Testing theories of labour market matching

Martyn Andrews
University of Manchester

Steve Bradley
Lancaster University

Dave Stott
Lancaster University

Richard Upward
University of Nottingham*

October 2002

Version submitted to

2003 Royal Economic Society Conference

*The authors thank The Economic and Social Research Council (under grant R000239142) for financial assistance. The data were kindly supplied by Lancashire Careers Service. The comments from participants at various presentations are gratefully acknowledged. These include the Economics departments at Leicester, Manchester, and Nottingham. Subsequent revisions to this paper will be available at `http://les1.man.ac.uk/ses/staff/ma/`.

Abstract

This paper estimates a model of two-sided search using micro-level data for a well-defined labour market. It examines the assumption of random matching and contrasts it with the stock-flow (or non-random) matching model of Coles and collaborators. Given a dataset of contacts, matches, and complete labour-market histories for both sides of the market, we estimate hazard functions for both (unemployed) job-seekers and vacancies. For job-seekers, the tests adds the stock of new vacancies to a standard job-seeker hazard which itself depends on the stocks of vacancies and unemployed. Our tentative results find very weak evidence of stock-flow matching.

Keywords: two-sided search, random matching, hazards
New JEL Classification: C41, E24, J41, J63, J64

Address for Correspondence:

Dr. M.J. Andrews
School of Economic Studies
University of Manchester
Manchester, M13 9PL

Email: `martyn.andrews@man.ac.uk`
Phone: +44-(0)161-275-4874

1 Introduction

Search theory is becoming the dominant paradigm in explaining micro- and macro- labour-market phenomena, see Mortensen & Pissarides' recent (1998, 1999) surveys. Examples include the "flows approach" to the study of labour markets and whether movements in the Beveridge Curve inform discussions as to whether the labour market has become more or less effective over time (see, for example, Blanchard & Diamond (1989) Blanchard & Diamond (1992) and Davis & Haltiwanger (1999)). Other policy issues are often discussed in a search framework, see, for example, Manning's recent (2001) discussion on labour-market interventions and Marimon & Zilibotti (2000) on whether worksharing policy can reduce unemployment. Of course, search theory has been central in modelling the impact of benefits on unemployment duration (including Lancaster (1979), Meyer (1990), and many studies in between) and what causes long-term unemployment (Jackman, Layard & Pissarides 1989).

It is self-evident that search models are richer if they incorporate two-sided search, that of workers searching for vacant jobs and of employers searching for unemployed workers (or possibly workers seeking to change jobs). Pissarides' (2000) text (originally published in 1990) is seminal, and might be contrasted with Burdett & Wright (1998). However, both models, and others like them, incorporate many of the same basic structures and assumptions. Burdett & Coles's (1999) survey identifies four key assumptions of search/matching (SM) literature. The three that are relevant to labour markets are (i) *Poisson arrival rates*, that is, the process that generates contacts between employers and workers is a Poisson; (ii) *random matching*, that is, if an employer contacts a worker, it is assumed his identity is a random draw from all possible workers; and (iii) there exists an *encounter function*. To quote Burdett & Coles (1999) directly: "deep in the heart of all SM models is . . . an encounter function . . . which relates the numbers of encounters per unit of time [contacts] as a function of [stocks of] unemployed workers and vacancies". In other words, the encounter function is a production function that generates a flow of contacts. If all contacts lead to matches (hires), it is also the matching function. In the real world, and also in more and more theoretical models that have heterogeneous agents, not all contacts lead to matches and so the probability of a match is also modelled.

These three assumptions are deeply embedded in the SM literature. It is therefore important to establish whether they are observed in real world data. Most evidence is based on aggregate/time series data, and estimates the matching function. Unfortunately such data are not ideal. It is clearly much better to have micro-level data that distinguish contacts and matches, which is why very few studies have been able to estimate separate contact and matching probability functions. The same applies to the assumption of non-random matching. Non-random matching is almost exclusively associated with Melvyn Coles and collaborators, who provides a very persuasive alternative view as to how agents search and match with other. This is the *stock-flow matching* model.[1] But, again, it is self-evident that micro-level data are required, where we are able to observe who contacts whom, and who eventually matches with whom.

The data we have at our disposal, for a well-defined market where we observe *both* sides of the market, are ideal for examining these hypotheses. They refer to the youth labour market in Lancashire, are collected as an administrative dataset used by Lancashire Careers Service (LCS), between 1988 and 1992. We observe every contact between those employers and job seekers who used the Careers Service during the sample period, and for each contact we observe whether or not a successful match is made (approximately one-third of all matches in the youth labour market). Job seekers in this database are not just the unemployed, but also those in work, on a training scheme or in further education or at school. A wide range of covariates for job seekers and vacancies are observed. We also observe the actual day on which hires are made, and so we can compute daily durations for both the unemployed and vacancies. Nonetheless, these high frequency agent–level data are superior to those hitherto used for investigating SM assumptions, especially stock–flow matching. In this particular paper, we estimate hazard functions for both (unemployed) job-seekers and vacancies, using the same sample of matches. Decomposing the hazard function into a matching probability and the arrival rate of applicants is left to further research.

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 in to an estimable statistical model in Section 3. In Sections 4 and 5, we describe in some detail the data described immediately above and how they can be used to construct the key variables in the stock-flow matching model. Section 6 sets out the econometric methodology and in Section 7 we discuss our results. Section 8 concludes.

2 Theoretical framework

In this section we explain how the predictions of the stock–flow matching model are translated into specific econometric hypotheses. 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

[1] The best exposition is Coles & Smith (1998), but also see Coles & Petrongolo (2001).

$\lambda()$ 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 worker finds an employer acceptable and an employer finds a worker acceptable. In some two-sided search models $\mu(\theta)$ is an increasing function in slack markets and then becomes a decreasing function in tighter markets.

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.

There is a large microeconometric literature that has estimated the hazard out of unemployment using unemployment duration data,[2] but there is far less evidence for vacancies.[3] Search in a stationary environment predicts that the hazard is constant, although most estimated hazards show declining hazards. This is thought to be due to either some form of negative duration dependence or unmodelled unobserved heterogeneity. Assuming the latter can be controlled for using appropriate econometric techniques (see below), duration dependence can arise either because the arrival rate of suitable offers falls and/or the matching probability falls, as seen in decomposing the hazard in (1) above.[4] Other microeconometric studies do not estimate either hazard directly. Some have estimated the hiring function $\delta(U, V)$ directly[5] or the matching probability[6] or better still, have de-

[2]See van den Berg (1999, Footnote 1) for a recent list of contributions and surveys.

[3]See, for example, van Ours & Ridder (1991, 1992, 1993), Barron, Berger & Black (1997), Burdett & Cunningham (1998), and Russo & van Ommeren (1998), Andrews, Bradley & Upward (2001*a*).

[4]See van Ours (1990) for vacancies and van den Berg (1990) for unemployment.

[5]See Lindeboom, van Ours & Renes (1994), Anderson & Burgess (1997), and Broersma & van Ours (1999).

[6]See Teyssière (1996) and Andrews, Bradley & Upward (2001*b*).

composed the hiring function into λ and μ (see equation 3).[7] However, the great majority of empirical work on the hiring function has used aggregate time-series data.[8]

The important feature of the random matching model is that it is a model that explicitly allows for search/congestion externalities, which cannot be eliminated by price adjustments. By contrast, there is no congestion in Coles & Smith's stock-flow matching model, as workers are able to search all of the market in a short period of time, as are employers of workers. Unemployment and vacancies persist because suitable partners were not available on this first search of the market, and so workers/employers have to wait for new opportunities to flow into the market at a later date.

We now present a formal, albeit simplified, version of the stock-flow matching model to explain how the key predictions differ from the model above. Time is made of up discrete periods and agents arrive randomly, at a flow rate of u for job seekers and v for vacancies. In what follows, the possibility that two or more agents can arrive in a given period can be ignored. As above, the matching probability is μ. In some periods, a single job seeker will enter the market and will examine the stock of 'old' vacancies $\bar{V}$ ('old' in that they were in the market in the previous period). This job seeker *either* does not match with any of the stock of vacancies with probability $(1-\mu)^{\bar{V}}$ (and the stock of 'old' unemployed $\bar{U}$ increases by one in the next period) *or* he matches with one of the vacancies (and the stock of 'old' $\bar{V}$ decreases by one in the next period). Because of discounting, there is no stock-stock matching between $\bar{U}$ and $\bar{V}$; had there been gains to trade, pairs would have matched in an earlier period. There is no flow-flow matching because two agents cannot arrive together (but see below).

Thus the per-period flow of job seekers out of the marketplace is made up of two types. The first type are the new job seekers, who arrive with flow u and exit with probability $1-(1-\mu)^{\bar{V}}$. The second type are the old unemployed job seekers, who may match with new vacancies, the latter arriving with flow v. The arrival rate per unemployed job seeker is $v/\bar{U}$ (the analogue of λ^w above) and the matching probability is $1-(1-\mu)^{\bar{U}}$ rather than μ above, (the product of which is the analogue of $h^w = \lambda^w \mu$). Aggregating over all $\bar{U}$ gives an outflow rate of $v[1-(1-\mu)^{\bar{U}}]$. Adding the two flow types together gives

$$\delta(\bar{U}, \bar{V}, u, v) = u[1-(1-\mu)^{\bar{V}}] + v[1-(1-\mu)^{\bar{U}}]. \tag{4}$$

Identical considerations for the vacancy outflow lead to exactly the same expression. Equation (4) is the stock-flow matching analogue of (3) above. It has increasing returns to scale in $\bar{U}$ and $\bar{V}$, but is non-homogeneous. However, the more important difference is that it depends on the inflow rates u and v as well as the stocks $\bar{U}$ and $\bar{V}$, where the

[7]See van Ours & Lindeboom (1996).

[8]See Petrongolo & Pissarides (2001) for a comprehensive survey.

stock of job-seekers and vacancies in the random matching model are given by:

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

For Coles & Smith this is the first testable implication of stock-flow matching.

The second testable implication concerns the hazards. Job seekers who match immediately are only in the market for one period. Their hazard of exit $1 - (1 - \mu)^{\bar{V}}$ is much bigger than $(v/\bar{U})[1 - (1 - \mu)^{\bar{U}}]$ as $v \ll \bar{V}$ and $u \ll \bar{U}$ (and assuming that $u = v$ and $\bar{U} = \bar{V}$ in steady state). The old job seekers remain in the market for much longer on average, with average duration $(\bar{U}/v)[1 - (1 - \mu)^{\bar{U}}]^{-1}$ periods. This implies a step-wise hazard for both job seekers and vacancies. Also, the 'high' hazard for new job-seekers depends on u and $\bar{V}$ whereas the 'low' hazard for old job seekers depends on v and $\bar{U}$. This is the third testable implication. A fourth testable implication is that the matching probability for those who fail to match immediately should actually be invariant to duration in the market—older agents leave less quickly because they were unlucky, not because they become less 'attractive'. However, the assumption of no flow-flow matching is made only for mathematical convenience, and in general we would expect the hazards/outflow rates to also depend on v/u because of standard congestion arguments, as in the random matching model above.

As noted above, there is almost no evidence on the stock–flow matching model, unlike for random matching. Coles & Smith (1998) present estimates of $h^w = \delta(U, V, u, v)/U$ using monthly aggregate time-series Job Centre data between 1987 and 1995, where they observe U stratified by grouped duration, total V, monthly inflows u and v and outflows δ, also stratified by grouped duration. Their findings are strongly supportive of the theory. Gregg & Petrongolo (1997) use similar data and come to similar conclusions. Coles & Petrongolo (2001) have a recent interesting innovation to these two tests, using similar data.

In the rest of this paper, we estimate worker and employer hazards to see which of the random matching or stock-flow matching models are better supported by the micro-level data collected from Lancashire Careers Service. Estimates of the aggregate matching and contact functions is left for future research.

3 A statistical model of non-random matching

In this section, we develop an estimable statistical model that incorporates most of the features and predictions discussed above. Testable parametric restrictions that make the random matching model a special case of the non-random matching model are a key feature of this model. However, Coles & Smith's (1998) theory is amended to allow for matches between old job-seekers and old vacancies.

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.

It is the matching probabilities, *conditional* on contacting, that are different between types of pair. These are given by[9]

$$\begin{aligned}
&\mu_{11} \text{ if new job seeker, new vacancy}\\
&\mu_{12} \text{ if new job seeker, old vacancy}\\
&\mu_{21} \text{ if old job seeker, new vacancy}\\
&\mu_{22} \text{ if old job seeker, old vacancy.}
\end{aligned}$$

This allows the possibility that old-old matches can take place, even if there is stock-flow matching, but with a much lower probability. Note that new-new matches might be as likely as both types of old/new matches.[10] Random matching is a special case when

$$H_0 : \mu_{11} = \mu_{12} = \mu_{21} = \mu_{22} \quad (= \mu, \text{ say}), \tag{5}$$

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

$$\delta_{11} = \mu_{11}\frac{uv}{UV}\lambda(U, V) = a\mu_{11}uvU^{\alpha-1}V^{\beta-1} \tag{6}$$

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

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

$$\delta_{22} = \mu_{22}\frac{\bar{U}\bar{V}}{UV}\lambda(U, V) = a\mu_{22}\bar{U}\bar{V}U^{\alpha-1}V^{\beta-1}. \tag{9}$$

where each δ_{ij} is the average number of matches of each type per period. Multiplying $\lambda(U, V)$ by $uv/UV, \ldots, \bar{U}\bar{V}/UV$ splits the average number by type, which is then multiplied by the matching probability. Note that old-old contacts are relatively *very* frequent

[9] We make use of this subscript i, j notation throughout: i always refers to job-seekers and "1" always means new.

[10] Coles & Petrongolo (2001) allow for one-sided stock-flow matching, which is their *efficiency wage* model. One can model this by specifying $\mu_{12} \neq \mu_{21}$, or $\theta \neq 1$, or both.

