, L. & van Ours, J. (1999), 'Job searchers, job matches and the elasticity of matching', *Labour Economics* **6**, 77–93.

Burdett, K. & Coles, M. (1999), 'Long-term partnership formation: marriage and employment', *Economic Journal* **109**, F307–334.

Burdett, K., Coles, M. & van Ours, J. (1994), Temporal aggregation bias in stock-flow models. CEPR Discussion Paper No. 967.

Burdett, K. & Wright, R. (1998), 'Two-sided search with nontransferable utility', *Review of Economic Dynamics* **1**, 220–45.

Coles, M. & Petrongolo, B. (2003), A test between unemployment theories using matching data. Mimeo, July.

Coles, M. & Smith, E. (1998), 'Marketplaces and matching', *International Economic Review* **39**, 239–55.

Gregg, P. & Petrongolo, B. (2005), 'Stock-flow matching and the performance of the labour market', *European Economic Review*. Forthcoming.

Heckman, J. & Singer, B. (1984), 'Econometric duration analysis', *Journal of Econometrics* **24**, 63–132.

Petrongolo, B. & Pissarides, C. (2001), 'Looking into the black box: a survey of the matching function', *Journal of Economic Literature* **39**, 390–431.

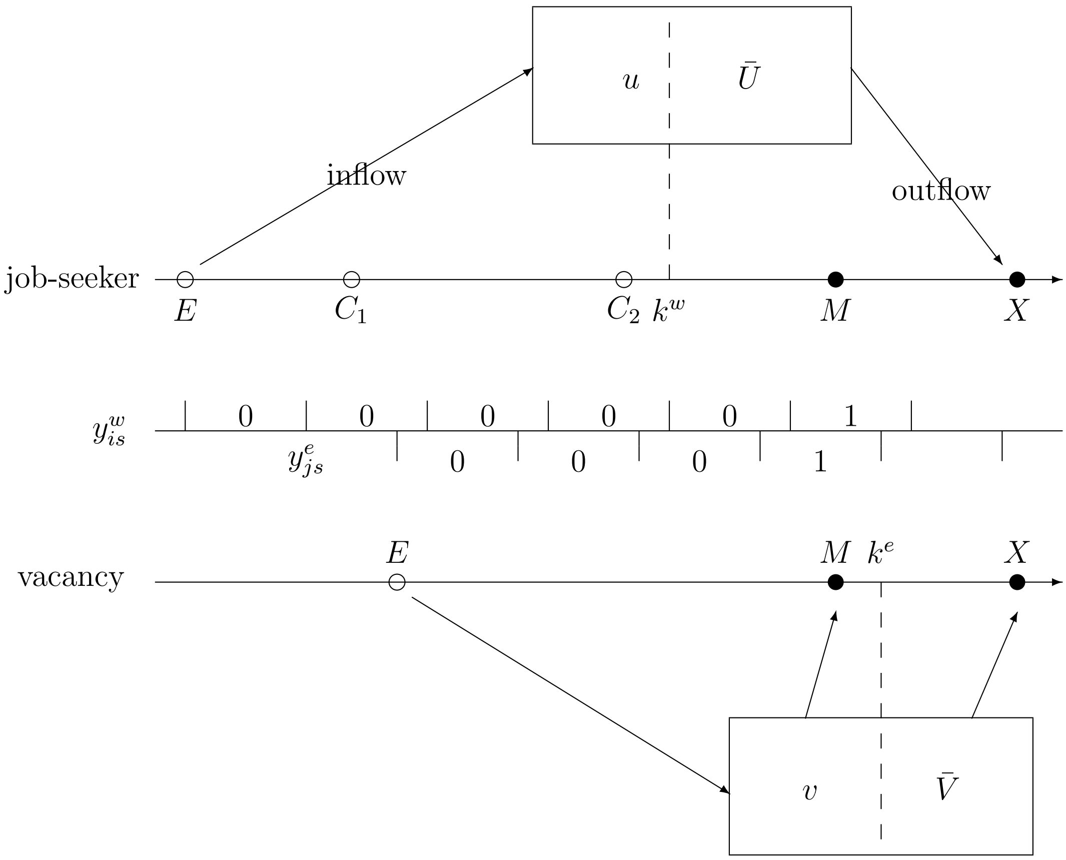
Petrongolo, B. & Pissarides, C. (2005), 'Scale effects in markets with search', *Economic Journal*. Forthcoming.

Pissarides, C. (2000), *Equilibrium Unemployment Theory*, second edn, Basil Blackwell, Oxford.

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

Wooldridge, J. (2002), *Econometric analysis of cross section and panel data*, MIT Press.

