Edinburgh School of Economics

Discussion Paper Series

Number 12

Bayesian inference in models based on equilibrium search theory

Gary Koop (University of Edinburgh)

Date
August 1999

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

Bayesian Inference in Models Based on Equilibrium Search Theory*

GARY KOOP
Department of Economics, University of Edinburgh,
Edinburgh, EH8 9JY, U.K.
E-mail: G.Koop@ed.ac.uk

September 1998
Revised: August 1999

The equilibrium search model represents a substantial advance over previous models of job search, but its empirical performance fails in some crucial ways. Hence, flexible extensions are called for. In this paper, we consider three such extensions in areas where Bayesian methods can be used to great advantage. These are: i) A specification were Bayesian priors are used to center a model over the restrictions implied by economic theory; ii) A model involving a nonlinear production function where a closed form expression for the likelihood function does not exist; and iii) A model which allows for unobserved heterogeneity. We show how posterior simulation methods can be used for empirical analysis for all three models. The paper includes an empirical exercise involving the school-to-work transitions of Canadian and U.S. school leavers.

*I would like to thank two anonymous referees, Chris Ferrall, Richard Paap, Dale Poirier and Mark Steel for helpful comments.

1 Introduction

The goal of finding models which are both based in rigorous economic theory and fit the data is an important, albeit elusive, one. Labor economists have developed many structural models of firm and worker behavior in an attempt to explain observed wage distributions and unemployment durations. For instance, two empirical regularities have been noted by labor economists: 1) Wage densities are non-degenerate, even for groups of individuals that are homogeneous in terms of their observed characteristics; and 2) Unemployed workers rarely reject jobs that are offered to them. Empirical equilibrium search models have been developed to explain these regularities (see, among others, Ridder and van den Berg, 1997, van den Berg and Ridder, 1998, or van den Berg, 1999). These models posit that firms and individuals act optimally (in a sense which we will make clear below) and that markets are in a steady state equilibrium. Unfortunately, such models do not fit the data well in important ways. For instance, although the equilibrium search model does allow for a non-degenerate wage density, it is increasing in w. Typically, empirical wage densities are roughly bell-shaped. This can be seen in Figure 1 which plots the empirical distributions for the Canadian and U.S. data sets which we will use in this paper. Hence, equilibrium search models are considered by some an empirical failure. In light of this, the question arises as to the nature of future research strategies to develop models which better fit the observed data. Three possibilities come to mind. First, models which are centered over the restrictions implied by economic theory can be used. Second, more complicated theoretical economic models can be developed. Third, unobserved heterogeneity or measurement error can be appealed to. In this paper we argue that Bayesian methods are ideal for all these possibilities. We illustrate each using a particular model choice, but the techniques developed are valid more generally.

We begin with an extension to the standard equilibrium search model where workers are allowed to depart from optimal behavior. We use a prior to center the model over worker optimality. We carry out a prior sensitivity analysis to consider three cases where i) Worker optimality is imposed; ii) Departures from optimality are assumed to be small; and iii) Large departures from optimality can potentially exist. We show how this model can be used to test for worker optimality and learn about unidentified parameters. Classical econometric approaches which do not allow for prior information typically either dogmatically impose optimality or assume departures from optimality can be of any magnitude. Furthermore, testing of optimality using likelihood ratio tests is impossible (see Lancaster, 1997, and Koop and Poirier, 1997) . Hence, we argue for the usefulness of Bayesian methods with this model.[1]

We then extend the equilibrium search model to allow for a nonlinear production function. This particular extension implies that a closed form expression for the likelihood function does not exist. Nevertheless, we show how Bayesian simulation methods can be used to empirically analyze this model.[2] The non-existence of a closed form expression for the likelihood occurs for many possible theoretical extensions. For instance, Nielsen and Rosholm (1999) extend the equilibrium search model to allow for firm-specific investments in training and obtain a likelihood function with this characteristic. They use approximate methods to estimate their model, but the methods developed in this paper can be used to carry out an exact likelihood analysis. Hence, Bayesian methods provide powerful tools for the analysis of complicated theoretical extensions of the standard equilibrium search model.

We conclude by considering a model with unobserved heterogeneity. We find such models less interesting since unobserved heterogeneity is difficult to model in anything other than an ad hoc fashion. As discussed in Ridder and van den Berg (1997, section 7) unobserved heterogeneity can always be called upon to fit any observed wage distribution.[3] However, models with unobserved heterogeneity or measurement error have been estimated by many researchers and it is worthwhile establishing that Bayesian estimation of such models can be carried out in a straightforward manner. In fact, these extensions typically imply the presence of an individual effect which can be considered as latent data. Simple Markov Chain Monte Carlo (MCMC) methods with data augmentation can be developed for what would otherwise be complicated econometric problems. Furthermore, if unobserved heterogeneity exists, one would hope that it is not too large. After all, if individuals are extremely different it is questionable to fit a structural model to a cross-section or panel of individual data. Bayesian priors are a logical avenue for allowing for heterogeneity but ensuring that it is not too large.

Throughout this paper, we use Bayesian methods for both estimation and testing since they seem well-suited for handling the extensions of the equilibrium search model that are likely to be of empirical interest. However, a common criticism of traditional Bayesian methods for testing is that they involve the comparison of completely specified models. In the

[1] A previous version of this paper (available at http://www.ed.ac.uk/~gkoop/) considered another model where departures from optimizing behaviour on the part of firms were allowed, but the model was centered over firm optimality.

[2] The methods developed in this paper can also be used to do maximum likelihood estimation of this model.

[3] For instance, in a perfectly competitive market individual wages equal marginal value product. By allowing the latter to vary across individuals in an unrestricted manner that is not observed by the econometrician, we can fit any empirical wage density precisely in the context of an equilibrium model.

job search literature such a criticism is highly relevant since the researcher is often interested in some measure of how well a model fits the data in an absolute sense. That is, he/she does not want to compare an existing model to a completely specified alternative one. Meng (1994) discusses a Bayesian method of evaluating the fit of a given model without reference to an alternative. It bears some similarity with the classical method of bootstrapping a test statistic and attempts to answer the question: "How likely is it that the observed data came from the model under consideration?". This so-called posterior predictive p-value method has been rarely used in econometrics. We apply it here and find that it is a useful tool for evaluating and understanding the properties of equilibrium search models.

The paper is organized with one section for each of the three extensions. An additional section describes posterior predictive P-values. Subsequently, we use these methods in the context of an empirical exercise using Canadian and U.S. data taken from Ferrall (1997). Results are encouraging in that Bayesian methods seem to offer useful theoretical and empirical insights into model development and are computationally feasible.

2 The Basic Equilibrium Search Model

2.1 Derivation of Likelihood Function

Various types of equilibrium search models have been developed and/or estimated in the literature (see, among many others, van den Berg and Ridder, 1998, Christensen and Kiefer, 1997, Kiefer and Neumann, 1993, Ridder and van den Berg, 1997, Mortensen, 1990, Burdett, 1990, Nielsen and Rosholm, 1999, van den Berg, 1999). In this section, we do not repeat the theoretical derivation offered in these and other papers. Rather, we survey the basic setup and offer some intuition for a few key results. The likelihood function we end up with is identical to that used in Christensen and Kiefer (1997).

Suppose we have data on i=1,..,N individuals who initially experience a unemployment spell of duration d_i followed by an employment spell of duration j_i. When employed, the individual receives a wage of w_i.[4] We assume all spells are complete to simplify the discussion. Incomplete spells can be handled in the standard way.[5]

We assume a fixed, homogeneous population of workers and firms. The ratio of the measure (number) of workers to firms is m. Workers are initially unemployed and job offers arrive as events in a Poisson process at a rate of λ_0. Once workers are employed they can

[4] This sort of data is often available for school leavers. If other types of data are available, then somewhat different likelihood functions apply.