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.

The aggregate matching function sums the four δ_{ij}s. Under H_0, this aggregate matching function is given by

$$\delta = \mu \frac{[uv + u\bar{V} + \bar{U}v + \bar{U}\bar{V}]}{UV} \lambda(U, V) = \mu\lambda(U, V), \tag{10}$$

that is, generates Equation (3) above, except that here μ is no longer a function of labour-market tightness. The reason is that any effects of U and V via $\mu(U, V)$ cannot be identified separately from $\lambda(U, V)$. For the same reason, we set $a = 1$ in equations (6–9) because a cannot be separately identified from μ_{ij}.

The corresponding hazard functions are given by:

$$h^w_{11}(u, \bar{U}, v, \bar{V}) \equiv \delta_{11}/u = \mu_{11} \frac{uv}{UV} \lambda(U, V)/u = \mu_{11} v U^{\alpha-1} V^{\beta-1} \tag{11}$$
$$h^w_{12}(u, \bar{U}, v, \bar{V}) \equiv \delta_{12}/u = \mu_{12} \frac{u\bar{V}}{UV} \lambda(U, V)/u = \mu_{12} \bar{V} U^{\alpha-1} V^{\beta-1} \tag{12}$$
$$h^w_{21}(u, \bar{U}, v, \bar{V}) \equiv \delta_{21}/\bar{U} = \mu_{21} \frac{\bar{U}v}{UV} \lambda(U, V)/\bar{U} = \mu_{21} v U^{\alpha-1} V^{\beta-1} \tag{13}$$
$$h^w_{22}(u, \bar{U}, v, \bar{V}) \equiv \delta_{22}/\bar{U} = \mu_{22} \frac{\bar{U}\bar{V}}{UV} \lambda(U, V)/\bar{U} = \mu_{22} \bar{V} U^{\alpha-1} V^{\beta-1} \tag{14}$$

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).

Notice two things. First, $h^w_{22}/h^w_{12} = \mu_{22}/\mu_{12}$ and $h^w_{21}/h^w_{11} = \mu_{21}/\mu_{11}$. This means that the job seeker's hazard to old employers will drop sharply when the job seeker becomes old if $\mu_{12} \gg \mu_{22}$ but that the shape of the job seeker's hazard to new employers may or may not fall because we have no *a priori* view about whether $\mu_{11} \lesseqgtr \mu_{21}$. This stepwise shape in the old job seeker hazard was noted in Section 2 above. Second, the hazard to old employers will be much higher than to new employers simply because $\bar{U} \gg u$.

The easiest way to proceed is to specify the logarithms of each of u, $\bar{U}$, v, $\bar{V}$ as covariates. We now add across competing risks:

$$h^w_{1.}(u, \bar{U}, v, \bar{V}) \equiv h^w_{11} + h^w_{12} = (\delta_{11} + \delta_{12})/u \tag{15}$$
$$h^w_{2.}(u, \bar{U}, v, \bar{V}) \equiv h^w_{21} + h^w_{22} = (\delta_{21} + \delta_{22})/\bar{U}. \tag{16}$$

The first equation is the job seeker hazard when the job seeker is new and the second equation is when the job seeker is old. (In fact, this model is estimated as a single regression where the four covariates are interacted with two dummy variables: one for when the job seeker is new and one for when the job seeker is old.) All of the above is repeated for employer hazards:

$$h^e_{.1}(u,\bar{U},v,\bar{V}) \equiv h^e_{11} + h^e_{21} = (\delta_{11}+\delta_{21})/v \tag{17}$$
$$h^e_{.2}(u,\bar{U},v,\bar{V}) \equiv h^e_{12} + h^e_{22} = (\delta_{12}+\delta_{22})/\bar{V}. \tag{18}$$

To interpret the estimates obtained from this log-linear specification in $u, \bar{U}, v, \bar{V}$, consider the hazard for old job seekers $h^w_{2.}$,

$$\log h^w_{2.} = \log(\mu_{21}v + \mu_{22}\bar{V}) + (\alpha-1)\log U + (\beta-1)\log V,$$

and differentiate:

$$\begin{aligned}
\frac{\partial \log h^w_{2.}}{\partial \log u} &= (\alpha-1)\frac{u}{U} & \frac{\partial \log h^w_{2.}}{\partial \log v} &= \frac{\mu_{21}v}{\mu_{21}v+\mu_{22}\bar{V}} + (\beta-1)\frac{v}{V}\\
\frac{\partial \log h^w_{2.}}{\partial \log \bar{U}} &= (\alpha-1)\frac{\bar{U}}{U} & \frac{\partial \log h^w_{2.}}{\partial \log \bar{V}} &= \frac{\mu_{22}\bar{V}}{\mu_{21}v+\mu_{22}\bar{V}} + (\beta-1)\frac{\bar{V}}{V}.
\end{aligned} \tag{19}$$

Adding together the estimates for $\log u$ and $\log \bar{U}$ gives $\alpha-1$ and similarly adding together the estimates for $\log v$ and $\log \bar{V}$ gives β. Similar expressions apply for $h^w_{1.}$, $h^e_{.1}$ and $h^e_{.2}$, but are not shown. From the estimates on $\log v$ and $\log \bar{V}$ one can solve for μ_{22}/μ_{21} twice, using sample averages for v and $\bar{V}$. In practice these are identical, providing the identity $V \equiv v + \bar{V}$ holds.

The effect of there being more job seekers in the market lowers the exit hazard for the old job seekers. If the increase were all new job seekers the effect on the hazard would be $(\alpha-1)u/U$ whereas if the increase were old job seekers it would be $(\alpha-1)\bar{U}/U$, which is much bigger. This is simply a composition effect as there are $\bar{U}/u$ times more old job seekers looking for vacancies than new job seekers. (Each has the same effect, but expressed as an elasticity, the old "do better".) In fact, an increase in the number of new job seekers can be decomposed into two effects. The first is a negative effect, $-u/U$, as more new job seekers means less chance of bumping into a vacancy (market is slacker), but this is offset partially by a second effect, there being more contacts, $\alpha u/U$.

There are analogous effects from an increase in the number of vacancies on the market. The first effect is that more contacts occur, ie $\beta v/V$ if new and $\beta\bar{V}/V$ if old. The second is the effect of new/old vacancies on the exit probability, given a contact occurs. For new vacancies, this component of the partial derivative is

$$\frac{\mu_{21}v}{\mu_{21}v+\mu_{22}\bar{V}} - \frac{v}{V}$$

which is positive if $\mu_{22} < \mu_{21}$. For old vacancies, this second effect is

$$\frac{\mu_{22}\bar{V}}{\mu_{21}v+\mu_{22}\bar{V}} - \frac{\bar{V}}{V}$$

which is negative if $\mu_{22} < \mu_{21}$ (and equal and opposite to the expression immediately above it). The fact that this is negative delivers the key prediction of the stock-flow matching model, that $\frac{\partial \log h^w_{2.}}{\partial \log \bar{V}}$ is smaller than it would be under random matching. However, there is no guarantee that $\frac{\partial \log h^w_{2.}}{\partial \log \bar{V}}$ is exactly zero (it clearly depends on $\bar{V}/v$ and μ_{22}/μ_{21}), even under pure stock-flow matching. This is because of the random nature of the contact function: one extra old job seeker entering the market affects the exit probability for old job seekers even if they cannot match ($\mu_{22} = 0$).[11] It is also for this reason that each of the 4 hazards depends on each of the 4 covariates, unlike the prediction noted in Section 2 above.

To emphasise, the terms involving μs only have an effect if $\mu_{22} \neq \mu_{21}$, which it is under stock-flow matching. Otherwise, it doesn't matter whether one meets an old or new vacancy—the exit probability, given a contact, is unaffected.

Exactly the same considerations apply to the other three hazards $h^w_{1.}$, $h^e_{.1}$ and $h^e_{.2}$, and, in particular, to $\frac{\partial \log h^e_{.2}}{\partial \log \bar{U}}$.

There is a better interpretation of the model when it is reparameterised so that the covariates are u, U, v, and V, again all in logarithms. Continuing with the old job-seeker hazard as an example:

$$\begin{aligned} \frac{\partial \log h^w_{2.}}{\partial \log U} &= \alpha - 1 & \frac{\partial \log h^w_{2.}}{\partial \log V} &= \frac{\mu_{22} V}{\mu_{21} v + \mu_{22} \bar{V}} + \beta - 1 \equiv \pi_1 \\ \frac{\partial \log h^w_{2.}}{\partial \log u} &= 0 & \frac{\partial \log h^w_{2.}}{\partial \log v} &= \frac{(\mu_{21} - \mu_{22}) v}{\mu_{21} v + \mu_{22} \bar{V}} \equiv \pi_2. \end{aligned} \tag{20}$$

An increase in the stock of unemployed job seekers has the familiar effect of $\alpha - 1$, and it does not matter whether the extra stock comprise old or new job seekers, because the extra effect from old job-seekers is zero. This is specification test of the particular statistical model we have adopted. If we are then able to drop $\log u$ from the specification, we then have the non-random matching model, which itself nests the random matching model. Three variables, $\log U$, $\log V$ and $\log v$, generate estimates of α, β and μ_{22}/μ_{21}. To obtain an estimate of β, one adds together the estimates on $\log V$ and $\log v$ (ie $\pi_1 + \pi_2 = \beta$). An estimate of μ_{22}/μ_{21} is given by

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

Part of the test of the random matching model is whether new vacancies have any effect on the hazard over and above that of all vacancies, ie whether v is significant and positive; it is clear that a test of $H_1 : \pi_2 = 0$ is equivalent to $H_1 : \mu_{21} = \mu_{22}$ because $\mu_{22}/\mu_{21} = 1$ if H_1 is true. The advantage of this approach is that we are able to test for stock-flow matching

[11] This might seem a weakness of this particular statistical matching model, but cannot be investigated unless separate data on contacts and matches is available. This is deferred to future research.

with a Wald test using a heteroscedastic robust (Huber-White) covariance matrix. In the previous parameterisation one would have to test stock-flow matching by comparing log-likelihoods, which is invalid in the presence of heteroscedasticity. (The reasons why we almost certainly have heteroscedasticity are discussed later.)

Using expressions similar to Equations (20–21), the new job-seeker hazard delivers estimates of α, β and μ_{12}/μ_{11}, and so one can test $H_2 : \mu_{11} = \mu_{12}$. Imposing H_1 and H_2 on equations (15–16) gives:

$$\log h^w_{1.} = \log \mu + (\alpha - 1) \log U + \beta \log V \tag{22}$$

$$\log h^w_{2.} = \log \mu + (\alpha - 1) \log U + \beta \log V \tag{23}$$

The actual random matching model, of course, merges 2 regressions into one by pooling $h^w_{1.}$ with $h^w_{2.}$:

$$\log h^w = \log \mu + (\alpha - 1) \log U + \beta \log V \tag{24}$$

Note that these 2 further restrictions are *not* part of the test: here we are testing whether α and β are the same across old and new variants (although it implicitly imposes the third equality in H_0).

Analogous considerations apply to employer hazards, giving the equivalent random matching model if all 6 equivalent restrictions hold:

$$\log h^e = \log \mu + \alpha \log U + (\beta - 1) \log V. \tag{25}$$

4 The data

The data we have at our disposal were described in the penultimate paragraph of the Introduction. In this first subsection we give some of the institutional background to the youth labour market in the UK in the late 1980s. In the following subsection we describe in some detail the information we observe. In Section 5, we define the empirical counterparts that are needed to test stock-flow matching, namely the old and new stocks U, $\bar{U}$, u, V, $\bar{V}$, and v above, and the flow of old and new matches, corresponding to δ_{ij} above.

4.1 Institutional Background

The collapse of the youth labour market in the UK in the early 1980s led to the introduction of the Youth Training Scheme (YTS) in 1983.[12] It has remained in place ever since,

[12]Fuller details are given in Andrews et al. (2001*b*), from which this subsection is taken.

albeit in several disguises. The YTS is not a homogeneous programme; it can be seen as a route to a wide variety of skilled occupations, or seen as a work-experience programme designed to mop up the excess supply of youth labour. Since its introduction, at the age of sixteen youths can choose between four labour-market activities: different types of YTS, continue their education, get a job or become unemployed. Employers can also choose whether to recruit youths via the YTS or directly into a job.

The Careers Service fulfills a similar role for the youth labour market as Employment Offices and Job Centres provide for adults. Its main responsibilities are to provide vocational guidance for youths and to act as an employment service to employers and youths. The latter includes a free pre-selection service for employers. Use of the Careers Service is voluntary for employers with job vacancies, whereas notification of YTS vacancies is compulsory, so that the government offer of a guaranteed place for all 16-17 year old youths can be monitored. Having notified the Careers Service of the type of vacancy—the occupation, the wage, a closing date for applications and selection criteria—job seekers are selected for interview. In other words, a contact is made. Either a match occurs or the pair each continue their search.

The data we use are the computerised records of the Lancashire Careers Service (LCS). The LCS holds records on *all* youths aged between 15 and 18, including those who are seeking employment. We observe every vacancy notified by employers to the Careers Service between March 1988 and June 1992. All YTS vacancies and about 30% of job vacancies are notified to the Careers Service. Job vacancies for which the Careers Service is not the method of search are not included in the data. Job vacancies require both high- and low-quality job seekers, and are representative of all entry-level jobs in the youth labour market. It follows that our data are representative of all job seekers, because we observe *all* contacts between notified job vacancies and job seekers. This is not an issue for YTS vacancies because all of them are notified to the Careers Service.

4.2 Observed data in the LCS database

Each contact, and therefore each match, in the labour market covered by the LCS data originates from a stock of job-seekers S and a stock of vacancies V. These decompose as follows:

Job seekers (S):

- Unemployed (U)
- (in) Jobs (N)

- (on) YT scheme (Y)
- School-leaver (F)

Each vacancy is filled by one of these types of job-seeker, or it is lapsed or it is censored (almost zero in these data).

