Matching Frictions, Efficiency Wages, and Unemployment in the USA and the UK*

James M. Malcomson
University of Oxford

Sophocles Mavroeidis
Brown University

This version: November 27, 2006
Preliminary: please do not quote

*We would like to give special thanks to Jens Larsen from the Bank of England, for his contribution to this project at an early stage. We thank Daron Acemoglu, Olivier Blanchard, Frank Kleibergen, Vasilios Symeonides, Jakob Madsen, Dale Mortensen, Chris Pissarides and Coen Teulings, as well as participants in numerous seminars, for comments, help and advice. Malcomson is grateful to the Leverhulme Foundation for financial support. Part of the research described in this paper was undertaken while Mavroeidis was a Research Fellow at the University of Amsterdam whose financial support is gratefully acknowledged.

Abstract

This paper combines matching frictions with efficiency wages to deter shirking in a model that is estimated for the USA and the UK to derive the underlying structural parameters. Methods robust to weak instruments are used to show that, for both countries, both matching frictions and efficiency wages play a significant role in enabling the model to fit the data even with non-prescriptive formulations for wage determination. The results indicate that adding an efficiency wage element to matching frictions may be a better way to fit the data than simply searching for an alternative wage formulation.

Keywords: Matching frictions, efficiency wages, unemployment, shirking, robust inference

JEL classification: E2, J3, J6

1 Introduction

Two theoretical approaches that have been widely used recently in discussions of unemployment are models of matching frictions, stemming from the work of Diamond (1982), Blanchard and Diamond (1989), Mortensen and Pissarides (1994) and Pissarides (1985), and shirking models of efficiency wages based on Shapiro and Stiglitz (1984). A number of recent contributions have calibrated or estimated tightly-specified aggregate formulations of the Mortensen-Pissarides matching model — see, for example, Cole and Rogerson (1999), Yashiv (2000), Hall (2005a), Shimer (2005) and Yashiv (2006).[1] These, however, typically find it hard to match aspects of the US data, at least with wage determination based on the standard Nash bargain of the original model, and this has generated the search for alternative wage determination procedures, see Hall (2005b) and Hall and Milgrom (2005).

This paper explores a different approach, asking whether it is more consistent with the data to add an efficiency wage element to the matching frictions, a natural step since the two approaches are complements, not substitutes. It does this by constructing a model that combines matching frictions with a tightly-specified shirking model of efficiency wages based on the extension of Shapiro and Stiglitz (1984) in MacLeod and Malcomson (1998) for which model parameters can be estimated empirically. The paper then estimates the combined model econometrically for the US and the UK. To address the concern in the earlier literature (see, for example, Bean (1994)) about the identification of aggregate time-series econometric models of this type (and particularly their wage equations), the paper uses empirical methods that are robust to weak identification. Specifically, it uses a novel method to construct confidence sets for inference purposes that are robust to weak instruments. The bottom line is that the data for both countries calls for the inclusion of an efficiency wage element in the model in addition to matching frictions even for non-prescriptive formulations of wage determination. This suggests that adding an efficiency wages to matching frictions may be a better way to fit the data than simply searching for an alternative wage formulation. The paper also provides an indication of the relative contributions of matching frictions and efficiency wages to long-run unemployment.

The model of matching frictions and vacancy creation used here is essentially an econometric specification of that in Mortensen and Pissarides (1994) applied recently to US data by Hall (2005b) and Shimer (2005). The model of efficiency wages is essentially that of Shapiro and Stiglitz (1984) as extended in MacLeod and Malcomson (1998). In addition to incorporating both frictional and efficiency wage unemployment, the model incorporates a further type of unemployment that can arise for the following reason. To sustain the efficiency wage equilibrium in the model of Shapiro and Stiglitz (1984) requires, as pointed out by Carmichael (1985), a mechanism to prevent wages from being bid down. That workers will shirk if it is in their interest to do so is not in itself sufficient for this because it is wages in the future that influence the incentives to shirk. Thus, when hiring an employee, a firm can reduce the starting wage to the point at which the employee is indifferent between taking the job and not taking the job without affecting incentives to shirk. But, if firms can do that, it becomes in each firm's interest to replace its current employees with new ones to take advantage of the low starting wage. Then employees

[1] There is also a growing literature applying disaggregated versions of the matching model with heteregeneous firms and employees to micro data. For a recent example, see Cahuc, Postel-Vinay, and Robin (2006).

have no incentive not to shirk because they will never receive the higher future wages required to deter shirking. Firms, anticipating this, will not hire them in the first place, so no employment occurs. Thus, as MacLeod and Malcomson (1998) show, the efficiency wage equilibrium cannot be sustained. MacLeod and Malcomson (1998) also show that an equilibrium with employment can be sustained by a market convention about the wage that is appropriate for the job. Firms adhere to the convention because employees either shirk or refuse a job if they do not. Employees adhere to the convention because firms either do not hire them or else fire them if they do not. Thus it is in both sides' interests to stick to the convention. Such a convention provides the mechanism necessary to prevent the bidding down of wages that destroys equilibrium with employment.

Conventions of this sort are not, however, restricted to sustaining the efficiency wage equilibrium of Shapiro and Stiglitz (1984). They can support any wage high enough to deter shirking and low enough to enable firms to make profits, as MacLeod and Malcomson (1998) show. The implications for the model in Shapiro and Stiglitz (1984) are illustrated in Figure 1(a). On the vertical axis is the wage w as a share of worker productivity p, on the horizontal axis the ratio of filled jobs to workers j. For $j = 1$, all workers are employed, so $1 - j$ corresponds to the unemployment rate. The upward sloping curve labelled NSC is the Shapiro-Stiglitz no-shirking condition, which gives the lowest wage that deters shirking for a given unemployment rate. It is upward sloping in employment because a higher wage is required to prevent shirking when unemployment is lower. A wage on or above that curve is sufficient to deter shirking at the corresponding unemployment rate. The downward sloping line labelled L^d is the labour demand, or *job creation*, curve specifying the maximum number of profitable jobs that can exist at a given wage. The appropriate market convention can sustain the Shapiro-Stiglitz equilibrium at the intersection of the two curves, point A in Figure 1(a). But other market conventions can sustain as equilibria any higher wage such as w^* (with the corresponding employment rate given by the labour demand curve at U^*) up to the level at which the labour demand curve cuts the vertical axis. Bargaining power of matched workers or trade unions may also raise wages above the minimum level required to prevent shirking but are not necessary for that. Whatever the reason for a wage above that corresponding to point A, it results in higher unemployment. We refer to such unemployment as *high wage unemployment.*

It is straightforward to add matching frictions to this framework. Such frictions reduce the profitability of creating a new job because that job may not be filled straightaway. They reduce profitability more as the ratio of jobs to workers increases, so the job creation curve becomes steeper. Moreover, some jobs remain vacant while finding a match so the number of filled jobs is less than the total number; the horizontal distance between the filled jobs and the job creation lines in Figure 1(b) corresponds to the number of jobs created at a given wage that remain vacant determined, as standard in the literature, by a matching function. The number of such vacancies increases as the unemployment rate is reduced because there are fewer unemployed workers with which to match, so the filled jobs line is steeper than the job creation line. But matching frictions leave the no-shirking condition unchanged. The resulting curves are all illustrated in Figure 1(b). With these changes taken into account, the underlying analysis of high wage unemployment remains largely unchanged. For a wage convention that sets the wage at w^* in Figure 1(b), jobs are created to the level on the job creation curve corresponding to that wage (point C) and the number of filled jobs is at the point on the filled jobs curve corresponding to that wage (point B). The equilibrium unemployment rate is thus U.

Which of the multiple equilibria comes about depends on the convention that de-

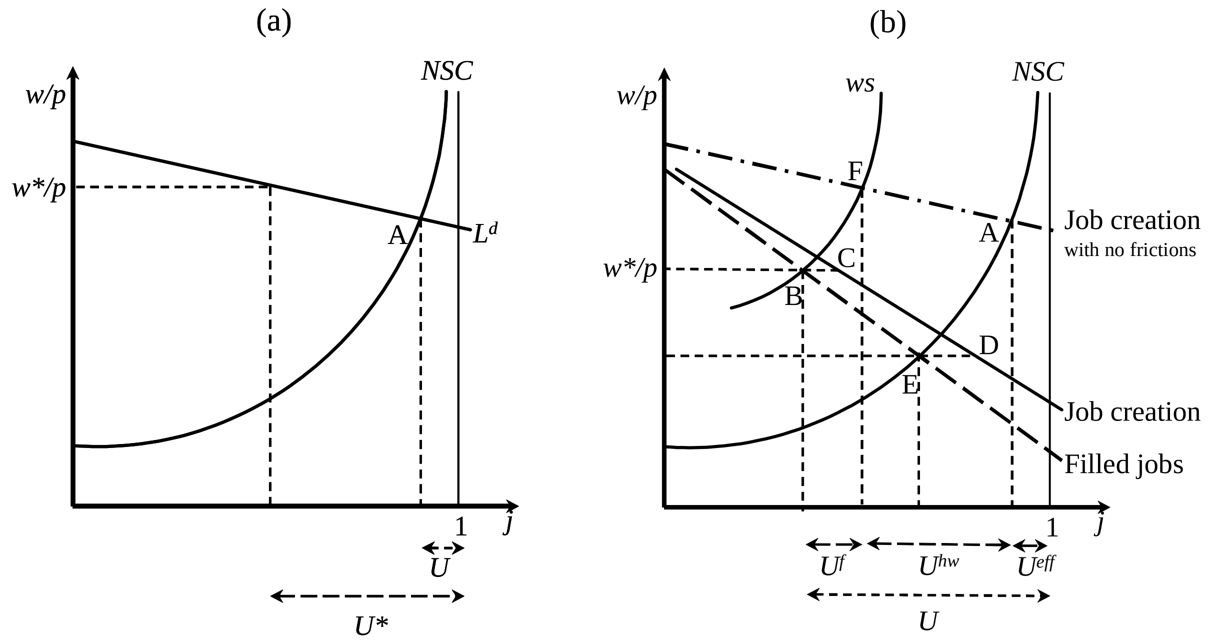

Figure 1: (a) Shapiro-Stiglitz model (b) Model with matching frictions

termines the wage. That is something external to the model. For empirical purposes, a natural way to specify it is via a statistically determined wage equation, or dynamic wage curve in the language of Blanchflower and Oswald (1994)—the convention in the model determines what the wage will be as a function of economic conditions which is exactly what an empirical wage equation does. In effect, the empirically determined wage equation acts as an equilibrium selection device, as in Hall (2005b). It can also take account of wages that are above the minimum level necessary to deter shirking because of trade union or insider bargaining power. When estimated along with the matching function and a dynamic version of the labour demand curve, it can be used to determine all the parameters of the model.

The extent to which unemployment results from matching frictions, efficiency wages and high wages, respectively can be measured in a way that is illustrated in Figure 1(b). The curve ws represents the wage share derived from a stationary representation of an empirically estimated wage equation. The long-run equilibrium wage selected by this curve is w^* with employment at B, the corresponding point on the filled jobs curve. The long-run unemployment rate is then given by U. Removing all matching frictions with everything else unchanged shifts the long-run equilibrium from point B to point F on the job creation curve with no frictions, so a measure of unemployment arising from matching frictions is given by U^f. Removing high wages (that is, reducing wages to the lowest level consistent with deterring shirking) while leaving everything else (including matching frictions) unchanged corresponds to making the wage curve identical to the no-shirking condition, as implicit in Shapiro and Stiglitz (1984). This shifts the long-run equilibrium from point B to point E, so a measure of unemployment arising from high wages is given by U^{hw}. Removing both matching frictions and high wages shifts the long-run equilibrium from point B to point A, leaving just efficiency wage unemployment U^{eff}. (An alternative measure of unemployment arising from matching frictions is the shift from E to A, and of that arising from high wages the shift from F to A, but in our calculations the differences turn out to be negligible.)

Figure 1 illustrates only long-run equilibria. For estimation, the specifications of the job creation equation, the wage equation and the no-shirking condition are explicitly dynamic. The first of these is specified by the condition that the expected cost of creating an additional vacancy equals the expected future profit from having an additional job to fill, taking account of the probability of filling it. Similarly the no-shirking condition recognizes that the incentive to provide effort depends on the path of future wages and the probability of obtaining an alternative job. The final equation in the model is the matching function. The economic specifications of the wage and job creation equations correspond directly to moment conditions, so a natural estimation procedure is the Generalized Method of Moments (GMM) initiated by Hansen (1982). Because of the concern with identification in models of this kind, we construct confidence sets for the long-run values of interest (the long-run unemployment rate and its various decompositions) using methods described in Stock, Wright, and Yogo (2002) that are robust to weak identification. As shown by Kleibergen and Mavroeidis (2006), these methods yield reliable inference without requiring any identification assumptions.

The model is estimated on data for the USA and the UK. For both countries we find that the data call for both matching frictions and efficiency wages — the parameters of the matching function are such as to enable us to reject the hypothesis that all vacancies are matched straightaway at the 0.1% level and the no-shirking condition is significantly above the workers' reservation wage. However, the relative contributions of matching frictions and efficiency wages to unemployment differs substantially between the two countries. For the US, of the long-run unemployment rate estimated at 5.9%, matching frictions account for 1.7%, high wages for 0.7%, and efficiency wages for 3.5%. For the UK, the long-run unemployment rate is estimated at 6.1%. But there, matching frictions account for only 0.1% (though still significantly different from zero), high wages for another 0.2%, and efficiency wages for 5.8%. In the estimation, we allow for considerable flexibility in the wage equation and, while the point estimates naturally differ for different specifications, the basic conclusion that matching frictions do not account for all long-run unemployment is highly robust. Even with wage determination not restricted to the standard Nash bargaining solution, the model needs more than just matching frictions to match the data well.

The paper is organized as follows. The next section describes the model and characterizes equilibrium. The following section provides details of the empirical implementation and the estimation procedure. This is followed by a description and discussion of the estimation results. Section 5 applies robust inference procedures to investigate long-run unemployment and its components. That is followed by a short conclusion.

2 Theory

2.1 The model

The model consists of risk-neutral workers and firms with a common discount factor δ_t at time t. A job may have one worker working a specified number of hours or no worker at all. A worker's utility in period t from being employed at total cost to the firm w_t and incurring effort e_t is $w_t\tau_t - c_te_t$, where τ_t is the ratio of take-home pay to the total cost of employment to the firm and c_t is the time-dependent disutility of effort measured in monetary terms. Effort takes one of two values, $e_t = 1$ (working) and $e_t = 0$ (shirking).

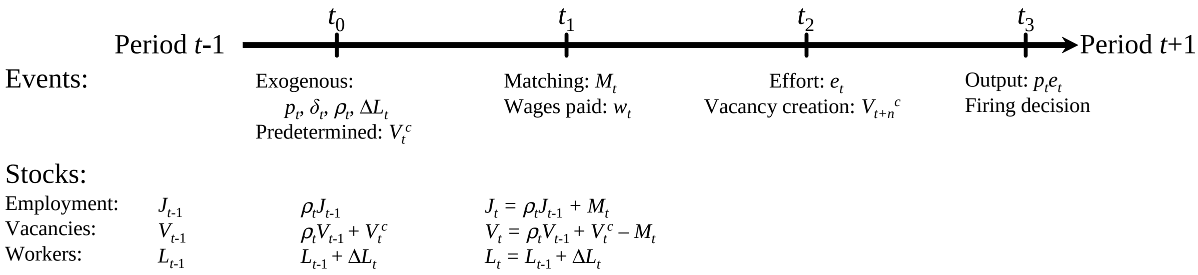

Figure 2: Timing of events in period t

The output received by the firm is $p_t e_t$, so its period t profit from employing a worker is $p_t e_t - w_t$. Monitoring by the firm is perfect but not verifiable in court, so a firm knows a worker's effort in its job but cannot make the wage conditional on that. As in Shapiro and Stiglitz (1984) it can, however, fire an employee who shirks.[2]

The timing of events for period t is shown in Figure 2. At the start of period t, the economy is characterized by the following stocks determined at $t-1$: J_{t-1} filled jobs and employed workers; V_{t-1} vacancies unfilled after matches in period $t-1$ have been formed; and L_{t-1} workers, of whom $L_{t-1} - J_{t-1}$ are unemployed. At t_0, four exogenous events occur. First, a common productivity p_t for all jobs producing in period t is observed. Second, the discount factor δ_t for receipts and payments at $t+1$ is observed. Third, a fraction $1 - \rho_t$ of the jobs filled in period $t-1$, and of unfilled vacancies at $t-1$, become unprofitable for exogenous reasons and are destroyed. Fourth, vacancies that firms decided at $t-n$ (with $n \geq 1$) to create for period t become available to be filled. Once vacancy creation has taken place, the stock of vacancies becomes $\rho_t V_{t-1} + V_t^c$. To keep a vacancy available for filling, a firm must incur a hiring cost ψ_t each period. The number of periods in advance n at which vacancy creation decisions for period t are made is given exogenously. Creating an additional vacancy incurs a capital cost that, discounted back to $t-n$ (when the decision to create the vacancy is made), is denoted Ψ_{t-n}. Thus, as recommended by Shimer (2005) for fitting US data, vacancies are a genuine state variable. The specifications of ψ_t and Ψ_t are determined empirically. Finally at t_0, labour supply increases exogenously by ΔL_t. All these events are public information.