[5] For instance, in a Bayesian analysis, an MCMC algorithm with data augmentation can be used in a straightforward way.

continue searching and new job offers arrive at a rate of λ_1. Firms lay off workers at a rate of δ. The distribution of wage offers is given by F(w), the corresponding density is denoted by f(w). The value of leisure to individuals is equal to b. When working, their productivity (i.e. their marginal value product) is p. We assume workers and firms know the values of all parameters and distributions.

To derive the likelihood function worker and firm behavior must be specified. We assume that unemployed workers set a reservation wage, r, and accept the first wage offer above this. Exactly how this reservation wage is chosen will be discussed below. Employed workers accept any new wage offer above their current wage.

Firm behavior is described by F(w), which is determined endogenously assuming that the job market is in a steady state equilibrium. Given that firms know the acceptance strategies of workers, the supply of labor to a firm who offers wage w can be derived and denoted by l(w) (see Ridder and van den Berg, 1997, equation (8)). Assuming a linear production function, the steady state profit flow is given by:

$$\pi(\mathrm{w}) = \ (\mathrm{p\text{-}w})\mathrm{l(w)}. \tag{1}$$

No firm will ever offer a wage below r, since no worker would accept it, hence we know the support of f(w) begins at r. Furthermore, in equilibrium all wage offers must yield the same profit flow (i.e. firms must have no incentive to change their offers). Finally, f(w), being a valid p.d.f., must integrate to one. Using these facts, it can be shown (Ridder and van den Berg, 1997, section 4.1) that:

$$\mathrm{f(w)}=\frac{\delta+\lambda_1}{2\lambda_1}\frac{1}{\sqrt{\mathrm{p\text{-}r}}\sqrt{\mathrm{p\text{-}w}}}, \tag{2}$$

for w∈[r,h]. This density is increasing in w, a point which we will return to later. The highest wage offered is given by:

$$\mathrm{h} = \beta^2\mathrm{r}+(1\text{-}\beta^2)\mathrm{p}, \tag{3}$$

where

$$\beta = \frac{\delta}{\delta+\lambda_1}.$$

To provide some intuition for how a model with homogeneous workers and firms can yield a non-degenerate wage distribution, consider first a firm which posts a low wage offer, w_L, each time a vacancy arises. Contrast this with a second firm which posts a high wage offer,

w_H, each time a vacancy arises. That is, $r<w_L<w_H<p$. One might naively think that the first firm is more profitable than the second, since profit per employee is larger in the former (i.e. $(p-w_L)>(p-w_H)$). However, this is not necessarily the case since this reasoning neglects the l(w) term in the profit function. It can be shown that, in equilibrium, $l(w_H)>l(w_L)$, so that the firm offering the higher wage becomes larger. Hence, each firm faces a trade-off: it can offer a low wage and stay small or offer a high wage and become large. At each point in the wage density in equation (2), this trade-off is perfectly balanced to ensure constant profit for each wage in the interval [r,h].

The previous reasoning hinges on the fact that $l(w_H)>l(w_L)$. To provide some intuition for why this is so, note i) The firm which offers the lower wage is more likely to lose employees to other firms offering higher wages (since on-the-job search is allowed in this model); ii) There is a fixed number of workers and firms in this market; and iii) The fact λ_0 and $\lambda_1$are not infinite implies search frictions exist (i.e. job offers do not arrive instantly). Combining these three facts it can be seen that low wage firms will tend to lose their employees more often and take longer to fill vacancies. Hence, in a steady state low wage firms will have fewer employees than high wage firms. Ridder and van den Berg (1997) and van den Berg (1999) are particularly good sources for additional intuition and more information on the theoretical derivations underlying the equilibrium search model.

Following Christensen and Kiefer (1987), the likelihood function can be calculated by noting that the joint p.d.f. of d_i, w_i and j_i for i=1,..,N can be written as: $f(d_i,w_i,j_i)=f(d_i)f(w_i)f(j_i|w_i)$ where $f(w_i)$ is given in (2),

$$f(d_i) = \lambda_0 \exp(-\lambda_o d_i), \tag{4}$$

and

$$f(j_i|w_i) = (\delta + \lambda_1)\frac{\sqrt{p-w_i}}{\sqrt{p-r}} \exp\left(-(\delta + \lambda_1)\frac{\sqrt{p-w_i}}{\sqrt{p-r}} j_i\right). \tag{5}$$

Given a random sample of d_i, w_i and j_i for i=1,..,N, the likelihood function is $L=\prod_{i=1}^{N} f(d_i)f(w_i)f(j_i|w_i)$

Summary

Parameters

- λ_0: Arrival rate of offers to unemployed workers.
- λ_1: Arrival rate of offers to employed workers.
- δ: Layoff rate.

- b: Value of leisure.
- p: Worker productivity.
- r: Reservation wage.
- h: Highest wage offer.
- m: Measure of workers relative to firms.

Data (observed for i=1,..,N individuals)

- d_i: Unemployment duration for individual i.
- w_i: Wage received by individual i.
- j_i : Employment duration for individual i.

To figure out the behavior of individuals, some objective function must be assumed. It is common to assume that workers maximize the expected value of future income.[6] This assumption we will refer to as implying optimal behavior on the part of the worker. It turns out that optimal unemployed individuals will accept the first job offer with wage above a reservation wage given by:

$$r = \gamma b + (1-\gamma) p, \tag{6}$$

where

$$\gamma = \frac{\left(1+\frac{\lambda_1}{\delta}\right)^2}{\left(1+\frac{\lambda_1}{\delta}\right)^2 + \left(\frac{\lambda_0}{\delta}-\frac{\lambda_1}{\delta}\right)\frac{\lambda_1}{\delta}}.$$

Note that Christensen and Kiefer (1997) begin by parameterizing the equilibrium search model *with optimality imposed* in terms of its structural parameters, $\theta = (\lambda_0, \lambda_1, \delta, b, p)$, but then proceed to work with an alternative parameterization, $(\lambda_0, \lambda_1, \delta, r, h)$. Equations (3) and (6) provide the links between these two alternatives. We do not impose worker optimality at this stage and, hence, will let $\phi = (\theta', r)' = (\lambda_0, \lambda_1, \delta, b, p, r)'$ denote the structural parameters.

[6] Note that, following most of the literature including Christensen and Kiefer (1997), we do not discount expected future income. If we were to allow for a constant discount rate, this would not affect the value of the likelihood function since this parameter is not identified, even if optimality is assumed. If b were known, then the discount rate would be identified if optimality were assumed. Koop and Poirier (1997) discuss learning about unidentified parameters in a different model, but the issues raised are similar. It should be noted in passing that lack of identification causes no problems for a Bayesian analysis with a proper prior.

Learning about these is an important goal of empirical analysis of the equilibrium search model. From a Bayesian point of view it is likely more easy to elicit priors for such structural parameters.

The assumptions underlying optimal behavior by workers (i.e. homogeneity, maximizing expected future income, complete knowledge of model) are undoubtedly too restrictive to be literally true. However, this does not necessarily mean that a model based on these assumptions is irrelevant for empirical analysis. Bayesian methods allow for economic theory to be used as a guide for prior elicitation and, for the equilibrium search model, we argue that this is potentially a very useful property. In the equilibrium search model, the optimal reservation wage is given in (6). Suppose that one feels that the reservation wage is probably not set optimally, but that deviations from optimality are likely fairly small. In particular, one feels

$$r \sim f_N(\gamma b + (1-\gamma)p, \mathrm{sdr}^2), \tag{7}$$

where $f_N(c, C)$ denotes the Normal distribution with mean c and variance C. In other words, the reservation wage is Normally distributed, centered over the optimal value given in (6). As the standard deviation of r, sdr, approaches zero the optimal reservation wage is imposed and the standard equilibrium search model used in Christensen and Kiefer (1997) is obtained. As sdr increases one becomes more and more uncertain about optimality. Intermediate values of sdr can be used to reflect the degree of confidence the researcher feels about the optimality conditions implied by economic theory. In the Bayesian paradigm, (7) can be considered prior information and incorporated in the posterior. Informally, this approach might be appropriate for the researcher who says: "I think it likely that optimality roughly holds" or "I think that optimality does not hold, but that deviations from optimality are reasonably small."