Vacancies (V):

- Job vacancy filled via LCS (J)
- Job vacancy not filled via LCS (J')
- YT vacancy filled via LCS (T)
- YT vacancy not filled via LCS (T')

Each job-seeker finds one of these types of vacancy, or she lapses ('out of the labour market', olm) or she is censored.[13] Thus all vacancies filled by LCS is defined as

$$V \equiv J + T$$

which, when added to those not filled by LCS, $J' + T'$, gives a total stock of filled vacancies equal to:

$$J + T + J' + T' \equiv V + J' + T'.$$

The primary unit of observation is a contact, ordered by calendar time, labelled $i = 1, \ldots$. The binary variable m_i takes the value unity if a match occurs. c_i is an analogous variable that is always unity. Associated with each contact is the identity of the job-seeker w and vacancy e (itself associated with an employer) and the day on which the contact occurred τ. Formally we define the set of triplets

$$\{(w, e, \tau)\} = \{i \mid \mathbf{W}(i) = w, \mathbf{E}(i) = e, \tau\},$$

where the variable $\mathbf{W}(i)$ maps each job-seeker into the contact, if any, she makes on day τ and similarly $\mathbf{E}(i)$ does the same for vacancies. From this triplet, we 'match in' various types of information. From w:

- the origin state of the job-seeker, and hence the stock of job-seekers S. This varies by day through the duration of the job-seeker's stay in his/her origin state, ie between dates $\tau - t^w$ and τ, where

[13] In some ways, a vacancy that lapses is the analogue of a job-seeker who exits 'out of the labour market'.

- t^w is the duration of the spell in S (measured in days);
- a vector of characteristics $\mathbf{x}^w$.

From e:

- the origin state of the vacancy, and hence the stock of vacancies V. This varies by day through the duration of the vacancy, ie between dates $\tau - t^e$ and τ, where
- t^e is the duration of the spell in V (measured in days);
- a vector of characteristics $\mathbf{x}^e$;
- the wage/training allowance ω.

For vacancies not filled by the Careers Service ($V' \equiv J' + T'$), we do not observe the information immediately above.

Thus for each contact/match, we observe the following vector of information:

$$(\tau, w, e, S, V, t^w, t^e, \mathbf{x}^w, \mathbf{x}^e, \omega).$$

All of the analysis in this paper is conducted at the level of individual matches, where typically the variable being modelled is the duration between matches for job-seekers and between matches for vacancies. In keeping with most of the existing literature, we could conduct aggregate analyses, where we would count the number of matches that occur in any period t. However, it is the case that there is no extra information contained in such analysis and so estimating aggregate matching functions generally gives similar, but less efficient, estimates and is therefore unnecessary.

Table 1 summarises, over the whole sample period, the total number of matches stratified by the origin state of both job-seeker and vacancy. In what follows, we do not model matches of job-seekers who are at school (F), those in jobs (N) or on training schemes (Y). Modelling those who are searching whilst at school will potentially bias the results towards stock-flow matching in that there will be left-censoring causing a spike at zero durations (Andrews, Bradley & Stott 2002). On the other hand, they are part of the same labour market and potentially compete for the same vacancies as do the unemployed, and so are included in the risk set for vacancy hazards. We do not model those on jobs or training schemes because we are not prepared to make arbitrary assumptions about whether they are involved in "on-the-job" search. Because we do not model N, Y and F job-seekers, this just leaves the two-types of match for which we observe information on both sides of the market.

Thus our analysis below is based on 2761 matches between job vacancies filled via the CS and unemployed job-seekers, and the 10416 matches between YT vacancies filled via the CS and unemployed job-seekers. These two totals are defined as follows:

$$n_1 = \sum_i m_i 1(U, J) = 2761 \qquad n_2 = \sum_i m_i 1(U, T) = 10416.$$

where the function $1(U, J)$ defines a dummy variable that is unity if the match is between an unemployed job seeker and a job vacancy and zero otherwise. $1(U, T)$ is similarly defined, but for a training vacancy. The 2761 U, J matches represent exits from both sides of the market, that is there are 2 hazards that can be estimated from this sample of matches, an unemployment hazard $h^w(U, J)$ and a job vacancy hazard $h^e(U, J)$. Similarly, 2 hazards can be estimated from 10416 U, T matches, an unemployment hazard $h^w(U, T)$ and a job vacancy hazard $h^e(U, T)$. To be able to estimate hazards from both sides of the market using identical exits are a unique feature of these data. The risk set for the job-vacancy hazard is 14148 LCS job-vacancy spells and the risk set for the YT-vacancy hazard is 36853 spells (see the rightmost column of Table 1.) Notice that the risk set for both unemployment hazards $h^w(U, J)$ and $h^w(U, T)$ is the same (34659 unemployed job-seeker spells) and suggests that we can estimate two more hazards for job and YT vacancies that are not filled via LCS, namely $h^w(U, J')$ and $h^w(U, T')$, in a competing risks framework.[14] The problem here is that we do not observe vacancy information T' and J' for these hazards.

[14]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.

Table 1: Total number of observations by match type

	Job seekers								
Vacancies	U	F	N	Y	S	lapsed	censd	end	total
LCS job vacancies (J)	2761	1573	644	579	5557	8407	0	184	14148
Non-LCS job vacancies (J')	8422	7396	1163	8935	25916	n/o	n/o	n/o	25916
LCS YT vacancies (T)	10416	7548	1344	0	19308	14764	0	2781	36853
Non-LCS YT vacancies (T')	848	1409	532	0	2789	n/o	n/o	n/o	2789
All LCS vacancies (V)	13177	9121	1988	579	24865	23171	0	2965	51001
All vacancies ($V+V'$)	22447	17926	3683	9514	53570	23171	0	2965	79706
Unemployed	0	0	6286	5095	11381				
Out of the labour market	6679	797	531	2027	10034				
Censored during sample	3095	0	10563	795	14453				
Censored at end of sample	2438	0	11064	5237	18739				
Total at risk	34659	18723	32127	22668	108177				

n/o means not observed. Blank cells refer to meaningless information.

5 Old and new stocks and flows

To estimate the statistical model, we need to decide how long a job-seeker or a vacancy is on the market before it changes from being 'new' to 'old', or in Coles and Smith's terminology, from 'flow' to 'stock'. Then the aggregate stocks of job-seekers and vacancies have to be disaggregated into those who are old and new. The point at which this happens is defined as k^w for job seekers and k^e for employers, and is measured in weeks. We refer to the first k^w and k^e weeks as the matching *window*.

5.1 The raw data

The data are organised into sequential binary response form (see, for example, Stewart 1996). For the vacancy [resp. job-seeker] hazard we define

$$y_{is} = \begin{cases} 00\ldots0001 & \text{if the vacancy [resp. job-seeker] exits to } U, J \text{ match} \\ 00\ldots00 & \text{otherwise} \end{cases}$$

where i indexes the individual vacancy [resp. job-seeker] and s indexes duration. Essentially we have an unbalanced panel of vacancies with t_i^e weekly observations for each vacancy, and an unbalanced panel of job-seekers with t_i^w weekly observations for each job seeker. Note that

$$\sum_i \sum_s y_{is} = \sum_i m_i 1(U, J) = n_1 = 2761.$$

for both worker and employer hazards. For vacancies, there are $14148 - 2761 = 11387$ spells when the final y_{is} is zero, whereas for unemployed job seekers, there are $34659 - 2761 = 31898$ spells (see Table 1).

5.2 Old and new flows

If, for example, $k^e = k^w = 4$ weeks, then the first 4 zeros correspond to when the vacancy or job seeker is "new", for which we define the following dummy variables: $1(s \leq k^e)$ and $1(s \leq k^w)$. The cross-tabulations given in Table 2 describe almost all there is to know about these data. Thus we define m_{11} as the number of matches between a new job-seeker, ie who has been unemployed for less than k^w days, and a new vacancy, ie one that has been open for less than k^e days:

$$m_{11} = \sum m_i 1(t^w \leq k^w) 1(t^e \leq k^e) 1(U, J)$$

These are Coles & Smith's *flow-flow* matches. Similarly

$$m_{22} = \sum m_i 1(t^w > k^w) 1(t^e > k^e) 1(U, J)$$

defines the number of *stock-stock* matches. The number of *stock-flow* matches are:

$$m_{12} = \sum m_i 1(t^w \leq k^w) 1(t^e > k^e) 1(U, J) \quad \text{and}$$
$$m_{21} = \sum m_i 1(t^w > k^w) 1(t^e \leq k^e) 1(U, J)$$

From the cross-tabulations, we can see there are $m_{11} = 420$ flow-flow matches, $m_{12} = 191$ and $m_{21} = 1467$ flow-stock matches, and $m_{22} = 683$ stock-stock matches. These four numbers total the $n_1 = 2761$ matches.

Table 2: Who matches who?

	new	old	total
Vacancies			
zeros	38653	97419	136072
censored (last obs of spell is 0)	53	131	184
exits to new job seeker (last . . . 1)	420	191	611
exits to old job seeker (last . . . 1)	1467	683	2150
Total	40593	98424	139017
Unemployed			
zeros	125456	371736	497192
censored (last obs of spell is 0)	1550	3983	5533
exits to new vacancy (last . . . 1)	420	1467	1887
exits to old vacancy (last . . . 1)	191	683	874
Total	127617	377869	505486

All of the above is repeated for the $n_2 = 10416$ matches between unemployed job seekers and training vacancies.

5.3 Old and new stocks

During a given week $t-1$, there is an inflow v^+_{t-1} into stock of vacancies V_{t-1}, and an outflow v^-_{t-1}, such that the stock at the beginning of week t is given by:

$$V_t = V_{t-1} + (v^+_{t-1} - v^-_{t-1}). \tag{26}$$

This equation disaggregates into expressions for job vacancies and training vacancies:

$$J_t = J_{t-1} + (j^+_{t-1} - j^-_{t-1})$$
$$T_t = T_{t-1} + (t^+_{t-1} - t^-_{t-1}).$$

This is the familiar identity that the change in the stock equals the net inflow. The job vacancy outflow is decomposed into

$$j_t^- = m_t(U, J) + m_t(N, J) + m_t(Y, J) + m_t(F, J) + l_t(J)$$

where $l_t(J)$ is the number of job vacancies which are lapsed or whose spell is censored. This applies only to vacancies that are filled through LCS, and there is another expression for training vacancies T. The vacancy stock data are a stock sample. In other words, all the components of Equation (26) are observed in the LCS data.

Similarly, during week $t-1$, there is an inflow u_{t-1}^+ into stock of unemployed U_{t-1}, and an outflow u_{t-1}^-, such that

$$U_t = U_{t-1} + (u_{t-1}^+ - u_{t-1}^-) \tag{27}$$

The outflow is decomposed into

$$u_t^- = m_t(U, J) + m_t(U, J') + m_t(U, T) + m_t(U, T') + l_t(U)$$

where $l_t(U)$ is the number of unemployed who 'lapse' (exit the labour market) or whose spell is censored.

Unfortunately, the unemployment data are a flow sample, which means that U_t is not observed. However, we observe job-seeker data for about three years before the sample period, and so U_t is built up recursively from the net inflow into unemployment $u_t^+ - u_t^-$ each period. Given that week $t = 1$ is in April 1988, this means that U_{-30} is set to zero. Another implication of having a flow sample is that for the first year (1988–89), the stock only refers to new entrants onto the market, namely the cohort of Year 11 leavers in 1988 (hereafter the '1988 cohort'). This comprises mainly 16-year-olds. For the second year (1989–90) the stock refers to both the 1988 and 1989 cohorts (mainly 16 and 17 year-olds). In a sense this does not matter, as the stocks still correspond to the flows. In other words, in the first year, $m_t(U, J)$, J and U all refer to the 1988 cohort; in the second year, $m_t(U, J)$, J and U all refer to the 1988 and 1989 cohorts; and only in the third year will the data refer to everybody in the youth labour market. See Figure 1.

Alternative official sources of unemployment and vacancy stocks are available but cannot be disaggregated into old and new stocks.[15] When we plot the NOMIS U-stocks (16/17 year-olds, monthly) versus LCS U-stocks (observed daily, but plotted at monthly intervals) over time, we can see this effect, where they basically coincide from 1989–90 onwards (Figure 2).[16] The other noticeable thing is the very close correspondence, even at the

[15]As these are from the Online Information Service (NOMIS), they are referred to as NOMIS data (`http://www.nomisweb.co.uk`). They originate from the Office of National Statistics.

[16]NOMIS data refer to 16–17 year-olds and 18+ year-olds, and so we cannot create series for 16–18 year-olds.

end of the sample, where one might expect the recursive nature of measurement error to have its largest effect. This is convincing evidence that our stocks are extremely well measured, and of course the LCS data, being job-seeker based, reflect the large inflow of school-leavers onto the market between April and June each year. The NOMIS data, being claimant-based, miss this feature of the data.

Each stock can be disaggregated into 'old' and 'new' as follows, using the stock of unemployed for illustration:

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

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 $U_{t-1} - u^-_{t-1}|U_{t-1}$. Comparing with (27) above, $u^-_{t-1} \equiv u^-_{t-1}|u^+_{t-1} + u^-_{t-1}|U_{t-1}$, that is, all those who exit during week $t-1$ *must* either be from the inflow in the same week u^+_{t-1} or from the stock at the beginning of the week U_{t-1}. Because the data are weekly, clearly $k^w = 1$ week in this example, but the above expression generalises for any window size k:

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

Analogous expressions for job and training stocks also exist. Notice that we adopt a different terminology to Coles and Smith: we refer to their 'flow' u^k_t as 'new stock' and their 'stock' $\bar{U}^k_t$ as 'old stock', corresponding to 'old flows' and 'new flows' that have already been defined in Section 5.2 above.