At t_1, firms with vacancies and unemployed workers create M_t new matches at agreed wage w_t and that wage is paid. Creating new matches requires search. The search friction is characterized by a matching function for which an empirical functional form is specified later. At t_2, workers decide the effort e_t to incur and firms decide how many vacancies to create for period $t+n$. Finally in period t, at t_3, firms with workers observe output $p_t e_t$ and decide whether to retain or fire their worker.

Employment in period t is the fraction of jobs in the previous period that are not destroyed, $\rho_t J_{t-1}$, plus newly matched vacancies M_t, so

$$J_t = \rho_t J_{t-1} + M_t. \tag{1}$$

Let $j_t = J_t / L_t$, $m_t = M_t / L_t$ and $l_t = L_t / L_{t-1}$. Then, divided by L_t, (1) becomes

$$j_t = \rho_t j_{t-1} / l_t + m_t. \tag{2}$$

[2]Formally, p_t is the total productivity (net of non-labour costs) of employing a worker in period t for optimal hours and non-labour inputs, and w_t the total labour cost of doing so.

This, with a specific functional form for the matching function, is one of the model equations that is estimated. The stock of vacancies at the end of period t, V_t, is the sum of vacancies at the outset of the period after destruction has taken place, $\rho_t V_{t-1}$, and newly created vacancies, V_t^c, minus matches M_t, so

$$V_t = \rho_t V_{t-1} + V_t^c - M_t. \tag{3}$$

Let $v_t^c = V_t^c / L_t$ denote the ratio of new vacancies to workers at t. In an equilibrium in which no workers actually shirk, the ratio of total vacancies to workers at the time of matching at t_1 is v_t given by

$$v_t = \frac{\rho_t V_{t-1} + V_t^c}{L_t} \tag{4}$$

$$= \rho_t \frac{v_{t-1} - m_{t-1}}{l_t} + v_t^c, \tag{5}$$

the second inequality following from manipulation of (3).[3]

Also in an equilibrium in which no workers actually shirk, the stock of unemployed workers seeking matches at t_1 consists of workers who were unemployed in the previous period, $L_{t-1} - J_{t-1}$, workers who were employed in the previous period but have lost their job, $(1-\rho_t) J_{t-1}$, and new workers, $\Delta L_t = L_t - L_{t-1}$, making $L_t - \rho_t J_{t-1}$ in total. Thus the *job-seeking rate at* t_1 is u_t given by

$$u_t = \frac{L_t - \rho_t J_{t-1}}{L_t} = 1 - \rho_t \frac{j_{t-1}}{l_t}. \tag{6}$$

2.2 Equilibrium

Equilibrium requires that the wage path is such that workers do not shirk and that, as long as new vacancies are created, the expected profit from creating an additional vacancy is zero. To see the implications of the second of these, denote by Π_t the expected present value of current and future profits at t_1 from having a job filled at wage cost w_t. This equals output net of wage costs at t, plus the expected present value of profits from period $t+1$ on, discounted by the discount factor at t and the probability that the relationship is not ended before production at $t+1$ because the job is destroyed for exogenous reasons. Thus

$$\Pi_t = p_t - w_t + E_t\left(\delta_t \rho_{t+1} \Pi_{t+1}\right), \text{ for all } t, \tag{7}$$

where E_t is the expectation operator conditional on information available at t_2. The probability of filling a vacancy at t conditional on the numbers of vacancies and matches is m_t/v_t. Hence, the present discounted value $\bar{\Pi}_t$ of having a vacancy available for matching at t is

$$\bar{\Pi}_t = -\psi_t + \frac{m_t}{v_t}\Pi_t + \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right), \text{ for all } t. \tag{8}$$

[3] From (4), $\rho_t V_{t-1} + V_t^c = v_t L_t$ which, used in (3), gives $V_t = (v_t - m_t) L_t$. Substitution of this for $t-1$ into (4) gives

$$v_t = \frac{\rho_t (v_{t-1} - m_{t-1}) L_{t-1} + V_t^c}{L_t} = \rho_t (v_{t-1} - m_{t-1}) \frac{L_{t-1}}{L_t} + \frac{V_t^c}{L_t},$$

which corresponds to (5).

The interpretation is as follows. The hiring cost ψ_t is incurred to keep the vacancy available for this period. With probability m_t/v_t, the vacancy is matched with a worker and yields expected future profit Π_t; with probability $1 - m_t/v_t$, it is not matched with a worker and, if not destroyed for exogenous reasons, remains available to be filled in period $t+1$. For an equilibrium in which (as in practice) vacancies are created in each period, firms decide at $t-n$ to create new vacancies v_t^c that become available to be filled in period t up to the level at which

$$E_{t-n}\left(\bar{\Pi}_t \prod_{j=1}^{n} \delta_{t-j}\rho_{t+1-j} - \Psi_{t-n}\right) = 0, \text{ for all } t, \tag{9}$$

where we use the convention that the expectation operator E_t applied to a variable at a date $t+i$ with $i \geq 1$ is taken over the joint distribution of the random variables at $t+1, \ldots, t+i$, and it is assumed that vacancies in the process of creation in period t also become unprofitable at the same rate $(1-\rho_t)$ as jobs already created and are thus abandoned. (This assumption is not an essential characteristic of the model but simplifies the presentation.) Of course, if it were the case that $\Pi_t < 0$, existing jobs at t would all be closed down and no vacancies filled, so there would be no employment. A sufficient condition to ensure $\Pi_t \geq 0$ is that $E_t\left(p_t - w_t\right) \geq 0$ for all t, although it is clearly not necessary that this hold in every period.

Equation (9) is the basis of the job creation line in Figure 1. As it stands, it is not suitable for empirical purposes because $\bar{\Pi}_t$ contains terms stretching into the infinite future. Applied to $t+n$, however, (9) can be used to replace terms further in the future than $t+n-1$ by the cost of creating vacancies at $t+n$ and $t+n+1$. The manipulations required to do this are given in Appendix A, which shows that, with the convention $\prod_{i=1}^{j} x_i = 1$ for $j = 0$ for any variable x_i, (9) can be re-written as

$$\begin{aligned}
E_{t-n}\Bigg\{ & \frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left[\sum_{j=0}^{n-1}\left(p_{t+j}-w_{t+j}\right)\left(\prod_{i=1}^{j}\delta_{t-1+i}\rho_{t+i}\right)\right. \\
& \left. + \frac{v_{t+n}}{m_{t+n}}\left[\Psi_t + \psi_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i} - \delta_t\rho_{t+1}\left(1-\frac{m_{t+n}}{v_{t+n}}\right)\Psi_{t+1}\right]\right] \\
& + \delta_{t-n}\rho_{t-n+1}\left[\left(1-\frac{m_t}{v_t}\right)\Psi_{t-n+1}\right] - \Psi_{t-n} - \psi_t\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\Bigg\} \\
& = E_{t-n}\left(z_{t,n}\right), \quad \text{for all } t,
\end{aligned} \tag{10}$$

where $z_{t,n}$ is a covariance term specified in (52) in Appendix A that depends on n. For $n=1$,

$$z_{t,1} = -\frac{m_t}{v_t}\delta_{t-1}\rho_t E_t\left[\frac{1}{m_{t+1}/v_{t+1}}\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1} - \Psi_t\right)\right],$$

which depends on the covariance of the excess profits from having a job available to be filled next period with the inverse of the probability of filling that job in that period. Under perfect foresight, $E_{t-n}z_{t,n} = 0$ necessarily. For other cases, that can be tested, at least in part, as a result of the over-identifying restrictions it implies.

Equation (10) is the *job creation equation* used for empirical analysis. Its interpretation is more straightforward when a vacancy becomes available to be filled the period after the decision to create it, that is $n = 1$. For $n = 1$ and $E_{t-1}(z_{t,1}) = 0$, (10) simplifies to

$$E_{t-1}\left\{\delta_{t-1}\rho_t\frac{m_t}{v_t}\left[p_t - w_t + \frac{v_{t+1}}{m_{t+1}}\left(\Psi_t + \psi_{t+1}\delta_t\rho_{t+1} - \left(1 - \frac{m_{t+1}}{v_{t+1}}\right)\delta_t\rho_{t+1}\Psi_{t+1}\right)\right]\right\}$$
$$+ E_{t-1}\left\{\delta_{t-1}\rho_t\left[-\psi_t + \left(1 - \frac{m_t}{v_t}\right)\Psi_t\right]\right\} = E_{t-1}(\Psi_{t-1}), \text{ for all } t. \quad (11)$$

The term on the right-hand side is the expected cost of creating a vacancy to become available in period t, as measured at $t-1$ when the decision to create the vacancy is made. The left-hand side gives the expected benefit from creating that vacancy. Consider first the final term in braces. The cost ψ_t has to be incurred to keep the vacancy available at t. With probability $1-m_t/v_t$ the vacancy will not be filled in period t, when it first becomes available. In that case, the expected future profits from having created the vacancy are just the same as if the vacancy had been created one period later, discounted by the factor $\delta_{t-1}\rho_t$ to allow for the costs having been incurred one period earlier and for the probability that there is one additional period for the vacancy to become unprofitable for exogenous reasons. By the equilibrium condition for vacancies that become available to fill in $t+1$, those expected future profits equal the expected cost of creating a vacancy for that period, $E_t(\Psi_t)$. Now consider the first term in braces. With probability m_t/v_t, the vacancy will be filled in period t. The terms multiplying that correspond to the expected future profits from filling it. These consist of the expected profits in period t itself, $p_t - w_t$, plus the expected future profits from $t+1$ on. These latter are the same as for a vacancy created one period later that becomes available for filling at $t+1$ and is filled immediately. By the equilibrium condition for vacancies that become available at $t+1$, these consist of the difference between the expected cost of creating the vacancy, $E_t(\Psi_t)$, and the expected profits if it is not filled, adjusted by the appropriate probabilities. The expected profits if it is not filled are, in turn, the same as those of having a vacancy become available one period later at $t+2$ which, by the equilibrium condition for vacancy creation for $t+2$, equals the expected cost of creation $E_{t+1}(\Psi_{t+1})$ discounted appropriately. The difference between (10) for $n > 1$ and (11) is that, to get the appropriate discount factors in the former, we have used the equilibrium condition for creating a vacancy n periods ahead, so we have also to add the series of appropriately discounted one-period profits $p_{t+j} - w_{t+j}$ from $t+1$ to $t+n-1$. Note that there is nothing here specific to an efficiency wage story. Essentially, (10) is a slightly generalized econometric specification of the equilibrium condition in Hall (2005b) and Shimer (2005) that there are zero profits to creating additional vacancies.

Now consider the implications of the equilibrium condition that workers do not shirk. For a worker in a match in period t, the expected present value W_t of deciding at t_2 not to shirk and staying with the firm consists of take-home pay less the disutility of effort in period t, $w_t\tau_t - c_t$, plus the expected future utility from not being dismissed for shirking. Thus,

$$W_t = w_t\tau_t - c_t + \delta_t E_t\left[\rho_{t+1}W_{t+1} + (1-\rho_{t+1})\bar{W}_{t+1}\right], \text{ for all } t, \quad (12)$$

where $\bar{W}_{t+1}$ is the expected present value of starting period $t+1$ unemployed, an event that happens with the probability $1-\rho_{t+1}$ that the job comes to an end for exogenous reasons. The probability that a worker unemployed at t_0 finds a job in the matching process at t_1 conditional on job seeking rate u_t and matching rate m_t is m_t/u_t. Hence, the present discounted value $\bar{W}_t$ of seeking a match at t is

$$\bar{W}_t = \frac{m_t}{u_t} W_t + \left(1 - \frac{m_t}{u_t}\right) \left(b_t + \delta_t E_t \bar{W}_{t+1}\right), \text{ for all } t, \tag{13}$$

where b_t is the utility received while unemployed in period t, including not only unemployment benefits but also utility (from, for example, home production) that would not be obtained from shirking while being employed. The right-hand side of (13) can be interpreted as follows. With probability m_t/u_t, the worker is hired at t and receives expected future utility W_t from being in a match. With probability $1 - m_t/u_t$ the worker is not hired at t and receives utility of b_t for period t plus the expected utility from starting the next period unemployed.

A worker in a match in period t will shirk unless the expected future utility, W_t, from not doing so is at least as great as that from shirking (with no disutility of effort), collecting the wage w_t in period t, but being fired and receiving the expected future utility $\delta_t E_t \bar{W}_{t+1}$ from starting period $t+1$ unemployed. Thus a necessary condition for the worker not to shirk, the *no-shirking condition* (NSC), is

$$W_t \geq w_t \tau_t + \delta_t E_t \bar{W}_{t+1}, \text{ for all } t. \tag{14}$$

Substitution for W_t from (12) and re-arrangement allows this condition to be written

$$\delta_t E_t \left[\rho_{t+1} \left(W_{t+1} - \bar{W}_{t+1}\right)\right] \geq c_t, \text{ for all } t. \tag{15}$$

The economic interpretation is that, with no wage penalty in the current period from shirking, the employee will shirk unless the discounted expected future gains to being employed over being unemployed, given that the employment will continue with probability only ρ_{t+1} even if the worker does not shirk, exceeds the disutility of effort. (Separation payments received by the worker can be thought of as increasing c_t.) With the use of (13) and (12) for date $t+1$, the left-hand side of (15) can be written

$$\begin{aligned}
&\delta_t E_t \left[\rho_{t+1} \left(W_{t+1} - \bar{W}_{t+1}\right)\right] \\
&= E_t \left\{\delta_t \rho_{t+1} \left[W_{t+1} - \frac{m_{t+1}}{u_{t+1}} W_{t+1} - \left(1 - \frac{m_{t+1}}{u_{t+1}}\right) \left(b_{t+1} + \delta_{t+1} E_{t+1} \bar{W}_{t+2}\right)\right]\right\} \\
&= E_t \left\{\delta_t \rho_{t+1} \left(1 - \frac{m_{t+1}}{u_{t+1}}\right) \left[W_{t+1} - \left(b_{t+1} + \delta_{t+1} E_{t+1} \bar{W}_{t+2}\right)\right]\right\} \\
&= E_t \left\{\delta_t \rho_{t+1} \left(1 - \frac{m_{t+1}}{u_{t+1}}\right) \left[\left(w_{t+1}\tau_{t+1} - c_{t+1} - b_{t+1}\right) + \delta_{t+1} E_{t+1} \rho_{t+2} \left(W_{t+2} - \bar{W}_{t+2}\right)\right]\right\}.
\end{aligned} \tag{16}$$

Similar use of (13) and (12) for dates $t+2$ on allows the no-shirking condition (15) to be written

$$E_t \left[\sum_{i=t+1}^{\infty} \left(w_i \tau_i - c_i - b_i\right) \prod_{j=t+1}^{i} \delta_{j-1} \rho_j \left(1 - \frac{m_j}{u_j}\right)\right] \geq c_t, \text{ for all } t. \tag{17}$$

For the formulation used here, the results in MacLeod and Malcomson (1998) imply that the no-shirking condition (15), and thus (17), is not only necessary for an equilibrium in which workers do not shirk but, together with a condition that firms make non-negative profits that is certainly satisfied by (10), is also sufficient.[4] Note that this applies even if workers do not have conventional bargaining power as a result of matching frictions or collective bargaining. (Worker bargaining power may, of course, also raise the equilibrium wage above the no-shirking condition.) Because (17) is an inequality, (17) and (10) do not determine unique equilibrium paths for wages and employment, merely restrictions on the set of permissible equilibrium paths. For stationary equilibria, these restrictions correspond to points in Figure 1 on the filled jobs line and to the left of E.

2.3 Equilibrium selection

Any paths that satisfy the job creation equation (10) and the no-shirking condition (17) are equilibrium paths with positive employment and some new vacancies created each period. Which of those paths is selected, and thus gives rise to an actual history, depends on the convention that determines the evolution of wages in the labour market, see MacLeod and Malcomson (1998) for the efficiency wage model and Hall (2005b) for the matching model. Because the convention is selecting among equilibria of the model, the model itself does not tell us more about it than that the path it selects must satisfy the equilibrium conditions. We can ensure the job creation equation is satisfied by representing the wage convention as the intersection between a wage equation and the job creation equation, as in Figure 1. All we then have to do is to ensure that the wage equation satisfies the no-shirking condition (17) if there are efficiency wages. If there are no efficiency wages, we want the wage equation to satisfy properties that are appropriate for a model with just matching frictions..