Formally, the model considered in this section maintains the assumption that workers use a constant reservation wage policy, but relaxes the assumption that r is optimally selected. Constant reservation wage models are an established reference point in the literature, so we feel it useful to maintain this assumption as we move one step away from optimality. For future reference, the prior used for this model is given by:

$$p(\phi) = p(\theta)p(r|\theta),$$

where $p(\theta)$ is discussed below and $p(r|\theta)$ is given by (7).

It is worth digressing briefly to discuss identification issues in this model. Note that the parameters m and b are unidentified – they do not appear in the likelihood function. In previous empirical analyses of the equilibrium search model (e.g. Christensen and Kiefer, 1997) the parameter m is ignored, but inferences about b are made by assuming optimality. That is, b is identified only through imposing (6), the optimal reservation wage equation. Without this restriction, the maximum value of the likelihood function is the same for a model with optimality imposed and one without it imposed. Hence, likelihood ratio tests cannot be used to test for worker optimality (i.e. the likelihood ratio will always be one regardless of the data). However, from a Bayesian point of view, b enters the prior for r and (7) imposes prior correlations between all of the parameters. Intuitively, the data allows us to learn about the identified parameters, and such learning can spill over to the unidentified one via the prior correlations. Furthermore, the Bayesian model incorporates both prior and likelihood and, hence, a posterior odds ratio for testing worker optimality can be calculated. Assuming prior odds are unity, the posterior odds ratio in favor of worker optimality can be written as:

$$P.O. = \frac{p(\mathrm{r}{=}\gamma \mathrm{b}{+}(1\text{-}\gamma)\mathrm{p}|\mathrm{Y})}{p(\mathrm{r}{=}\gamma \mathrm{b}{+}(1\text{-}\gamma)\mathrm{p})}, \tag{8}$$

where the numerator and denominator are the marginal posterior and prior, respectively, for r evaluated at the point implied by optimality. This is the well-known Savage-Dickey density ratio, and the prior used in this paper is such that it is an exact expression for the posterior odds ratio, not an approximation (see Verdinelli and Wasserman, 1995). In practice, we calculate $P.O.$ using kernel methods on posterior and prior simulator output to estimate the numerator and denominator. Computational methods to provide such simulator output are described in the Appendix.

3 The Equilibrium Search Model with Nonlinear Production Function

There are many ways we could extend the equilibrium search model to make it more realistic. Ridder and van den Berg (1997) discuss several such ways (e.g. allowing for nonparticipation by individuals, allowing for the layoff rate to vary with the wage, allowing for a nonlinear production function). However, they point out that one of the key problems with the equilibrium search model is that it implies f(w) is increasing in w, which appears to be counterfactual. Making reasonable assumptions about the extensions (e.g. that the production function is concave), Ridder and van den Berg (1997) show how all such extensions will not make things better (i.e. f(w) will still be increasing in w). In this section, we focus on nonlinearity in production as being perhaps the most promising of the several extensions considered by Ridder and van den Berg (1997). In particular, they argue that if the production function is allowed to be convex, then the resulting f(w) might fit the data better. Furthermore, the nonlinear production function equilibrium search model has not been analyzed except for very specific functional forms. We show here how Bayesian methods can be used to analyze a general version of this model.

We will assume the same setup as previously (i.e. the equilibrium search model where workers choose a constant reservation wage that is not necessarily optimal)[7] and that profits are calculated based on a nonlinear production function as:

$$\pi(w) = pl(w)^{\varepsilon} - wl(w). \tag{9}$$

The model discussed previously has $\varepsilon = 1$. Note that the production function is concave if $0<\varepsilon<1$ and convex if $\varepsilon > 1$. As before, no firm will offer a wage less than r since they know employees will always reject it. The expression for l(w) — which is determined by worker behavior and is, hence, unaffected by the extension to nonlinear production — is given in Ridder and van den Berg (1997, equation 8). In the basic equilibrium search model, to derive the wage distribution, we set $\pi(w) = \pi(r)$ and solve to obtain an expression for F(w). The latter is then differentiated to yield f(w). This strategy does not work here since there is, in general, no closed form solution for F(w). A closed form solution does exist for $\varepsilon = \frac{1}{2}$ and $\varepsilon = 2$ and these two special cases are discussed in Ridder and van den Berg (1997). However, to the best of our knowledge, no empirical analysis of the general case has been done. Using Ridder and van den Berg;'s expression for l(w) and setting $\pi(w) = \pi(r)$ we obtain:

$$w = \frac{(\delta + \lambda_1(1 - F(w)))^2}{\square_1 \square_2} \left[p \left(\frac{\square_1 \square_2}{(\delta + \lambda_1(1 - F(w)))^2} \right)^{\varepsilon} - p \left(\frac{\square_1}{\square_2} \right)^{\varepsilon} + r \frac{\square_1}{\square_2} \right], \tag{10}$$

where

$$\square_1 = \frac{m\delta\lambda_0}{\delta + \lambda_0},$$

and

$$\square_2 = \delta + \lambda_1.$$

[7]Note, in their analysis of the nonlinear production model, Ridder and van den Berg (1997) assume $\lambda_0 = \lambda_1$ which implies that the optimal reservation wage is b. Here we do not make this assumption and the optimal reservation wage is difficult to calculate, involving a complicated integral. In principle, however, we could impose worker optimality on the present model.

Note a further complication with the nonlinear production model is that the wage distribution depends on another structural parameter, m, the ratio of the measure (number) of workers to firms in the market. We denote the vector of structural parameters in this model as $\omega = (\phi', \varepsilon, m)'$.

Ridder and van den Berg (1997) discuss a similar model with $\varepsilon = \frac{1}{2}$ and point out that equations analogous to (10) are not enough to define the wage offer distribution. In particular, even for their simplified model, complicated nonlinear inequality restrictions involving the parameters must be imposed to ensure a valid, non-degenerate, wage distribution. In the more general model considered here, the imposition of these inequality restrictions analytically is impossible. An advantage of the Bayesian approach is that such inequality restrictions can easily be imposed through the prior. That is, we set the prior to zero for all parameter values that do not satisfy the restrictions. We use the same prior as for the basic equilibrium search model truncated to insure the restrictions hold. In practice, imposing the restrictions through the prior is equivalent to randomly drawing from the posterior ignoring the restrictions and discarding all draws which imply:

- $\pi(w) < 0$.
- $\min(w_i) < r$.
- $\max(w_i) > p$.
- $m < 0$.

It is not possible to solve (10) for F(w). However, (10) provides us with an expression for the inverse function of F(w) (i.e. $F^{-1}(w)$). Given knowledge of the inverse of the cumulative distribution function, it is simple to take random draws using the transformation method (see, e.g., Press, Flannery, Teukolsky and Vetterling, 1986, pp. 200-203).[8] Hence, even though f(w) does not have a known form, we can take random draws from it. Using the draws, we can use kernel methods to obtain f(w) and, hence, the likelihood function. This substantially increases the computational burden (i.e. at every draw from the posterior simulator, numerous draws from f(w) must be taken). However, our empirical results indicate the practicality of this approach. Many promising extensions of the equilibrium search model (e.g. Nielsen and Rosholm, 1999) yield a likelihood function with no closed form expression. It is this fact, perhaps, that has held up empirical work using theoretically rich extensions of the basic model. Hence, simulation methods such as those developed here for a particular

[8] This method is also known as the inverse CDF method.

extension, have great potential for advancing empirical work in job search. Precise details on prior hyperparameters and computation are given in the empirical section and the Appendix, respectively.

4 The Equilibrium Search Model with Heterogeneity

There are many ways of introducing heterogeneity into equilibrium search models. See, for instance, Ridder and van den Berg (1997) for a survey and Eckstein and Wolpin (1995) for a structural model which relies on heterogeneity. The Bayesian methods used in this paper should be suitable for estimating a wide range of such models. Here we focus on heterogeneity in the productivity of workers. In particular, we assume each worker has productivity p_i which is known to the worker and all firms. Other than this we assume the setup of basic equilibrium search model of Section 2. For simplicity, we assume workers set reservation wages optimally.