5.4 Old and new (raw) hazards

The total outflow, over the whole sample period, from job vacancies is (2761 in the data)

$$\begin{aligned} n_1 &= m_{11} + m_{12} + m_{21} + m_{22} \\ &= \frac{m_{11}}{v} v + \frac{m_{12}}{\bar{J}} \bar{J} + \frac{m_{21}}{v} v + \frac{m_{22}}{\bar{J}} \bar{J} \\ &= h^e_{11} v + h^e_{12} \bar{J} + h^e_{21} v + h^e_{22} \bar{J} \end{aligned}$$

The stocks of $V \equiv v + \bar{J}$ and $U \equiv u + \bar{U}$ are calculated by counting the "at risk" total in the sequential binary response form (see Table 2). In fact, if one just counts the zeros, this is exactly the same number as the aggregate stocks over the whole sample period. Dividing by the number of periods (221 weeks) gives the average stock.

Hence the raw vacancy hazard to new unemployed job seekers is given by:

$$h_{11}^e = 420/40593 = 0.0103$$
$$h_{12}^e = 191/98424 = 0.00194$$

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

$$h_{21}^e = 1467/40593 = 0.0361$$
$$h_{22}^e = 683/98424 = 0.00694$$

Thus

> average stock of new job vacancies = 40593/211 = 184, and
> average stock of old job vacancies = 98424/221 = 445.

Notice that the drop in the hazard for vacancies matching with old unemployed job seekers is $h_{22}^e/h_{21}^e = \mu_{22}/\mu_{21} = 0.192$ is perfectly consistent with stock-flow matching.

The total outflow, over the whole sample period, from unemployed job seekers is the same number of matches (2761), but is a different expression

$$\begin{aligned} n_1 &= m_{11} + m_{12} + m_{21} + m_{22} \\ &= \frac{m_{11}}{u}u + \frac{m_{12}}{u}u + \frac{m_{21}}{\bar{U}}\bar{U} + \frac{m_{22}}{\bar{U}}\bar{U} \\ &= h_{11}^e u + h_{12}^e u + h_{21}^e \bar{U} + h_{22}^e \bar{U} \end{aligned}$$

Hence the raw unemployment hazard to new job vacancies is given by:

$$h_{11}^e = 420/127617 = 0.00329$$
$$h_{21}^e = 1467/377869 = 0.00388$$

and the raw unemployment hazard to the old job vacancies is given by:

$$h_{12}^e = 191/127617 = 0.00150$$
$$h_{22}^e = 683/377869 = 0.00181.$$

Thus

> average stock of new unemployed = 127617/221 = 577, and
> average stock of old unemployed = 377869/221 = 1710.

Here the drop in the hazard for unemployed matching with old vacancies is $h_{22}^w/h_{12}^w = \mu_{22}/\mu_{12} = 1.208$. This, of course, is not consistent with stock-flow matching. However, remember that this subsection simply illustrates the data for an arbitrarily chosen four week window.

5.5 Which window-size?

It is tempting to suggest that stock-stock matches should be less common than the other three types of match; we explained in Section 3 why this need not be so. This is clearly not true when $k^w = k^e = 4$, and so the first issue that needs to be resolved is how we choose values of k^w and k^e so that the stock-flow matching model is given the best chance to work. Note that none of Coles & Smith (1998), Gregg & Petrongolo (1997), Coles & Petrongolo (2001) have this problem as they use monthly aggregated time-series data.

In Figure 3, we plot the raw baseline hazards for all four hazards that we seek to estimate later, namely $h^w(U,J)$, $h^e(U,J)$, $h^w(U,T)$, and $h^e(U,T)$. Although the data are weekly, we group weeks together into the following intervals because estimation is much quicker and this never has any effect on the estimates of the covariates. The intervals are the same as Coles & Smith: (0,1], (1,2], (2,4], (4,6], (6,8], (8,13], (13,26], (26,39], (39,52], (52, ∞] weeks. Also drawn are the step-wise hazard functions calculated for a 4 week window in the previous subsection.[17]

For unemployment, there is clear evidence of non-monotonicity, with each hazard rising sharply to a peak at 5/6 weeks, and then declining gradually. We interpret the sharp increase as job-seekers learning to search (visiting Careers Offices, completing application forms, learning interview techniques and so on); the subsequent decline partly represents the usual duration dependence. In short, from the job-seeker hazards, there is little evidence that the hazard declines rapidly at very short durations. However, the job vacancy hazard is quite different and does exhibit a rapidly declining hazard. The YT vacancy hazard has the same shape as the two unemployed job-seeker hazards, which might well be consistent with the fact that this market is supply constrained whereas the jobs market is very much the reverse. Two conclusions emerge. First, the behaviour of the (secondary) training market is quite different form the (primary) jobs market and it is unlikely that stock-flow matching is the appropriate paradigm, even if we find evidence in the jobs market. Hereafter, we estimate models for U,T matches, but only report them in an Appendix for comparison with our main set of results.

Second, it is difficult to see in Figure 3 where the optimal window size is. Hence, in Figure 6, we plot the numbers of stock-stock, stock-flow, and flow-flow matches against window size, but keeping $k^w = k^e$. The argument here is that it is the same search technology being used on both sides of the market, which implies that the window should be the same. It should also be the same for the training vacancies market. It is obvious

[17]The original data are daily, and are plotted in Figure 4 for unemployment hazards and Figure 5 for vacancy hazards. We actually plot all four unemployment hazards because they form a complete set of competing risks for an unemployed job-seeker. However, we do not estimate full models for $h^w(U,J')$ and $h^w(U,T')$ as J' and T' are unobserved.

that the number of flow-flow matches *must* increase 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 $m_{12} + m_{21}$ monotonically increases with window size, and then decreases monotonically. The fact the number of stock-flow matches is largest when the window size is about one month suggests that a useful starting place is to choose $k^w = k^e = 4$ (which, coincidentally, is the window size that Coles and Smith are restricted to in their data).

We experimented with various quasi-formal methods for trying to find optimal values of $k^e = k^w \neq 4$, by searching over other integer values of k. For example, two simple regressions reproduce the figures given in the two crosstabs in Table 2 and so we looked for the k that maximised their log-likelihood. Another technique was to choose k that maximised the drop in the old hazard for unemployed job seekers [resp job vacancies] when exiting to old job vacancies [resp old job seekers], ie jointly minimised h^e_{22}/h^e_{21} and h^w_{22}/h^w_{12}. None of these methods led to a consistent answer, and so our conclusion is that this search for the optimal window size is a chimera, and the appropriate strategy is to choose a small number of (k^w, k^e) pairs to see whether it makes any differences to the regression analyses, hazards, etc. In the current version of the paper, we only report results for $k^e = k^w = 4$ weeks (but see the Appendix for what happens when $k^e = k^w = 1$).

5.6 Size of labour market

The data cover the whole of Lancashire, a county in the United Kingdom that comprises 14 geographically distinct towns/cities (in fact, local authority districts, or LADs). The issue here is whether the stocks should vary by these 14 districts, being distinct labour markets, or whether the same value should be used irrespective of where in Lancashire the match takes place, or something in between. For the intermediate case, we grouped Lancashire into just 3 labour markets (West, Central and East), recognising that job-seekers can travel between certain towns when looking for work. When we specify just three “districts” in Lancashire, 96% of all matches take place between an unemployed job seeker and job vacancy from the same district. This number drops to 75% when Lancashire is treated as 14 LADs, which is convincing evidence that the 3 district specification is the best one. Throughout Huber/White standard errors correct for within labour-market correlations between job-seekers/vacancies. This also why we reparameterised the model so that our test of stock-flow matching is based on Wald rather than LR type tests.

Figures 7 and 8 plot old and new stocks of job vacancies and unemployed job seekers for the 3 LADs. As would be expected, the plots of new unemployed stocks is much more stationary than the old stock; the same is true for vacancy stocks. It is clear that the peaks in both new and old unemployed arise from young people leaving school in May/June each year (the so-called recruitment cycle). The seasonal pattern in vacancies is similar, but nowhere as pronounced, although it is noticeable that the stock of new vacancies tends to precede the months when school-leavers actually leave school.

6 Econometric methodology

The hazard for each week s and for each job-seeker i is modelled as follows. We assume proportional hazards and introduce a positive-valued random variable (or mixture) ϵ:

$$h_s^w(U_{is}, J_{is}, \epsilon_i^w) = \bar{h}_s^w \epsilon_i^w \exp(\mathbf{x}_{is}' \boldsymbol{\beta}^w)$$

$\bar{h}_s^w$ is the baseline hazard, and does not vary by i. $\varepsilon_i^w \equiv \log \epsilon_i^w$ has density $f_\varepsilon^w(\varepsilon^w)$, and is a job-seeker specific random effect. There are identical expressions for vacancy hazards, but with superscript e.

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

$$\begin{aligned} L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}) &= \int_{-\infty}^{\infty} \left[\prod_{s=1}^{t_i} h_s(\mathbf{x}_{is}', \varepsilon_i)^{y_{it}} [1 - h_s(\mathbf{x}_{is}', \varepsilon_i)]^{1-y_{it}} \right] f_\varepsilon(\varepsilon_i) d\varepsilon_i, \\ h_s(\mathbf{x}_{is}', \varepsilon_i) &= 1 - \exp[-\exp(\mathbf{x}_{is}' \boldsymbol{\beta} + \gamma_s + \varepsilon_i)]. \end{aligned}$$

where, for notational clarity, we have suppressed the superscript w, and so the same equation also applies to the employer hazard. Because of the proportional hazards assumption, the covariates affect the hazard via the complementary log-log link. The γ_ss 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 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) (see Section 5.5 above). This is because estimating models with unobserved heterogeneity proved to be too demanding of the data.[18] In the current set of results we use Gaussian mixing, with variance σ^2.[19]

The specification for $\mathbf{x}_{is}'$ was discussed at length in Section 3. To recap, we define a dummy variable for whether the spell index s is less than the window size $1(s \leq k)$, and its complement $1(s > k)$. This is then interacted with the covariates.

[18] We have investigated the effect of using weeks rather than days. The aggregation bias is minimal and models with daily baseline hazards simply cannot be estimated with unobserved heterogeneity.

[19] We have estimated models with non-parametric Heckman-Singer hazards, but the results are very similar. Gaussian mixing is much quicker to converge.

$$1(s \leq k) \log U,\ 1(s \leq k) \log u,\ 1(s \leq k) \log J,\ 1(s \leq k) \log j$$
$$1(s > k) \log U,\ 1(s > k) \log u,\ 1(s > k) \log J,\ 1(s > k) \log j$$

It is worth emphasising that both stocks vary by duration s and job-seeker/vacancy i, because they vary through calendar time and because each job-seeker/vacancy enters the market place at different calendar times. As just noted, instead of having just two dummies for the baseline hazard $1(s \leq k)$ and $1(s > k)$, we estimate the unrestricted version just discussed.

Temporal aggregation bias is an important issue in this literature, and is discussed at length by Burdett, Coles & van Ours (1994), Gregg & Petrongolo (1997) and Coles & Petrongolo (2001). 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 (2001) estimate matching functions using a quite sophisticated technique that deals with this problem. In our data this will not be problem as we observe weekly flows and stocks that also vary weekly; had we used daily stocks, the issue would completely disappear.[20]

7 Results

In Table 3 we report estimates of the three basic specifications, namely random matching and two types of random matching; the top block of three is without heterogeneity and the bottom block is with. The non-random matching model is reported in the first panel of both blocks. We interpret the results in the context of the statistical model developed in Section 3—see Equation (20) in particular. The implied estimates of α, β and the μ-ratios are also reported.

Looking at the models without unobserved heterogeneity, the first finding is that $\log u$ is not significant in the old job-seeker hazard, nor is $\log v$ significant in the old vacancy hazard. Both variables are significant is the new hazards. Thus our first specification test of the statistical model we have adopted in partially successful, and suggests that the appropriate non-random matching model is not one that drops all four of these variables.

In terms of classical matching elasticities α and β, the estimates are generally sensible, but showing a slight, but significant, degree of increasing returns to scale; this is particularly

[20] In future regressions, we will use daily data. Moreover, our value-added is that we can 'test' the procedures proposed by Coles & Petrongolo (2001) by aggregating the data into months, and thereby quantify the size of the bias.

strong for matches involving old vacancies using vacancy duration data, with $\widehat{\alpha} + \widehat{\beta} = 1.575$, although the standard error is bigger too.

Our simple way of testing for stock-flow matching is to see whether an increase in the number of new unemployed job vacancies [resp job seekers] significantly increases the exit probability for old unemployed job seekers [resp vacancies]. In the old job seeker hazard, $\frac{\partial \log h^w_{2.}}{\partial \log j} = 0.141$, but is not significant. This converts to a point estimate for $\mu_{22}/\mu_{21} = 0.621$, but one whose 95% confidence interval is sufficiently wide that it contains unity. In the old vacancy hazard, $\frac{\partial \log h^e_{.2}}{\partial \log u} = 0.088$, but this time it is significant. Said differently, the implied point estimate of $\mu_{22}/\mu_{12} = 0.724$ has a confidence interval that does *not* contain unity.

These results provide some evidence that the hazards drop slightly when a job seeker or vacancy becomes old when matching to an old agent on the other side of the market. The estimates are not precisely estimated, because this test relies on correlations between the stocks of market participants and the number of individual-level matches. The old and new stocks in our data are basically three time-series for each stock, one for the three districts in Lancashire — there is little cross-section variation in the data. However, the time-series variation is considerable because of the so-called recruitment cycle (Figures 7 and 8).