For the forward-looking rational expectations model used here, it is natural for the wage equation also to satisfy forward-looking rational expectations. This gives a clear identifying assumption that enables us to use appropriately lagged values of variables as instruments. To ensure that our wage equation satisfies our requirements, we use the specification

$$E_{t-2}\left[\delta_{t-1}\rho_t\theta_t\left(1-\frac{m_t}{u_t}\right)\left(\tau_t\frac{w_t}{p_t}-\frac{b_t}{p_t}\right)-h\left(x_t\right)\right]=0, \tag{18}$$

where $h\left(x_t\right)$ is a function of variables x_t not known at $t-2$. For a perfectly competitive labour market with a given number of homogeneous workers, the labour supply curve has a reverse-L shape. Thus, if there is unemployment ($m_t/u_t<1$) in a perfectly competitive market, the wage must be such that after-tax earnings $\tau_t w_t$ equal the utility b_t received while unemployed. That is consistent with (18) if, but only if, $h\left(x_t\right)\equiv 0$. If, however, there is no unemployment, the utility received while unemployed plays no role in wage determination — the wage just has to satisfy the labour demand curve (given by the job creation equation) when all workers are employed. That is consistent with (18) when $h\left(x_t\right)\equiv 0$ because, when there is no unemployment, $m_t/u_t=1$. With wage bargaining that arises from matching frictions, after tax earnings may be above b_t when $m_t/u_t<1$, which is consistent with (18) for $h\left(x_t\right)>0$. But the wage converges to the competitive wage as matching frictions go to zero, a property that should hold for *any* bargaining specification, not just the Nash bargain traditionally used in matching models. That is

[4] MacLeod and Malcomson (1998) show that there may (but need not) also exist equilibria with different contractual arrangements in which there is no efficiency wage unemployment even with $c_t>0$. That, however, requires vacancies to exceed unemployment sufficiently, which is not consistent with our data.

consistent with (18) if, but only if, $h(x_t) \to 0$ whenever $m_t/u_t \to 1$ or, more generally, as matching frictions go to zero.

To be consistent with an efficiency wage, the specification of the term $h(x_t)$ in (18) must ensure that the wage lies above the lowest wage that would satisfy the no-shirking condition (17) with equality at each date. We specify $h(x_t)$ as a mark-up on the disutility of work c_t/p_t that is a function of appropriate variables in the model. To implement that, we use equality in (17) to substitute for terms in $W_{t+1} - \bar{W}_{t+1}$ and $W_{t+2} - \bar{W}_{t+2}$ in (16) to get

$$\begin{aligned} c_t &= E_t \left\{ \delta_t \rho_{t+1} \left(1 - \frac{m_{t+1}}{u_{t+1}}\right) \left[\left(\underline{w}_{t+1} \tau_{t+1} - c_{t+1} - b_{t+1}\right) + c_{t+1} \right] \right\} \\ &= E_t \left[\delta_t \rho_{t+1} \left(1 - \frac{m_{t+1}}{u_{t+1}}\right) \left(\underline{w}_{t+1} \tau_{t+1} - b_{t+1}\right) \right], \end{aligned}$$

where $\underline{w}_{t+1}$ denotes the lowest wage at $t+1$ that will satisfy (15) at t. Writing this one period earlier, dividing by p_{t-1}, and using the definition $\theta_t = p_t/p_{t-1}$, we can write it as

$$E_{t-1} \left[\delta_{t-1} \rho_t \theta_t \left(1 - \frac{m_t}{u_t}\right) \left(\tau_t \frac{\underline{w}_t}{p_t} - \frac{b_t}{p_t} \right) \right] = \frac{c_{t-1}}{p_{t-1}}. \tag{19}$$

We assume c_t/p_t has a constant long-run value c/p but, since we do not wish to rule out changes in c_t/p_t completely, we permit changes that are iid deviations from the long-run value. Taking expectations on both sides of (19) conditional on period $t-2$, we then get a wage equation of the form in (18) if we specify

$$h(x_t) = \frac{c}{p}\left(1 + f(x_t)\right), \tag{20}$$

where $f(x_t)$ is some non-negative function of variables x_t known at t. We discuss the specification of f_t in detail in Section 3. To be consistent with our forward-looking specification, we exclude variables known at $t-2$. Because (18) is specified in terms of an expected value, it does not impose that $w_t \geq \underline{w}_t$ for all t but, for f_t non-negative, it holds in the long run. Moreover, it permits us to test whether the no-shirking condition binds at all t by testing whether $f_t = 0$ for all t. Finally, it can identify c/p and hence the location of the long-run no-shirking condition.

The wage equation (18) is different from the conventional log-linear wage equations that have a long tradition in the literature, for example, Layard, Nickell, and Jackman (2005) and Blanchard and Katz (1999). Such a wage equation will satisfy the no-shirking condition for sufficiently small disutility of effort c_t as long as it has the property that the wage goes to infinity as unemployment goes to zero. But it can identify only an upper bound on c_t/p_t and, hence, cannot determine the precise location of the no-shirking condition. Although theory does not provide a natural identifying assumption for conventional log-linear wage equations, we also estimate the model using one as a robustness check on results that do not depend on the precise location of the no-shirking condition. For appropriate selection of variables, this specification can encompass the wage bargaining in such matching models as Blanchard and Diamond (1989) and Pissarides (2000) and the wage curve of Blanchflower and Oswald (1994). We discuss the precise specification in Section 3.

3 Empirical implementation

The model consists of three equations: a matching equation, a job creation equation, and a wage equation. The Generalized Method of Moments (GMM) of Hansen (1982) is a natural estimation method for the job creation equation (10) and the wage equation (18) because their economic specifications correspond to moment conditions. As a system of equations, there are potential efficiency gains to estimating the equations jointly. Moreover, the hypotheses on long-run unemployment that we investigate imply cross-equation parameter restrictions that can be tested most naturally using a system approach. However, because the system is non-linear and there are a large number of potentially relevant variables and their lags that we do not wish to exclude *a priori*, the system approach is unwieldy — the collinearity between potentially relevant variables creates problems for convergence. So here we adopt the compromise of deriving a parsimonious specification that is satisfactory statistically by conducting preliminary analysis on each of the equations in the model individually. We then estimate the parameters of this parsimonious specification as a system (checking that the specification remains statistically satisfactory) and use that for conducting inference.

As discussed later, each equation in the model can be estimated using conditional moment restrictions of the form $E_s\varepsilon_t^i = 0$, where ε_t^i denotes the residuals of equation i and $s < t$. In particular, we show that $s = t - 1$ for the matching and conventional wage equations, $s = t - 2$ for the non-linear wage equation (18), and choose n such that $s = t - n$ for the job-creation equation (10). This type of moment condition implies $EZ_s^i\varepsilon_t^i = 0$ for any vector of instruments Z_s^i that contains variables known at time s. In other words, the set of admissible instruments for each equation is infinite. It is well-known that use of many instruments can have adverse effects on the finite sample properties of GMM estimators and tests. In particular, use of more instruments typically increases the finite sample bias of the estimators, especially if those additional instruments are poorly correlated with the endogenous variables they are instrumenting, see Stock, Wright, and Yogo (2002). Moreover, the power of the tests of over-identifying restrictions deteriorates, making it harder to discover any model mis-specification, see Mavroeidis (2005). We therefore assign a small number of instruments Z_t^i to each equation by including up to two lags of the variables that appear in that particular equation.

Commonly, a two-step GMM estimator is used for computational convenience. Two-step estimators are asymptotically efficient. However, a number of studies have shown that they have poor finite-sample properties under weak or many instruments (for example, suffer from large biases and size-distortions, see Stock, Wright, and Yogo (2002)). An alternative estimator proposed by Hansen, Heaton, and Yaron (1996) is the Continuously Updated GMM Estimator (CUE) defined as the minimizer with respect to an m-dimensional vector of parameters ϑ of the objective function

$$S(\vartheta) = T^{-1} f_T(\vartheta) V_{ff}^{-1}(\vartheta) f_T(\vartheta), \tag{21}$$

where T denotes the sample size, $f_t(\vartheta)$ is a K-dimensional moment function whose expectation $Ef_t(\vartheta)$ vanishes at the true value of the parameters, $f_T(\vartheta) = \sum_{t=1}^{T} f_t(\vartheta)$ are the corresponding sample moments and $V_{ff}(\vartheta) = \lim_{T\to\infty} var\left[T^{-1/2} f_T(\vartheta)\right]$ denotes their asymptotic variance matrix. We use the CUE because it has been recently shown to have better finite-sample properties than two-step estimators, see Newey and Smith (2004). Moreover, the available test statistics that are robust to failure of the identification assumption are based on the CUE objective function (21), see Stock and Wright (2000) and Kleibergen (2005) and Kleibergen and Mavroeidis (2006). Finally, to operationalize (21) we use the Newey and West (1987) heteroskedasticity and autocorrelation consistent (HAC) estimator of $V_{ff}(\vartheta)$, as suggested by Kleibergen (2005).

The concern about identification in aggregate time-series models of the type used here makes it important to use inference procedures that are robust to weak instruments. Weak identification implies that GMM estimators are inconsistent, that their distribution can be very different from the usual Normal approximation even in relatively large samples, and that conventional standard errors may underestimate the true uncertainty in the estimates. See, for example, Mavroeidis (2004). As a result, 95% confidence intervals derived by inverting a Wald test, such as the usual two-standard-error band about a point estimate, may be too narrow in the sense that the probability that they contain the true value of the parameter can be much less than 95%. So for testing hypotheses we employ, in addition to standard Wald tests, two further tests that are robust to weak instruments. One is the test proposed by Stock and Wright (2000), which is based on the fact that, under mild regularity conditions such as that $f_T(\vartheta)$ follows a central limit theorem and that a consistent estimator of $V_{ff}(\vartheta)$ exists, the GMM objective function (21) evaluated at the true value of ϑ is asymptotically distributed as χ^2 with K degrees of freedom, *irrespective* of whether ϑ is identified or not. This test is a generalization of a test that was originally proposed by Anderson and Rubin (1949) in the context of the linear instrumental variables regression model. We refer to it as the Anderson-Rubin-Stock-Wright (ARSW) test.

One potential difficulty with the interpretation of the ARSW test stems from the fact that it jointly tests the null hypothesis on the parameters of the model and the validity of the over-identifying restrictions, see Stock and Wright (2000). Thus, the test statistic may be large (and associated confidence sets may be tight) when the over-identifying restrictions are violated. We address that problem by testing separately the validity of the over-identifying restrictions using the Hansen (1982) test, which, when computed using the CUE is robust to weak identification, see Kleibergen and Mavroeidis (2006). Another possible weakness of the ARSW test is its lack of power when the model is heavily over-identified, which reinforces the case for using a small number of instruments.

The second identification-robust test we use is that proposed by Kleibergen (2005). Kleibergen derives a particular orthogonal decomposition of the ARSW statistic that overcomes the aforementioned weaknesses of the ARSW test. Kleibergen shows that the ARSW statistic $S(\vartheta)$ can be decomposed into two asymptotically orthogonal components called $K(\vartheta)$ and $J(\vartheta)$. The former is a quadratic form involving the derivative of $S(\vartheta)$ w.r.t. ϑ, and in large samples, it has a χ^2 distribution with degrees of freedom equal to the number of parameters. Thus, a test based on that statistic is a particular type of Lagrange multiplier test, so we refer to it as the KLM test. The statistic $J(\vartheta) = S(\vartheta) - K(\vartheta)$ is interpretable as a test of the over-identifying restrictions at the point ϑ. See Appendix B for formal definitions and further details on the estimation methods.

3.1 Data

The data we use and the construction of variables are described in Appendix C. Here we provide a brief summary and discuss some of the more important issues.

Wherever possible, we have used standard time-series data available from the OECD. Employment and unemployment are the quarterly averages of the monthly series reported in OECD Economic Indicators, the former measured in heads (not hours). Labour force is measured as the sum of unemployment and employment. Productivity and wages are constructed from National Accounts data. Productivity is measured by GDP in fixed prices divided by employment in heads, and wages by compensation of all employees in fixed prices, gross of employment-related taxes imposed on both employers and employees, divided by employment in heads. Inflation is, as mentioned, measured by log changes in the GDP deflator and the CPI is an inclusive consumer price index. The tax measures included in the wedge can be found in the OECD National Accounts. The sample covers the last four decades for the US and last two decades for the UK. Because we use quarterly data, we specify the discount factor as

$$\delta_t = \frac{1}{1 + r_t/4}, \tag{22}$$

where r_t as the *annualised gross* real interest rate.

Data for vacancy stocks and flows, where available, are obtained directly from national sources. At the level of aggregation of this model, there are no data series equivalent to ρ_t. A series that accords with the definitions in the model can be calculated from data on vacancy stocks and flows by combining (1) and (3):

$$\rho_t = \left(J_t + V_t - V_t^c\right) / \left(J_{t-1} + V_{t-1}\right). \tag{23}$$

That is what we have used for the UK. For the US, no vacancy flow data is available. Separations data that can be used to construct a series for job destruction $(1-\rho_t)$ directly is available from the Job Openings and Labor Turnover Survey (JOLTS) but only from December 2000. So we adopted the time-honoured practice discussed by Blanchard and Diamond (1990) of constructing a series for job destructions from the number of short-term unemployed, in our case (because we are using quarterly data) those with spells shorter than 14 weeks. Moreover, if the increase in the labour force all goes through the unemployment pool first, then this increase should be subtracted from the short-term unemployed before calculating the job destruction rate. We adjusted the data for this, though the effect on the calculated ρ_t is very small. We also made an adjustment for direct job-to-job flows using the procedure suggested in Shimer (2005) based on the idea that, on average, a worker losing a job has half a period to find a new one before being recorded as unemployed. In our notation, the formula is

$$\rho_t = \frac{\text{short-term unemployment rate}_t - \text{increase in labor force}_t}{j_{t-1}\left(1 - \frac{1}{2}\frac{m_t}{u_t}\right)}.$$

Use of (2) and (6) respectively to substitute for m_t and u_t in this enables us to solve for a series for ρ_t that is consistent with the model.[5] We scaled the resulting series to the mean level of matches in the JOLTS data over the period for which that is available. With data on ρ_t, a series for V_t^c can be constructed from (23) as

$$V_t^c = V_t + J_t - \rho_t\left(V_{t-1} + J_{t-1}\right).$$

[5]Even with the adjustment suggested by Shimer (2005), the measure of separations does not include workers moving directly from jobs to self-employment or leaving the labour force but it is not clear how to allow for that.

3.2 Matching function

It is conventional to estimate matching functions as a relationship between matches, vacancies and unemployment. It is, however, more consistent with the theory to replace vacancies by our measure of the vacancy rate v_t and unemployment by our measure of the job-seeking rate u_t because these correspond to the numbers of vacancies and workers respectively who are seeking matches at the time matching takes place. For the form of the function, we adopt the Cobb-Douglas formulation used widely in the literature. There is, however, considerable empirical evidence of serial correlation with that formulation, see Petrongolo and Pissarides (2001). So, to avoid mis-specification of the short-run dynamics biasing our estimates, we use a partial adjustment model to account for those dynamics. Thus, the empirical version of the matching function takes the form

$$\ln m_t = \lambda_m \ln m_{t-1} + (1-\lambda_m)(\ln \alpha_0 + \alpha_1 \ln v_t + \alpha_2 \ln u_t) + SRD + \varepsilon_t^m, \qquad 0 < \alpha_0, \alpha_1, \alpha_2 \leq 1, \quad (24)$$

where ε_t^m is a structural shock to the matching rate that satisfies $E_{t-1}\varepsilon_t^m = 0$ and SRD denotes additional terms needed to account for short-run dynamics, determined empirically so that the disturbance ε_t^m is serially uncorrelated. Our preliminary single-equation estimation indicates that SRD should include $\Delta \ln u_{t-1}$ for the UK, and both $\Delta \ln v_t$ and $\Delta \ln u_t$ (but with coefficients that sum to one) for the US. For convenience in referring to the long-run parameters in this, we define the parameter vector $\alpha = (\alpha_0, \alpha_1, \alpha_2)$. The assumption $E_{t-1}\varepsilon_t^m = 0$ implies that we can estimate α, λ_m and $\alpha_{\Delta \ln u}$ by GMM using lags of $\ln m_t, \ln v_t$ and $\ln u_t$ as instruments.