Figures



E	Start of spell	M	Match (end of search)
C	Contact (no match)	X	Exit (end of spell)
k^w	Job-seeker becomes old	k^e	Vacancy becomes old

Figure 1: Stock-flow matching and sequential binary response form

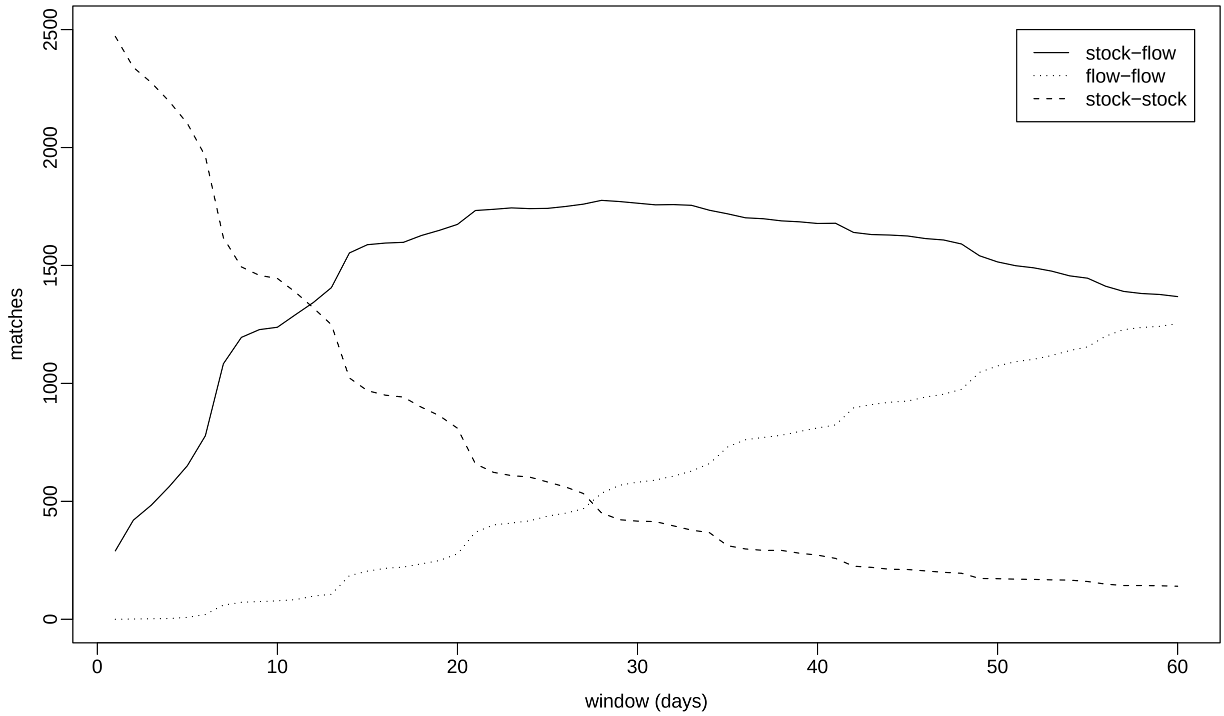
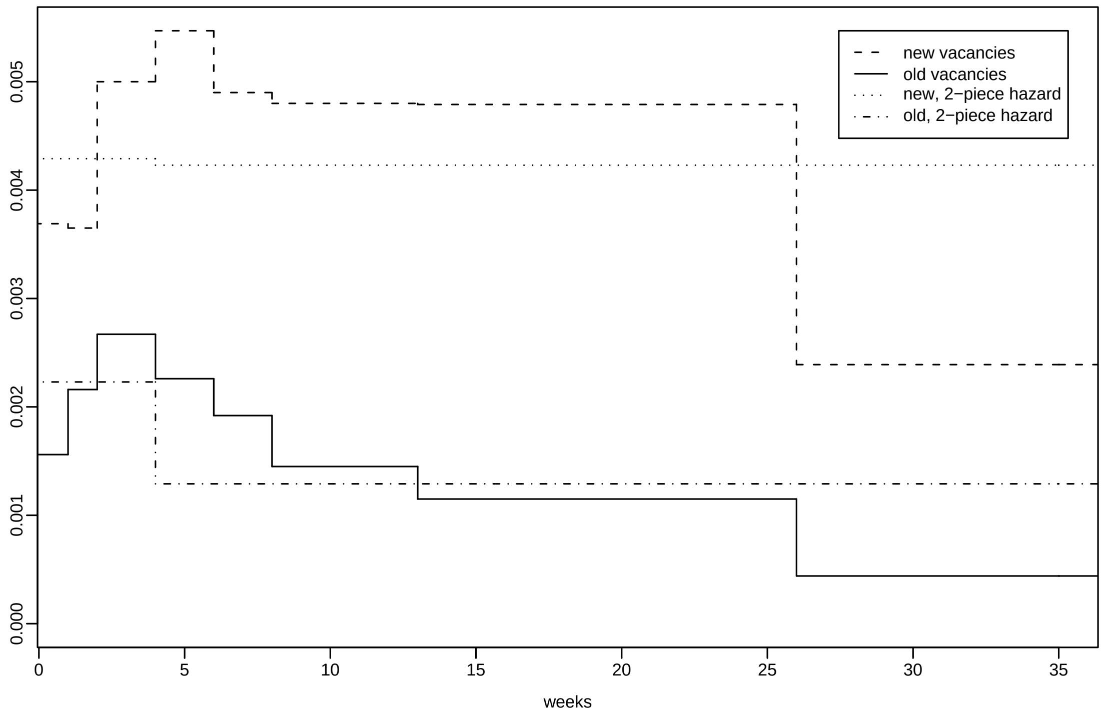