Table 3: Estimated hazards for unemployed job-seekers and job vacancies, non-random and random matching models with and without unobserved heterogeneity, 4-4 window*

	job-seeker, h^w		vacancies, h^e		
	new, $h^w_{1.}$	old, $h^w_{2.}$	new, $h^e_{.1}$	old, $h^e_{.2}$	mean[a]
Without unobserved heterogeneity					
$\log u$	-0.118 (0.055)	-0.092 (0.102)	-0.140 (0.079)	0.088 (0.040)	192
$\log U$	-0.169 (0.030)	-0.299 (0.130)	0.683 (0.138)	0.796 (0.279)	759
$\log j$	0.415 (0.157)	0.141 (0.101)	-0.301 (0.098)	-0.029 (0.188)	58
$\log J$	0.057 (0.234)	0.381 (0.075)	-0.072 (0.101)	-0.280 (0.244)	216
α, β	0.713, 0.472	0.609, 0.522	0.543, 0.627	0.884, 0.691	
$\alpha+\beta$	1.185 (0.079)	1.131 (0.049)	1.170 (0.023)	1.575 (0.229)	
	$\mu_{12}/\mu_{11} = 0.275$	$\mu_{22}/\mu_{21} = 0.621$	$\mu_{21}/\mu_{11} = 1.944$	$\mu_{22}/\mu_{12} = 0.724$[b]	
	[0.008]	[0.163]	[0.077]	[0.029]	
Log likelihood	-16741.0		-12115.1		
$\log U$	-0.281 (0.057)	-0.371 (0.102)	0.544 (0.081)	0.892 (0.273)	759
$\log J$	0.336 (0.127)	0.480 (0.040)	-0.281 (0.066)	-0.302 (0.189)	216
$\alpha+\beta$	1.055 (0.101)	1.108 (0.064)	1.264 (0.024)	1.590 (0.210)	
Log likelihood	-16754.1		-12137.1		
$\log U$	-0.350 (0.086)		0.641 (0.130)		759
$\log J$	0.451 (0.059)		-0.289 (0.089)		216
$\alpha+\beta$	1.100 (0.050)		1.353 (0.063)		
Log likelihood	-16756.5		-12150.4		
With unobserved heterogeneity					
$\log u$	-0.113 (0.060)	-0.086 (0.048)	-0.234 (0.058)	-0.215 (0.061)	192
$\log U$	-0.176 (0.092)	-0.321 (0.057)	0.940 (0.078)	1.383 (0.102)	759
$\log j$	0.424 (0.111)	0.146 (0.055)	-0.181 (0.082)	-0.084 (0.100)	58
$\log J$	0.048 (0.099)	0.393 (0.051)	-0.170 (0.077)	-0.299 (0.105)	216
Variance (σ^2)	0.415 (0.096)		3.954 (0.297)		
α, β	0.711, 0.472	0.593, 0.539	0.706, 0.649	1.168, 0.617	
$\alpha+\beta$	1.184 (0.080)	1.132 (0.048)	1.355 (0.069)	1.785 (0.090)	
	$\mu_{12}/\mu_{11} = 0.267$	$\mu_{22}/\mu_{21} = 0.611$	$\mu_{21}/\mu_{11} = 3.994$	$\mu_{22}/\mu_{12} = 3.328$	
	[0.000]	[0.008]	[0.000]	[0.000]	
Log likelihood	-16730.1		-11619.2		
$\log U$	-0.283 (0.059)	-0.388 (0.038)	0.682 (0.052)	1.123 (0.070)	759
$\log J$	0.335 (0.063)	0.494 (0.034)	-0.283 (0.052)	-0.356 (0.073)	216
$\alpha+\beta$	1.052 (0.073)	1.106 (0.046)	1.399 (0.063)	1.768 (0.087)	
Variance (σ^2)	0.414 (0.096)		3.832 (0.293)		
Log likelihood	-16743.2		-11637.4		
$\log U$	-0.363 (0.032)		0.842 (0.047)		759
$\log J$	0.461 (0.030)		-0.314 (0.047)		216
$\alpha+\beta$	1.098 (0.039)		1.528 (0.058)		
Variance (σ^2)	0.400 (0.095)		3.847 (0.294)		
Log likelihood	-16746.2		-11652.7		
Observations	505486		139017		

*Estimates based on 2761 matches (1887 to new vacancies and 874 to old vacancies, 611 to new unemployed and 2150 to old unemployed) between 34659 unemployed job-seeker spells (26114 job-seekers) and 14148 LCS job vacancies (9555 orders).

[a] Unlogged means are not the same as in Section 5.4 as they are weighted averages across 3 LADs.

[b] The μ-ratios calculated from Equation (21) and analagous expressions. We do not report standard errors, as the μ-ratios are not Normally distributed. By definition, p-values are the same as for underlying parameter estimates.

It is worth emphasising that our test has nothing to do with shape of agents' baseline hazards. We think that this is a correct test of stock-flow matching for the following reason. One can conceive of the data being generated in one of two ways. First, the four μs are the same (random matching) in which case the estimates of μ_{22}/μ_{21} and μ_{22}/μ_{12} would both be insignificantly different from unity and the estimated hazards would be flat. The second possibility is where the true μ_{22} is much lower than the other 3 μs (stock-flow matching), in which case the two tests would be rejected and the hazards would drop when the agents become old. If we observe non-flat hazards in the data, but the tests are not rejected, it must be that the hazards are not flat for other reasons (duration dependence, unobserved heterogeneity, institutional features such as benefits). This is why the estimates are different from the raw baseline hazards earlier (Figure 3), where, recall, $\widehat{\mu_{22}/\mu_{12}} = 1.208$ and $\widehat{\mu_{22}/\mu_{21}} = 0.192$.

To investigate this further, we re-estimated the models using Gaussian unobserved heterogeneity, which are reported in the bottom half of Table 3. The effect on the job-seeker hazards is minimal, but is quite strong on the other side of the market. Repeating the above calculations shows that there is no longer evidence of stock-flow matching, $\widehat{\mu_{22}/\mu_{12}} = 3.328$, and in fact one rejects $H_1 : \mu_{12} = \mu_{22}$ in favour of μ_{22} being bigger, not smaller, than μ_{12}. Another effect of modelling the unobserved heterogeneity is that now the baseline hazard is flatter (Figure 9), which is consistent with the movement in the estimate of μ_{22}/μ_{12} between the models with and without heterogeneity, and also suggests that the sharp fall in the vacancy hazards in the first month is due to unobservables and *not* stock-flow matching (Figure 5). However, more needs doing here, as we have a rich set of covariates from both sides of the market that might be added to these regressions.

Also notice that our data come from the different sides of the same market, whose only relationship with each other is that the number of exits coincide. So do the results concur? The slightly disappointing conclusion is perhaps not: the matching elasticities tend to be bigger when using vacancy data. Moreover, one can obtain estimates of any μ-ratio from both sides of the market. For example, from the top block of Table 3 we can get two different estimates of μ_{22}/μ_{11}

$$\widehat{\mu_{22}/\mu_{11}} = (\widehat{\mu_{12}/\mu_{11}})(\widehat{\mu_{22}/\mu_{12}}) = 0.275 * 0.724 = 0.199$$
$$\widehat{\mu_{22}/\mu_{11}} = (\widehat{\mu_{22}/\mu_{21}})(\widehat{\mu_{21}/\mu_{11}}) = 0.621 * 1.944 = 1.207.$$

It looks as if the two estimates are different, although it is difficult to actually test whether this is so (a bit a like a cross-equation in simultaneous equations models). The same is repeated for the bottom block of Table 3:

$$\widehat{\mu_{22}/\mu_{11}} = (\widehat{\mu_{12}/\mu_{11}})(\widehat{\mu_{22}/\mu_{12}}) = 0.267 * 3.328 = 0.889$$
$$\widehat{\mu_{22}/\mu_{11}} = (\widehat{\mu_{22}/\mu_{21}})(\widehat{\mu_{21}/\mu_{11}}) = 0.611 * 3.994 = 0.970.$$

The second and third panels in Table 3 report estimates of the classical random matching model, all of which, look perfectly consistent with the existing literature (except for the increasing returns to scale). They obviously do not differ much from the corresponding stock-flow matching models as we only find weak evidence of favour of the latter.

Finally, in the Appendix, we report corresponding estimates for a 1 week window. By definition, the number of matches involving new agents must fall (Figure 6) as do the old stocks $\bar{U}$ and $\bar{V}$. The results are not at all convincing, with standard errors much higher compared with the 4-week windows, and the corresponding point estimates are therefore less plausible. In particular, the estimates on the μ ratios from the vacancy hazards are particularly disappointing.

8 Conclusion

In this paper we report preliminary estimates of job-seeker and employer 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 the classical random matching 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 suggests that job seekers find it harder to match to old vacancies once they become old themselves. Exactly the reverse applies to the old vacancy hazard, where the test examines the effect of the stock of new job seekers. The test does not examine whether the vacancy hazard or job seeker hazards fall at certain durations, because this can happen for other reasons.

Our tentative results find very weak evidence of stock-flow (or non-random) matching.

References

Anderson, P. & Burgess, S. (1997), Empirical matching functions: estimation and interpretation using disaggregate data, Working Paper 5001, National Bureau of Economic Research, February.

Andrews, M., Bradley, S. & Stott, D. (2002), 'Matching the demand for and supply of training in the school-to-work transition', *Ecconomic Journal* **112**, C201–19.

Andrews, M., Bradley, S. & Upward, R. (2001*a*), Employer search, vacancy duration, and skill shortages: an analysis of vacancies in the youth labour market, Discussion paper, University of Nottingham, June.

Andrews, M., Bradley, S. & Upward, R. (2001*b*), 'Estimating the probability of a match using micro-economic data for the youth labour market', *Labour Economics* **8**, 335–57.

Barron, J., Berger, M. & Black, D. (1997), Employer search, training and vacancy duration. Upjohn Institute for Employment Research, Kalamazoo: Michigan.

Blanchard, O. & Diamond, P. (1989), 'The Beveridge Curve', *Brookings Papers on Economic Activity* **1**, 1–60.

Blanchard, O. & Diamond, P. (1992), 'The flows approach to labor markets', *American Economic Review* **82**, 354–59.

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. & Cunningham, E. (1998), 'Toward a theory of vacancies', *Journal of Labor Economics* **16**, 445–78.

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

Coles, M. & Petrongolo, B. (2001), A test between unemployment theories using matching data. University of Essex, ILR, Discussion Paper No. 01/64.

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

Davis, S. & Haltiwanger, J. (1999), Gross job flows, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3B, Elsevier, Amsterdam, chapter 41, pp. 2711–805.

Gregg, P. & Petrongolo, B. (1997), Random or non-random matching? Implications for the use of the UV curve as a measure of matching effectiveness. CEP Discussion Paper No. 348.

Jackman, R., Layard, R. & Pissarides, C. (1989), 'On vacancies', *Oxford Bulletin of Economics and Statistics* **51**, 377–94.

Lancaster, T. (1979), 'Econometric methods for the duration of unemployment', *Econometrica* **47**, 939–56.

Lindeboom, M., van Ours, J. & Renes, G. (1994), 'Matching employers and workers: an empirical analysis of the effectiveness of search', *Oxford Economic Papers* **46**, 45–67.

Manning, A. (2001), Monopsony and the efficiency of labour-market interventions. CEP Discussion Paper No. 514.

Marimon, R. & Zilibotti, F. (2000), 'Employment and distributional effects of restricting working time', *European Economic Review* **44**, 1291–1326.

Meyer, D. (1990), 'Unemployment insurance and unemployment spells', *Econometrica* **58**, 757–82.

Mortensen, D. & Pissarides, C. (1998), Job reallocation, employment fluctuations and unemployment differences, *in* M. Woodford & J. Taylor, eds, 'Handbook of Macroeconomics', North-Holland, Amsterdam, chapter 18, pp. 1171–228.

Mortensen, D. & Pissarides, C. (1999), New developments in models of search in the labor market, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3B, Elsevier, Amsterdam, chapter 39, pp. 2567–627.

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

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

Russo, G. & van Ommeren, J. (1998), 'Recruitment methods and vacancy duration', *Bulletin of Economic Research* **50**, 155–66.

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

Teyssière, G. (1996), 'Matching processes in the labour market: an econometric study', *Labour Economics* **2**, 421–35.

van den Berg, G. (1990), 'Search behaviour, transitions to non-participation and the duration of unemployment', *Economic Journal* **100**, 842–865.

van den Berg, G. (1999), 'Empirical inference with equilibrium search models of the labour market', *Economic Journal* **109**, F283–306.

van Ours, J. (1990), An empirical analysis of employers' search, *in* J. Hartog, G. Ridder & J. Theeuwes, eds, 'Panel Data and Labor Market Studies', North-Holland, Amsterdam, pp. 191–214.

van Ours, J. & Lindeboom, M. (1996), "Seek and ye shall find": an empirical analysis of the matching of job seekers and vacancies. Paper presented at Labour Market Changes and Income Dynamics conference.

van Ours, J. & Ridder, G. (1991), 'Cyclical variations in vacancy durations and vacancy flows: an empirical analysis', *European Economic Review* **35**, 1143–1155.

van Ours, J. & Ridder, G. (1992), 'Vacancies and the recruitment of new employees', *Journal of Labor Economics* **10**, 138–155.

van Ours, J. & Ridder, G. (1993), 'Vacancy durations: search or selection?', *Oxford Bulletin of Economics and Statistics* **55**, 187–198.

Appendix A

In this appendix we report what happens when the matching window is reduced to one week from four, and the corresponding regressions for matches between unemployed job-seekers and training vacancies.