3.3 Job creation equation

The second equation to be estimated empirically is the job creation equation (10). For empirical purposes it is convenient to scale (10) by p_{t-n} and to define the new variables

$$\widetilde{\delta}_t = \delta_{t-1}\rho_t\theta_t$$

$$\widetilde{\delta}_{t,j} = \prod_{i=1}^{j} \widetilde{\delta}_{t+i} = \widetilde{\delta}_{t+1} \ldots \widetilde{\delta}_{t+j}, \text{ for } j \geq 1, \text{ and } \widetilde{\delta}_{t,0} = 1,$$

where $\theta_t = p_t/p_{t-1}$, as before. The variable $\widetilde{\delta}_t$ has the interpretation of a one-period-ahead effective discount factor, while $\widetilde{\delta}_{t,j}$ is a j-period ahead effective discount factor, in both cases allowing for the probability of job destruction and productivity growth. Since we have no data for the term appearing on the right-hand side of equation (10), we make the over-identifying assumption (which we test) that its conditional expectation is zero, and derive the following estimable specification of the job creation equation:

$$E_{t-n}\left\{ \frac{m_t}{v_t}\widetilde{\delta}_{t-n,n} \left[\sum_{j=0}^{n-1} \left(1 - \frac{w_{t+j}}{p_{t+j}}\right) \widetilde{\delta}_{t,j} + \frac{v_{t+n}}{m_{t+n}} \left[\frac{\Psi_t}{p_t} + \frac{\psi_{t+n}}{p_{t+n}}\widetilde{\delta}_{t,n} - \widetilde{\delta}_{t+1}\left(1 - \frac{m_{t+n}}{v_{t+n}}\right)\frac{\Psi_{t+1}}{p_{t+1}} \right] \right] + \left(1 - \frac{m_t}{v_t}\right)\widetilde{\delta}_{t-n+1}\frac{\Psi_{t-n+1}}{p_{t-n+1}} - \frac{\Psi_{t-n}}{p_{t-n}} - \frac{\psi_t}{p_t}\widetilde{\delta}_{t-n,n} \right\} = 0, \text{ for all } t. \quad (25)$$

To estimate this equation we need to model the cost of creating vacancies Ψ_t and the hiring cost ψ_t. In the very long run, the unemployment rate in the countries discussed here is untrended, so we need to ensure that the model can reproduce this. For the existence of a stationary long-run equilibrium, the costs Ψ_t and ψ_t must trend with worker productivity so that Ψ_t/p_t and ψ_t/p_t are stationary. In addition to this requirement, our model for Ψ_t is motivated by the following considerations. First, the costs of creating a job ready to be filled at time t may be incurred at any time from $t-n$ (when the decision to create it is made) to t. Second, we allow for the possibility that vacancy creation costs may also depend on the number of vacancies created at any time during period $t-n$ to t. This could be due to externalities or economies of scale in job creation. To capture these considerations, we use the specification

$$E\left(\frac{\Psi_{t-n}}{p_{t-n}}\right)=\sum_{j=0}^{n}\widetilde{\delta}_{t-n,j}\left(\gamma_0+\gamma_{1,j}v_{t-n+j}^{c}\right). \tag{26}$$

Because the coefficients $\gamma_{1,j}$, $j=1,...,n$ are not, in fact, significantly different from zero for either country, we set them to zero and use the parsimonious specification

$$E\left(\frac{\Psi_{t-n}}{p_{t-n}}\right)=\sum_{j=0}^{n}\widetilde{\delta}_{t-n,j}\gamma_0+\gamma_1 v_{t-n}^{c}. \tag{27}$$

For the hiring cost ψ_t we adopt the simple specification

$$\frac{\psi_t}{p_t}=\gamma_h. \tag{28}$$

Using (27) and (28) in (25), we derive the empirical version of the job creation equation

$$\begin{aligned}
&\frac{m_t}{v_t}\widetilde{\delta}_{t-n,n}\left\{\sum_{i=0}^{n-1}\left(1-\frac{w_{t+i}}{p_{t+i}}\right)\widetilde{\delta}_{t,j}+\frac{v_{t+n}}{m_{t+n}}\left[\sum_{j=0}^{n}\widetilde{\delta}_{t,.j}\gamma_0+\gamma_1 v_t^c+\gamma_h\widetilde{\delta}_{t,n}\right.\right.\\
&\quad\left.\left.-\widetilde{\delta}_{t+1}\left(1-\frac{m_{t+n}}{v_{t+n}}\right)\left(\sum_{j=0}^{n}\widetilde{\delta}_{t+1,.j}\gamma_0+\gamma_1 v_{t+1}^c\right)\right]\right\}\\
&\quad+\left(1-\frac{m_t}{v_t}\right)\widetilde{\delta}_{t+1-n}\left(\sum_{j=0}^{n}\widetilde{\delta}_{t-n+1,.j}\gamma_0+\gamma_1 v_{t-n+1}^c\right)\\
&\qquad\qquad-\sum_{j=0}^{n}\widetilde{\delta}_{t-n,.j}\gamma_0+\gamma_1 v_{t-n}^c-\gamma_h\widetilde{\delta}_{t-n,n}=\varepsilon_t^{jc},\quad E_{t-n}\varepsilon_t^{jc}=0,\quad \text{for all } t.
\end{aligned} \tag{29}$$

For convenience, we refer to the parameters of this equation by the vector $\gamma=(\gamma_0,\gamma_1,\gamma_h)$.

Despite its complicated appearance, Equation (29) is linear in the parameters γ for every choice of the integer n, and can be estimated by linear GMM using variables known at date $t-n$ as instruments. Note that the error process ε_t^{jc} is *not* a mean innovation process and, in particular, it may exhibit serial correlation up to order $2n-1$ without invalidating the model.

To render equation (29) estimable we need to specify n. Since there are no theoretical arguments for any particular choice of n, and since for any choice of n we lose $2n$ observations in the sample, we set n as the smallest value for which the moment conditions

$E_{t-n}\varepsilon_t^{jc} = 0$ are satisfied and the residuals do not exhibit autocorrelation beyond lag $2n-1$. Our preliminary single-equation estimation indicates that $n = 2$ for the UK and $n = 3$ for the US.

For the UK, the coefficients γ_0 and γ_1 in the job creation cost are significant. However, the coefficient of hiring costs γ_h is insignificantly different from zero, so we impose that restriction on the model. For the US, the results are the opposite, namely γ_h is highly significant, while γ_0 and γ_1 are not significantly different from zero. (In fact, the point estimate for those parameters is slightly negative). Hence, we impose the restriction that job creation costs in the US are zero. The validity of those restrictions is tested using both Wald and identification-robust tests, see Table 5 in Appendix D.

We also tested the assumption that the right-hand side of (10) is zero using the Hansen test of over-identifying restrictions and found no evidence against the validity of that assumption. Finally, our model predicts that the residuals ε_t^{jc} should not exhibit serial correlation beyond lag $2n$. We tested this implication using the test proposed by Cumby and Huizinga (1992), and found no evidence against the null of no excess serial correlation (see Table 5 in Appendix D).

3.4 Non-linear wage equation

Our primary wage equation is (18). We start by using this to test whether wage determination is consistent with bargaining when matching frictions are the sole source of unemployment, with no efficiency wages. In that case, as explained in Section 2.3, the wage equation should converge to a reverse-L shaped labour supply curve ($h(x_t) \to 0$) as unemployment arising from matching frictions goes to zero. As unemployment goes to zero, m_t/u_t goes to one, so we can test the restriction on $h(x_t)$ using the parametric model

$$h(x_t) = h_1\left(\frac{m_t}{u_t}\right) + \beta_z' z_t, \tag{30}$$

where z_t is a vector of variables that are not known at time $t-2$, in accordance with our forward-looking rational expectations identification restriction. A necessary condition for $h(x_t) \to 0$ as $m_t/u_t \to 1$ is $h_1(1) = 0$ and $\beta_z = 0$. The simplest test of that hypothesis is to specify $h_1(m_t/u_t) = \beta_m \ln(m_t/u_t)$ and include only a constant in z_t. For robustness, we considered alternative parametrizations of $h_1(m_t/u_t)$ with the property that $h_1(1) = 0$. To capture potential non-linearity in h_1, we considered polynomials in $\ln(m_t/u_t)$ and $(m_t/u_t - 1)$. We also tried replacing m_t/u_t by v_t/u_t, a measure of labour market tightness widely used in bargaining models with matching frictions, which also goes to one as matching frictions go to zero. In every case, the restriction $\beta_z = 0$ was resoundingly rejected by all tests (at significance level less than 0.1%). This also rules out the hypothesis that the labour market is perfectly competitive because that requires $h(x_t) = 0$ for all t, as explained in Section 2.3.

In view of this result, we turn to the specification of $h(x_t)$ in (20) that combines matching frictions with efficiency wages. To model $f(x_t)$ in (20), we use a parametric function of relevant variables x_t that satisfies the necessary non-negativity constraint

$$f(x_t) = \exp(\beta' x_t). \tag{31}$$

With this specification of f, (18) can be estimated by GMM using variables known at time $t-2$ as instruments.[6]

[6] The formulation in (18) allows us to incorporate a disutility component to working that is not avoided

The parameter vector β is not identified separately from c/p in (18) if x_t contains only a constant but it is straightforward to test whether this identification problem arises. The null hypothesis can be formulated as $f(x_t) \equiv 0$ in (20) and tested using the Hansen test of over-identifying restrictions. Under the null hypothesis, the model reduces to a linear regression of

$$Y_t \equiv \widetilde{\delta}_t \left(1 - \frac{m_t}{u_t}\right) \left(\tau_t \frac{w_t}{p_t} - \frac{b_t}{p_t}\right) \tag{32}$$

on a constant and implies that any variable known at $t-2$ must be uncorrelated with Y_t. A test of over-identifying restrictions in this case is simply an F-test of exclusion restrictions for any set of variables known at time $t-2$. This test rejects very strongly (at significance levels less than 0.1%) even when only Y_{t-2} is used as an instrument, indicating that an identification problem due to $f(x_t)$ being a constant does not arise. An additional implication of this test is that the no-shirking condition cannot bind at all t because a necessary (but *not* sufficient) condition for this is that $f(x_t)$ is constant, a rejection of the basic efficiency wage model of Shapiro and Stiglitz (1984) in which the the wage always corresponds to a point on the no-shirking condition.[7]

In accordance with our identification assumption that the wage equation should be forward-looking, we exclude from x_t any variables that are known at time $t-2$. Rational expectations then imply that we can use all those variables as instruments. In an over-identified model, the validity of this assumption is testable using the Hansen test and a test of residual autocorrelation. Thus, the regressors x_t in (31) include no more than the first lag of the other variables that appear in the wage equation (18), that is, the variables in (32). Our choice of other regressors satisfies the "payoff relevance" criterion that, in a forward-looking model, wages in the long run should be determined by only those variables that affect the payoffs of the two parties from a wage agreement and their payoffs in the case of failure to reach an agreement. For the firm, those payoffs are Π_t and $\bar{\Pi}_t$ given by (7) and (8). For the worker, the relevant payoffs are W_t and $\bar{W}_t$, given by (12) and (13). It follows from inspection of those equations that the only payoff relevant variable in addition to those that appear in (32), namely $\delta_t, \rho_t, \theta_t, b_t/p_t$, τ_t and $1 - m_t/u_t$, is m_t/v_t.[8] However, preliminary estimation indicated that the coefficient on the regressor m_t/v_t (or alternatively v_t/u_t) is insignificant and that the results do not change significantly if this variable is excluded from the model. In a highly non-linear model like this one, over-parametrization causes problems with convergence, in addition to the usual loss in efficiency. Therefore, we use a parsimonious specification of $f(x_t)$ that contains only those variables that are significant and that passes the Hansen and residual autocorrelation tests. The specification of the wage equation that fits the data

---

by shirking (and thus not captured by c) in the form of a constant added to b_t/p_t. Since, however, we never estimated a value for this constant significantly different from zero, we do not complicate the exposition by incorporating it formally.

[7] This conclusion does not depend on the assumption that the deviations of c_t/p_t from its long-run value are iid. Allowing c_t/p_t to deviate from its long-run value c/p by some finite-order unobserved moving average process, we still find evidence that the no-shirking condition cannot bind at all t.

[8] Note that, by (28), ψ_t/p_t is constant, and the cost Ψ_{t-n} of creating a vacancy has already been incurred at the time a vacancy becomes available for matching and so is no longer payoff relevant.

best was found to be

$$\widetilde{\delta}_t\left(1-\frac{m_t}{u_t}\right)\tau_t\left(\frac{w_t}{p_t}-\frac{b_t}{p_t\tau_t}\right)$$
$$=\frac{c}{p}\left[1+\exp\left(\beta_0+\beta_u\ln\frac{U_{t-1}}{u_{t-1}}+SRD\right)\right]+\widetilde{\varepsilon}_t^w,\qquad E_{t-2}\widetilde{\varepsilon}_t^w=0,\tag{33}$$

where SRD includes $\ln b_{t-1}/p_{t-1}$ and $\ln\tau_{t-1}$ in deviations from their long-run values. For the UK, SRD also includes $\Delta\ln U_t/u_t$ and $\Delta\ln\tau_t$ to ensure $\widetilde{\varepsilon}_t^w$ is not autocorrelated beyond lag 1.[9]

3.5 Conventional log-linear wage equation

In addition to the non-linear wage equation (18), we use a standard log-linear specification to check the robustness of our results. We use as the dependent variable $\ln y_t$, where y_t is defined by

$$y_t=\frac{w_t}{p_t}-\frac{b_t}{p_t\tau_t}.\tag{34}$$

Thus y_t is the wage share w_t/p_t in excess of the unemployment benefit $b_t/(p_t\tau_t)$ before tax. To avoid imposing the impact of benefits on wages, we also investigate whether $b_t/(p_t\tau_t)$ enters significantly separately on the right-hand side.

As standard in the literature, we use a partial adjustment model to account for the short-run dynamics in $\ln y_t$. Our choice of potentially relevant regressors is guided by the literature cited in Section 2.3 and by the "payoff relevance" discussed in the previous section. The variables relevant to the firm's payoff (other than the wage share w_t/p_t that is being determined) are δ_t,ρ_t,θ_t and m_t/v_t. The only additional variables that enter the relevant payoffs for the worker (apart from the disutility of effort c_t which is unobserved) are b_t/p_t, τ_t and m_t/u_t. It is, however, conventional to use the unemployment rate $U_t\equiv 1-j_t$ in wage equations and also to include variables for inflation surprises $\Delta INFL_t$ and union density ud_t. The definitions (2) and (6) imply that $m_t/u_t=1-U_t/u_t$ so we can capture the effect of changes in m_t/u_t by changes in U_t/u_t. Moreover, matching models typically measure labour market tightness by the ratio of vacancies to unemployment corresponding to our v_t/u_t rather than in terms of the two probabilities of matching corresponding to m_t/v_t and m_t/u_t.[10] Since the effect of labour market tightness on the wage is important for the impact of matching frictions on unemployment, we want to avoid an idiosyncratic formulation inconsistent with the formulation in the matching literature. For these reasons we use, as a reasonable encompassing specification, the form

$$\ln y_t=\lambda_w\ln y_{t-1}+(1-\lambda_w)\left(\widetilde{\beta}_0+\widetilde{\beta}_U\ln U_t+\widetilde{\beta}_u\ln\frac{U_t}{u_t}+\widetilde{\beta}_v\ln\frac{v_t}{u_t}\right)+SRD+\varepsilon_t^w,\tag{35}$$

where the tilde distinguishes the parameters from those of the non-linear wage equation (33), ε_t^w is a structural disturbance that is assumed to be an innovation with respect to past information and SRD denotes additional terms for short-run dynamics containing

[9] The coefficients of the $\Delta\ln U_t/u_t$ and $\Delta\ln\tau_t$ are opposite in sign and not significantly different in magnitude, so this restriction has been imposed in estimation and only one of them is reported in the table.

[10] The hypothesis that the variables $\ln(m_t/v_t)$ and $\ln(m_t/u_t)$ enter the model with coefficients of equal magnitude and opposite sign could not be rejected by any test at very high levels of significance.

the variables $(\delta_t, \rho_t, \theta_t, b_t/p_t, \tau_t, \Delta INFL_t,\ ud_t)$ in deviation from their long-run levels, as well as lags of $\Delta \ln y_t$. Such terms are included up to the point where ε_t^w is serially uncorrelated. Note that b_t/p_t and τ_t affect the long-run wage share because they are included in y_t by the definition (34). The specification (35) permits $\ln U_t, \ln u_t$ and $\ln v_t$ to affect the long-run wage share independently while keeping separate the term in $\ln v_t/u_t$ that is typically used in matching models.

As with the other equations, we pare down the list of candidate regressors to a parsimonious specification for system estimation using preliminary single-equation estimation. That analysis indicates that the model should include up to four lags of $\Delta \ln y_t$ for the US, and two lags of $\Delta \ln y_t$ for the UK. Tests of exclusion restrictions using the ARSW and KLM tests that are robust to weak instruments establish that the variables $\delta_t, \rho_t, \theta_t, b_t/p_t, \Delta INFL_t$ and ud_t can be excluded from SRD for the US and all these plus τ_t for the UK. (Details are available from the authors on request.)

As regards the long-run wage share, the regressors $\ln U_t$, $\ln (U_t/u_t)$ and $\ln (v/u)$ are highly correlated, causing the coefficients $\widetilde{\beta}_U$ $\widetilde{\beta}_u$ and $\widetilde{\beta}_v$ to be imprecisely estimated. In fact, $\widetilde{\beta}_U$ is not statistically significant when $\ln (U_t/u_t)$ is in the model, so we set it to zero. Moreover, even in the parsimonious formulation with all the above restrictions imposed, the coefficient $\widetilde{\beta}_v$ is not significantly different from zero (at over 40% level of significance). This is corroborated by both the ARSW test and the KLM test that are robust to weak instruments (see Table 6 in Appendix D for details). In the final specification of the conventional wage equation, we therefore set $\widetilde{\beta}_v = \widetilde{\beta}_U = 0$ to arrive at the following parsimonious model

$$\ln y_t = \lambda_w \ln y_{t-1} + (1 - \lambda_w) \left(\widetilde{\beta}_0 + \widetilde{\beta}_u \ln \frac{U_t}{u_t} \right) + SRD + \varepsilon_t^w. \tag{36}$$

Matching frictions then affect the wage through the variable U_t/u_t which equals $1-m_t/u_t$.