Figure 2: Stock-flow counts by window size, $k^w = k^e$

(a) job−seeker hazard, hw; search duration



(b) vacancy hazard, he; search duration

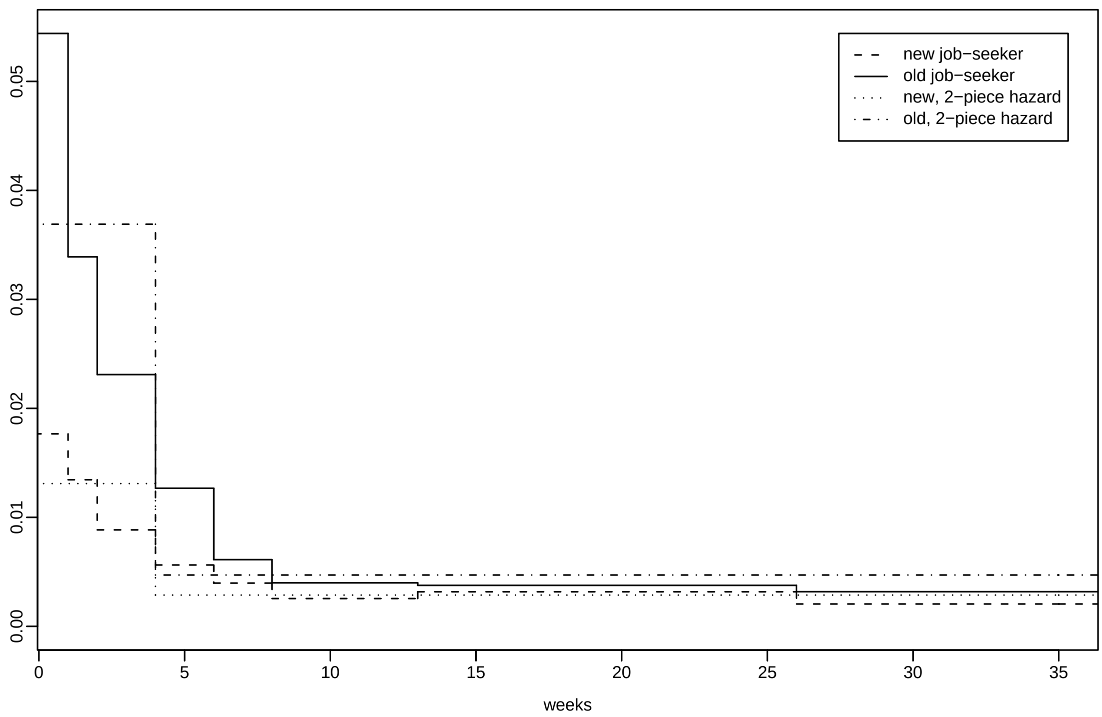
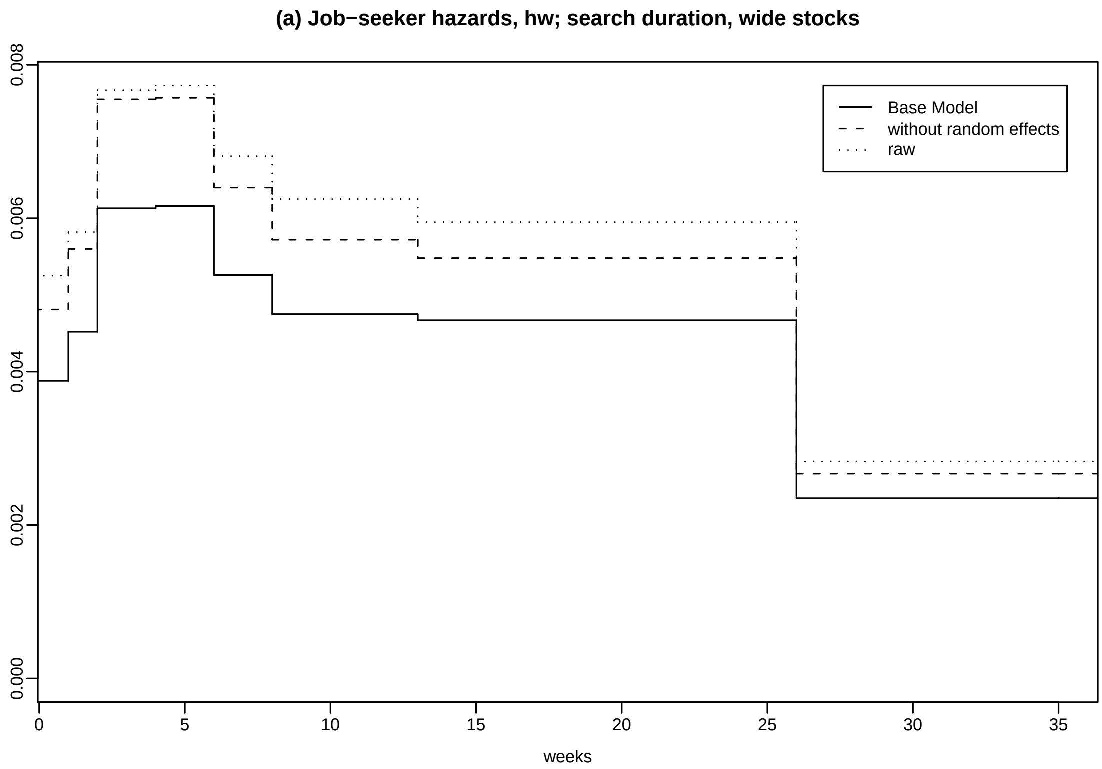