Table A.1: Estimated hazards for unemployed job-seekers and YT vacancies, non-random and random matching models with and without unobserved heterogeneity, 4-4 window*

	job-seeker, h^w		vacancies, h^e		
	new, $h^w_{1.}$	old, $h^w_{2.}$	new, $h^e_{.1}$	old, $h^e_{.2}$	mean[a]
Without unobserved heterogeneity					
$\log u$	0.479 (0.074)	-0.007 (0.121)	0.437 (0.342)	-0.149 (0.080)	192
$\log U$	-0.656 (0.094)	0.046 (0.180)	0.689 (0.470)	0.981 (0.107)	759
$\log t$	-0.009 (0.056)	-0.027 (0.012)	-0.126 (0.102)	0.026 (0.051)	206
$\log T$	0.779 (0.338)	0.906 (0.192)	0.134 (0.358)	-0.003 (0.245)	1486
α, β	0.823, 0.770	1.039, 0.879	1.126, 1.008	0.832, 1.023	
$\alpha+\beta$	1.592 (0.327)	1.918 (0.226)	2.134 (0.485)	1.854 (0.276)	
	$\mu_{12}/\mu_{11} = 1.069$	$\mu_{22}/\mu_{21} = 1.234$	$\mu_{21}/\mu_{11} = 0.246$	$\mu_{22}/\mu_{12} = 2.052$[b]	
	[0.866]	[0.030]	[0.202]	[0.064]	
Log likelihood	-48248.6		-54117.5		
$\log U$	-0.141 (0.062)	0.027 (0.065)	1.247 (0.112)	0.855 (0.075)	759
$\log T$	1.145 (0.338)	0.860 (0.190)	0.266 (0.365)	-0.028 (0.260)	1486
$\alpha+\beta$	2.005 (0.391)	1.888 (0.211)	2.512 (0.476)	1.826 (0.254)	
Log likelihood	-48376.2		-54190.2		
$\log U$	-0.028 (0.050)		0.936 (0.054)		759
$\log T$	0.938 (0.217)		0.057 (0.298)		1486
$\alpha+\beta$	1.910 (0.258)		1.993 (0.314)		
Log likelihood	-48392.3		-54258.2		
With unobserved heterogeneity					
$\log u$	0.479 (0.033)	-0.007 (0.025)	0.001 (0.051)	-0.308 (0.021)	192
$\log U$	-0.656 (0.045)	0.046 (0.030)	1.290 (0.078)	1.648 (0.037)	759
$\log t$	-0.009 (0.012)	-0.027 (0.008)	-0.112 (0.027)	0.037 (0.008)	206
$\log T$	0.779 (0.062)	0.906 (0.034)	0.411 (0.095)	-0.006 (0.057)	1486
Variance (σ^2)	0.000 (0.001)		1.274 (0.051)		
α, β	0.823, 0.770	1.039, 0.879	1.291, 1.299	1.340, 1.031	
$\alpha+\beta$	1.592 (0.056)	1.918 (0.033)	2.590 (0.086)	2.371 (0.051)	
	$\mu_{12}/\mu_{11} = 1.069$	$\mu_{22}/\mu_{21} = 1.234$	$\mu_{21}/\mu_{11} = 0.996$	$\mu_{22}/\mu_{12} = 14.463$	
	[0.428]	[0.000]	[0.991]	[0.000]	
Log likelihood	-48248.6		-51797.0		
$\log U$	-0.141 (0.029)	0.026 (0.022)	1.211 (0.050)	1.301 (0.028)	759
$\log T$	1.146 (0.052)	0.865 (0.032)	0.278 (0.075)	0.089 (0.051)	1486
$\alpha+\beta$	2.005 (0.050)	1.891 (0.032)	2.488 (0.072)	2.390 (0.048)	
Variance (σ^2)	0.016 (0.034)		1.285 (0.052)		
Log likelihood	-48376.1		-51904.9		
$\log U$	-0.030 (0.018)		1.282 (0.026)		759
$\log T$	0.946 (0.028)		0.135 (0.046)		1486
$\alpha+\beta$	1.915 (0.027)		2.417 (0.042)		
Variance (σ^2)	0.034 (0.035)		1.281 (0.053)		
Log likelihood	-48391.8		-51908.2		
Observations	505486		962263		

*Estimates based on 10416 matches (1746 to new vacancies and 8670 to old vacancies, 3023 to new unemployed and 7393 to old unemployed) between 34659 unemployed job-seeker spells (26114 job-seekers) and 36853 LCS YT vacancies (4346 orders).

[a] Unlogged means are not the same as in Section 5.4 as they are weighted averages across 3 LADs.

[b] The μ-ratios calculated from Equation (21) and analagous expressions. We do not report standard errors, as the μ-ratios are not Normally distributed. By definition, p-values are the same as for underlying parameter estimates.

Table A.2: Estimated hazards for unemployed job-seekers and job vacancies, non-random and random matching models with and without unobserved heterogeneity, 1-1 window*

	job-seeker, h^w		vacancies, h^e		
	new, $h^w_{1.}$	old, $h^w_{2.}$	new, $h^e_{.1}$	old, $h^e_{.2}$	mean[a]
Without unobserved heterogeneity					
$\log u$	-0.237 (0.125)	-0.102 (0.093)	-0.180 (0.088)	-0.054 (0.106)	52
$\log U$	0.342 (0.170)	-0.313 (0.129)	0.662 (0.120)	0.752 (0.255)	759
$\log j$	0.270 (0.126)	0.214 (0.084)	-0.004 (0.062)	-0.063 (0.082)	19
$\log J$	-0.314 (0.218)	0.348 (0.092)	-0.271 (0.044)	-0.254 (0.150)	216
α, β	1.105, -0.044	0.585, 0.562	0.482, 0.725	0.698, 0.683	
$\alpha+\beta$	1.061 (0.125)	1.148 (0.028)	1.207 (0.086)	1.381 (0.088)	
	$\mu_{12}/\mu_{11}=0.192$	$\mu_{22}/\mu_{21}=0.244$	$\mu_{21}/\mu_{11}=-0.815$	$\mu_{22}/\mu_{12}=3.965$[b]	
	[0.032]	[0.010]	[0.041]	[0.609]	
Log likelihood	-16719.4		-12136.7		
$\log U$	0.159 (0.099)	-0.371 (0.094)	0.526 (0.069)	0.702 (0.183)	759
$\log J$	-0.213 (0.192)	0.469 (0.058)	-0.278 (0.024)	-0.293 (0.124)	216
$\alpha+\beta$	0.946 (0.112)	1.098 (0.055)	1.248 (0.058)	1.409 (0.079)	
Log likelihood	-16745.0		-12146.5		
$\log U$	-0.350 (0.086)		0.641 (0.130)		759
$\log J$	0.451 (0.059)		-0.289 (0.089)		216
$\alpha+\beta$	1.100 (0.050)		1.353 (0.063)		
Log likelihood	-16756.5		-12150.4		
With unobserved heterogeneity					
$\log u$	-0.236 (0.136)	-0.100 (0.032)	-0.209 (0.066)	-0.194 (0.038)	52
$\log U$	0.340 (0.193)	-0.328 (0.041)	0.830 (0.086)	1.103 (0.062)	759
$\log j$	0.270 (0.215)	0.217 (0.034)	0.012 (0.075)	-0.007 (0.046)	19
$\log J$	-0.315 (0.202)	0.359 (0.036)	-0.294 (0.078)	-0.327 (0.062)	216
Variance (σ^2)	0.410 (0.095)		3.977 (0.300)		
α, β	1.104, -0.045	0.572, 0.576	0.621, 0.718	0.909, 0.666	
$\alpha+\beta$	1.059 (0.245)	1.147 (0.042)	1.339 (0.092)	1.575 (0.066)	
	$\mu_{12}/\mu_{11}=0.192$	$\mu_{22}/\mu_{21}=0.241$	$\mu_{21}/\mu_{11}=-0.656$	$\mu_{22}/\mu_{12}=-0.729$	
	[0.209]	[0.000]	[0.002]	[0.000]	
Log likelihood	-16708.6		-11627.4		
$\log U$	0.158 (0.161)	-0.386 (0.033)	0.643 (0.069)	0.928 (0.052)	759
$\log J$	-0.213 (0.168)	0.481 (0.030)	-0.279 (0.066)	-0.326 (0.054)	216
$\alpha+\beta$	0.945 (0.220)	1.095 (0.040)	1.364 (0.084)	1.602 (0.065)	
Variance (σ^2)	0.410 (0.095)		3.810 (0.290)		
Log likelihood	-16734.2		-11645.8		
$\log U$	-0.363 (0.032)		0.842 (0.047)		759
$\log J$	0.461 (0.030)		-0.314 (0.047)		216
$\alpha+\beta$	1.098 (0.039)		1.528 (0.058)		
Variance (σ^2)	0.400 (0.095)		3.847 (0.294)		
Log likelihood	-16746.2		-11652.7		
Observations	505486		139017		

*Estimates based on 2761 matches (888 to new vacancies and 1873 to old vacancies, 75 to new unemployed and 2686 to old unemployed) between 34659 unemployed job-seeker spells (26114 job-seekers) and 14148 LCS job vacancies (9555 orders).

[a]Unlogged means are not the same as in Section 5.4 as they are weighted averages across 3 LADs.

[b]The μ-ratios calculated from Equation (21) and analagous expressions. We do not report standard errors, as the μ-ratios are not Normally distributed. By definition, p-values are the same as for underlying parameter estimates.