Since firms know p_i, there is a different wage density for every individual:

$$f(w_i)=\frac{\delta+\lambda_1}{2\lambda_1}\frac{1}{\sqrt{p_i\text{-}r_i}\sqrt{p_i\text{-}w_i}}, \tag{11}$$

for $w_i \in[r_i,h_i]$, where

$$r_i = \gamma b+(1\text{-}\gamma)p_i, \tag{12}$$

and

$$h_i = \beta^2 r_i+(1\text{-}\beta^2)p_i. \tag{13}$$

The distribution of unemployment durations is unaffected by this change:

$$f(d_i) = \lambda_0 \exp(-\lambda_o d_i), \tag{14}$$

but the conditional distribution of employment durations becomes:

$$f(j_i|w_i) = (\delta+\lambda_1)\frac{\sqrt{p_i\text{-}w_i}}{\sqrt{p_i\text{-}r_i}}\exp\left(-(\delta+\lambda_1)\frac{\sqrt{p_i\text{-}w_i}}{\sqrt{p_i\text{-}r_i}}j_i\right). \tag{15}$$

Taking the product of (11), (14) and (15) over a random sample of individuals yields the likelihood function. However, if we were to place no further structure on the p_is we would have a model with more parameters than data points. Hence, we assume the p_i are i.i.d. random variables drawn from a common distribution. Here we take:

$$p_i \sim f_N(p, P), \tag{16}$$

although, of course, many other possibilities exist (for instance, Bontemps, Robin and van den Berg, 1995, assume log Normality).[9]

With this specification the structural parameters are $\varphi = (\phi', P)'$. P is a crucial parameter since it controls the amount of heterogeneity that is allowed for. As P goes to zero, the equilibrium search model without heterogeneity is obtained. An advantage of the Bayesian approach is that we can specify a prior for P which allocates a great deal of weight to small values of P. An exponential distribution has a mode at zero and seems a logical choice. Noting that the exponential is the Gamma with two degrees of freedom we take:

$$P \sim f_G(\mu_P, 2), \tag{17}$$

where $f_G(a, b)$ denotes the Gamma distribution with mean a and b degrees of freedom. The hyperparameter μ_P can be used to increase or decrease the degree of heterogeneity allowed for in the model.

Further details on the prior and computational methods are discussed in the empirical section and the Appendix, respectively. It is worth noting that MCMC methods with data augmentation methods can easily allow for posterior analysis of any parameter in this model, including the individual p_is. Inference on such individual effects can be very difficult from a classical econometric perspective due to the incidental parameters problem.

5 Posterior Predictive P-Values

The typical Bayesian method of model comparison is the posterior odds ratio, which is the relative probability of two completely specified models. There are some cases where the researcher is interested in investigating the performance of a model in some absolute sense, not relative to a specific alternative model. One method of doing this is the posterior predictive p-value approach discussed in Gelman and Meng (1996) or Meng (1994). This approach is disliked by purist Bayesians since it violates the likelihood principal. Nevertheless, it does seem to be a good way of gaining some idea about how well a model fits the data.

[9] This model is in equilibrium in the sense that, for a given individual, every feasible wage offer will yield the same profit to the firm. However, it is possible that this common level of profit varies over individuals. Hence, firms might prefer to make offers to some individuals over others. Imposing the restriction that $\pi(\mathrm{w}_i)$ is the same for all i and all feasible w complicates the analysis greatly. Such a model is a topic for further research. Another way of surmounting this problem, would be to assume all workers have a common reservation wage. This model would have all workers being equally profitable, but would not reflect optimal behavior. Another alternative model would have firms knowing p and P, but not knowing the individual p_is.

To motivate the posterior predictive p-value approach, let M indicate a model (i.e. a likelihood function and a prior), Y^* a matrix of observable data (i.e. Y^* is a random matrix), Y the observed data (i.e. it is not random) and let $g(Y^*)$ be some function of interest of the data. $p(g(Y^*)|M,Y)$ summarizes everything our model says about $g(Y^*)$ after seeing the data. In other words, it tells us the types of data sets that our model can generate. For the observed data we can calculate g(Y). If g(Y) is in the extreme tails of $p(g(Y^*)|M,Y)$, then model M cannot do a good job of explaining g(Y) (i.e. g(Y) is not the sort of data characteristic that can plausibly be generated by the model). Formally, we can obtain tail area probabilities in a manner similar to classical p-value calculations. In particular, we calculate the probability of a model yielding a data set with more extreme properties than that actually observed (i.e. analogous to a two-tailed classical p-value).

$p(g(Y^*)|M,Y)$ can be calculated using simulation methods. That is, as shown in Gelman and Meng (1996), it can be written as:

$$p(g(Y^*)|M,Y)=\int p(g(Y^*)|M,\eta)p(\eta|M,Y)d\eta,$$

where η contains the model's parameters (e.g. $\eta = \phi$ for the basic equilibrium search model). The term $p(\eta|\ M,Y)$ is the posterior and the Appendix describes how to simulate from it for each of the models in the paper. We can simulate from $p(g(Y^*)|M,\eta)$ by noting that it merely involves simulating artificial data from the model for a given parameter value. Hence, we randomly draw parameter values from the posterior, given these parameter values we simulate data, Y^r, using the likelihood, we then calculate $g(Y^r)$ and repeat for r=1,..,R. Given $g(Y^r)$ for r=1,..,R we use kernel methods to build up $p(g(Y^*)|M,Y)$. Details on how Y^r for r=1,..,R can be simulated for each model are given in the Appendix.

The posterior predictive p-value approach requires the selection of a function of interest, g(.). The exact choice of g(.) will vary depending on the empirical application. For the equilibrium search model, the key predictive implications relate to the wage distribution and, hence, we consider four different g(Y)'s relating to the first four moments of the wage distribution.[10] In particular,

- $g_1(w)$ is the mean of the wages, $\overline{w}$.
- $g_2(w)$ is the standard deviation of wages.
- $g_3(w)$ is a measure of skewness of wages given by:

[10]Since, in the data sets used in this paper, all wages lie in a bounded interval, these moments will exist.

$$g_3(\mathrm{w}) = \frac{\sqrt{N}\sum_{i=1}^{N}(\mathrm{w}_i - \overline{\mathrm{w}})^3}{\left[\sum_{i=1}^{N}(\mathrm{w}_i - \overline{\mathrm{w}})^2\right]^{\frac{3}{2}}}.$$

- $g_4(\mathrm{w})$ is a measure of excess kurtosis of wages given by:

$$g_4(\mathrm{w}) = N\frac{\sum_{i=1}^{N}(\mathrm{w}_i - \overline{\mathrm{w}})^4}{\left[\sum_{i=1}^{N}(\mathrm{w}_i - \overline{\mathrm{w}})^2\right]^2} - 3.$$

6 Empirical Analysis of Canadian and U.S. Data

In this section we use the models above and computational techniques developed in the Appendix in an empirical analysis of two data sets. This data is used in Ferrall (1997) and is described in detail there.[11] The Canadian data includes 17-24 year old school leavers in 1986-87 and is taken from the Labour Market Activity Survey. Ferrall then selected U.S. data from the National Longitudinal Survey of Youth to match the Canadian data as closely as possible. To ensure that the data set is as homogeneous as possible we include only those individuals with no post-secondary schooling. After deleting observations with incomplete spells or missing wage data, we have data on d_i, w_i and j_i for N=224 Canadians and N=165 Americans.[12] Spells are measured in weeks and wages are weekly in the local currency. Table 1 provides some summary statistics.

Table 1: Summary Statistics

	Mean	St.Dev.	Minimum	Maximum
U.S.A.				
d	7.91	13.24	0.00	58.00
w	152.67	79.63	11.20	424.00
j	24.62	18.95	1.00	77.00
Canada				
d	10.51	16.51	0.00	74.00
w	212.46	99.86	45.00	592.00
j	23.84	19.23	1.00	77.00

[11] The data is available at ftp://www.amstat.org/jbes/View/97-2-APR/ferrall.