4 Estimation results

System estimates for both countries are reported in Table 1, with conventional standard errors in parentheses. As a specification test, we report the Hansen (1982) test of overidentifying restrictions based on system estimates, with p-values in square brackets. The validity of the over-identifying restrictions is not rejected at over 20% significance level for either country. We also performed the J-test on the single-equation estimates for each equation and failed to reject at over 20% level for each of the equations in each country, see Appendix D. We found no evidence of serial correlation in the residuals of each equation, $\varepsilon_t^m, \varepsilon_t^{jc}$, ε_t^w and $\tilde{\varepsilon}_t^w$ beyond what is implied by the model. Thus the system of model equations seem consistent with the data. The two different wage specifications make remarkably little difference to the parameter estimates of the matching equation and the job-creation equation, so we do not distinguish between them in our discussion of those two equations. Note that all of the structural parameters of interest are significantly different from zero at the usual 5% level using t and ARSW and KLM tests.

In the matching equation, the elasticities α_1 with respect to the vacancy rate and α_2 with respect to the job seeking rate are highly significant for both countries. The former is higher for the UK than the US, the latter lower. In both cases, matching frictions play a statistically significant role. We investigate this formally by testing the hypothesis

$$H_0 : \alpha_0 = \alpha_1 = 1, \alpha_2 = 0.$$

Table 1: System estimates for alternative specifications

	US		UK	
Specification	(1)	(2)	(1)	(2)
Matching function				
$\log(\alpha_0)$	-0.94 (0.06)	-0.95 (0.05)	-1.00 (0.15)	-0.97 (0.21)
α_1	0.19 (0.03)	0.17 (0.03)	0.60 (0.05)	0.60 (0.07)
α_2	0.55 (0.02)	0.57 (0.02)	0.23 (0.04)	0.24 (0.05)
λ_m	0.72 (0.03)	0.72 (0.03)	0.58 (0.05)	0.67 (0.05)
$a_{\Delta lnu}$	-0.49 (0.04)	-0.48 (0.04)	0.45 (0.11)	0.32 (0.10)
Job-creation equation				
γ_h	1.97 (0.06)	1.99 (0.06)	-	-
γ_0	-	-	1.88 (0.09)	1.91 (0.09)
γ_1	-	-	-33.05 (6.39)	-36.52 (5.70)
Non-linear wage equation				
c/p	0.11 (0.01)	-	0.23 (0.003)	-
β_0	0.56 (0.18)	-	2.90 (0.88)	-
β_u	2.11 (0.37)	-	15.29 (3.19)	-
$b_{b/p}$	-0.44 (0.09)	-	-3.86 (0.75)	-
b_τ	4.28 (0.85)	-	2.18 (2.44)	-
$b_{\Delta lnU}$	-	-	29.95 (6.32)	-
Log-linear wage equation				
$\widetilde{\beta}_0$	-	-0.91 (0.10)	-	-0.65 (0.03)
$\widetilde{\beta}_u$	-	-0.38 (0.10)	-	-0.51 (0.10)
λ_w	-	0.90 (0.02)	-	0.82 (0.05)
$\widetilde{b}_{\Delta lny_1}$	-	-0.23 (0.09)	-	-0.31 (0.06)
$\widetilde{b}_{\Delta lny_2}$	-	-0.05 (0.06)	-	-0.21 (0.09)
$\widetilde{b}_{\Delta lny_3}$	-	-0.15 (0.05)	-	-
$\widetilde{b}_{\Delta lny_4}$	-	0.22 (0.06)	-	-
$\widetilde{b}_\tau$	-	0.03 (0.01)	-	-
Hansen Test	14.19 [0.29]	11.62 [0.48]	12.25 [0.43]	12.70 [0.39]

The model is estimated by Continuously Updated GMM with Newey-West Weight matrix. Standard errors in parentheses, p-values in square brackets. Specification (1) uses the wage equation derived as a mark-up over the no-shirking condition. Specification (2) uses a conventional log-linear wage equation. $n = 3$ for the US and $n = 2$ for the UK. Sample for US: 1961Q2 - 2001Q2; for UK: 1981Q1 - 2000Q2. The number of over-identifying restrictions is 12 in all cases.

These restrictions imply that $m/v = 1$ in the long-run so that all vacancies are filled straightaway, as would be the case if there were no matching frictions and all unemployment was the result of efficiency or high wages. The Wald, ARSW and KLM tests all reject the above hypothesis at significance levels of less than 0.1% in both countries. However, contrary to what is assumed by Hall (2005b), Shimer (2005) and others, the matching function for the US appears to have decreasing returns to scale in the long run because the hypothesis $\alpha_1 + \alpha_2 = 1$ is resoundingly rejected against the alternative $\alpha_1 + \alpha_2 < 1$ by all three of the Wald, ARSW and KLM tests. So, although our estimates are well within the ranges in the literature surveyed by Petrongolo and Pissarides (2001), we checked the sensitivity of our results to the adjustments we made to the data and to the estimation methods we used in a number of ways. Specifically, we estimated the matching function using vacancies and unemployment with no adjustment to allow for unrecorded vacancies and direct job-to-job flows of employees, using a time trend rather than an error correction mechanism, and using OLS. In all these cases, the estimates of the elasticities of matches with respect to both vacancies and unemployment differed by less than one percentage point from our system estimates in Table 1 and the hypothesis of constant returns was resoundingly rejected by all the three tests we used. For only one test we tried was this not true. That was when we did not scale the data on separations (calculated using the method suggested by Shimer (2005)) to the same mean level as the separations measured by JOLTS for the period that the JOLTS data are available. In that case the sum of the elasticities is close to 1 and the test for constant returns easily accepted. Since, however, the JOLTS data are widely considered to be the best indicator of the magnitude of separations for the US, it seems more appropriate to use estimates based on the scaled data.

In the job creation equation, the coefficient γ_h corresponds to hiring costs that are incurred each period a vacancy is available for matching, while γ_0 and γ_1 relate to costs of creating a vacancy incurred over the n periods prior to it becoming available for matching. The coefficient γ_1 in the job creation cost is negative for the UK, indicating that there are economies of scale to creating vacancies. We are somewhat sceptical that the data can actually distinguish between these two types of costs in creating a vacancy. However, what really matters for our purposes is the shape of the job-creation equation, not whether that shape arises from the capital cost of creating a vacancy or the current cost of keeping it open.

The most interesting parameter estimate from the non-linear wage equation is that for c/p. The numbers in the table for this are to be interpreted as the proportion of a worker's output that is required to deter shirking. If the efficiency wage element in the model was negligible, they should be close to zero. The point estimates are 0.11 and 0.23 for the US and the UK respectively. Both are highly significantly different from zero. So the data strongly supports the inclusion of the efficiency wage element in the model in addition to matching frictions. The parameters β_0 and β_u, together with the value of c/p, determine the location and slope of the long-run wage share equation. The point estimates imply that, in terms of Figure 1(b), the wage curve for the US lies below that for the UK, implying a lower equilibrium wage for any given level of unemployment, other things equal. This is reflected in our point estimates of the long-run equilibrium wage share in the two countries which is 0.68 in the US versus 0.77 for the UK. Finally, the positive value of β_u indicates that the mark-up of wages over the minimum necessary to deter shirking is increasing in the unemployment rate, meaning that the wage curve in Figure 1(b) slopes less steeply than the NSC. Interestingly, the estimated mark-up f at the long-run equilibrium is 0.23 for the US but only 0.01 for the UK.

In the conventional log-linear wage equation, the adjustment coefficient λ_w is higher in the US than in the UK, in line with the conventional wisdom discussed by Blanchard and Katz (1999). The difference in the estimates of $\widetilde{\beta}_u$ implies that the wage equation is also somewhat steeper in the UK than in the US.

5 Implications for long-run unemployment

It is clear from the results in Table 1 that both the matching friction and efficiency wage elements in the model are statistically highly significant. But it is not obvious from them how important those elements are in terms of the economic interest in their impact on unemployment. An obvious metric for this is their impact on the long-run level of unemployment. We assess that in this section. Specifically, we derive point estimates and confidence intervals for the long-run unemployment rate U and its components: the components attributable to matching frictions, to high wages, and to efficiency wages. We denote these components by U^f, U^{hw} and U^{eff} as illustrated in Figure 1. Estimation and inference on each of these can be done in the same way as for U.

We use *long run* to refer to values taken when all shocks are zero and all variables are either constant or in appropriate constant ratios, indicated without subscripts. The exogenously determined variables with constant long-run values include the job destruction rate ρ_t, the discount factor δ_t, the labour force growth rate l_t, the growth rate of productivity $\theta_t \equiv p_t/p_{t-1}$, the tax wedge τ_t, and the flow utility from unemployment as a proportion of the productivity b_t/p_t. The constant long-run values of the variables m_t, v_t, v_t^c, j_t, u_t and w_t/p_t are determined endogenously by the model.

The long-run parameters $\theta, \rho, \delta, l, \tau$ and b/p are determined exogenously to the model, but their values are needed to draw inferences on long-run unemployment. Estimating these jointly with the other parameters gives serious problems of convergence. Since this makes the resulting confidence sets unreliable, we use a two-step procedure. We first estimate the long-run values of the exogenous variables by their sample averages.[11] We then keep these parameters fixed at their unrestricted point estimates when doing inference on long-run unemployment. As a result, our reported confidence sets do not take account of the uncertainty in estimating $\theta, \rho, \delta, l, \tau$ and b/p. The estimates of those parameters are reported in Table 2. The numbers are broadly similar across the two countries, with the notable exception of the job destruction rate $1 - \rho$, which is twice as high in the US as in the UK.

5.1 Derivation of long-run unemployment rates

Long-run equilibrium is characterized by the long-run versions of the matching equation (24), the job creation equation (29) with the empirical form of the job creation cost (27), and one of the wage equations (33) or (36). With the non-linear wage equation (33) and

[11] The variables τ_t and b_t/p_t appear to be trending in our sample. However, since they are restricted to lie between 0 and 1, they cannot be trending in the long run. In order to use values that reflect current economic conditions, we estimate τ and b/p using the average of the last 12 quarters in the sample.

Table 2: Long-run values of exogenous variables

Parameter	US	UK
δ	0.992	0.987
θ	1.005	1.005
ρ	0.896	0.956
l	1.005	1.001
τ	0.681	0.618
b/p	0.065	0.081

recalling that $U/u = 1 - m/u$, these can be written

$$m = \alpha_0 v^{\alpha_1} u^{\alpha_2}; \tag{37}$$

$$\frac{w}{p} = 1 - \frac{1-\delta\rho\theta}{m/v}\left[\frac{\psi}{p} + \frac{\Psi(v^c)}{p}\frac{1-\delta\rho\theta(1-m/v)}{(\delta\rho\theta)^n}\right], \tag{38}$$

$$\text{where } \frac{\psi}{p} = \begin{cases} 0, & \text{for the UK;} \\ \gamma_h, & \text{for the US;} \end{cases} \tag{39}$$

$$\text{and } \frac{\Psi(v^c)}{p} = \begin{cases} \gamma_0 \frac{1-(\delta\rho\theta)^{n+1}}{1-\delta\rho\theta} + \gamma_1 v^c, & \text{for the UK;} \\ 0, & \text{for the US;} \end{cases} \tag{40}$$

$$\frac{w}{p} = \frac{c}{p}\left[1 + e^{\beta_0}\left(\frac{U}{u}\right)^{\beta_u}\right]\left(\delta\rho\theta\tau\frac{U}{u}\right)^{-1} + \frac{b}{p\tau}. \tag{41}$$

With the conventional log-linear wage equation (36), (41) is replaced by

$$\frac{w}{p} = e^{\beta_0}\left(\frac{U}{u}\right)^{\widetilde{\beta}_u} + \frac{b}{p\tau}. \tag{42}$$

The long-run values of v, v^c and u are linked to m/v and $j = 1 - U$ through the identities

$$v^c = \frac{1-U}{m/v}\left(1 - \frac{\rho}{l}\right)\left[1 - \frac{\rho}{l}\left(1 - \frac{m}{v}\right)\right] \tag{43}$$

$$v = \frac{(1-U)(1-\rho/l)}{m/v} \tag{44}$$

$$u = 1 - \frac{\rho}{l}(1-U). \tag{45}$$

It is evident that the long-run values of the endogenous variables are simply functions of the long-run parameters of the model and so is the long-run unemployment rate U. This is defined implicitly by the intersection of the long-run job creation and wage curves given by (38) and (42), or alternatively (41), as in Figure 1. In other words, the system of equations (37), (38), and (41) or (42), together with the identities (43) to (45), define U implicitly as a function of all the long-run structural parameters α, γ and β or $\widetilde{\beta}$, and the long-run values of exogenous variables, denoted here by $(\delta\rho\theta, \tau, \rho/l, b/p)$. We can define the components of U illustrated in Figure 1 in an analogous way. U^f is difference between U and the level of unemployment that would arise if m/v were 1. Similarly, U^{hw} is the difference between U and the level of unemployment that would arise if the mark-up of wages over the minimum necessary to deter shirking, f in (33) or $e^{\beta_0}(U/u)^{\beta_u}$ in (41), were equal to 0. Finally, U^{eff} is what remains if both $m/v = 1$ and $f = 0$. Clearly, U^{hw} and U^{eff} are only identified from the non-linear wage equation (41).

Denote by $g(\vartheta)$ the implicit 4-dimensional function that maps the parameters of the model ϑ to $\left(U, U^f, U^{hw}, U^{eff}\right)$. Point estimates of U, U^f, U^{hw} and U^{eff} are obtained simply by evaluating the function $g(\cdot)$ at the CUE of ϑ.

5.2 Inference on long-run unemployment

Point estimates alone, however, are of limited use. To make inferences about the magnitude of the long-run unemployment rates U, U^f, U^{hw} and U^{eff}, we need to quantify the uncertainty surrounding those point estimates. We can construct 95% confidence sets for each parameter by inverting a particular test, that is, by collecting all the points of U that are not rejected by that test at the 5% level of significance. One approach is to derive asymptotic standard errors using the delta method and construct approximate 95% level confidence intervals by the usual two-standard-error bands about the point estimate, see Appendix D for details.

Wald-based confidence sets are potentially problematic because they are not robust to weak identification. As explained in Stock, Wright, and Yogo (2002), when instruments are weak the asymptotic standard errors cannot be estimated consistently, nor is the distribution of any Wald statistic approximately χ^2. As a result, confidence sets with nominal 95% coverage rate may contain the true value of the parameter much less often than 95% (i.e., they could be too tight). Therefore, to address concerns about identification, we also derive identification-robust confidence sets by inverting the ARSW and KLM tests, see Appendix B for details.

5.3 Long-run unemployment and its components

Tables 3 and 4 report point estimates, standard errors and three alternative 95% confidence intervals for U, U^f, U^{hw} and U^{eff} for the US and the UK respectively. Estimates are derived using both the non-linear and conventional log-linear specifications of the wage equation, though only for the former can we identify U^{hw} and U^{eff}. The Wald-based confidence intervals are symmetric about the point estimate by construction but the other two confidence intervals are not. This is standard. Confidence intervals derived by inverting tests (for example, likelihood ratio or score tests) are asymmetric except in very special cases (for example, when they are numerically equivalent to Wald confidence intervals). So, the asymmetry of the intervals reported here has nothing inherently to do with their being robust to weak identification.

For the non-linear wage equation in the UK, the Wald confidence bounds for U and U^f are only marginally tighter than the ARSW and KLM confidence bounds. Since the Wald bounds are not robust to weak identification but the other two are, this indicates U and U^f are well-identified. But even for the other cases, the ARSW and KLM bounds are sufficiently tight to provide valuable economic information.

For the US, the two different specifications of the wage equation result in essentially identical point estimates (5.9% and 5.8%) for the long-run unemployment rate (see the first column in Table 3), though the conventional log-linear wage equation gives slightly tighter confidence bounds, at least when computed by methods robust to weak identification. These are remarkably close to the sample average unemployment rate of 5.8%. For the UK, the non-linear wage equation gives a similar point estimate for the long-run

Table 3: Estimates and 95% confidence bounds for long-run unemployment and its components, US

	Unemployment rate			
Specification	total	frictional	high wage	efficiency wage
Non-linear wage eq.				
Point estimate	5.9%	1.7%	0.7%	3.5%
Standard error	0.1%	0.2%	0.2%	0.4%
Wald [min, max]	[5.6, 6.2]	[1.4, 2.1]	[0.3, 1.1]	[2.6, 4.3]
ARSW [min, max]	[4.1, 6.9]	[0.3, 2.3]	[0.6, 2.8]	[2.1, 5.6]
KLM [min, max]	[4.1, 6.9]	[0.7, 2.1]	[0.6, 1.4]	[2.4, 5.6]
Log-linear wage eq.				
Point estimate	5.8%	2.6%	-	-
Standard error	0.1%	0.5%	-	-
Wald [min, max]	[5.6, 6.2]	[1.7, 3.6]	-	-
ARSW [min, max]	[5.4, 6.4]	[0.8, 4.5]	-	-
KLM [min, max]	[5.7, 6.2]	[1.5, 3.8]	-	-

Standard errors are computed using the Delta method. Confidence bounds are reported in square brackets. ARSW refers to the Anderson-Rubin-Stock-Wright test, KLM refers to the KLM test.