Figure 3: Raw job-seeker and vacancy hazards split by old and new; $k^w = k^e = 4$



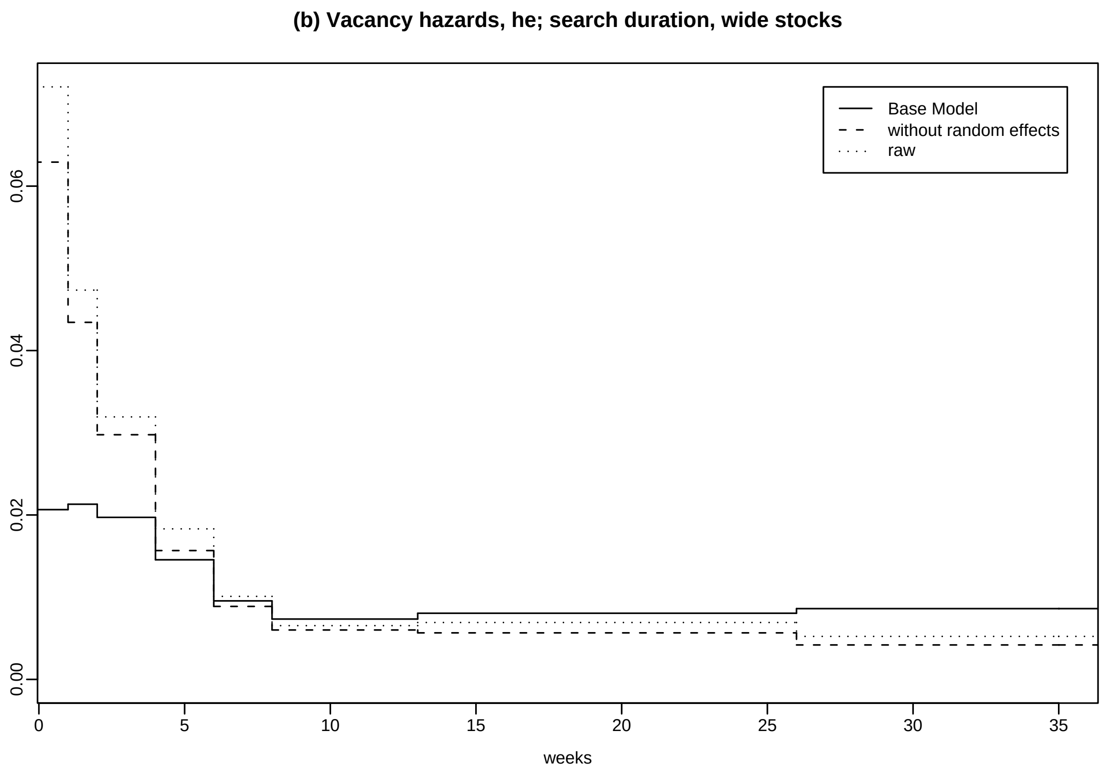


Figure 4: Conditional and raw baseline hazards