[12] We also deleted four outliers in the Canadian and one in the U.S. data set with suspiciously high weekly wages. We suspect that the individuals involved either misreported wages or were very different from the other individuals. If the latter were the case, then the assumption of homogeneity would be violated. Furthermore, some of the workers included were part-time. It is unclear whether they should be treated as homogeneous with full-time workers. In a more substantive empirical analysis than the present one, such data issues would be investigated more carefully.

6.1 Discussion of Prior

Bayesian analysis is based on the posterior distribution which is proportional to the likelihood times prior. It is common in Bayesian analysis to use priors that combine nicely with the likelihood function (e.g. natural conjugate priors). The computational methods developed in the Appendix mean that there is no particular benefit to using such priors of convenience — any prior can be used in a straightforward manner. We choose one particular prior for illustrative purposes. Note, of course, that a myriad of other prior choices, including noninformative, can be made if the researcher wishes.

Given we have already specified $r \sim f_N(\gamma b + (1-\gamma)p, \mathrm{sdr}^2)$, for the basic equilibrium search model we need only specify $p(\theta)$ which we do as:

$$p(\theta) = p(\lambda_0)p(\lambda_1|\lambda_0)p(\delta)p(\mathrm{b})p(\mathrm{p}).$$

We let the positive parameters have Gamma distributions and the others Normal distributions. That is, *a priori,*

- $\lambda_0 \sim f_G(\mu_{\lambda_0}, \nu_{\lambda_0})$
- $\lambda_1 \sim f_G(\mu_{\lambda_1}, \nu_{\lambda_1})$
- $\delta \sim f_G(\mu_\delta, \nu_\delta)$
- b$\sim f_N(\mathrm{b}_0,\mathrm{B}_0)$
- p$\sim f_N(\mathrm{p}_0,\mathrm{P}_0)$.

Of course, by letting degrees of freedom in the Gamma go to zero and variance in the Normal go to infinity we can get priors which are noninformative relative to the data.

For r, we carry out a prior sensitivity analysis by setting sdr=0.01,1.0 or 100 as indicated below. Since r is measured in dollars, by setting sdr=0.01 we are almost imposing worker optimality (i.e. this prior says workers choose the optimal reservation wage to within a cent or two). The value sdr=1 is perhaps more reasonable (i.e. it implies workers choose optimally to within a dollar or two). A noninformative prior is implied by sdr=100.

For the other parameters, our general prior elicitation strategy for is to make reasonable guesses about prior means and then be relatively noninformative by selecting suitable values for the other hyperparameters. We elicit the prior means by reasoning that job offers might come to the unemployed every 10 weeks and layoffs every 50 weeks. Furthermore, job offers to the employed probably are much more infrequent (e.g. perhaps 5 times more infrequent) than job offers to the unemployed. These values imply $\mu_{\lambda_0} = 0.10$, $\mu_{\lambda_0} = \frac{\lambda_0}{5}, \mu_\delta = 0.02$, and we set $\nu_{\lambda_0} = \nu_{\lambda_1} = \nu_\delta = 3$ indicating a large degree of prior uncertainty. Loosely speaking, we are assuming the prior contains as much information as would a sample of N=3 observations. Note that this prior has the reasonable implication that λ_0 and λ_1 are correlated with one another. For b we select $b_0 = 0$ and $B_0 = 100$. These are relatively noninformative values which allow the value of leisure to be negative. The latter characteristic is reasonable if work experience is valuable for future career prospects.

We find it harder to elicit a prior for p and, hence, draw on a small amount of data information to aid in our selection of prior mean. In particular, we set λ_0, λ_1 and δ to their prior means, r to the minimum observed wage and h to the maximum wage. Then we assume worker optimality and use equations (3) and (6) to back out a value for p_0. For the U.S. this strategy implies $p_0 = 562.67$. For Canada the corresponding figure is 774.36. We set $P_0 = 200^2$ indicating a noninformative prior for this parameter.

For the model with nonlinear production function, we use the same prior for the parameters in θ with the inequality restrictions outlined in Section 3 imposed. It is possible, but difficult to calculate the optimal reservation wage. Hence, for simplicity, we assume $r \sim f_N(r_0, R_0)$ and set r_0 equal to the minimum observed wage and $R_0 = 2.0$. For the new parameters we assume, *a priori,* $\varepsilon \sim f_N(1, \varepsilon_0)$ and $m \sim f_N(m_0, M_0)$. We set $\varepsilon_0 = 1.0$, $m_0 = 1.0$ and $M_0 = 0.5$.

For the model with unobserved heterogeneity we assume, additionally, that $\mu_P = 100$, thus allowing for a moderately large degree of heterogeneity.

6.2 Empirical Results

Table 2 gives posterior means and standard deviations for all of the parameters of all the models for the two data sets. Gaps in the table indicate either that the relevant parameter was not in the model (e.g. ε in the basic model), varied across individuals (e.g. r in the heterogeneous model) or was unidentified and *a priori* independent of all other parameters (e.g. b in the nonlinear production function model).

Table 2: Posterior Properties of Parameters for Various Priors

	sdr=.01		sdr=1		sdr=100		Nonlin		Hetero	
	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
U.S.										
λ_0	0.117	0.008	0.100	0.008	0.115	0.009	0.126	0.009	0.133	0.006
λ_1	0.015	0.002	0.016	0.003	0.014	0.002	0.010	0.002	0.018	0.002
δ	0.040	0.002	0.037	0.003	0.033	0.003	0.037	0.002	0.023	0.002
b	-597	30	-520	51	-201.93	88	—	—	-752	125
p	1221	95	1149	141	1156	125	609	79.5	620	65.3
r	7.74	3.61	7.95	3.55	8.21	2.79	10.30	1.69	—	—
h	578	2.42	575	2.77	574	2.75	—	—	—	—
ε	—	—	—	—	—	—	1.36	0.17	—	—
m	—	—	—	—	—	—	1.02	0.26	—	—
P	—	—	—	—	—	—	—	—	15.6	2.63
Can.										
λ_0	0.096	0.006	0.086	0.006	0.092	0.006	0.095	0.004	0.137	0.004
λ_1	0.011	0.002	0.013	0.002	0.012	0.002	0.015	0.005	0.018	0.001
δ	0.034	0.002	0.037	0.003	0.035	0.003	0.033	0.005	0.020	0.001
b	-560	56.0	-424	41	-163	83	—	—	-813	36.1
p	1314	93	1288	137	1300	130	821	87	570	15.9
r	42.2	2.90	42.8	2.21	42.6	2.34	44.1	0.71	—	—
h	594	2.15	594	1.89	594	2.11	—	—	—	—
ε	—	—			—	—	1.59	0.19	—	—
m	—	—	—	—	—	—	0.60	0.33	—	—
P	—	—	—	—	—	—	—	—	12.0	0.19

Notes to table: Columns labelled "sdr=" refer to the basic equilibrium search model with different priors on the reservation wage. The columns labelled "Nonlin" and "Hetero" refer to the equilibrium search model with nonlinear production function and heterogeneity in worker productivity, respectively.

Overall, the results seem reasonable and our Bayesian methods seem practical for empirically relevant data sets such as those considered here.[13] A few points are worth noting only briefly:

- Canadian and U.S. results are quite similar. This is not surprising given the similarities in summary statistics notable in Table 1.
- Results are fairly robust across models. This is especially true for λ_0 and $\lambda_1 + \delta$. Since these are largely determined by the duration data, j and d, rather than wage density, this is not surprising.

[13] The working paper version of this paper, available at http://www.ed.ac.uk/˜gkoop/, contains extensive analysis involving artificial data sets which confirms the computational practicality and reliability of Bayesian methods.

- The value of leisure seems to be negative for both countries. Given that school leavers are not eligible for unemployment insurance, and work experience is beneficial for future job prospects (or unemployment eligibility), this result is plausible.

- There is strong evidence of nonlinearity in the production function for both data sets. Estimates of ε are both greater than one indicating a convex production function.