Table 4: Estimates and 95% confidence bounds for long-run unemployment and its components, UK

	Unemployment rate			
Specification	total	frictional	high wage	efficiency wage
Non-linear wage eq.				
Point Estimate	6.1%	0.11%	0.2%	5.8%
Standard Error	0.2%	0.02%	0.1%	0.2%
Wald [min, max]	[5.6, 6.5]	[0.07, 0.15]	[0.0, 0.4]	[5.5, 6.2]
ARSW [min, max]	[5.6, 6.8]	[0.07, 0.27]	[0.04, 1.5]	[4.3, 6.4]
KLM [min, max]	[5.7, 6.5]	[0.09, 0.20]	[0.04, 3.2]	[1.8, 6.2]
Log-linear wage eq.				
Point estimate	7.3%	0.2%	-	-
Standard error	0.5%	0.04%	-	-
Wald [min, max]	[6.4, 8.2]	[0.1, 0.3]	-	-
ARSW [min, max]	[5.9, 9.2]	[0.2, 0.3]	-	-
KLM [min, max]	[6.4, 8.3]	[0.1, 0.3]	-	-

Standard errors are computed using the Delta method. Confidence bounds are reported in square brackets. ARSW refers to the Anderson-Rubin-Stock-Wright test, KLM refers to the KLM test.

unemployment rate (6.1%) as for the US. The conventional log-linear wage equation, however, gives a higher point estimate of 7.3%. Moreover, only the ARSW confidence bounds include the point estimate for the non-linear wage equation, which gives us less confidence in our estimates of the long-run unemployment rate for the UK than for the US. The point estimates for both UK wage equations are substantially below the sample average unemployment rate of 9%, which would however seem implausibly high for the long-run unemployment rate under current conditions. Over the same sample period as for the US, the average unemployment rate for the UK was 5.8%, which seems a more plausible level for the long-run unemployment rate and is remarkably close to that estimated using the non-linear wage equation. It is reassuring that the model estimates the long-run unemployment rate under current conditions at a plausible level despite that being substantially below the sample average.

The second columns of Tables 3 and 4 give values for the component due to matching frictions, U^f. For the US it is estimated at about 1.7% using the non-linear age equation (33). For the log-linear wage equation (36) the effect, at 2.6%, is rather bigger, though the confidence sets all include the point estimate for the non-linear wage equation. In both cases, it is statistically significant, although the confidence bounds (particularly the robust confidence bounds) indicate that it cannot be pinned down very precisely. For the UK, the point estimates of the component due to matching frictions are much smaller, 0.11% with the non-linear wage equation and 0.2% with the log-linear one. But, with both wage equations, the confidence bounds on the point estimates are very tight so that, despite being small, the point estimates are still significantly different from zero.

The third columns of Tables 3 and 4 give the component due to high wages, U^{hw}. That can be estimated only with the non-linear wage equation. The point estimates for the two countries are not greatly different (0.7%.for the US and 0.2% for the UK), but the robust (ARSW and KLM) confidence bounds indicated that it is not very precisely estimated.

The final columns of Tables 3 and 4 give values for the efficiency wage component of unemployment, the component remaining when both matching frictions and high wages are removed. Again, that can be estimated only with the non-linear wage equation. For the US, the point estimate is a substantial 3.5%, for the UK an even more substantial 5.8%. In both countries, it is larger than the other two components taken together. Even the widest confidence sets indicate that around 2% long-run unemployment can be attributed to efficiency wages in both countries. For the UK, however, the much wider confidence bounds for the KLM than for the ARSW test for both high wages and efficiency wages are suggestive of a known spurious decline in power of the KLM test against alternatives that correspond to points of inflection or local minima of the GMM objective function, see Kleibergen (2005). In such cases, the ARSW confidence bounds are more informative. But, whatever the reason, the issue arises only with respect to distinguishing between high wage and efficiency wage unemployment. All three tests agree that the difference between their sum and the total, which corresponds to the frictional component, is accurately estimated and very small. There is also another slight caveat here. Since our measure of efficiency wage unemployment is the residual after removing the other components, systematic measurement error in unemployment rates will affect it. Nonetheless, the magnitude and significance of our results indicates that there really does seem to be a need for both matching frictions and efficiency wages to account for unemployment in the US and the UK.

The approach used here is only one of the possible ways to measure the impact of matching frictions and efficiency wages on unemployment, both of which we know are statistically significant from the results in Section 4. But it suffices for giving an indication of their relative magnitudes. One alternative is to measure the impact of matching frictions by the shift from E to A in Figure 1(b). That can be done only with the non-linear wage equation but, for that equation, the difference turns out to be negligible.

6 Conclusion

In this paper, we have constructed and estimated econometrically for two countries (the USA and the UK) a model that incorporates both matching frictions and efficiency wages to deter shirking. The matching friction element is essentially an econometric specification of that in Mortensen and Pissarides (1994) calibrated recently to US data by Hall (2005b), Hall (2005a) and Shimer (2005). The model of efficiency wages is essentially that of Shapiro and Stiglitz (1984) as extended in MacLeod and Malcomson (1998). The model is sufficiently tightly specified to enable the estimation to recover the underlying model parameters. That permits the data to determine the extent to which unemployment is the result of matching frictions, of efficiency wages, and of high wages (that is, wages above the minimum level required to deter shirking).

Mindful of the concern there has been in the literature about the identification of aggregate time series models of the type used here, we have used empirical methods that are robust to weak instruments. To our knowledge, this is the largest model to which these identification-robust methods have so far been applied. At a methodological level, the paper demonstrates three things. First, it shows that inference methods robust to weak instruments can be used effectively in economic models of the type estimated here. Second, the rather small differences we find between the confidence intervals based on Wald statistics and those based on the robust Anderson-Rubin-Stock-Wright and Kleibergen statistics suggest that the concerns about identification in such models have been somewhat over-played. Third, it demonstrates that the model itself, combining as it does both matching frictions and efficiency wages, can be used effectively to recover the underlying structural parameters.

The main conclusion we draw from the results of the analysis is that both matching frictions and efficiency wages play a significant role in enabling the model to fit the data. Using as a metric of their economic magnitude their contributions to the long-run unemployment rate, we find that matching frictions have a bigger effect in the US than in the UK, where (though small) they are still significant. In contrast, efficiency wages have a bigger effect in the UK than in the US. But in both countries, the contribution of efficiency wages to long-run unemployment is substantial, with point estimates of more than half the total. Given the non-prescriptive nature of our specification of wage determination, the results suggest that adding efficiency wages to matching frictions may be a better way to fit the data than simply searching for an alternative wage formulation.

References

Anderson, T. W., and H. Rubin (1949): "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Ann. Math. Statistics*, 20, 46–63.

BEAN, C. R. (1994): "European Unemployment: A Survey," *Journal of Economic Literature*, 32(2), 573–619.

BLANCHARD, O., AND L. F. KATZ (1999): "Wage Dynamics: Reconciling Theory and Evidence," *American Economic Review*, 89(2), 69–74.

BLANCHARD, O. J., AND P. DIAMOND (1989): "The Beveridge Curve," *Brookings Papers on Economic Activity*, 1, 1–60.

BLANCHARD, O. J., AND P. DIAMOND (1990): "The Cyclical Behavior of the Gross Flows of U.S. Workers," *Brookings Papers on Economic Activity*, (2), 85–143.

BLANCHFLOWER, D. G., AND A. J. OSWALD (1994): *The Wage Curve*. MIT Press, Cambridge, MA and London.

CAHUC, P., F. POSTEL-VINAY, AND J.-M. ROBIN (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74(2), 323–364.

CARMICHAEL, H. L. (1985): "Can Unemployment Be Involuntary?: Comment," *American Economic Review*, 75(5), 1213–1214.

COLE, H. L., AND R. ROGERSON (1999): "Can the Mortensen-Pissarides Matching Model Match the Business-Cycle Facts?," *International Economic Review*, 40(4), 933–959.

CUMBY, R. E., AND J. HUIZINGA (1992): "Testing the autocorrelation structure of disturbances in Ordinary Least Squares and Instrumental Variables regressions," *Econometrica*, 60(1), 185–195.

DIAMOND, P. A. (1982): "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 90(5), 881–894.

HALL, R. E. (2005a): "Employment Efficiency and Sticky Wages: Evidence from Flows in the Labor Market," *Review of Economics and Statistics*, 87(3), 397–407.

——— (2005b): "Employment Fluctuations with Equilibrium Wage Stickiness," *American Economic Review*, 95(1), 50–65.

HALL, R. E., AND P. R. MILGROM (2005): "The Limited Influence of Unmployment on the Wage Bargain," Stanford University, Department of Economics.

HANSEN, L. P. (1982): "Large Sample properties of Generalized Method of Moments estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite Sample properties of Some alternative GMM estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

KLEIBERGEN, F. (2005): "Testing parameters in GMM without assuming that they are identified," *Econometrica*, 73(4), 1103–1123.

KLEIBERGEN, F. AND S. MAVROEIDIS (2006): "Testing subsets of structural parameters in GMM without assuming identification," Mimeo, Brown University, USA.

KLEIBERGEN, F., AND R. PAAP (2003): "Generalized Reduced Rank Tests using the Singular Value Decomposition," Discussion paper 003/4, Tinbergen Institute, Amsterdam.

LAYARD, R., S. NICKELL, AND R. JACKMAN (2005): *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford University Press, Oxford, reissued edn.

MACLEOD, W. B., AND J. M. MALCOMSON (1998): "Motivation and Markets," *American Economic Review*, 88(3), 388–411.

MADSEN, J. B. (1998): "General Equilibrium Macroeconomic Models of Unemployment: Can They Explain the Unemployment Path in the OECD?," *Economic Journal*, 108(448), 850–867.

MANNING, A. (1993): "Wage Bargaining and the Philips Curve: The Identification and Specification of Aggregate Wage Equations," *Economic Journal*, 103(416), 98–118.

MAVROEIDIS, S. (2002): "Econometric Issues in Forward-Looking Monetary Models," DPhil thesis, Oxford University, Oxford.

——— (2004): "Weak identification of forward-looking models in monetary economics," *Oxford Bulletin of Economics and Statistics*, 66(Supplement), 609–635.

——— (2005): "Identification issues in forward-looking models estimated by GMM with an application to the Phillips Curve," *Journal of Money, Credit and Banking*, 37(3), 421–449.

MORTENSEN, D. T., AND C. A. PISSARIDES (1994): "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3), 397–415.

NEWEY, W. K. AND R. J. SMITH (2004): "Higher order properties of GMM and Generalized Empirical Likelihood estimators," *Econometrica*, 72(1), 219–255.

NEWEY, W. K., AND K. D. WEST (1987): "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55(3), 703–708.

NEWEY, W. K., AND K. D. WEST (1994): "Automatic Lag Selections in Covariance Matrix Estimation," *Review of Economic Studies*, 61(3), 631–653.

PETRONGOLO, B., AND C. PISSARIDES (2001): "Looking Into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 39(2), 390–431.

PISSARIDES, C. A. (1985): "Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages," *American Economic Review*, 75(4), 676–690.

——— (2000): *Equilibrium Unemployment Theory*. MIT Press, Cambridge, MA, 2nd edn.

SHAPIRO, C., AND J. E. STIGLITZ (1984): "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, 74(3), 433–444.

SHIMER, R. (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95(1), 25–49.

Stock, J., J. Wright, and M. Yogo (2002): “GMM, weak instruments, and weak identification,” *Journal of Business and Economic Statistics*, 20, 518–530.

Stock, J. H., and J. H. Wright (2000): “GMM with weak identification,” *Econometrica*, 68(5), 1055–1096.

West, K. D. (1997): “Another heteroskedasticity- and autocorrelation-consistent covariance matrix estimator,” *Journal of Econometrics*, 76, 171–191.

Yashiv, E. (2000): “The Determinants of Equilibrium Unemployment,” *American Economic Review*, 90(5), 1297–1322.

——— (2006): “Evaluating the Performance of the Search and Matching Model,” *European Economic Review*, 50, 909–936.

Appendix A Derivation of job creation equation (10)

Write (8) as

$$\frac{m_t}{v_t}\Pi_t = \bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right), \text{ for all } t,$$

so

$$\Pi_t = \frac{1}{m_t/v_t}\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right)\right], \text{ for all } t.$$

Use this in (7) to write

$$\begin{aligned}&\frac{1}{m_t/v_t}\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right)\right] = p_t - w_t \\ &+ E_t\left\{\delta_t \rho_{t+1} \frac{1}{m_{t+1}/v_{t+1}}\left[\bar{\Pi}_{t+1} + \psi_{t+1} - \left(1 - \frac{m_{t+1}}{v_{t+1}}\right) E_{t+1}\left(\delta_{t+1} \rho_{t+2} \bar{\Pi}_{t+2}\right)\right]\right\}, \text{ for all } t.\end{aligned} \tag{46}$$

Now use (46) forwarded one period to substitute for the term in square brackets on the right-hand side to get

$$\begin{aligned}&\frac{1}{m_t/v_t}\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right)\right] = p_t - w_t + E_t\left\{\delta_t \rho_{t+1}\left[p_{t+1} - w_{t+1}\right.\right. \\ &\left.\left.+ E_{t+1}\left[\delta_{t+1}\rho_{t+2}\frac{1}{m_{t+2}/v_{t+2}}\left[\bar{\Pi}_{t+2} + \psi_{t+2} - \left(1 - \frac{m_{t+2}}{v_{t+2}}\right) E_{t+2}\left(\delta_{t+2}\rho_{t+3}\bar{\Pi}_{t+3}\right)\right]\right]\right]\right\}, \text{ for all } t.\end{aligned}$$

Since $\delta_t \rho_{t+1}$ is known at the time expectations are taken at $t+1$, we can write this as

$$\begin{aligned}&\frac{1}{m_t/v_t}\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right)\right] = p_t - w_t + E_t\left\{\delta_t \rho_{t+1}\left(p_{t+1} - w_{t+1}\right)\right. \\ &\left.+\delta_t\rho_{t+1}\delta_{t+1}\rho_{t+2}\frac{1}{m_{t+2}/v_{t+2}}\left[\bar{\Pi}_{t+2} + \psi_{t+2} - \left(1 - \frac{m_{t+2}}{v_{t+2}}\right) E_{t+2}\left(\delta_{t+2}\rho_{t+3}\bar{\Pi}_{t+3}\right)\right]\right\}, \text{ for all } t.\end{aligned}$$

Now use (46) forwarded two periods to again substitute for the term in square brackets on the right-hand side to get

$$\begin{aligned}&\frac{1}{m_t/v_t}\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right)\right] \\ &= p_t - w_t + E_t\left\{\delta_t\rho_{t+1}\left(p_{t+1} - w_{t+1}\right) + \delta_t\rho_{t+1}\delta_{t+1}\rho_{t+2}\left\{p_{t+2} - w_{t+2}\right.\right. \\ &\qquad \left.\left.+\left\{E_{t+2}\left[\delta_{t+2}\rho_{t+3}\frac{1}{m_{t+3}/v_{t+3}}\left[\bar{\Pi}_{t+3} + \psi_{t+3} - \left(1 - \frac{m_{t+3}}{v_{t+3}}\right) E_{t+3}\left(\delta_{t+3}\rho_{t+4}\bar{\Pi}_{t+4}\right)\right]\right]\right\}\right\}\right\},\end{aligned}$$

or, since $\delta_t\rho_{t+1}\delta_{t+1}\rho_{t+2}$ is known at the time expectations are taken at $t+2$ and $\delta_{t+2}\rho_{t+3}$ and m_{t+3}/v_{t+3} are known at the time expectations are taken at $t+3$,

$$\begin{aligned}&\frac{1}{m_t/v_t}\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right)\right] \\ &= p_t - w_t + E_t\left\{\delta_t\rho_{t+1}\left(p_{t+1} - w_{t+1}\right) + \delta_t\rho_{t+1}\delta_{t+1}\rho_{t+2}\left(p_{t+2} - w_{t+2}\right)\right. \\ &\qquad \left.+\ \delta_t\rho_{t+1}\delta_{t+1}\rho_{t+2}\delta_{t+2}\rho_{t+3}\frac{1}{m_{t+3}/v_{t+3}}\left[\bar{\Pi}_{t+3} + \psi_{t+3} - \left(1 - \frac{m_{t+3}}{v_{t+3}}\right)\delta_{t+3}\rho_{t+4}\bar{\Pi}_{t+4}\right]\right\}, \text{ for all } t.\end{aligned}$$