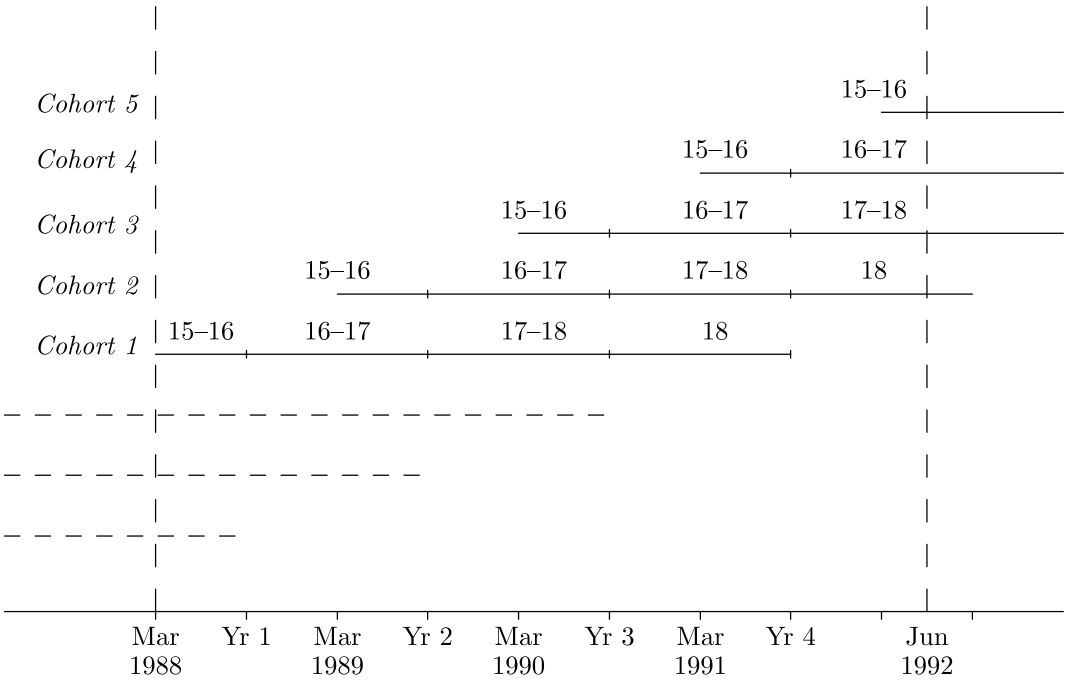

Figure 1: The job-seeker data are a flow sample

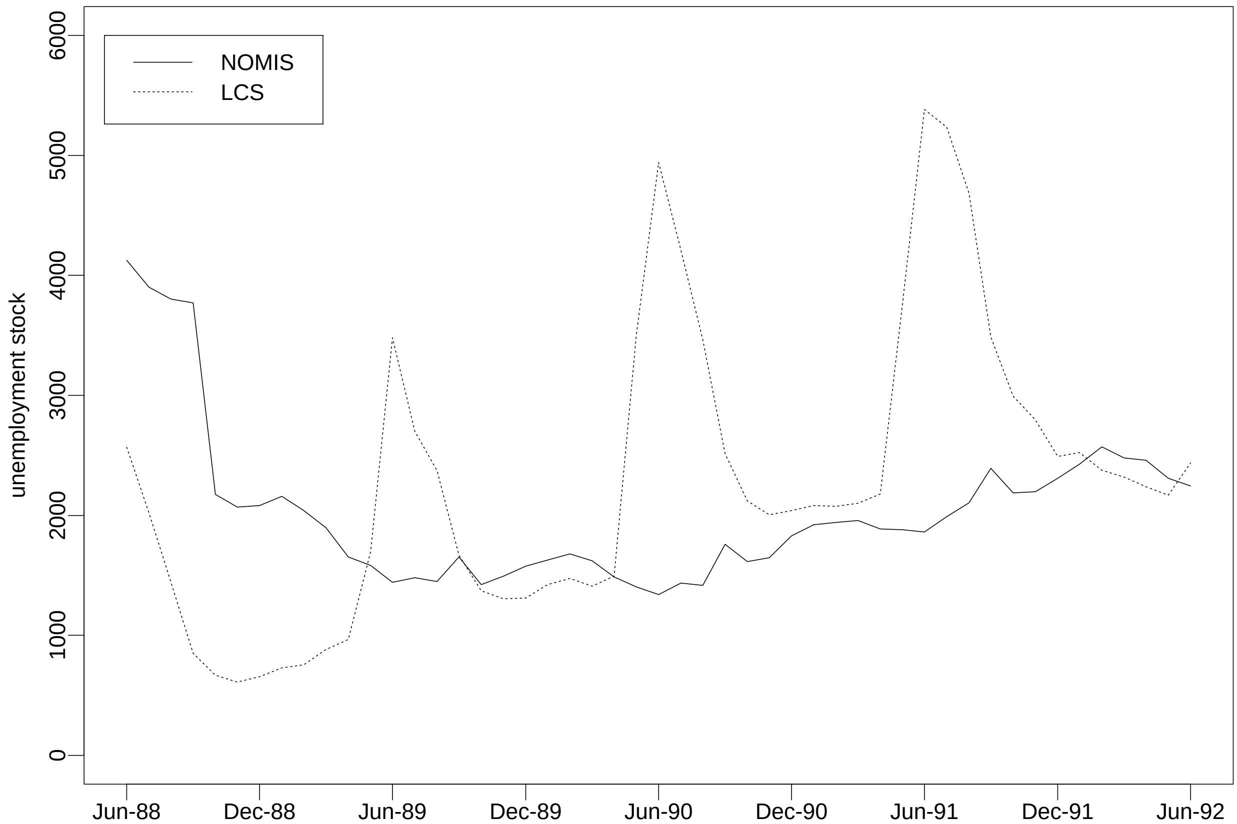

Figure 2: NOMIS and LCS unemployment stocks for 16 and 17 year-olds

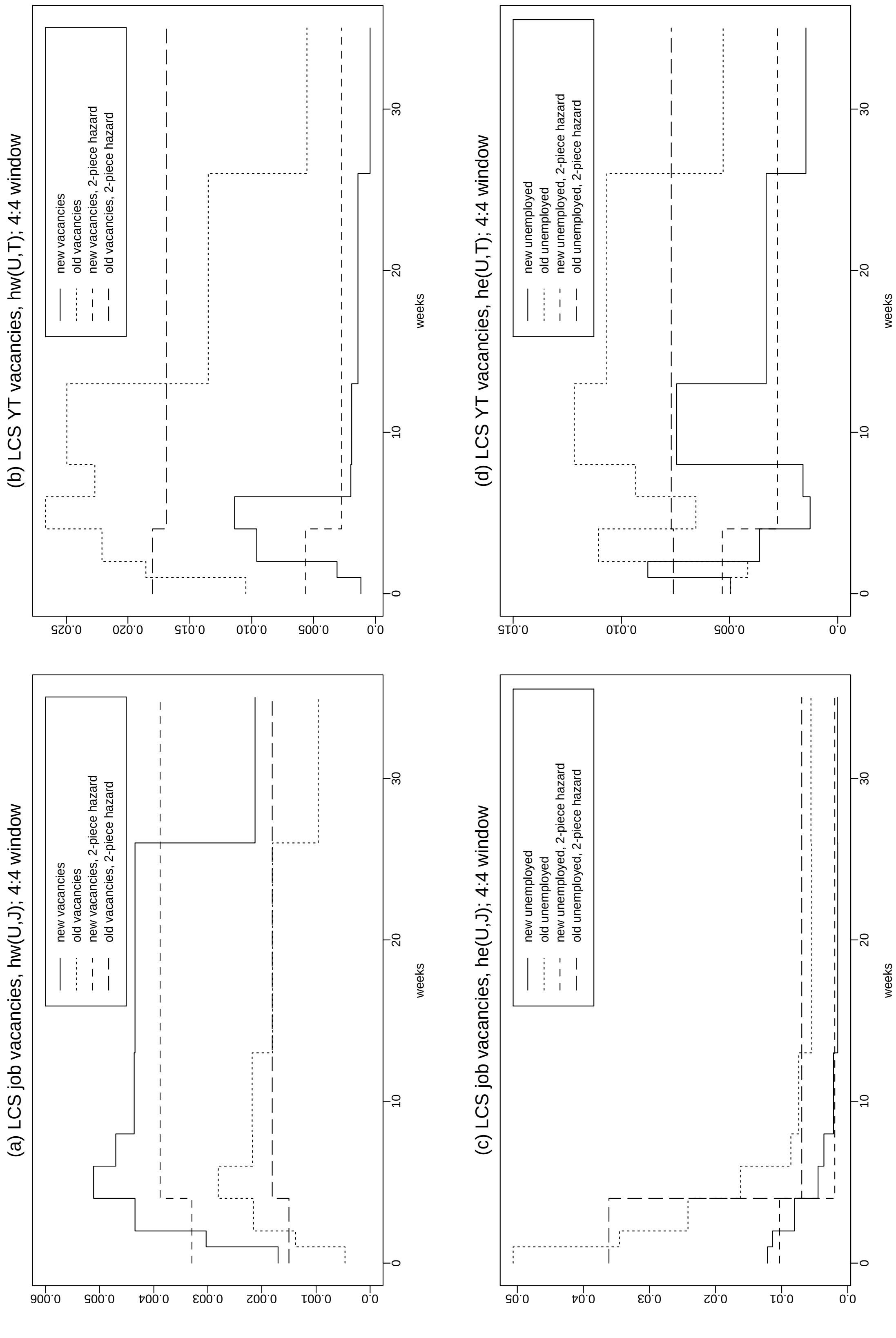

Figure 3: Raw unemployment and vacancy hazards split by old and new

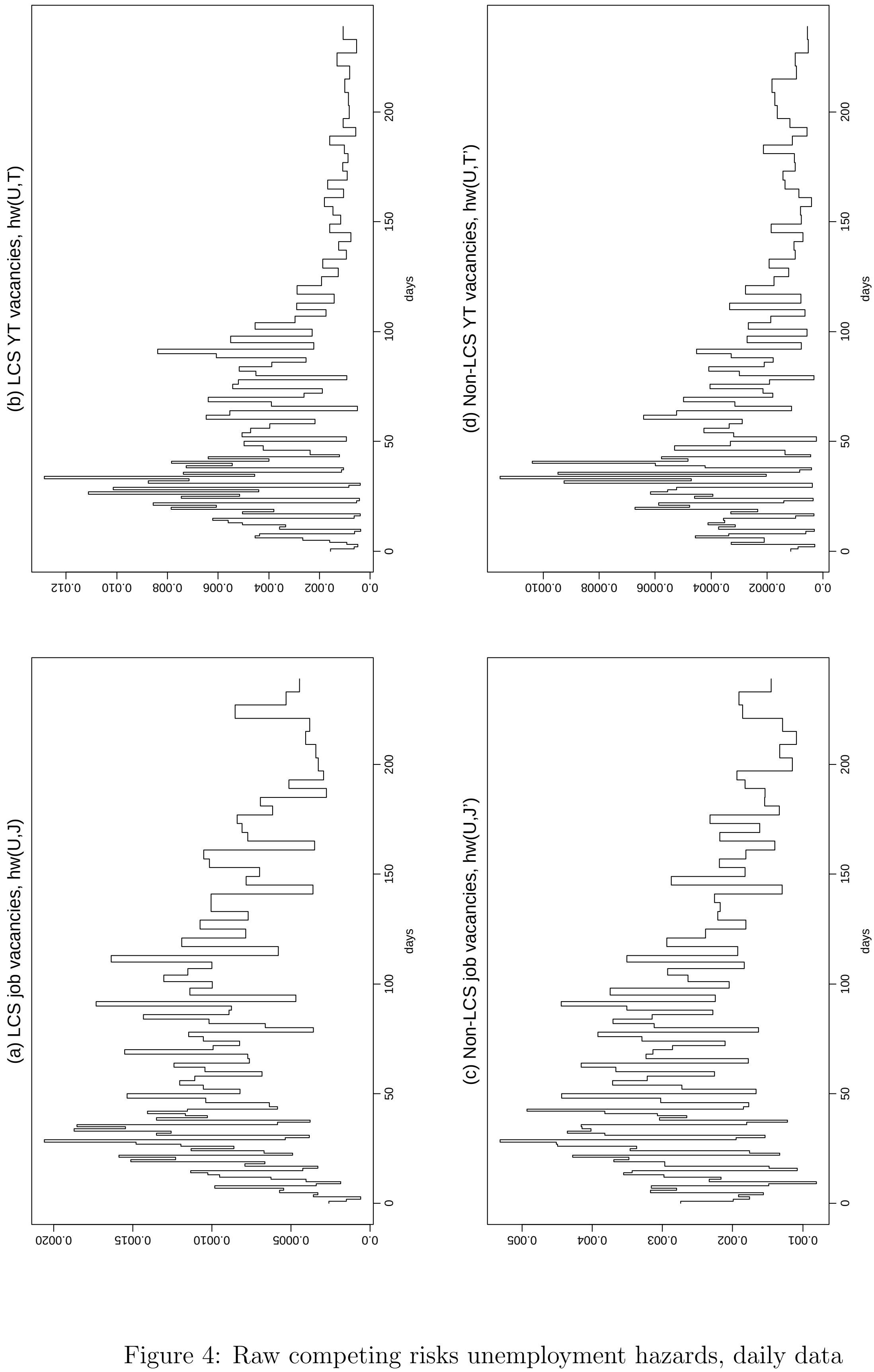

Figure 4: Raw competing risks unemployment hazards, daily data

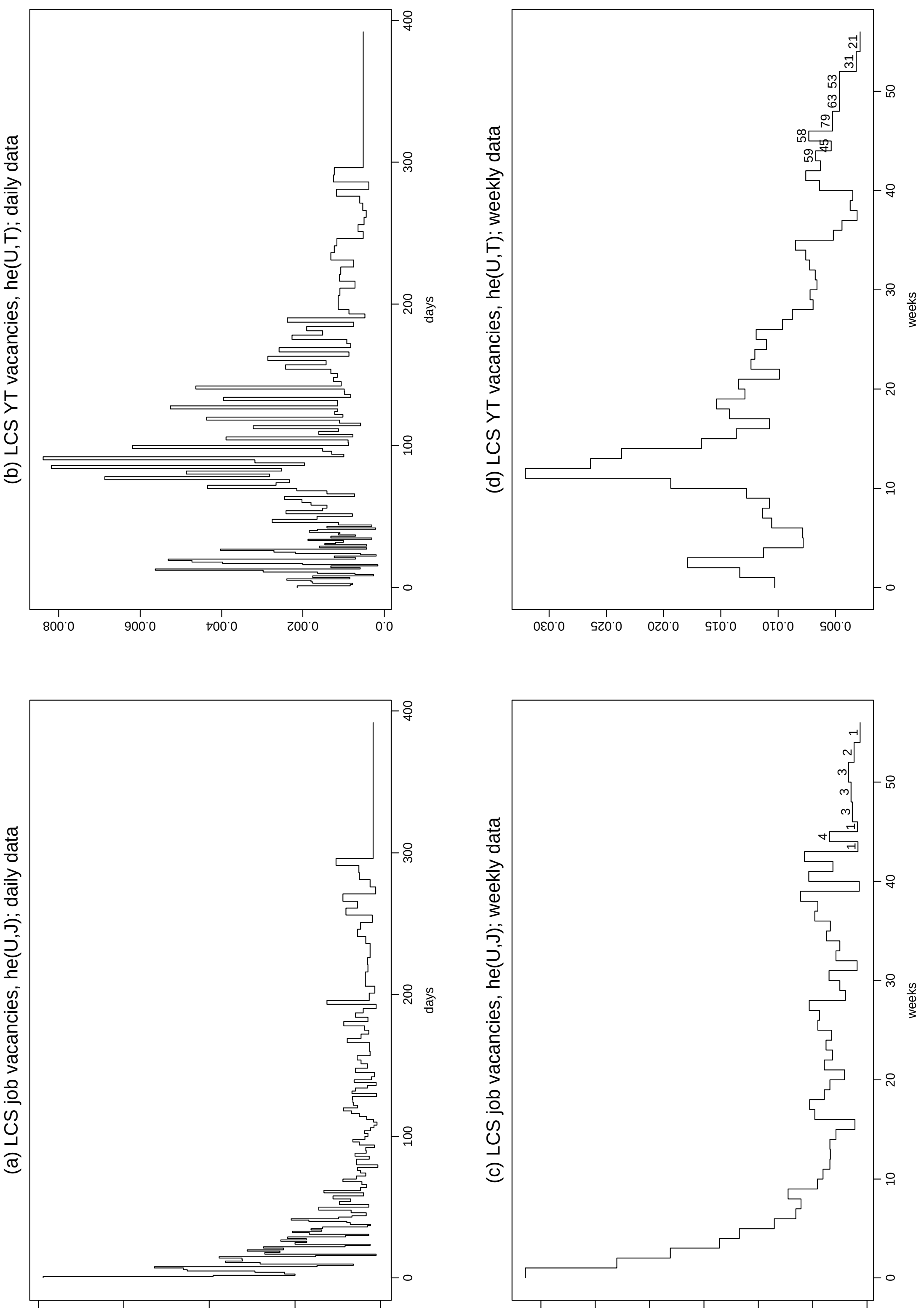