- There is also substantial evidence of heterogeneity in productivity for both data sets.

It is worthwhile to discuss the results of our prior sensitivity analysis and tests for worker optimality in the basic equilibrium search model in more detail. Looking at the left hand side of Table 2, it can be seen that a crucial difference between results for the three priors lies in the posterior for b. Another key difference lies in the posterior odds. For Canada, the posterior odds in favor of worker optimality are 0.0069, 0.8769 and 0.9890 for sdr=100.0, 1.0 and 0.01, respectively. For the USA, these posterior odds are 2.4×10^{-7}, 1.02 and 1.07, respectively. To understand these results, consider Figure 2 which plots the prior for b along with the posterior for each of the three values of sdr for the USA (the comparable figure for Canada is similar). Remember that b is not identified and learning about it only occurs through prior beliefs about whether workers are behaving optimally (see equation (7)). The parameters λ_0, λ_1 and δ are quite precisely estimated, largely from the duration data. Since the minimum and maximum wage provide a great deal of information about r and h, all the parameters of the model have fairly small standard deviations. Hence, when worker optimality is nearly imposed through sdr=0.01, b is quite well determined at the value implied by the reservation wage equation (i.e. most of the posterior probability is located in the region [-400,-700]). However, we have deliberately chosen a prior for b which is centered in a very different region than the data plus reservation wage equation implies (i.e. most of the prior probability is in the region [-200,200]). When we are very uncertain about whether workers are acting optimally (sdr=100), the optimal reservation wage equation is of little use in learning about b and the posterior in this case is located near the prior. The prior with sdr=1.0 is an intermediate case. This figure illustrates how different degrees of confidence in the implications of economic theory can result in different degrees of learning about nonidentified structural parameters such as b. The example here is purely illustrative but, in a serious empirical study it is likely that the researcher would be elicit priors for both b and r in such a way that substantial learning about b occurs even though it is not identified.

The posterior odds compare the equilibrium search model[14] of Section 2 to that model with optimality imposed. Since, as sdr goes to zero, optimality is imposed, the case sdr=0.01 is essentially equivalent to the model with optimality imposed. Hence, a posterior odds ratio of roughly 1.0 is to be expected (i.e. they are basically the same model and so are equally plausible). Intuition for the strong evidence against worker optimality when sdr=100 is provided by Figure 2 and the formula for the posterior odds ratio given in (8). The numerator of the posterior odds ratio, $p(\mathrm{r}=\gamma\mathrm{b}+(1-\gamma)\mathrm{p}|\mathrm{Y})$, becomes very small when sdr=100. Loosely speaking, the prior for b we are using is "bad" in that it disagrees with the value implied by the data plus worker optimality (see Figure 2). When sdr=100, the posterior for b strongly reflects the prior and, hence, is also "bad". This means the posterior for r (which is near the minimum observed wage) is very different from the posterior for γb+(1-γ)p and the numerator of (8) is close to zero. So, in essence, this model provides evidence against worker optimality since its prior is "bad" and the "bad" prior has a great influence on posterior results. As our prior beliefs about worker optimality are strengthened (e.g. sdr=1.0 or 0.01) this "bad" prior for b is overruled by the data[15] and the posterior for r becomes closer to γb+(1-γ)p. Hence, these latter models provide more support for worker optimality. This interpretation of tests for optimality as reflecting the reasonableness of the prior for the unidentified parameter (here b) is stressed in Koop and Poirier (1997). In the latter paper, more examples are given involving both "good" and "bad" priors and the reader is referred there for more details about statistical aspects of testing involving unidentified parameters.

Table 3 present posterior predictive p-values[16] and the message of this table is quite clear: None of these models do a good job of fitting the wage density, but the nonlinear production function and heterogeneous models do a much better job that the basic equilibrium search model.

Table 3: Posterior Predictive p-values

[14] Note that, in Bayesian econometrics, a model is a likelihood function and a prior.

[15] Note that, in the limiting case where worker optimality is imposed (i.e. sdr=0), the prior for b plays no role whatsoever.

[16] Since the different priors used in the basic equilibrium search model have little effect on posterior predictive p-values, we present results only for the case where sdr=0.01 and near optimality is imposed.

	Mean	St.Dev.	Skew	Kurtosis
U.S.				
Basic	3.1×10^{-4}	4.7×10^{-3}	9.9×10^{-5}	2.5×10^{-4}
Nonlinear	0.22	0.84	3.1×10^{-4}	9.8×10^{-6}
Hetero.	0.02	0.01	3.4×10^{-3}	3.2×10^{-3}
Canada				
Basic	7.5×10^{-4}	2.1×10^{-4}	3.0×10^{-4}	2.6×10^{-4}
Nonlinear	0.18	0.49	1.2×10^{-3}	3.6×10^{-4}
Hetero.	3.4×10^{-3}	1.8×10^{-3}	1.3×10^{-3}	6.0×10^{-4}

Notes to table: The rows labelled "Basic", "Nonlinear" and "Hetero" refer to the basic equilibrium search model with optimality imposed, and the ones with nonlinear production function and heterogeneity in worker productivity.

How should we interpret the results in Table 3? From a statistical point of view, it is obvious what is happening. The empirical wage densities (see Figure 1) are bell-shaped and cannot be fit well by the increasing wage density implied by the basic equilibrium search model (see equation (2)). The extra parameters in the nonlinear model can be used to make the wage density decreasing or increasing and, hence, it fits the data better. But since it cannot allow for a bell-shaped wage distribution it does not fit this apect of the wage data. The model with heterogeneity implicitly tries to fit a bell-shaped distribution with N different increasing distributions. Furthermore, the parameters δ and λ_1 are held constant for all individuals so, loosely speaking, a better fit can be achieved only by adjusting the upper and lower bounds of the distribution. Results indicate that this is a hard thing to do if only a moderate amount of heterogeneity is allowed for.

This section is meant solely as an illustration of Bayesian techniques and model development strategies. Hence, we do not find this failure to be disturbing. Furthermore, to quote Ridder and van den Berg (1997), page 103, "... consistency with all known empirical regularities is not expected. Firms and workers are not identical, and some allowance of this fact must be needed to obtain an acceptable description of the data". The challenge for future work is to incorporate such heterogeneity in the context of an optimizing theoretical model. The models developed by Mortensen (1998) and Nielsen and Rosholm (1999), allow for firm specific human capital through investments in training and, hence, imply heterogeneity in productivity which is endogenous rather than ad hoc. These papers move us another step towards the goal of creating models which are both based in rigorous economic theory and fit the data. The techniques developed in the present paper (i.e. posterior simulation algorithms, use of the transformation method to numerically evaluate the likelihood function, data augmentation, posterior predictive p-values and posterior odds for testing optimality) should be of great use in the empirical analysis of such models.

7 Conclusions

This paper has extended the existing literature in the following ways:

1. Bayesian methods have been developed for the analysis of equilibrium search models and extensions thereof.

2. These methods have been shown to be computationally feasible and reliable.

3. The logic of Bayesian analysis, which forces the researcher to specify a prior, leads to the development of models which are more flexible than the equilibrium search model but are centered over it. In essence, we have considered ways of approximately imposing optimality or allowing departures from optimality to exist but ensuring that such departures are small.

4. Methods are developed to test for optimal behavior.

5. The posterior predictive p-value approach of Meng (1994) has been shown to be a useful way of measuring the ability of equilibrium search model to fit data.

6. A complete Bayesian analysis of the equilibrium search model with nonlinear production function is carried out. This model has received little empirical attention since a closed form expression for the wage distribution does not exist. However, we show how this lack does not preclude Bayesian analysis.

7. A model with unobserved heterogeneity is developed and estimated using data augmentation methods. We argue that Bayesian methods are well-suited for such models since priors can be used to control the degree of heterogeneity that is allowed for.

The methods are applied to Canadian and U.S. data sets. Our Bayesian methods work well, but indicate strong rejections of the basic equilibrium search model and any of the extensions considered in this paper.