With the convention $\prod_{i=1}^{j} x_i = 1$ for $j = 0$, we can write in general for any $n \geq 1$

$$
\begin{aligned}
&\frac{1}{m_t/v_t}\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t \rho_{t+1} \bar{\Pi}_{t+1}\right)\right] \\
&= E_t\left\{\sum_{j=0}^{n-1} (p_{t+j} - w_{t+j}) \left(\prod_{i=1}^{j} \delta_{t-1+i}\rho_{t+i}\right)\right. \\
&\quad + \left(\prod_{i=1}^{n} \delta_{t-1+i}\rho_{t+i}\right) \frac{1}{m_{t+n}/v_{t+n}} \left[\bar{\Pi}_{t+n} + \psi_{t+n}\right. \\
&\quad + \left.\left. - \left(1 - \frac{m_{t+n}}{v_{t+n}}\right)\left(\delta_{t+n}\rho_{t+1+n}\bar{\Pi}_{t+1+n}\right)\right]\right\}, \qquad \text{for all } t.
\end{aligned}
$$

Multiply this through by $\frac{m_t}{v_t}\left(\prod_{j=1}^{n} \delta_{t-j}\rho_{t+1-j}\right)$ and take expectations at $t-n$ to write

$$
\begin{aligned}
&E_{t-n}\left\{\left(\prod_{j=1}^{n} \delta_{t-j}\rho_{t+1-j}\right)\left[\bar{\Pi}_t + \psi_t - \left(1 - \frac{m_t}{v_t}\right) E_t\left(\delta_t\rho_{t+1}\bar{\Pi}_{t+1}\right)\right]\right\} \\
&= E_{t-n}\left\{\frac{m_t}{v_t}\left(\prod_{j=1}^{n} \delta_{t-j}\rho_{t+1-j}\right) E_t\left\{\sum_{j=0}^{n-1} (p_{t+j} - w_{t+j})\left(\prod_{i=1}^{j} \delta_{t-1+i}\rho_{t+i}\right)\right.\right. \\
&+ \frac{1}{m_{t+n}/v_{t+n}}\left[\left(\prod_{i=1}^{n} \delta_{t-1+i}\rho_{t+i}\right)\left(\bar{\Pi}_{t+n} + \psi_{t+n}\right)\right. \\
&+ \left.\left.\left. - \left(1 - \frac{m_{t+n}}{v_{t+n}}\right)\left(\prod_{i=1}^{n} \delta_{t-1+i}\rho_{t+i}\right)\delta_{t+n}\rho_{t+1+n}\bar{\Pi}_{t+1+n}\right]\right\}\right\}, \qquad \text{for all } t.
\end{aligned}
$$

Since terms in m_t/v_t, δ_t and ρ_t are known when expectations are taken at time t, this can be written with rearranged product terms

$$
\begin{aligned}
&E_{t-n}\left\{\left(\prod_{j=1}^{n} \delta_{t-j}\rho_{t+1-j}\right)\left(\bar{\Pi}_t + \psi_t\right) - \left(1 - \frac{m_t}{v_t}\right)\delta_{t-n}\rho_{t+1-n}\left(\prod_{j=1}^{n} \delta_{t+1-j}\rho_{t+2-j}\right)\bar{\Pi}_{t+1}\right\} \\
&= E_{t-n}\left\{\frac{m_t}{v_t}\left(\prod_{j=1}^{n} \delta_{t-j}\rho_{t+1-j}\right)\left[\sum_{j=0}^{n-1} (p_{t+j} - w_{t+j})\left(\prod_{i=1}^{j} \delta_{t-1+i}\rho_{t+i}\right)\right.\right. \\
&+ \frac{1}{m_{t+n}/v_{t+n}}\left(\prod_{i=1}^{n} \delta_{t-1+i}\rho_{t+i}\right)\left(\bar{\Pi}_{t+n} + \psi_{t+n}\right) \\
&+ \left.\left. - \left(\frac{1}{m_{t+n}/v_{t+n}} - 1\right)\delta_t\rho_{t+1}\left(\prod_{i=1}^{n} \delta_{t+i}\rho_{t+1+i}\right)\bar{\Pi}_{t+1+n}\right]\right\}, \qquad \text{for all } t,
\end{aligned}
$$

which, since $\delta_{t-1}\rho_t$ and m_t/v_t are known at the time expectations are taken at t, can also be written as

$$E_{t-n}\left\{\left(\bar{\Pi}_t+\psi_t\right)\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}-\delta_{t-n}\rho_{t+1-n}E_{t+1-n}\left[\left(1-\frac{m_t}{v_t}\right)\bar{\Pi}_{t+1}\prod_{j=1}^{n}\delta_{t+1-j}\rho_{t+2-j}\right]\right\}$$
$$=E_{t-n}\left\{\frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left[\sum_{j=0}^{n-1}\left(p_{t+j}-w_{t+j}\right)\left(\prod_{i=1}^{j}\delta_{t-1+i}\rho_{t+i}\right)\right.\right.$$
$$+E_t\left(\frac{1}{m_{t+n}/v_{t+n}}\left(\bar{\Pi}_{t+n}+\psi_{t+n}\right)\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}\right)$$
$$\left.\left.-\delta_t\rho_{t+1}E_{t+1}\left(\left(\frac{1}{m_{t+n}/v_{t+n}}-1\right)\bar{\Pi}_{t+1+n}\prod_{i=1}^{n}\delta_{t+i}\rho_{t+1+i}\right)\right]\right\},\ \text{for all } t.$$

or

$$-E_{t-n}\left\{\left(\bar{\Pi}_t+\psi_t\right)\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}-\delta_{t-n}\rho_{t+1-n}E_{t+1-n}\left[\left(1-\frac{m_t}{v_t}\right)\bar{\Pi}_{t+1}\prod_{j=1}^{n}\delta_{t+1-j}\rho_{t+2-j}\right]\right.$$
$$-\frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left[\sum_{j=0}^{n-1}\left(p_{t+j}-w_{t+j}\right)\left(\prod_{i=1}^{j}\delta_{t-1+i}\rho_{t+i}\right)\right.$$
$$+E_t\left(\frac{1}{m_{t+n}/v_{t+n}}\left(\bar{\Pi}_{t+n}+\psi_{t+n}\right)\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}\right)$$
$$\left.\left.-\delta_t\rho_{t+1}E_{t+1}\left(\left(\frac{1}{m_{t+n}/v_{t+n}}-1\right)\bar{\Pi}_{t+1+n}\prod_{i=1}^{n}\delta_{t+i}\rho_{t+1+i}\right)\right]\right\}=0,\ \text{for all } t. \tag{47}$$

Recall that

$$E_t\left(x_{t+n},y_{t+n}\right)-E_t\left(x_{t+n}\right)E_t\left(y_{t+n}\right)=cov_t\left(x_{t+n},y_{t+n}\right).$$

Moreover, note that

$$\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}=\prod_{j=1}^{n}\delta_{t+n-j}\rho_{t+n+1-j} \tag{48}$$

and define z_t, z'_t, z''_{t+1-n} by

$$
\begin{aligned}
z_t &\equiv cov_t\left(\frac{1}{m_{t+n}/v_{t+n}}, \bar{\Pi}_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i} - \Psi_t\right) \\
&= E_t\left[\frac{1}{m_{t+n}/v_{t+n}}\left(\bar{\Pi}_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i} - \Psi_t\right)\right] \qquad (49)\\
z'_t &\equiv cov_t\left(\frac{1}{m_{t-1+n}/v_{t-1+n}}, \bar{\Pi}_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i} - \Psi_t\right) \\
&= E_t\left[\frac{1}{m_{t-1+n}/v_{t-1+n}}\left(\bar{\Pi}_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i} - \Psi_t\right)\right] \qquad (50)\\
z''_{t+1-n} &\equiv cov_{t+1-n}\left(\frac{m_t}{v_t}, \bar{\Pi}_{t+1}\prod_{i=1}^{n}\delta_{t+1-j}\rho_{t+2-j} - \Psi_{t+1-n}\right) \\
&= E_{t+1-n}\left[\frac{m_t}{v_t}\left(\bar{\Pi}_{t+1}\prod_{i=1}^{n}\delta_{t+1-j}\rho_{t+2-j} - \Psi_{t+1-n}\right)\right], \qquad (51)
\end{aligned}
$$

where in each case the equality follows because, by (9) and (48), the product of the expectations is zero. Obviously z_t, z'_t and z''_t belong to the t-dated information set, so they are functions of variables known at t. Then

$$
\begin{aligned}
E_t\left(\frac{1}{m_{t+n}/v_{t+n}}\bar{\Pi}_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}\right) &= E_t\left(\frac{1}{m_{t+n}/v_{t+n}}\Psi_t\right) + z_t \\
E_t\left(\frac{1}{m_{t-1+n}/v_{t-1+n}}\bar{\Pi}_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}\right) &= E_t\left(\frac{1}{m_{t-1+n}/v_{t-1+n}}\Psi_t\right) + z'_t \\
E_{t+1-n}\left(\frac{m_t}{v_t}\bar{\Pi}_{t+1}\prod_{i=1}^{n}\delta_{t+1-j}\rho_{t+2-j}\right) &= E_{t+1-n}\left(\frac{m_t}{v_t}\Psi_{t+1-n}\right) + z''_{t+1-n}
\end{aligned}
$$

and, with the use of (9), (47) can be written

$$
\begin{aligned}
&- E_{t-n}\Bigg\{\Psi_{t-n} + \psi_t\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j} - \delta_{t-n}\rho_{t-n+1}\Bigg[E_{t-n+1}\left(\left(1-\frac{m_t}{v_t}\right)\Psi_{t+1-n}\right) - z''_{t+1-n}\Bigg] \\
&- \frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\Bigg[\sum_{j=0}^{n-1}(p_{t+j}-w_{t+j})\left(\prod_{i=1}^{j}\delta_{t-1+i}\rho_{t+i}\right) \\
&+ E_t\frac{1}{m_{t+n}/v_{t+n}}\left(\Psi_t + \psi_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}\right) \\
&+ z_t - \delta_t\rho_{t+1}E_{t+1}\left(\left(\frac{1}{m_{t+n}/v_{t+n}} - 1\right)\Psi_{t+1}\right) + z'_{t+1}\Bigg]\Bigg\} = 0, \text{ for all } t,
\end{aligned}
$$

or, noting which expressions can be moved inside expectations,

$$
\begin{aligned}
&- E_{t-n}\left\{\Psi_{t-n}+\psi_t\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}-\delta_{t-n}\rho_{t-n+1}\left[\left(1-\frac{m_t}{v_t}\right)\Psi_{t-n+1}\right]\right.\\
&-\frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left[\sum_{j=0}^{n-1}(p_{t+j}-w_{t+j})\left(\prod_{i=1}^{j}\delta_{t-1+i}\rho_{t+i}\right)\right.\\
&\left.\left.+\frac{1}{m_{t+n}/v_{t+n}}\left(\Psi_t+\psi_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}\right)-\delta_t\rho_{t+1}\left(\frac{1}{m_{t+n}/v_{t+n}}-1\right)\Psi_{t+1}\right]\right\}\\
&= E_{t-n}\left\{\delta_{t-n}\rho_{t-n+1}z''_{t+1-n}-\frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left(z_t+z'_{t+1}\right)\right\}, \qquad \text{for all } t.
\end{aligned}
$$

This can be re-arranged as

$$
\begin{aligned}
E_{t-n}&\left\{\frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left[\sum_{j=0}^{n-1}(p_{t+j}-w_{t+j})\left(\prod_{i=1}^{j}\delta_{t-1+i}\rho_{t+i}\right)\right.\right.\\
&\left.+\frac{v_{t+n}}{m_{t+n}}\left[\Psi_t+\psi_{t+n}\prod_{i=1}^{n}\delta_{t-1+i}\rho_{t+i}-\delta_t\rho_{t+1}\left(1-\frac{m_{t+n}}{v_{t+n}}\right)\Psi_{t+1}\right]\right]\\
&\left.+\delta_{t-n}\rho_{t-n+1}\left[\left(1-\frac{m_t}{v_t}\right)\Psi_{t-n+1}\right]-\Psi_{t-n}-\psi_t\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right\}\\
&= E_{t-n}\left\{\delta_{t-n}\rho_{t-n+1}z''_{t+1-n}-\frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left(z_t+z'_{t+1}\right)\right\}, \text{ for all } t.
\end{aligned}
$$

This corresponds to (10) for $z_{t,n}$ defined as

$$
z_{t,n}=\delta_{t-n}\rho_{t-n+1}z''_{t+1-n}-\frac{m_t}{v_t}\left(\prod_{j=1}^{n}\delta_{t-j}\rho_{t+1-j}\right)\left(z_t+z'_{t+1}\right). \tag{52}
$$

For $n=1$, we have

$$
z_{t,1}=\delta_{t-1}\rho_t z''_t-\frac{m_t}{v_t}\delta_{t-1}\rho_t\left(z_t+z'_{t+1}\right).
$$

With the definitions (49)–(51), this can be written

$$
\begin{aligned}
z_{t,1}&=\delta_{t-1}\rho_t E_t\left[\frac{m_t}{v_t}\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1}-\Psi_t\right)\right]\\
&-\frac{m_t}{v_t}\delta_{t-1}\rho_t\left[E_t\left(\frac{1}{m_{t+1}/v_{t+1}}\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1}-\Psi_t\right)\right)+E_t\left(\frac{1}{m_t/v_t}\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1}-\Psi_t\right)\right)\right]\\
&=\delta_{t-1}\rho_t\frac{m_t}{v_t}E_t\left[\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1}-\Psi_t\right)\right]\\
&-\frac{m_t}{v_t}\delta_{t-1}\rho_t E_t\left(\frac{1}{m_{t+1}/v_{t+1}}\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1}-\Psi_t\right)\right)+E_t\left(\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1}-\Psi_t\right)\right)\\
&=-\frac{m_t}{v_t}\delta_{t-1}\rho_t E_t\left[\frac{1}{m_{t+1}/v_{t+1}}\left(\bar{\Pi}_{t+1}\delta_t\rho_{t+1}-\Psi_t\right)\right],
\end{aligned}
$$

the intermediate equality following because m_t/v_t is known at the time expectations are taken at t and the final line from (10). This is as specified in the text.

Under perfect foresight, (9) implies

$$\bar{\Pi}_t \prod_{j=1}^{n} \delta_{t-j}\rho_{t+1-j} - \Psi_{t-n} = 0, \text{ for all } t,$$

so $z_t = z_t' = z_t'' = 0$ and, hence from (52), $z_{t,n} = 0$, as claimed in the text.

Appendix B Inference Methods

The KLM statistic is the quadratic form

$$K(\vartheta) = T^{-1} f_T(\vartheta)' V_{ff}(\vartheta)^{-1/2} P_{V_{ff}(\vartheta)^{-1/2} D_T(\vartheta)} V_{ff}(\vartheta)^{-1/2} f_T(\vartheta) \tag{53}$$

where $P_X = X(X'X)^{-1} X'$ for any matrix X and $D(\vartheta)$ depends on $\partial f_T(\vartheta)/\partial\vartheta$ and $\partial V_{ff}(\vartheta)/\partial\vartheta$ which results in a score statistic, see Kleibergen (2005, Eq. (16)). The test based on comparing $K(\vartheta)$ to critical values of the $\chi^2(p)$ distribution is the KLM test, where p is the number of parameters. Another useful test statistic is $J(\vartheta) = S(\vartheta) - K(\vartheta)$, for $S(\vartheta)$ defined in (21). In large samples, this is independent of $K(\vartheta)$ and distributed as $\chi^2(k-p)$.

For hypotheses involving subsets or functions of the parameters, generally denoted by $g(\vartheta) = 0$, the ARSW, KLM and J tests are performed as follows. First, derive the restricted CUE $\widetilde{\vartheta}$, by minimizing $S(\vartheta)$ in (21) subject to $g(\vartheta) = 0$. Kleibergen and Mavroeidis (2006) show that $S\left(\widetilde{\vartheta}\right)$ is asymptotically bounded by $\chi^2(k-p+r)$, where r is the number of restrictions to be tested. So the subset ARSW test is derived by comparing $S\left(\widetilde{\vartheta}\right)$ to the requisite quantile of the $\chi^2(k-p+r)$. Similarly, $K\left(\widetilde{\vartheta}\right)$ is bounded by a $\chi^2(r)$ and $J\left(\widetilde{\vartheta}\right)$ by a $\chi^2(k-p)$, and the KLM and J tests are derived analogously. None of these statistics require any identification assumptions on ϑ. If any element of ϑ happens to be poorly identified, the resulting confidence sets are expected to be wide, see Kleibergen and Mavroeidis (2006) for further details.

B.1 Confidence sets for long run unemployment

In section 5.2 we defined $g(\vartheta)$ as the transformation from the structural parameters ϑ to $\left(U, U^f, U^{hw}, U^{eff}\right)$. Let $\hat{G}$ denote the Jacobian of this transformation with respect to ϑ evaluated at the estimated value $\hat{\vartheta}$. Then, the asymptotic variance matrix of $\left(U, U^f, U^{hw}, U^{eff}\right)$ can be estimated by $\hat{G}\hat{V}_\vartheta\hat{G}'$, where $\hat{V}_\vartheta$ is a consistent estimate of the variance of $\hat{\vartheta}$. For any linear combination of the elements of $\left(U, U^f, U^{hw}, U^{eff}\right)$, denoted by a four-dimensional vector e, the asymptotic standard error can be computed as $\sqrt{e'\hat{G}\hat{V}_\vartheta\hat{G}'e}$. For example, for the standard error of U we set $e = (1, 0, 0.0)'$. The confidence interval of plus/minus two standard errors about the point estimate is a Wald confidence interval.

The derivation of the ARSW and KLM confidence set for U involves minimizing the GMM objective function (21) subject to the restriction that $g_1(\vartheta) = U_0$ to derive the restricted estimate $\widetilde{\vartheta}_0$. The restricted minimum of the objective function, $S\left(\widetilde{\vartheta}_0\right)$, is the

ARSW statistic and it is asymptotically bounded by a $\chi^2(K-p+1)$ random variable *irrespective* of whether U (or any other parameter) is identified or not, as explained in the previous section. The K statistic is then computed by the formula (53) evaluated at $\widetilde{\vartheta}_0$. It is common for such confidence sets to be disjoint. Because we are interested mainly in the smallest and largest value of U that is consistent with the data at a given level of significance, we report here only the boundaries of each confidence set. The precision with which those bounds are computed can be increased by making the grid of values of U finer. The same procedure can be applied to derive one-dimensional confidence sets for the other parameters U^f, U^{hw} and U^{eff}.

In implementing the above method of inverting the ARSW and KLM tests, we faced some computational difficulties arising from the fact that the transformation g from the original parameters ϑ to $\left(U, U^f, U^{hw}, U^{eff}\right)$ is highly non-linear and that the model involves a large number of unknown parameters. (To our knowledge, this is the largest model to which these identification-robust methods have been applied so far). The most common problem we encountered was lack of convergence of the restricted CUE estimator. To overcome this difficulty without resorting to iterative methods, we used a mixture of numerical optimization and grid search methods. The procedure is as follows. Define U^{nf} as the level of unemployment without frictions and U^{nhw} as the level of unemployment without high wages. Instead of considering a one-dimensional grid of points for the parameter of interest, say U between 0 and 10%, we considered a four-dimensional grid for the vector $\left(U, U^{nf}, U^{nhw}, U^{eff}\right)$, subject to the admissibility restrictions. For every value of $\left(U_0, U_0^{nf}, U_0^{nhw}, U_0^{eff}\right)$ in the grid, we computed the restricted CUE of ϑ subject to the four restrictions $g(\vartheta) = \left(U_0, U_0^{nf}, U_0^{nhw}, U_0^{eff}\right)$ using a derivative-based method. Since the number of unrestricted parameters is smaller than before, the CUE converged much more readily. Then, to find the minimum of the objective function subject to a single restriction, say $U = U_0$, we used grid search over the remaining three parameters U^f, U^{hw}, U^{eff}. Because this procedure involves grid search in four dimensions, it is computationally expensive when a high degree of precision is required. In order to increase the precision we took a two-step approach. We first set a relatively large grid step (0.5%) to identify the region of the parameter space that is clearly inconsistent with the data. Then, we refined our grid search focusing on the remaining region of the parameters using a smaller grid step.

Appendix C Data

To satisfy identities in the employment data, account must be taken of the self-employed. We treat them as an exogenously given proportion of the labour force. Government jobs provide matches and so need to be to be taken account of in the matching function. But these cannot reasonably be expected to be determined by the profit criteria underlying the job creation equation (10), so we treat them as exogenously determined. To be consistent with that, the productivity and wage measures are constructed from National Accounts data for the business sector only.

The tax measures and benefit replacement ratios are based on OECD data (National Accounts, Main Economic Indicators, International Financial Statistics) and ILO data (Yearbook). We have obtained this data directly from Jakob Madsen. For a detailed description of the construction of tax and benefits variables, see Madsen (1998). Tax and

Table 5: Preliminary tests on the job creation equation

	US			UK		
Specification	(0)	(1)	(2)	(0)	(1)	(2)
Parameters						
γ_0	-0.19 (0.17)	-	0.50 (0.03)	1.12 (0.71)	-	1.87 (0.10)
γ_1	-1.49 (2.46)	-	2.15 (1.02)	-39.7 (11.8)	-	-34.6 (6.9)
γ_h	2.76 (0.57)	1.95 (0.07)	-	2.16 (2.07)	3.90 (0.13)	-
Tests						
Wald	-	2.32 [0.31]	23.48 [0.00]	-	20.96 [0.00]	1.09 [0.30]
ARSW	-	8.02 [0.24]	124.27 [0.00]	-	16.92 [0.01]	3.71 [0.59]
KLM	-	2.44 [0.30]	103.51 [0.00]	-	5.86 [0.05]	1.12 [0.29]
Diagnostics						
Hansen $\chi^2(5)$	5.09 [0.28]	8.02 [0.24]	124.27 [0.00]	2.21 [0.70]	16.92 [0.01]	3.71 [0.59]
Ser Corr $\chi^2(5)$	4.48 [0.48]	4.39 [0.49]	4.76 [0.45]	1.58 [0.90]	1.66 [0.89]	1.63 [0.90]

Instruments include lags of $w/p, \tilde{\delta}$ and v^c. CUE-GMM with Newey-West Weight matrix. Sample for US: 1961 (2) - 2001 (2); for UK: 1981 (1) - 2000 (2).
Diagnostics: Hansen-Sargan test of overidentifying restrictions; Cumby and Huizinga (1992) test of residual autocorrelation from lags $2n$ to $2n+4$.

benefit data are only available at lower frequency (annual). Benefits and taxes are likely to be adjusted on an infrequent basis, so extrapolation appears inappropriate, and any given data point has simply been stretched to cover four periods; a data point observed in 1993 is assumed to be constant over the four quarters in 1993.

Vacancy data for the UK is obtained from Office for National Statistics: Labour Market Trends (Vacancy creation: "Unfilled vacancies at UK Job centres". Vacancy stock: "Inflow of vacancies at UK job centres"). The data used to construct series for job destruction for the US, the number of unemployed with spells shorter than 14 months, are from the Current Population Survey.

Appendix D Preliminary estimation and tests

Table 5 presents single-equation estimates of the job-creation equation for the US and the UK. It is clear that the structural parameters γ_0, γ_1 and γ_h cannot be accurately estimated in the unrestricted specification, column (0) in Table 5. In both countries, the standard errors of the estimated coefficients are large, and in fact, in the US, job creation costs are estimated to be slightly negative. Therefore, we consider two alternative specifications in which we set to zero either the job creation costs (specification 1) or the job hiring costs (specification 2). We perform the Wald, ARSW and KLM tests of these two specifications against the unrestricted model for each country, and we find that γ_h is significantly different from zero in the US but not in the UK, and conversely, γ_0, γ_1 are different from zero in the UK but not in the US. We impose those restrictions hereafter.

Also reported in Table 5 are two specification tests. The Hansen test of overidentifying restrictions is a standard specification test for models estimated by GMM. (The test statistic is equal to the value of the objective function (21) evaluated at the CUE $\hat{\vartheta}$). This is effectively a test of the identifying assumption made when we set the right-hand

Table 6: Preliminary estimation and tests on the wage equation

	US			UK		
Specification	(0)	(1)	(2)	(0)	(1)	(2)
Parameters						
β_0	-0.91 (0.24)	-0.93 (0.21)	-0.87 (0.08)	-0.58 (0.35)	-0.58 (0.34)	-0.69 (0.05)
β_u	-0.99 (0.97)	-0.39 (0.20)	-0.33 (0.08)	-0.39 (0.83)	-0.48 (0.46)	-0.62 (0.16)
β_v	-0.08 (0.19)	-0.05 (0.16)	-	0.08 (0.23)	0.06 (0.19)	-
β_U	0.21 (0.33)	-	-	-0.02 (0.14)	-	-
λ_w	0.89 (0.04)	0.87 (0.03)	0.88 (0.03)	0.89 (0.05)	0.89 (0.05)	0.88 (0.04)
$b_{\Delta lny_1}$	-0.23 (0.11)	-0.20 (0.10)	-0.21 (0.10)	-0.18 (0.09)	-0.18 (0.08)	-0.18 (0.08)
$b_{\Delta lny_2}$	-0.05 (0.07)	-0.03 (0.07)	-0.04 (0.06)	-0.25 (0.10)	-0.25 (0.10)	-0.26 (0.10)
$b_{\Delta lny_3}$	-0.15 (0.06)	-0.14 (0.05)	-0.15 (0.05)	-	-	-
$b_{\Delta lny_4}$	0.19 (0.07)	0.19 (0.07)	0.19 (0.07)	-	-	-
b_τ	0.03 (0.03)	0.05 (0.02)	0.05 (0.02)	-	-	-
Tests						
Wald	-	0.41 [0.52]	0.52 [0.77]	-	0.02 [0.90]	0.12 [0.94]
ARSW	-	1.84 [0.87]	1.93 [0.93]	-	9.48 [0.22]	9.61 [0.29]
KLM	-	0.53 [0.47]	0.62 [0.73]	-	0.02 [0.88]	0.17 [0.92]
Diagnostics						
Hansen	1.27 [0.87]	1.84 [0.87]	1.93 [0.93]	9.46 [0.15]	9.48 [0.22]	9.61 [0.29]
Ser. Corr.	4.11 [0.53]	5.00 [0.42]	5.11 [0.40]	8.58 [0.13]	8.59 [0.13]	7.16 [0.21]

The dependent variable is $y = \log(w/p - b/p/\tau)$. Instruments include lags of y, $\log(U/u)$, $\log(v/u)$, $\log(U)$ and $\log \tau$. CUE-GMM with Newey-West Weight matrix. Standard errors in parentheses, p-values in square brackets. Tests of specifications (1) and (2) against nesting model (0). Sample for US: 1961 (2) - 2001 (2); for UK: 1981 (1) - 2000 (2).
Diagnostics: Hansen-Sargan test of overidentifying restrictions; Cumby and Huizinga (1992) test of residual autocorrelation from lags 1 to 5.

side of equation (10) to zero. For both the unrestricted specifications and the chosen restricted specifications (specification 1 for the US and specification 2 for the UK), the Hansen test does not reject the validity of our over-identifying assumption at over 20% level of significance. This conclusion is robust to increasing the instrument set. The tests of residual autocorrelation reported here are those proposed by Cumby and Huizinga (1992), using the West (1997) estimator of the weighting matrix, see Mavroeidis (2002) for details.

Appendix E A log-linear bargaining wage equation

With the log-linear wage equation (36), matching frictions alone can never account for all unemployment as long as $\beta_u \neq 0$. With matching frictions the sole source of unemployment, theory implies that unemployment goes to zero as matching frictions become negligible. This holds whether wages are determined by the traditional Nash Bargaining Solution used by Blanchard and Diamond (1989), Mortensen and Pissarides (1994) and Pissarides (2000) or alternative bargaining procedures such as those in Hall (2005b) and

Hall and Milgrom (2005). But, with $\beta_u \neq 0$, the right-hand side of (36) goes to (plus or minus) infinity as $U_t \to 0$, resulting in a non-feasible wage share. To check whether this seriously biases the conclusions we draw for the log-linear wage equation formulation, we also tried a different form (referred to as the *bargaining wage equation*) that does not have this property. In particular, instead of setting $\beta_v = \beta_U = 0$ in (35), we set $\beta_u = \beta_U = 0$, resulting in the specification:

$$\ln y_t = \bar{\lambda}_w \ln y_{t-1} + \left(1 - \bar{\lambda}_w\right) \left(\bar{\beta}_0 + \bar{\beta}_v \ln \frac{v_t}{u_t}\right) + SRD + \bar{\varepsilon}_t^w. \tag{54}$$

(We use a bar to distinguish the parameters of the bargaining wage equation from the parameters of (36).) Since in the absence of matching frictions $v_t/u_t = 1$, (54) is not inconsistent with matching frictions accounting for all unemployment. Indeed, it provides a straightforward way to test whether that is the case, at least in the long run. With no matching frictions, the long-run wage share from (54) is given by $\ln y = \bar{\beta}_0$. For long-run unemployment to be zero in the absence of matching frictions, the long-run job creation equation (38) would have to be satisfied at that wage share when the probability of a vacancy being matched at each date, m/v, is one.

System estimates for the bargaining specification (54) are given in Table 7, with standard errors in parentheses. As a specification test, we report the Hansen (1982) test of overidentifying restrictions based on the system estimates, with p-values in square brackets. Compared to the specification (36) reported in Table 1, the p-value of Hansen test drops from 0.48 for the US to 0.19 the bargaining specification. For the UK, the fit of the two specifications is very similar. The formulation (54) fits the data somewhat less well than (36).[12]

The final three rows of Table 7 report test statistics for the restriction that in the long run unemployment goes to zero as matching frictions become negligible, with p-values in square brackets. As noted above, the restriction corresponds to the long-run job creation equation (38) being satisfied at the wage share implied by (54) with $m/u = m/v = 1$. The long-run wage share is then given by $\ln y = \bar{\beta}_0$ or, in view of (34),

$$\frac{w}{p} = \frac{b}{p\tau} + e^{\bar{\beta}_0}.$$

For the US, given that $\gamma_0 = \gamma_1 = 0$, the long-run job creation equation with $m/v = 1$ is

$$\frac{w}{p} = 1 - (1 - \delta\rho\theta)\,\gamma_h.$$

The hypothesis that these are both satisfied by the same value of w/p is thus

$$H_{US} : 1 - (1 - \delta\rho\theta)\,\gamma_h = \frac{b}{p\tau} + e^{\bar{\beta}_0}. \tag{55}$$

For the UK, given that $\gamma_h = 0$, the corresponding hypothesis is

$$H_{UK} : 1 - \frac{1 - \delta\rho\theta}{(\delta\rho\theta)^n} \left[\frac{1 - (\delta\rho\theta)^{n+1}}{1 - \delta\rho\theta}\gamma_0 + \gamma_1\left(1 - \frac{\rho}{l}\right)\right] = \frac{b}{p\tau} + e^{\bar{\beta}_0}. \tag{56}$$

[12] The GMM criterion function is larger for (54) than for (36), albeit not significantly at the usual 5% level. Moreover, when both $\ln(U/u)$ and $\ln(v/u)$ are included in a single nesting model, the latter is the less significant regressor (in the sense that the hypothesis that it can be removed can be accepted with higher confidence). This holds for both t and the identification-robust tests.

Table 7: System estimates for bargaining wage equation

	US	UK
Matching function		
$\log(\alpha_0)$	-0.96 (0.05)	-0.92 (0.20)
α_1	0.16 (0.03)	0.61 (0.07)
α_2	0.56 (0.02)	0.25 (0.05)
λ_m	0.72 (0.03)	0.65 (0.05)
$a_{\Delta lnu}$	-0.48 (0.04)	0.31 (0.10)
Job-creation equation		
γ_h	1.99 (0.06)	-
γ_0	-	1.87 (0.08)
γ_1	-	-32.7 (5.54)
Bargaining wage equation		
$\bar{\beta}_0$	-0.49 (0.03)	-0.28 (0.05)
$\bar{\beta}_v$	0.32 (0.12)	0.19 (0.04)
$\bar{\lambda}_w$	0.93 (0.02)	0.86 (0.04)
$\bar{b}_{\Delta lny_1}$	-0.23 (0.09)	-0.32 (0.06)
$\bar{b}_{\Delta lny_2}$	-0.07 (0.06)	-0.19 (0.09)
$\bar{b}_{\Delta lny_3}$	-0.15 (0.05)	-
$\bar{b}_{\Delta lny_4}$	0.20 (0.07)	-
$\bar{b}_\tau$	0.01 (0.01)	-
Hansen test	16.02 [0.19]	15.26 [0.23]
Tests for H_{US} & H_{UK}		
Wald	22.32 [0.00]	8.57 [0.00]
ARSW	20.79 [0.08]	49.61 [0.00]
KLM	4.88 [0.03]	34.37 [0.00]

The model is estimated by Continuously Updated GMM with Newey-West Weight matrix. Standard errors in parentheses, p-values in square brackets. $n = 3$ for the US and $n = 2$ for the UK. Sample for US: 1961Q2 - 2001Q2; for UK: 1981Q1 - 2000Q2. The number of over-identifying restrictions is 12 in all cases. Hypotheses specified in text. ARSW refers to the Anderson-Rubin-Stock-Wright test, KLM refers to the KLM test. p-values in square brackets.

We report in Table 7 the test statistics for those hypotheses using the Wald, ARSW and KLM tests. It is clear that the hypothesis is overwhelmingly rejected for the UK. For the US, the only test under which one might even consider accepting the hypothesis is ARSW which, although like the KLM test robust to weak instruments, is less powerful.