Figure 5: Raw vacancy hazards, daily & weekly data

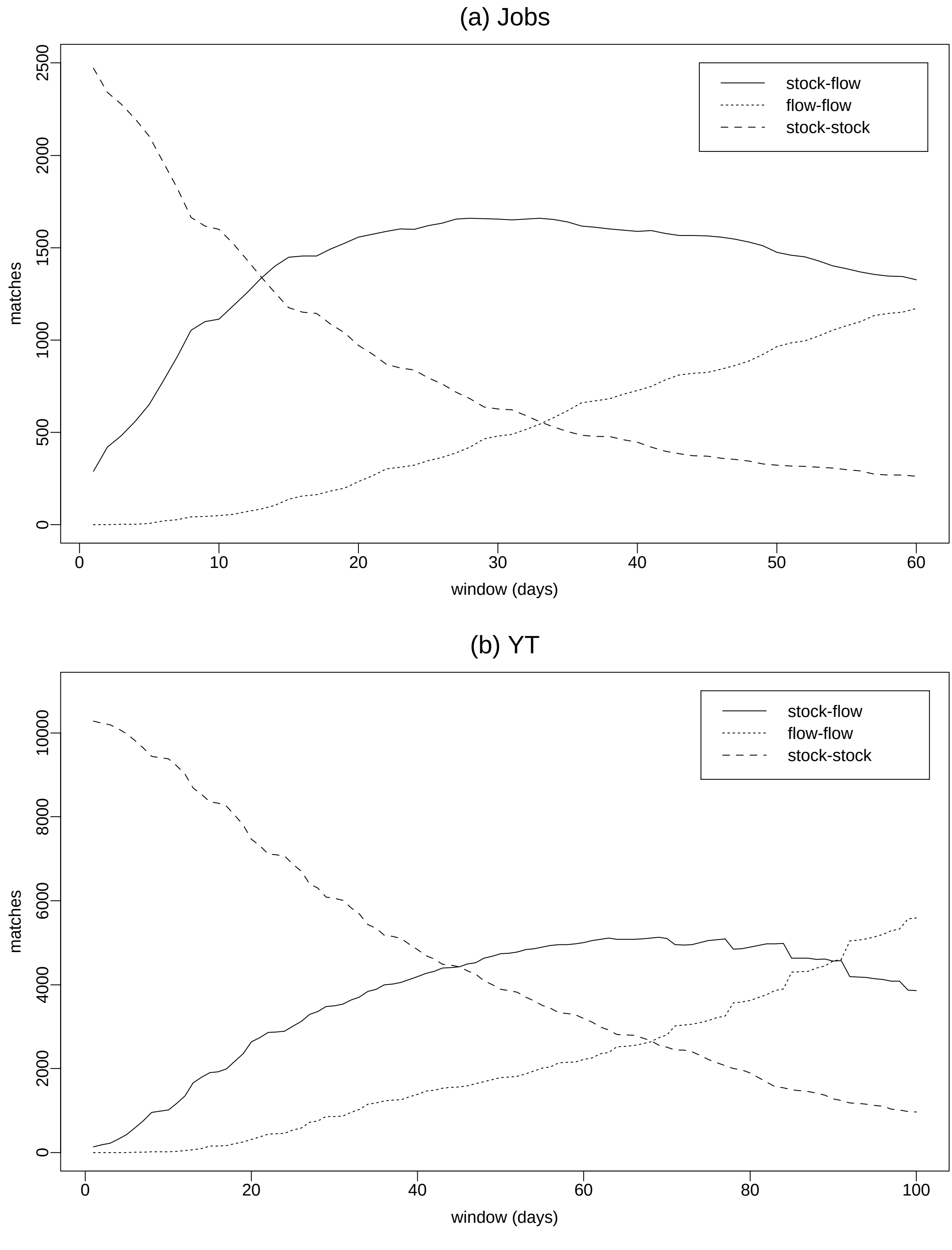

Figure 6: Stock-flow counts by window size

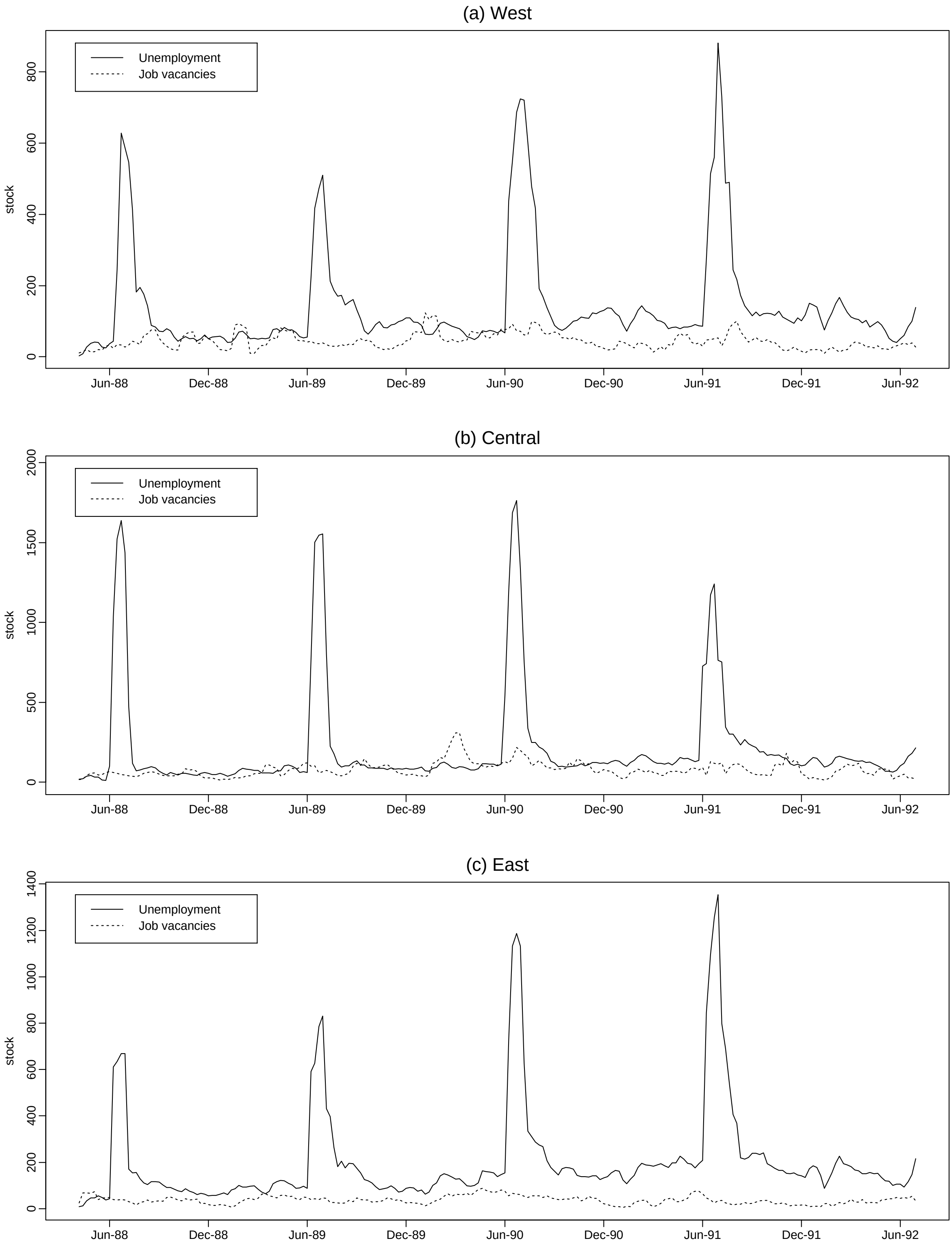

Figure 7: New and old vacancy stocks for 3 labour markets; 4-4 window

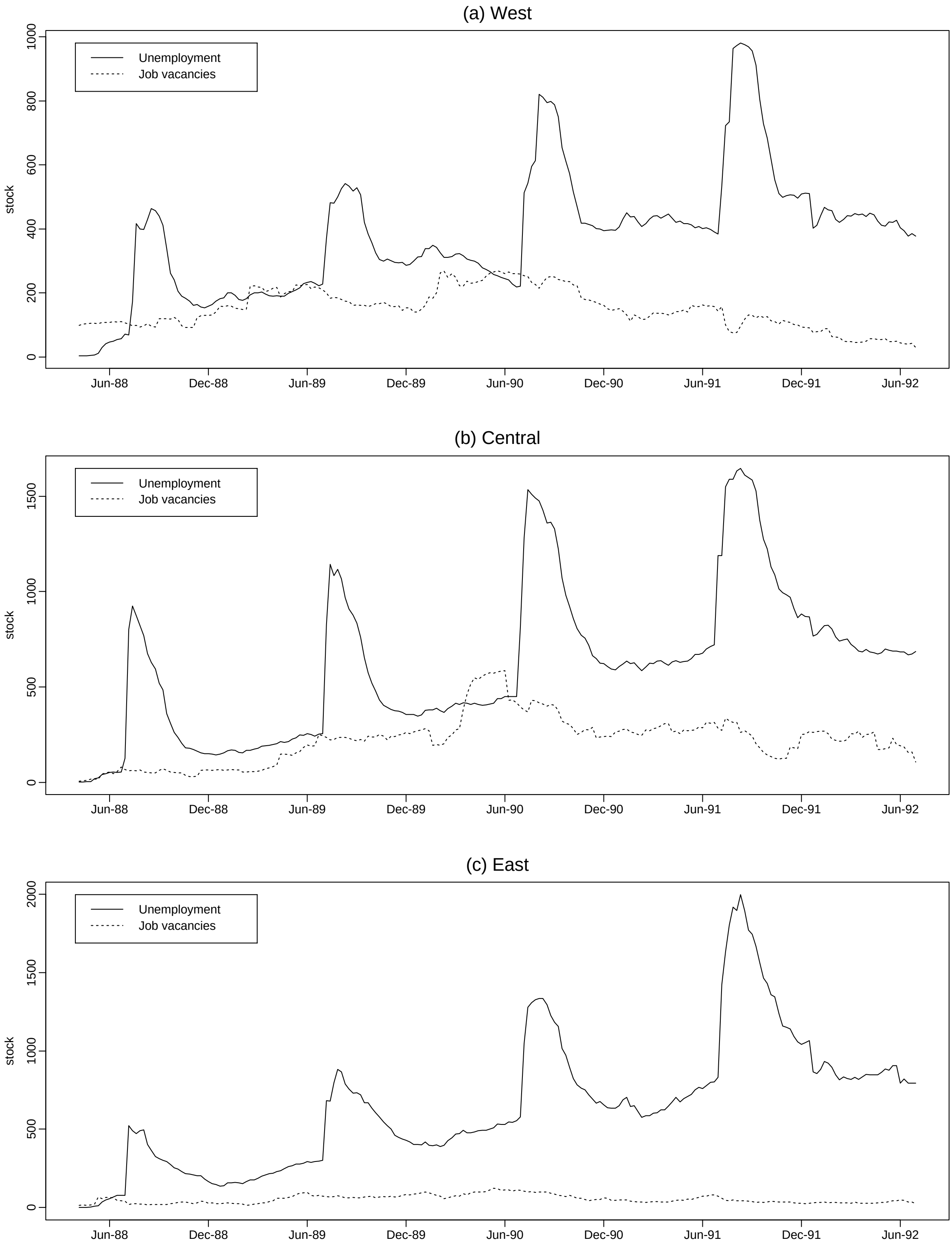

Figure 8: New and old unemployment stocks for 3 labour markets; 4-4 window

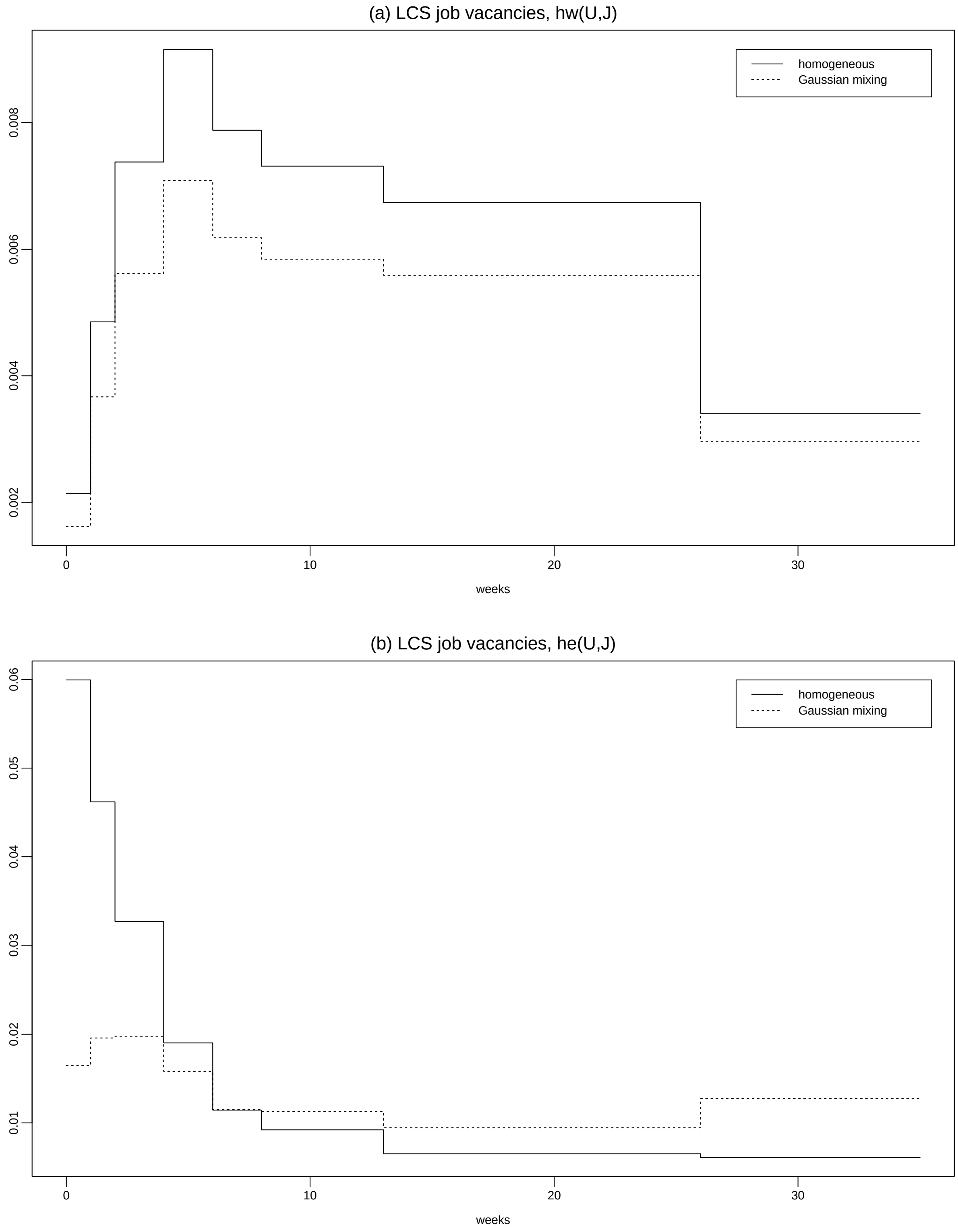

Figure 9: Non-random matching unemployment and vacancy hazards; 4-4 window