8 Appendix: Computational Methods

It is now standard for Bayesian inference to be carried out by simulating the posterior (see Geweke, 1999, for an overview of computational Bayesian methods). Bayesian estimation of non-equilibrium job search models has been carried out by Lancaster (1997), Kiefer and Steel (1998) and Koop and Poirier (1997). These papers use simple models which allow for a particularly convenient way of drawing from the posterior (e.g. Gibbs sampling, which draws repeatedly from the posterior conditionals). Unfortunately, the empirical equilibrium search model has a non-standard form and the posterior conditionals are such that a simple Gibbs sampler cannot be set up. Hence, a more general simulation algorithm such as importance sampling or Metropolis-Hastings (Chib and Greenberg, 1995) must be used.

8.1 The Basic Equilibrium Search Model

There are many possible algorithms which could be used. In this paper, we use a random walk chain Metropolis-Hastings algorithm. Christensen and Kiefer (1997) provide a detailed analysis of the likelihood function of this model parameterized in terms of $\lambda_0, \lambda_1, \delta$,r,h. They point out that maximum likelihood estimation of the model is simple since the MLE's for r and h are superconsistent and the profile likelihood for the remaining parameters is globally concave. In other words, the likelihood function is well-behaved and unimodal. This suggests that the random walk chain Metropolis-Hastings algorithm which wanders through the parameter space is likely to work well (i.e. it won't miss modes or get stuck).

The goal is to draw a sample, $\phi^1, ..., \phi^S$ from the posterior, p(ϕ|Y), where Y is the data (i.e. Y contains d_i, w_i and j_i for i=1,..,N). Given such a sample, posterior moments, etc. of the parameters can be calculated. Since the exact form of p(ϕ|Y) does not simplify, we do not produce it here (i.e. it simply is proportional to the product of (2), (4) and (5) times the prior given above). The Metropolis-Hastings algorithm works by taking an arbitrary starting value, ϕ^0, and then drawing ϕ^{*s} for s=1,...,S from a candidate-generating density q($\phi^{*s}|\phi^{s-1}$). The random walk chain Metropolis algorithm defines q(.|.) such that:

$$\phi^{*s} = \phi^{s-1} + z,$$

where z is an increment random variable. We choose z $\sim f_N(0, V_z)$.[17] The algorithm works as follows:

- Choose an arbitrary starting value ϕ^0.
- Repeat for s=1,..,S.

[17]Note that the prior (and hence posterior) implicitly puts zero weight on non-positive values of λ_0, λ_1 and δ. The likelihood (and hence posterior) places zero weight on values of r greater than the minimum observed wage and h less than the maximum observed wage. Hence, candidate draws in any of these regions are accepted with probability 0.

- Generate ϕ^{*s} from $q(\phi^{*s}|\phi^{s-1})$.
- Generate u from the Uniform [0,1] distribution.
- If u□ $\frac{p(\phi^{*s}|Y)}{p(\phi^{s-1}|Y)}$ then $\phi^s = \phi^{*s}$.
- Else $\phi^s = \phi^{s-1}$.

Note that taking one draw can be done very quickly. Virtually all that is required is to draw from the Normal distribution and evaluate the prior times likelihood at the draw. This enables us to set S very large (i.e. S=50,000) to obtain accurate results. A big advantage of the random walk chain Metropolis algorithm over possible alternatives (e.g. importance sampling, independence chain Metropolis-Hastings or rejection sampling) is that it is not important that q(.|.) approximate $p(\theta|Y)$. However, it is important that the random walk chain Metropolis-Hastings algorithm move around through the parameter space of the posterior. This is achieved through careful selection of V_z. In practice, we set $V_z = c \times var(\theta|Y)$, where $var(\theta|Y)$ is the posterior covariance matrix calculated using initial experimental runs. Chib and Greenberg (1995) suggest that V_z should be selected so that between 0.25 and 0.50 of the candidate draws are accepted. We experiment with different values of c to insure an acceptance rate in this region (usually $c = 0.25$ works well).

The calculation of posterior predictive p-values requires simulation of artificial data from $p(g(Y^*)|M,\phi)$. This can be done quite simply by first drawing a wage from (2) using rejection methods, then an unemployment duration from the exponential distribution given in (4), then a job duration conditional on the drawn wage from the exponential distribution given in (5).

8.2 The Equilibrium Search Model with Nonlinear Production Function

To obtain a Metropolis-Hastings algorithm for the extension to nonlinear production function, we need to evaluate the likelihood function at each replication. The only component of the likelihood function, $L=\prod_{i=1}^{N} f(d_i)f(w_i)f(j_i|w_i)$, which is affected by the extension to nonlinearity in production is $f(w_i)$. This does not have a closed-form expression, but we can simulate R artificial values from this density.[18] These R wages can be used to approximate f(w). In practice, we use R=1,000 artificially drawn wages and use a nonparametric kernel algorithm to create an approximation for f(w). This seems to provide a reasonably accurate

[18] This strategy is also used in the posterior predictive p-value calculation.

approximation in a moderate amount of computer time. Of course, if computer time is not a problem R=10,000 or more can be chosen.

The algorithm can be summarized as follows:

- Choose an arbitrary starting value ω^0.
- Repeat for s=1,..,S.
- Generate ω^{*s} from q($\omega^{*s}|\omega^{s-1}$).
- Generate w^r from f(w;ω^{*s}) for $r = 1, .., R$ using the transformation method.
- Use w^r for $r = 1, .., R$ to approximate f(w) and, hence, p(ω^{*s}|Y)
- Generate u from the Uniform [0,1] distribution.
- If u□ $\frac{\text{p}(\omega^{*s}|\text{Y})}{\text{p}(\omega^{s-1}|\text{Y})}$ then $\omega^s = \omega^{*s}$.
- Else $\omega^s = \omega^{s-1}$.

One other subtlety arises relating to the prior. The Metropolis-Hastings algorithm requires the evaluation of the kernel of the posterior (i.e. the prior times likelihood) at each replication. In practice, we evaluate the prior at each replication, *ignoring the restrictions on the prior described in Section 3*. But this does not matter. To see this, consider the untruncated Normal prior for m evaluated at a replication and note that it is proportional to the prior for a truncated Normal evaluated at the same point. The factor of proportionality will be the same for every replication and can be omitted (it cancels in the acceptance probability calculation step). The same considerations imply to the other, more complicated restrictions. Of course, when we simulate w^r for r=1,...,R, we do impose the restrictions (i.e. we discard all draws which violate the restrictions), but when calculating the acceptance probability they can safely be ignored. However, when calculating the posterior odds ratio using the Savage-Dickey density ratio we do have to be careful with the integrating constant in the prior. Hence, we do not calculate posterior odds for this model. At some computational cost, it would be possible to do so by simulating from the prior as well as the posterior and using kernel methods to approximate prior and posterior.

8.3 The Equilibrium Search Model with Heterogeneity

In order to simulate random draws from the posterior distribution, we use a Metropolis-Hastings with data augmentation algorithm. That is, note that, if we knew what the p$_i$'s were, the model would simplify enormously. In fact, the posterior conditional on the p_is, $p(\varphi, P|\text{Data}, p_i$ for i=1,..,N), is exactly the same as the posterior for the basic equilibrium search model times the prior for P given in (17). Hence, the random walk chain Metropolis-Hastings algorithm we developed above can be used (trivially modified to include the exponential prior for P). Furthermore, $p(p_i|\text{Data}, \varphi, P)$ can be simulated in a simple way using (16). Formally, the algorithm is:

- Choose arbitrary starting values φ^0, P^0.
- Repeat for s=1,..,S.
- Generate p_i from $f_N(p^{s-1}, P^{s-1})$ for i=1,..,N.
- Generate φ^s, P^s from $p(\varphi, P|\text{Data}, p_i^s$ for i=1,..,N) using the algorithm for the basic equilibrium search model (augmented to include the prior for P).

9 References

Bontemps, C., Robin, J. and van den Berg, G. (1995). "Equilibrium search with on-the-job search," manuscript, CREST-INSEE, Vrie Universiteit Amsterdam.

Burdett, K. (1990). "Empirical wage distributions: A new framework for labor market policy analysis," in Hartog, J., Ridder, G. and Theeuweis, J. (eds.), **Panel Data and Labor Market Studies**. North-Holland, Amsterdam.

Chib, S. and Greenberg, E. (1995). "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, 49, 327-335.

Christensen, B. and Kiefer, N. (1997). "Inference in nonlinear panel models with partially missing observations: The case of the equilibrium search model," *Journal of Econometrics*, 79, 201-219.

Eckstein, Z. and Wolpin, K. (1995). "Duration to first job and the return to schooling: Estimates from a search-matching model," *Review of Economic Studies*, 62, 263-286.

Ferrall, C. (1997). "Unemployment insurance eligibility and the school-to-work transition in Canada and the United States," *Journal of Business and Economic Statistics*, 15, 115-129.

Gelman, A. and Meng, X.-L. (1996). "Model checking and model improvement," pages 189-201 in W. Gilks, S. Richardson and D. Spiegelhalter (eds.), **Markov Chain Monte Carlo in Practice**, Chapman and Hall, London.

Geweke, J. (1999). "Using simulation methods for Bayesian econometric models: Inference, development, and communication" (with discussion and rejoinder), *Econometric Reviews*, 18, 1-126.

Kiefer, N. and Neumann, G. (1993). "Wage dispersion with homogeneity: The empirical equilibrium search model," in H. Bunzel, P. Jensen and N. Westergard-Nielsen (eds.), **Panel Data and Labour Market Dynamics**. Elsevier Science Publishers, Amsterdam.

Kiefer, N. and Steel, M.F.J. (1998). "Bayesian analysis of the prototypical search model," *Journal of Business and Economic Statistics*, 16, 178-186.

Koop, G. and Poirier, D. (1997). "Testing for optimality is stationary job search models based on reservation wages," manuscript, Dept. of Economics, Univ. of Edinburgh.

Lancaster, T. (1997). "Exact structural inference in job-search models," *Journal of Business and Economic Statistics*, 15, 165-179.

Meng, X.-L. (1994). "Posterior predictive p-values," *Annals of Statistics*, 22, 1142-1160.

Mortensen, D.T. (1998). "Equilibrium unemployment with wage posting: Burdett-Mortensen meet Pissarides," Working paper 98-14, Centre for Labour Market and Social Research, University of Aarhus.

Mortensen, D.T. (1990). "Equilibrium wage distribution: A synthesis," in Hartog, J., Ridder, G. and Theeuweis, J. (eds.), **Panel Data and Labor Market Studies**. North-Holland, Amsterdam.

Nielsen, M. and Rosholm, M. (1999). "Wages, training, and job turnover in a search-matching model," manuscript, Department of Economics, University of Aarhus.

Press, W., Flannery, P., Teukolsky, S. and Vetterling, W. (1986). **Numerical Recipes: The Art of Scientific Computing**. Cambridge University Press, Cambridge.

Ridder, G. and van den Berg, G. (1997). "Empirical equilibrium search models," Chapter 4 **Advances in Economics and Econometrics: Theory and Applications**, Proceedings of the Seventh World Congress of the Econometric Society, S. Kreps and K. Wallis (eds.). Cambridge University Press, Cambridge.

Van den Berg, G. (1999). "Empirical inference with equilibrium search models of the labour market," *Economic Journal*, 109, F283-F307.

Van den Berg, G. and Ridder, G. (1998) "An empirical equilibrium search model of the labor market," *Econometrica*, 66, 1183-1222.

Verdinelli, I. and Wasserman, L. (1995). "Computing Bayes factors using a generalization of the Savage-Dickey density ratio," *Journal of the American Statistical Association*, 90, 614-618.

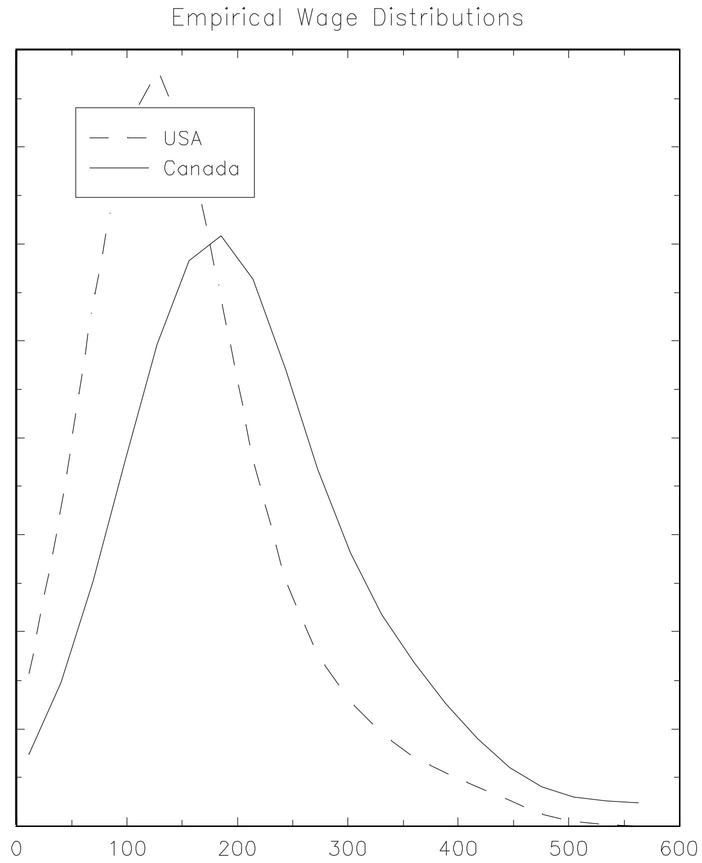
Empirical Wage Distributions
USA
Canada
0
100
200
300
400
500
600

Figure 1:

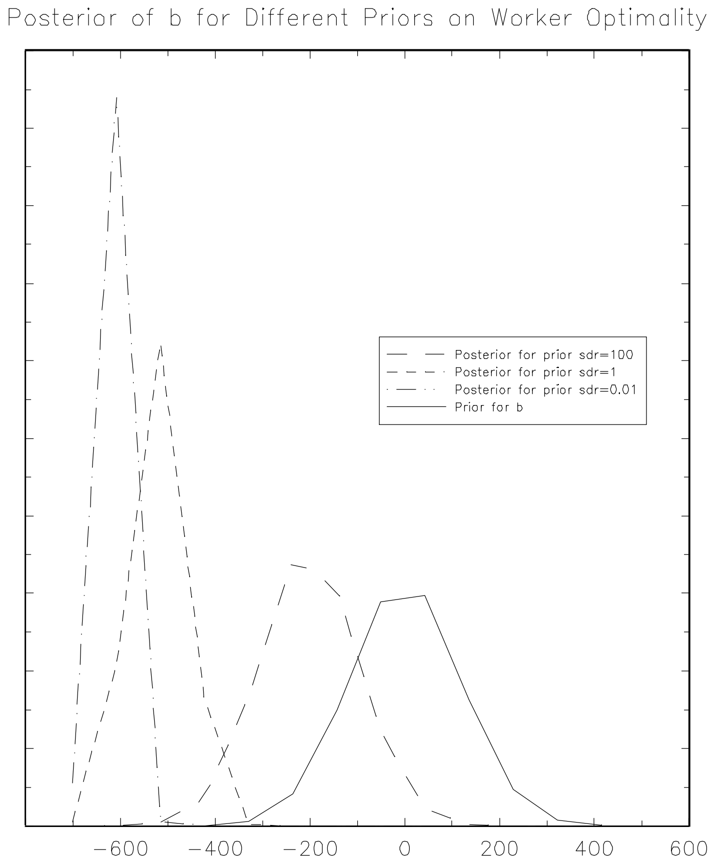
Posterior of b for Different Priors on Worker Optimality
Posterior for prior sdr=100
Posterior for prior sdr=1
Posterior for prior sdr=0.01
Prior for b
−600
−400
−200
0
200
400
600

Figure 2: