

# Parametric and Nonparametric Inference in Equilibrium Job Search Models\*

GARY KOOP

*Department of Economics, University of Leicester,  
Leicester, LE1 7RH, U.K.*

*E-mail: Gary.Koop@leicester.ac.uk*

October 2003

Equilibrium job search models allow for labor markets with homogeneous workers and firms to yield nondegenerate wage densities. However, the resulting wage densities do not accord well with empirical regularities. Accordingly, many extensions to the basic equilibrium search model have been considered (e.g. heterogeneity in productivity, heterogeneity in the value of leisure, etc.). It is increasingly common to use nonparametric forms for these extensions and, hence, researchers can obtain a perfect fit (in a kernel smoothed sense) between theoretical and empirical wage densities. This makes it difficult to carry out model comparison of different model extensions. In this paper, we first develop Bayesian parametric and nonparametric methods which are comparable to the existing non-Bayesian literature. We then show how Bayesian methods can be used compare various nonparametric equilibrium search models in a statistically rigorous sense.

---

\*I would like to thank Mark Steel for helpful comments.

# 1 Introduction

Most potential policy initiatives in labor markets (e.g. minimum wage legislation) have effects on both workers<sup>1</sup> and employers. Hence, it is strongly desirable that policy analysis be carried out in the context of an equilibrium model which takes into account the interaction between worker and firm behavior. This consideration has motivated the development of a large literature on equilibrium job search models (see, among many others, van den Berg and Ridder, 1998, Ridder and van den Berg, 1997, Bontemps, Robin and van den Berg, 1999, Mortenson, 1990, 1998, Burdett, 1990 and Kiefer and Neumann, 1993). Details of the equilibrium search model will be provided in the next section, suffice it to note here that it assumes a fixed number of homogeneous workers and firms with full information about all the parameters in the model. Workers and firms interact in a market where jobs arrive and are destroyed at random exogenous rates. Assuming workers maximize the expected value of their future income stream and firms maximize expected steady state profit flow, it turns out that a non-degenerate wage offer distribution exists with density given by:

$$f(w_i) \propto \frac{1}{\sqrt{p - w_i}},$$

for  $w_i \in [r, h]$  where  $w_i$  is the wage offer to individual  $i$  for  $i = 1, \dots, N$ ,  $p$  is the productivity of workers (i.e. their marginal value product),  $r$  their reservation wage and  $h$  is the upper bound on the wage distribution. Further details will be provided in the next section, at this point the key thing to note is that the wage offer density is increasing in  $w$ . However, empirical wage offer densities are invariably roughly bell-shaped, but skewed to the right. Accordingly, a conflict exists. On the one hand, the equilibrium search model is structural and, hence, well-suited for policy analysis. On the other hand, it does a very poor job of fitting the data.

---

<sup>1</sup>Throughout this paper we use the term "worker" to refer to individuals, regardless of whether they are employed or unemployed.

In response to this conflict, many of the recent papers in the literature add a new component to the basic equilibrium search model (see, among others, Bontemps, Robin and van den Berg, 1999, 2000, Mortenson, 1998, Nielsen and Rosholm, 1999, Koning, Ridder and van den Berg, 1995). This component might be unobserved heterogeneity in productivity or in workers' value of leisure, measurement error in wages or an economic concept like an investment in training function. From a statistical point of view, the wage offer density can be written as  $f(w_i|\theta_1, h(\theta_2))$ , where  $\theta_1$  and  $\theta_2$  are parameter vectors and  $h(\cdot)$  is the appropriate new component. This approach can be criticized on the ground that economic theory rarely is a reliable guide to the choice of functional form for  $h(\cdot)$ . For instance, Koning, Ridder and van den Berg (1995) assume  $h(\cdot)$  is log-Normal in a model which allows for heterogeneity in productivity. The choice of log-Normality does not have any basis in economic theory. One of the chief advantages of equilibrium search models over earlier partial search models (see, e.g., Lancaster, 1997) was that the former used a wage offer distribution which was derived using an assumption of rational firm behavior, whereas the latter merely made assumptions of convenience (e.g. the wage density is assumed to be log-Normal or exponential). Merely adding a new *ad hoc* component such as  $h(\cdot)$  into the equilibrium search model could be interpreted as a regressive step away from structural and back towards reduced form modelling. Nevertheless, the use of a parametric form for  $h(\cdot)$  implies a parametric likelihood function and, thus, estimation, testing and model comparison can be done using either Bayesian or non-Bayesian likelihood-based paradigms.

Recent work (e.g. Bontemps, Robin and van den Berg, 1999, 2000 and Nielsen and Rosholm, 1999), perhaps in realization that economic theory is rarely a guide to choice of  $h(\cdot)$ , does not specify a functional form for it. Instead nonparametric kernel smoothing algorithms are used. In essence, kernel methods are used to fit  $f(\cdot)$  (or the earnings density which is a simple transformation of  $f(\cdot)$ ) and then the implied form of  $h(\cdot)$  required to provide this perfect fit can be obtained. Since  $f(\cdot)$  is usually the key element of the likelihood function, this method can be loosely interpreted as saying "Let us fit the likelihood

function perfectly<sup>2</sup>, and then see what it implies  $h(\cdot)$  must look like.” Such an approach has the advantage of not making an *ad hoc* functional form assumption for  $h(\cdot)$ , but also has drawbacks. For instance, it attributes all of the departures from the basic equilibrium search model to a particular source. One paper might say all such departures are due to heterogeneity in productivity, another might attribute them to measurement error, another to effects relating to investment in training, etc. It is difficult to choose between different models since all fit the likelihood function perfectly. In the previous paragraph, the parametric approach was criticized since it involved an *ad hoc* functional form assumption. The nonparametric approach does not suffer from such a criticism, but does still depend on strong assumptions that are difficult to test (e.g. that all departures from the basic equilibrium search model are due to unobserved heterogeneity in productivity).

These considerations motivate the present paper. In previous work (Koop, 2001), it was argued that Bayesian methods provided a useful tool for analyzing the equilibrium search model and computational methods for doing so were developed. One contribution of the present paper is similar to this. That is, methods of Bayesian inference for the extensions of the equilibrium search model described in the previous paragraphs are developed. A particular focus is the development of a Bayesian nonparametric approach which is comparable to the non-Bayesian nonparametric approach of Bontemps, Robin and van den Berg (2000). There is a sizeable statistical literature on Bayesian nonparametric methods, but relatively little of this has made it over to the field of econometrics (exceptions include Campolieti, 1997 and Ruggiero, 1994). In the present paper, Bayesian nonparametric methods are developed for various extensions of the basic equilibrium search model.

A second contribution of the paper derives from the fact that all of the models in the paper, including the nonparametric ones, involve a well-defined probability density for the data. This implies that standard statistical tools for model comparison<sup>3</sup> can be calculated.

---

<sup>2</sup>Here, and throughout the paper, by "perfect" fit we mean only in a kernel smoothed sense.

<sup>3</sup>The posterior model probability (i.e. the probability that a given model generated the data) is the

Using the Bayesian approach, two nonparametric models (i.e. with  $h(\cdot)$ s arising from different sources of heterogeneity) can both fit the data perfectly yet be compared using formal statistical methods. The paper describes how to do this and how to interpret the model comparison measures which are calculated. Loosely speaking, the Bayesian approach forces the researcher to specify what a reasonable form for  $h(\cdot)$  might be, but they do not impose this form. If the data force large departures from this reasonable form, Bayesian tools of model comparison will indicate less support for the model. We argue that this formalizes what researchers do informally. For instance, suppose the researcher believes that heterogeneity in productivity may exist, but is relatively small and unimodal. However, the nonparametric distribution exhibits large dispersion and is multi-modal. In this case, the researcher might reject the model with heterogeneity in productivity. The Bayesian methods outlined below are a statistically coherent formalization of the sensible but informal actions of the researcher.<sup>4</sup>

The paper is organized as follows: The following section outlines the basic equilibrium search model. The third section derives equilibrium search models with heterogeneity introduced in various parametric ways. The fourth section is similar to the third, but heterogeneity is introduced in various nonparametric ways. The fifth section contains an empirical application using the data set of Bowlus, Kiefer and Neumann (1995). Computational algorithms are provided in the Appendix.

---

main tool of Bayesian model comparison. The posterior odds ratio is the ratio of two posterior model probabilities. The posterior odds ratio is also the ratio of the marginal likelihoods for the two models (i.e. the Bayes factor) times the prior odds ratio.

<sup>4</sup>It is worth stressing that one reaction to the literature discussed in this paper might be that the researcher should just go out and get better data (e.g. matching employer/employee data). This is a sensible reaction. The point of the present paper is to argue that, insofar as such data is not available and one is interested in working with the sorts of (widely used) models described here, it is good to use a formal statistical approach such as the one adopted here.

## 2 The Basic Equilibrium Search Model

The basic equilibrium search model assumes 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  $\lambda_0$ . Once workers are employed they can continue searching and new job offers arrive at a rate of  $\lambda_1$ . Firms lay off workers at a rate of  $\lambda_2$ . 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. Here we provide only the final results and a few details of the derivation. Complete details, along with extensive intuition, are given in many of the papers listed at the beginning of the Introduction. Ridder and van den Berg (1997) includes a great deal of survey material which is particularly useful.

If workers maximize the expected value of future income, then unemployed workers will set a reservation wage,  $r$ , and accept the first wage offer above it. The optimal reservation wage is given by:

$$r = \xi b + (1 - \xi)p, \tag{2.1}$$

where

$$\xi = \frac{\left(1 + \frac{\lambda_1}{\lambda_2}\right)^2}{\left(1 + \frac{\lambda_1}{\lambda_2}\right)^2 + \left(\frac{\lambda_0}{\lambda_2} - \frac{\lambda_1}{\lambda_2}\right) \frac{\lambda_1}{\lambda_2}}.$$

Employed workers accept any 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  $w$  every time a vacancy arises can be derived and denoted by  $l(w)$  (see, e.g., Ridder and van den Berg, 1997, equation (8)). Assuming a linear production function, the steady state profit flow of a firm offering wage  $w$  is given by:

$$\pi(w) = (p - w)l(w).$$

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). Using these facts, it can be shown that:

$$f(w) = \frac{\lambda_2 + \lambda_1}{2\lambda_1} \frac{1}{\sqrt{p-r}\sqrt{p-w}}, \quad (2.2)$$

for  $w \in [r, h]$ . The highest wage offered is given by:

$$h = \beta^2 r + (1 - \beta^2)p, \quad (2.3)$$

where

$$\beta = \frac{\lambda_2}{\lambda_2 + \lambda_1}.$$

The contributions of this paper all relate to the wage offer distribution, as opposed to the duration of unemployment or employment spells. Accordingly, we simplify the analysis by assuming we only observe  $w_i$  for  $i = 1, \dots, N$  where  $w_i$  is the wage offer accepted by initially unemployed worker  $i$ . Identification issues relating to the structural parameters,  $\lambda_0, \lambda_1, \lambda_2, b, p$  are often of great interest in models of search (see Koop and Poirier, 2002 and Koop, 2001). However, to focus the paper, we ignore these issues and work only with the identified parameters which directly affect the shape of the wage density,  $p$  and  $r$ . We

also define  $\tau = \frac{\lambda_2}{\lambda_1}$ ,  $\gamma = \ln(r)$  and  $\pi = \ln(p)$ .<sup>5</sup> Accordingly, the likelihood function for  $W = (w_1, \dots, w_N)'$  is given by:

$$L(W; p, r) = \prod_{i=1}^N f(w_i), \quad (2.4)$$

where

$$f(w_i) = \frac{1 + \tau}{2\tau} \frac{1}{\sqrt{(p-r)}\sqrt{p-w_i}} I(B), \quad (2.5)$$

and  $I(B)$  is an indicator function for the bounds of the wage distribution. That is,  $B$  is defined by  $r < w_i < h$  and  $h = \beta^2 r + (1 - \beta^2)p$  with  $\beta = \frac{1}{1+\tau}$ . Note that all the parameters enter  $B$ .

Bayesian inference can be carried out by specifying a prior and using a Markov Chain Monte Carlo (MCMC) algorithm for posterior simulation. Details are given in the Appendix.

## 3 Parametric Extensions of the Equilibrium Search Model

### 3.1 Heterogeneity in Productivity

The basic equilibrium search model assumes homogeneity of workers and firms. Heterogeneity can be introduced in many ways. Perhaps the most common is to allow for heterogeneity in productivity. Here, following much of the literature (e.g. Koning, Ridder and van den Berg, 1995 or Bontemps, Robin and van den Berg, 1999), we assume heterogeneity

---

<sup>5</sup>Formally,  $\tau$  is identified. However, without data on durations, identification only occurs through the bounds of the wage density. In practice, this means identification is weak and, accordingly, we simply set  $\tau = 2$  which is roughly what is typically observed. An earlier version of this paper treated  $\tau$  as an unknown parameter, and all results on model comparison were qualitatively the same to those given in the present version.



in productivity arises since the labor market is segmented into many smaller homogeneous markets. Each of these homogeneous markets has a different productivity.<sup>6</sup> Following Koning, Ridder and van den Berg (1995), it is assumed that productivity is log-Normally distributed. That is, the log-productivities of the workers,  $\pi_i$  for  $i = 1, \dots, N$ , are i.i.d. draws from the Normal distribution:

$$p(\pi_i) = f_N(\mu_\pi, \sigma_\pi^2), \quad (3.1)$$

where  $\mu_\pi, \sigma_\pi^2$  are unknown parameters to be estimated and  $f_N(c, C)$  denotes the Normal density with mean  $c$  and variance  $C$ . This parametric distribution for productivity can, as in Koning, Ridder and van den Berg (1995), be treated as part of the likelihood function. Alternatively (and equivalently), it can be interpreted as a hierarchical prior for  $\pi_i$ . We find it more natural to use the latter interpretation and, hence, refer to (3.1) as a prior.

The likelihood function is as in the basic model, except for adding a subscript. That is, (2.5) becomes

$$f(w_i) = \frac{1 + \tau}{2\tau} \frac{1}{\sqrt{(p_i - r)}\sqrt{p_i - w_i}} I(B), \quad (3.2)$$

where  $p_i = \exp(\pi_i)$  and  $B$  is defined by  $r < w_i < h_i$  and  $h_i = \beta^2 r + (1 - \beta^2)p_i$ . Note that the assumption of a reservation wage common to all workers,  $r$ , is retained. This can be justified if unemployed workers do not know, *a priori*, which market they will be working in and, hence, all use  $\mu_\pi$  to select their reservations wages.

Computational methods for Bayesian inference of this model are given in the Appendix. Heuristically, conditional on knowing  $\pi_i$  for  $i = 1, \dots, N$ , the posterior for the remaining

---

<sup>6</sup>In the jargon of the literature, we are assuming "between market" rather than "within market" heterogeneity. The latter is much more difficult to deal with theoretically. The interested reader is referred to Bontemps, Robin and van den Berg (1999) or Koning, Ridder and van den Berg (1995) for more discussion of this issue.

parameters is essentially the same as for the basic search model. Furthermore, the posterior for  $\pi_i$  is relatively easy to work with (see Appendix). Hence, an MCMC algorithm with a data augmentation step can be used to efficiently carry out Bayesian inference. In contrast, classical analysis (see, e.g., Koning, Ridder and van den Berg, 1995) involves integrating over the productivity distribution at each iteration, which greatly adds to the computational burden.

### 3.2 Measurement Error in Wages

The wage data used in empirical analyses of job search models is invariably taken from surveys. The measurement error which undoubtedly arises from individuals misrepresenting their wages could explain why the basic equilibrium search model fits empirical wage distributions so poorly, although intuitively it seems unlikely to do so. Remember, the basic job search model implies an increasing wage density when empirical wage densities are roughly bell-shaped. For measurement error to explain this divergence it must be of a form that is somewhat unusual. That is, measurement error must work in such a way to convert an observed bell-shaped wage density into one where the true wages imply an increasing wage density. Nevertheless, measurement error is often mentioned as being important in labor market data sets. For this reason, and in order to illustrate our econometric methodology, we present an equilibrium search model which includes measurement error.

Let  $w_i^*$  be the true, unobserved, wage of individual  $i$  and assume:

$$w_i^* = w_i + \varepsilon_i, \tag{3.3}$$

where  $\varepsilon_i$  is measurement error, assumed to be independent of  $w_i$ , and i.i.d. with density

$$p(\varepsilon_i) = f_N(\mu_\varepsilon, \sigma_\varepsilon^2). \tag{3.4}$$

Note that  $\mu_\varepsilon$  is allowed to be non-zero, so a systematic bias in reporting of wages is allowed

for.

Assuming the basic equilibrium search model holds for the true wages, a likelihood function can be derived by replacing  $w_i$  by  $w_i^*$  in (2.5) and using (3.3) to write:

$$f(w_i) = \frac{1 + \tau}{2\tau} \frac{1}{\sqrt{(p-r)}\sqrt{p-w_i-\varepsilon_i}} I(B), \quad (3.5)$$

where  $B$  is defined by  $r < w_i + \varepsilon_i < h$ . Analogous to the model with heterogeneity in productivity, we can treat (3.4) either as a part of the likelihood functional or as part of a hierarchical prior. If we do the latter,  $\varepsilon_i$  for  $i = 1, \dots, N$ , can be interpreted as parameters with prior given by (3.4). Hence, (3.5) and (2.4) give the likelihood function for the observed data,  $w_i$ , for  $i = 1, \dots, N$ .

Computational methods for Bayesian inference of this model are given in the Appendix. The informal motivation of these are similar to the model with heterogeneity in productivity and an MCMC algorithm with data augmentation can be derived.

### 3.3 Other Models

In order to focus on the econometric methodology, we do not consider any further extensions to the basic equilibrium search model. However, it is worthwhile noting in passing that the issues discussed in this paper hold for a wide range of models which introduce observation-specific heterogeneity in some way (e.g. a model where individuals are heterogeneous with respect to their value of leisure and, hence, their reservation wage). Furthermore, the Bayesian computational tools developed here will, with some modification, be suitable with many sorts of extensions to the basic job search model. It is also possible to work with models using several extensions of the basic model at one (e.g. a model with both heterogeneity in productivity and measurement error) by extending the tools of the present paper in a simple, obvious way.

## 4 Nonparametric Extensions of the Equilibrium Search Model

### 4.1 Heterogeneity in Productivity

Although rarely used in the econometrics literature (exceptions include Ruggiero, 1996 and Campolieti, 1997), various methods which can be called "nonparametric" exist in the Bayesian statistics literature. The most common of these approaches uses mixtures of Dirichlet processes to model a distribution with unknown form (see among many others, Antoniak, 1974, Ferguson, 1973, West, Muller and Escobar, 1994 and Escobar and West, 1995). In this section, we derive and motivate this approach for the case of the equilibrium search model with unobserved heterogeneity in productivity. Computational methods for Bayesian inference are given in the Appendix. For notational convenience, we will suppress the dependence of the wage distribution below on all parameters other than  $\pi$ .

The equilibrium search model with unobserved heterogeneity in productivity implies a wage distribution for a firm with log-productivity  $\pi_i$ , which we denote by  $F(w_i|\pi_i)$ , where

$$F(w_i|\pi_i) = \frac{1 + \tau}{\tau} \left( 1 - \sqrt{\frac{p_i - w_i}{p_i - r}} \right) I(B), \quad (4.1)$$

where  $p_i = \exp(\pi_i)$  (i.e. this is the integral of the wage density given in equation 3.2). In the preceding section, we assumed  $\pi_i$  came from a known distribution (i.e. its density was given by  $p(\pi_i) = f_N(\mu_\pi, \sigma_\pi^2)$ ). In the present section, we assume  $\pi_i$  comes from a distribution,  $G(\pi_i)$ , with unknown form and use Bayesian nonparametric methods. From a Bayesian viewpoint,  $G(\cdot)$  is best interpreted as a prior. However, it can equally be interpreted as part of the likelihood and the non-Bayesian may prefer to interpret it that way.

Formally, we assume  $G(\cdot)$  is a Dirichlet process. A precise definition of the Dirichlet process is given in any of the citations mentioned at the beginning of this section. Here we give an intuitive motivation of how Dirichlet processes can be used to carry out non-

parametric Bayesian inference. In standard Bayesian inference, one begins with a prior distribution for the *unknown parameters* in the model and then updates using data information to obtain a posterior. In nonparametric Bayesian inference, one begins with a prior for the *unknown distribution*,  $G(\cdot)$ , and then updates with data information to obtain a posterior. Let  $G_0$  be this base prior distribution which is the prior expectation for  $G(\cdot)$ .<sup>7</sup> Here, we let  $G_0$  be Normal with mean  $\mu_\pi$  and variance  $\sigma_\pi^2$ . In other words, our prior for the nonparametric model is that it is the same as the parametric model of Section 3.1. The next element in the Bayesian nonparametric analysis involves the researcher specifying the prior parameter,  $\alpha$ , which is a scalar precision parameter which represents the weight of our belief in the prior  $G_0$ . The exact role of  $\alpha$  will be clarified below. Formally, we have, for  $i = 1, \dots, N$ :

$$w_i | \pi_i \sim F(w_i | \pi_i),$$

$$\pi_i \sim G(\cdot)$$

$$G | \alpha \sim D(\alpha G_0),$$

where " $\sim$ " denotes "is distributed as", and the last equation is notation for "G is a random distribution generated by a Dirichlet process with base measure  $\alpha G_0$ ".

Antoniak, 1974 (or the other citations in this section) give more details about precisely what the previous sentence means. However, its implications are best understood by considering what the Dirichlet process prior implies for certain *conditional* prior distributions. It can be shown that:

$$\pi_i | \pi^{(i)} \sim \alpha a_{N-1} G_0 + a_{N-1} \sum_{j=1, j \neq i}^N \delta(\pi_j), \quad (4.2)$$

---

<sup>7</sup>See Escobar and West (1995) for a discussion (and further citations) that argue that this "prior" plays a role that is similar to the kernel in nonparametric kernel smoothing algorithms.

where  $\pi^{(i)} = (\pi_1, \dots, \pi_{i-1}, \pi_{i+1}, \dots, \pi_N)'$ ,  $\delta(\pi_j) = 1$  if  $\pi_i = \pi_j$  ( $=0$  otherwise) and  $a_{N-1} = \frac{1}{\alpha + N - 1}$ . In other words, *a priori*,  $\pi_i$  is either drawn from  $G_0$  (in this paper, the Normal distribution) or randomly chosen from the other realized values,  $\pi^{(i)}$ . Equation (4.2) also makes clear the role of  $\alpha$ . As  $\alpha$  approaches zero, the first term on the right hand side of (4.2) vanishes and the conditional prior becomes purely nonparametric. That is,  $\pi_i | \pi^{(i)}$  approaches a discrete distribution with points of support given by the other realized values. Equivalently, (4.2) becomes a histogram of the other realized values,  $\pi^{(i)}$ . In contrast, as  $\alpha$  approaches infinity the prior becomes more and more parametric and, in the limit,  $\pi_i | \pi^{(i)} \sim G_0(\pi_i)$ . In the present case, we approach the model of Section 3.1 with productivity being log-Normally distributed across individuals.

The conditional posteriors analogous to (4.2) provide additional intuition. It can be shown that these are:

$$\pi_i | \pi^{(i)}, W \sim q_{i0} G_{i0} + \sum_{j=1, j \neq i}^N q_{ij} \delta(\pi_j), \quad (4.3)$$

where

$$g_{i0} = f(w_i | \pi_i) g_0(\pi_i), \quad (4.4)$$

$$q_{i0} \propto \alpha \int f(w_i | \pi_i) g_0(\pi_i) d\pi_i, \quad (4.5)$$

$$q_{ij} \propto f(w_i | \pi_j), \quad (4.6)$$

$\sum_{l=0}^N q_{il} = 1$  and  $g_{i0}(\cdot)$ ,  $g_0(\cdot)$  and  $f(\cdot)$  are densities corresponding to the distributions  $G_{i0}(\cdot)$ ,  $G_0(\cdot)$  and  $F(\cdot)$ , respectively.

Note first that (4.3), like (4.2) is a mixture of a parametric and nonparametric part. The parametric part is a posterior obtained by updating the base prior via the likelihood

function for one observation,  $w_i$  (i.e. it is the posterior we would have obtained using the parametric model of Section 3.1 with one data point). The nonparametric part involves merely the other realizations,  $\pi^{(i)}$ . Secondly, the weight placed on the parametric part is given by (4.5). Note that this is proportional to the marginal density of the realized data point  $w_i$  (i.e. the marginal likelihood which would have obtained in Section 3 if we only had one data point). Loosely speaking, this measures how likely the observed data point could have come from the parametric model. If  $q_{i0}$  is high, then more weight is placed on the parametric model. Secondly, the prior parameter,  $\alpha$ , also enters (4.5). As discussed above, as  $\alpha$  goes to zero less and less weight is placed on the parametric part of the model. As  $\alpha$  goes to infinity, the model goes to the purely parametric model of Section 3. Thirdly, the weight placed on any of the other realizations in  $\pi^{(i)}$  is given by (4.6). This is the wage density assuming a log-productivity of  $\pi_j$  evaluated at the  $i^{th}$  observation. Hence, this is a measure of how plausible it is that  $w_i$  is generated from the wage density calculated using  $\pi_j$ . Classical nonparametric kernel algorithms can be interpreted as fitting a density at a point using weighted averages of nearby data points. Note that this is exactly what the Bayesian method is doing. The weights in the local averaging are given by  $q_{ij}$  which can be interpreted as a measure of how "close" or "plausible"  $\pi_j$  is for each  $j \neq i$ . Fourthly, for fixed  $\alpha$ , as  $N$  goes to infinity the parametric part of (4.3) will receive less and less weight. Hence, asymptotically the approach becomes increasingly nonparametric. Fifthly, (4.3) can be interpreted as a smoothed histogram where the smoothing is provided by  $G_{i0}$ .

Equations (4.3)-(4.6) can be used as part of an MCMC algorithm for carrying out nonparametric Bayesian inference. Hence, the previous material is all one needs to know in order to understand and implement Bayesian nonparametric methods. However, in practice, we use a more efficient algorithm outlined in West, Muller and Escobar (1994) (see Appendix for more details).

## 4.2 Measurement Error

The derivation of a model with measurement error having a distribution of unknown form proceeds along similar lines to the previous subsection. In particular, the equilibrium search model with measurement error implies a wage distribution for a firm given by

$$F(w_i|\varepsilon_i) = \frac{1 + \tau}{\tau} \left( 1 - \sqrt{\frac{p - w_i - \varepsilon_i}{p - r}} \right) I(B). \quad (4.7)$$

We assume  $\varepsilon_i$  comes from a distribution,  $G(\varepsilon_i)$ , with unknown form and model  $G(\cdot)$  using a Dirichlet process. Hence, we have

$$w_i|\varepsilon_i \sim F(w_i|\varepsilon_i),$$

$$\varepsilon_i \sim G(\cdot)$$

$$G|\alpha \sim D(\alpha G_0),$$

where  $G_0$  is given by (3.4). With these choices (and with  $\pi$ s replaced by  $\varepsilon$ s), the conditional priors and posteriors are as in equations (4.2) through (4.6) and the same intuitive motivation applies. Further details about Bayesian inference for this model are given in the Appendix.

## 4.3 Interpretation of Bayesian Nonparametric Methods

The Bayesian nonparametric approach allows us to fit the wage distribution almost perfectly in the same way as the nonparametric kernel approaches which are increasingly popular in the empirical job search literature (see, e.g., Bontemps, Robin and van den Berg, 1999, 2000 and Nielsen and Rosholm, 1999). However, because the Bayesian approach is based on a well-defined probabilistic model, all the usual tools of Bayesian model comparison



and testing can be used. In particular, we can calculate the marginal likelihood for the models described in this section and use them to calculate Bayes factors comparing a nonparametric model to a parametric alternative. As is standard in Bayesian analyses, the marginal likelihood will reflect the fit of the model, but will also include a reward for parsimony. Of course, the nonparametric model will fit the data better than corresponding parametric alternative (e.g. the nonparametric model with heterogeneity in productivity will fit better than the parametric model with heterogeneity in productivity). However, the reward for parsimony means that the more complicated nonparametric model may not have the higher marginal likelihood. In essence, Bayesian methods provide us with a formal measure, rooted in probability theory, of whether the added flexibility of the nonparametric model is of sufficient worth.

Furthermore, Bayesian methods allow for the comparison of two different nonparametric models. In other words, they allow for a valid statistical comparison of two models, both of which might fit the data almost perfectly. In the context of the empirical job search literature, where there are many different ways of improving data fit (e.g. heterogeneity in productivity or reservation wage, measurement error, etc.), the ability of Bayesian methods to allow for formal model comparison is an enormous advantage. Hence, it is useful to explain what a Bayes factor comparing two nonparametric models will be capturing. Traditionally, most authors have focussed on the fact that Bayes factors contain a reward for parsimony (i.e. all else being equal, the Bayes factor supports the more parsimonious model) and a reward for fit (i.e. all else being equal, the Bayes factor supports the model which fits best). However, Bayes factors also reward the model which evinces a higher degree of compatibility between prior and likelihood function.<sup>8</sup> If we are comparing two nonparametric models, the first two of these rewards will tend to be quite similar and, hence, the Bayes factor will likely be driven by the third. Thus, the choice between two

---

<sup>8</sup>See, for instance, Judge, Griffiths, Hill, Lutkepohl and Lee, 1985, pp. 128-131 for a discussion of these issues in the context of the Normal linear regression model.

nonparametric models will come down to which one has least conflict between prior and data information. We would argue that this is a very sensible basis for model comparison. Remember that a key component of the prior is a distributional assumption about the form of heterogeneity. With heterogeneity in productivity, the base prior,  $G_0$ , says productivity is log-Normally distributed. In the other model, the base prior says measurement error is Normally distributed.

Furthermore, the values selected for prior hyperparameters can be chosen to make statements like "heterogeneity in productivity is likely to be fairly small", or "measurement error will likely have mean zero", etc. A Bayes factor comparing the nonparametric model with heterogeneity in productivity to the one with measurement error will support the model where the data evidence is most consistent with these prior assumptions. So, for instance, if we were to find the distribution of productivity to be very different from log-Normal, while measurement error was almost Normal, then the model with measurement error would be supported. We argue that this is a very sensible basis (perhaps the only basis) for comparison of models which fit perfectly. Furthermore, it formalizes a sensible informal practice. If one were to estimate a nonparametric model and find the results looked odd, this would be informal evidence against the model. The Bayesian approach puts this reasonable practice on a firm statistical footing.

## 5 Empirical Results

The data used in this paper are taken from Bowlus, Kiefer and Neumann (1995) who provide details about the data (see also Kiefer and Neumann, 1993).<sup>9</sup> Briefly, the data consists of N=697 weekly wages for white high school graduates taken from the NLSY and reported in constant 1982 US dollars. All wages are for the first job taken after graduation

---

<sup>9</sup>The data were obtained from the *Journal of Applied Econometrics* data archive at <http://qed.econ.queensu.ca:80/jae/>.

which lasts for more than 2 weeks and involves more than 20 hours per week. Details of the relatively noninformative priors used along with MCMC algorithms for posterior computation are given in the Appendix. The mean, median, standard deviation, minimum and maximum of the wages are 185.38, 175.00, 71.69, 73.92 and 364.72, respectively.

Table 1 contains point estimates and standard deviations of all parameters in each of the five models (i.e. the basic model plus two models with heterogeneity in productivity and two with measurement error). Table 2 contains the logs of the Bayes factors comparing each pair of models

For the models with heterogeneity in productivity, a very large amount of heterogeneity is present. For instance, the point estimates for the parametric model imply a 95% interval for log-productivity of [5.127,6.161]. If we take the exponential of the upper and lower bounds to obtain a rough idea of what this interval implies for productivity we obtain [\$168.51,\$473.90]. In other words, in order extend the basic equilibrium search model to obtain a better fit for the wage density, we have to allow for productivity of unskilled workers in different markets to vary from around \$150 to \$500 per week.

For the nonparametric model with heterogeneity in productivity,  $\alpha$  is very large relative to  $N$ . This indicates that the parametric part of the model is doing quite well at fitting the wage density and the nonparametric part of the model does not add much (see equations 4.2 and 4.5).

<b>Table 1: Posterior Means of Parameters</b> <b>(Posterior standard deviations in parentheses)</b>					
	Basic	Hetero-Param.	Hetero-Non.	Error-Param.	Error-Non.
$r$	70.921 (2.878)	72.806 (1.115)	72.875 (1.042)	71.993 (1.994)	78.971 (11.59)
$\gamma$	4.261 (0.042)	4.288 (0.016)	4.289 (0.015)	4.276 (0.029)	4.359 (0.142)
$p$	401.81 (0.487)	—	—	333.98 (18.66)	403.30 (11.82)
$\pi$	5.996 (0.001)	—	—	5.809 (0.054)	5.998 (0.032)
$h$	365.11 (0.331)	—	—	304.87 (16.54)	331.19 (18.77)
$\mu_\pi$	—	5.644 (0.019)	5.643 (0.018)	—	—
$\sigma_\pi$	—	0.264 (0.017)	0.265 (0.017)	—	—
$\mu_\varepsilon$	—	—	—	9.781 (6.834)	16.61 (6.517)
$\sigma_\varepsilon$	—	—	—	37.213 (5.073)	31.195 (4.116)
$\alpha$	—	—	5,669.8 (1245.8)	—	28.145 (42.969)

Notes to Table 1: The headings in columns 2-6 refer to the basic equilibrium search model (section 2), the parametric model with heterogeneity in productivity (section 3.1), the nonparametric model with heterogeneity in productivity (section 4.1), the parametric model with measurement error in wages (section 3.2) and the nonparametric model with measurement error in wages (section 4.2), respectively.

The model with Normal measurement error finds  $\sigma_\varepsilon$  to be very large. Interestingly, the estimated upper bound on the wage distribution is less than the highest observed wage. The reason for this anomaly is that the  $\varepsilon_i$ s for the higher wages in the sample are very negative. Remember that equation (3.4) implies that measurement errors are distributed

Normally, independent of each other and the data. If we calculate the posterior means of the  $\varepsilon_i$ s they strongly seem to violate the latter assumption. In fact, the correlation between the observed wages and the errors is -.88! That is, to fit the wage density (see equation 3.5) the model wants to use  $\varepsilon_i$  to adjust low wages upward and high wages downward. Hence, fit considerations imply measurement error should be highly correlated with wages. This is at odds with our assumption that measurement error is independent of the wages. The tension between these two imperatives yields a model which fits a bit better than the basic model, but the conflict between prior and data information implies that Bayes factors give little support for the model with measurement error (see Table 2). In other words, for measurement error to explain the poor fit of the basic model, it must be the case that high wage individuals are consistently massively over-reporting their wages while low wage individuals are under-reporting. Since this does not seem reasonable (where "reasonableness" is measured by the prior) the Bayes factors massively reject the model.

The nonparametric model with measurement error in wages indicates that large departures from the base prior are required in order to fit the data (Remember that the base prior,  $G_0$ , has measurement error being i.i.d. Normal). One indication of this is that the posterior for  $\alpha$  indicates that it is quite small (relative to N). Hence, more weight is being placed on the nonparametric part of the model (see equation 4.5).

These findings are reinforced in Figures 1 and 2, which plot the empirical wage density<sup>10</sup> along with the wage densities for each of our models evaluated at the posterior means of the parameters. To be precise, for the parametric model with measurement error the wage density is given in (3.5). We plug posterior means for  $r$  and  $p$  into this formula. For each of a grid of possible values for  $w$ , we take the posterior means of  $\mu_\varepsilon, \sigma_\varepsilon^2$  and repeatedly simulate  $\varepsilon$ s and evaluate (3.5) at each simulated draw and average the result. This allows us to plot the wage density with  $\varepsilon$  integrated out. A similar strategy is used for the other parametric model. For the nonparametric models we use (4.2), evaluated at the posterior

---

<sup>10</sup>The empirical wage density is a simple histogram using 20 bins of equal size.

mean for  $\alpha$ , to integrate out  $\varepsilon$  or  $\pi$ .

These figures confirm that the basic equilibrium search model implies an increasing wage density and, thus, offers a very poor fit for the empirical wage density. In contrast, both of the models with heterogeneity in productivity appear to offer a very good fit (in a smoothed sense). That is, the fitted wage densities for these models look like a nonparametrically kernel smoothed version of the empirical wage density. However, the nonparametric model offers very little improvement in fit over the parametric one.

Adding Normal measurement error to the equilibrium search model improves fit somewhat, but the fit is not nearly as good as that provided by the models with heterogeneity in productivity. Allowing for measurement error in a nonparametric fashion seems to offer only slight improvements in fit. The reason for the apparent poor performance of the nonparametric measurement error model is partly due to the way we have constructed Figure 2. Note that the observed wages are not used (i.e. we are integrating out  $\varepsilon$  at each grid point for the wages). But, in order to fit better, the wages and the errors have to be correlated. The Dirichlet process implicitly does this through equation (4.6) of the posterior (i.e. it groups clusters of workers with similar wages as having the same measurement error). However, since we do not use the data in our figures, we do not use (4.6). Intuitively, Figures 1 and 2 are not well-designed to show the posterior fit of the nonparametric models and the apparent lack of a perfect fit (in a smoothed sense) by the nonparametric model in Figure 2 is to be expected.

Nevertheless, it is the case that the fit of the nonparametric model with measurement error is not as good as a kernel smoothed nonparametric estimate would provide. To see why the relatively poor fit of this nonparametric model occurred, we investigated what sort of measurement errors would be required for the wage density from the model with measurement error to match perfectly with the empirical wage density. To be precise, we carried out the following experiment. Firstly, we used the raw wage data and constructed a kernel smoothed estimate of the wage density. Secondly, we took equation (3.5) evaluated

at the posterior means for  $p$  and  $r$ , and found the values of the  $\varepsilon_i$ s which would cause (3.5) to match exactly with the kernel smoothed density from the first step.

The resulting values of the  $\varepsilon_i$ s were truly bizarre. For instance, the standard deviation of the  $\varepsilon_i$ s required to fit the wage density perfectly was 519,205. If the data is ordered so that individual 1 has the lowest wage and individual N the highest, then the correlation between  $\varepsilon_i$  and  $\varepsilon_{i+1}$  is 0.99996. It seems that even Dirichlet mixtures of Normals have trouble fitting a distribution with such properties when any sort of prior information is allowed for. By letting the priors for  $\alpha$ ,  $\mu_\varepsilon$  and  $\sigma_\varepsilon$  become almost completely noninformative, we could obtain a better fit for the nonparametric model. However, such noninformative prior choices would tend to yield Bayes factors which always favored the parametric model. In short, regardless of whether we make reasonable prior choices based on what we expect measurement error to look like (as we have done in this paper), or use a noninformative prior, we will always come to the conclusion that measurement error is not a good way of extending the basic equilibrium search models.

The previous discussion can be formalised by examining Table 2. As frequently happens with Bayes factors, results are very strong.<sup>11</sup> That is, there is a clear ranking of models in terms of the posterior support they receive: 1) the parametric model with heterogeneity in productivity, 2) the nonparametric model with heterogeneity in productivity, 3) the basic model, 4) the nonparametric model with measurement error, and 5) the parametric model with measurement error. The former two models fit the data very well and the assumption of log-Normality for the productivity distribution seems to match the data quite well (i.e. the prior and likelihood are in accordance). Accordingly these two models obtain strong support over the basic equilibrium search model which implies a counterfactual increasing wage density.

---

<sup>11</sup>Very loosely speaking, log Bayes factors often behave like the difference of log-likelihoods with additional terms reflecting the priors. The log-likelihoods of the models used in this paper are extremely different from one another, and the priors give a strong reward for parsimony. These considerations help drive the strong findings of Table 2.

The models with measurement error do very poorly relatively to the basic model, even though they fit better. The reason for this was outlined above. Briefly, in order to improve fit, the model wants values of the errors which are very much at odds with the assumption that errors should be independent of the wages. Hence, if one believes that measurement error should be i.i.d. Normal (as the hierarchical prior for the parametric model implies) or departures from i.i.d. Normality are not too large (as the prior for the nonparametric model implies), then one should strongly reject measurement error as an explanation for the failure of the equilibrium search model. The nonparametric model with measurement error beats the comparable parametric model due to its improved fit and, more importantly, the fact that its prior is more flexible. Hence, the conflict between prior and data information is less for the nonparametric than the parametric model.

	Basic	Hetero-Param.	Hetero-Non.	Error-Param.	Error-Non.
Basic	0.0				
Hetero-Param.	298	0.0			
Hetero-Non.	228	-70	0.00		
Error-Param.	-3361	-3659	-3589	0.00	
Error-Non.	-146	-444	-374	3215	0.00

## 6 Conclusions

In this paper, we have developed Bayesian methods for parametric and nonparametric inference in various equilibrium search models. We have argued that a problem with the existing literature is that there are numerous extensions to the basic model which, along with nonparametric methods, can fit the observed wage density perfectly. This opens up the risk that the empirical job search literature will develop in an unsatisfactory manner, where each researcher has his or her own extension on the basic model and all researchers fit the data perfectly. We have argued that Bayesian methods are a logical solution to



the problem. Bayesian parametric or nonparametric methods allow for Bayes factors comparing competing models to be calculated. Bayes factors, apart from the usual goodness of fit and reward for parsimony considerations, contain a reward for coherence between data and prior information. Furthermore, the "priors" typically used in the job search literature are hierarchical in nature so that a non-Bayesian would interpret them as part of the likelihood (e.g. we use as a prior that productivity is log-Normally distributed, however Koning, Ridder and van den Berg, 1995 treat an identical assumption as part of the likelihood function). In the context of comparing models which all fit well, this reward for coherence between prior and data information can potentially be very important. That is, the researcher is required to specify what he/she thinks are sensible properties for the extension under consideration (e.g. measurement error should be roughly Normal, heterogeneity in productivity should not be too large, productivity should increase with training but diminishing returns to training should exist, etc.) and departures from the prior are evidence against the model. This gives a formal basis for the informal strategy of looking at the results one obtains and seeing if they look reasonable.

An empirical example using the data set from Bowlus, Kiefer and Neumann (1995) shows the practicality and usefulness of our approach. Three parametric models (i.e. the basic model and models with heterogeneity in productivity and measurement error) and two non-parametric models (i.e. one with heterogeneity in productivity and one with measurement error) were estimated. Before estimation, we argued that the models with heterogeneity in productivity are much more reasonable than the others. Our empirical results bore out this conjecture. For instance, we were able to provide strong statistical evidence against the models with measurement error — even the nonparametric one which, in theory, could have fit the data perfectly. In essence, it is possible to fit any wage density by allowing for measurement error. However, in the present empirical example, this "measurement error" has to be of such a strange form that it is highly questionable that it truly is measurement error.

The example used in the present paper is purely illustrative. That is, we deliberately chose one reasonable extension to the basic model (i.e. heterogeneity in productivity) and one unreasonable one (i.e. measurement error), in order to illustrate our statistical methods. However, as the theoretical job search literature develops new extensions to the basic equilibrium search model, there will undoubtedly be an increasing number of potential models all of which seems reasonable, *a priori*. The methods developed in this paper provide a statistically rigorous basic for choosing among such competing theories.

## 7 References

Antoniak, C. (1974). "Mixtures of Dirichlet processes with applications to nonparametric problems," *Annals of Statistics*, 2, 1152-1174.

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

Bontemps, C., Robin, J. and van den Berg, G. (1999). "An empirical equilibrium job search model with search on the job and heterogeneous workers and firms," *International Economic Review*, 1039-1074..

Bontemps, C., Robin, J. and van den Berg, G. (2000). "Equilibrium search with continuous productivity dispersion: Theory and non-parametric estimation," *International Economic Review*, forthcoming.

Bowlus, A., Kiefer, N. and Neumann, G. (1995). "Estimation of equilibrium wage distributions with heterogeneity," *Journal of Applied Econometrics*, 10, S133-S152.

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

Campolieti, M. (1997). "Bayesian estimation of discrete duration models," unpublished PhD dissertation, University of Toronto.

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.

Escobar, M. and West, M. (1995). "Bayesian density estimation using mixtures," *Journal of the American Statistical Association*, 90, 577-588.

Ferguson, T. (1973). "A Bayesian analysis of some nonparametric problems," *Annals of Statistics*, 1, 209-230.

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.

Gelfand, A. and Dey, D. (1994). "Bayesian model choice: Asymptotics and exact calculations," *Journal of the Royal Statistical Society, Series B*, 56, 501-514.

Geweke, J. (1999). "Using Simulation Methods for Bayesian Econometric Modelling: Inference, Development and Communication," *Econometric Reviews*, 18, 1-74.

Judge, G., Griffiths, W., Hill, R.C., Lutkepohl, H. and Lee, T.-C. (1985). **The Theory and Practice of Econometrics** (second edition). John Wiley and Sons, New York.

Kiefer, N. and Neumann, G. (1993). "Wage dispersion with homogeneity: The empirical equilibrium search model," in H. Bunzel, P. Jensen and N. Westergaard-Nielsen (eds.), **Panel Data and Labor 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.

Koning, P., Ridder, G. and van den Berg, G. (1995). "Structural and frictional unemployment in an equilibrium search model with heterogeneity," *Journal of Applied Econo-*

*metrics*, 10, S133-S152.

Koop, G. (2001). "Bayesian inference in models based on equilibrium search theory," *Journal of Econometrics*, 102, 311-338.

Koop, G. and Poirier, D. (2002). "Testing for optimality in stationary job search models," *The Econometrics Journal*, forthcoming.

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

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

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

Nielsen, M. and Rosholm, M. (1999). "Wages, training and turnover in a search-matching model," Centre for Labor Market and Social Research, University of Aarhus, manuscript.

Ridder, G. and van den Berg, G. (1997). "Empirical equilibrium search models," Chapter 4 in **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.

Poirier, D. (1995). **Intermediate Statistics and Econometrics**. The MIT Press: Cambridge.

Ruggiero, M. (1994). "Bayesian semiparametric estimation of proportional hazards models," *Journal of Econometrics*, 62, 277-300.

West, M., Muller, P and Escobar, M. (1994). "Hierarchical priors and mixture models," in **Aspects of Uncertainty**, P. Freeman and A.F.M. Smith (eds.). John Wiley and Sons, New York.

## 8 Appendix: Bayesian Inference and Computation

Inference in all models is accomplished by simulating the posterior using Markov Chain Monte Carlo (MCMC) methods. Model comparison is done using Bayes factors (i.e. the ratio of marginal likelihoods for two competing models). For the non-Bayesian reader, definitions and explanations of all the technical terms used in this Appendix are given in Geweke (1999).

### 8.1 The Basic Equilibrium Search Model

The notation used in the paper becomes simpler if we work in terms of the logs of the reservation wage,  $\gamma = \ln(r)$ , and productivity,  $\pi = \ln(p)$ . Let  $\theta = (\gamma, \pi)'$  indicate the parameters in the model. The likelihood function is given in equations (2.4) and (2.5). Given the non-conventional nature of the likelihood, there is no particular need to choose any particular class of priors to facilitate computation. We stress that researchers can use virtually any prior (including non-informative ones) they wish and the methods developed here will apply. In the present paper, we assume  $p(\theta) = p(\gamma)p(\pi)$ , where *a priori*,  $p(\gamma) = f_N(r_0, R_0)$ ,  $p(\pi) = f_N(p_0, P_0)$ . By letting the variance in the Normal go to infinity we can get priors which are noninformative relative to the data. Given that the empirical example is largely illustrative and the contribution of the present paper does not relate to prior elicitation, we make relatively noninformative choices for prior hyperparameters. Of course, in a more serious empirical exercise more care should be taken with prior elicitation and a prior sensitivity analysis should be carried out. Noting that we are using weekly wages for recent high school graduates, setting  $r_0 = 5$  and  $p_0 = 6$  is reasonable, and setting  $R_0 = P_0 = 2$  makes the prior fairly flat relative to what we expect the data to indicate.

Koop (2001) uses a random walk chain Metropolis-Hastings algorithm (see Chib and Greenberg, 1995) to carry out Bayesian inference in this model. In the present paper we use a slightly more efficient MCMC algorithm which involves drawing sequentially from the

conditional posteriors:  $p(\pi|W, \gamma)$  and  $p(\gamma|W, \pi)$ . To be precise, we have

$$p(\gamma|W, \pi) \propto p(\gamma) \frac{1}{(\exp(\pi) - \exp(\gamma))^{\frac{N}{2}}} I(B). \quad (\text{A.1})$$

To draw from this density we use rejection methods. In particular, we use  $p(\gamma)I(B)$  as a source density kernel and use the fact that  $\frac{1}{(\exp(\pi) - \exp(\gamma))^{\frac{N}{2}}}$  is increasing in  $\gamma$  over the interval  $I(B)$  to obtain the appropriate bound necessary to calculate the acceptance probability.

Next we have

$$p(\pi|W, \gamma) \propto p(\pi) \frac{1}{(\exp(\pi) - \exp(\gamma))^{\frac{N}{2}}} \prod_{i=1}^N \frac{1}{(\exp(\pi) - w_i)^{\frac{1}{2}}} I(B). \quad (\text{A.2})$$

To draw from this density we also use rejection methods. In particular, we use  $p(\pi)I(B)$  as a source density kernel and use the fact that  $\frac{1}{(\exp(\pi) - \exp(\gamma))^{\frac{N}{2}}} \prod_{i=1}^N \frac{1}{(\exp(\pi) - w_i)^{\frac{1}{2}}}$  is decreasing in  $\pi$  over the interval  $I(B)$  to obtain the appropriate bound necessary to calculate the acceptance probability.

In order to calculate the marginal likelihood necessary for calculating Bayes factors, we use the modified harmonic mean approach first suggested by Gelfand and Dey (1994), as implemented in Geweke (1999). This involves a truncation parameter labelled  $p$  in Geweke (1999, Section 4.3) which we choose to be 0.95. Experimentation indicates that marginal likelihood is not sensitive to choice of  $p$ .

## 8.2 Parametric Extensions of the Equilibrium Search Model

### 8.2.1 Heterogeneity in Productivity

The model and methods are similar to those outlined for the basic model above and we use the same prior for common parameters. In particular, the posterior conditional for  $\gamma$  is identical to (A.1) except that  $p$  and  $\pi$  now vary over individuals and, hence, have a subscript  $i$  and  $I(B)$  is altered by the fact that  $h$  now varies with  $i$ .

The likelihood function is given by (2.4) and (3.2). The first stage in the hierarchical prior is given by (3.1) (i.e.  $p(\pi_i) = f_N(\mu_\pi, \sigma_\pi^2)$  where  $\mu_\pi, \sigma_\pi^2$  are unknown parameters to be estimated). The second stage in the hierarchy is given by  $p(\mu_\pi) = f_N(p_0, P_0)$  and  $p(\sigma_\pi^{-2}) = f_G(\mu_0, \nu_0)$  where  $f_G(a, b)$  denotes the Gamma density with mean  $a$  and  $b$  degrees of freedom (see Poirier, 1995, page 100). By letting the degrees of freedom in the Gamma go to zero we can obtain a prior which is noninformative relative to the data. Note that we choose the same prior for  $\mu_\pi$  as we did for  $\pi$  in the previous model (although, of course, we could have chosen a different prior). We set  $\mu_0 = 1.0$  and  $\nu_0 = 1.0$ , relatively noninformative choices which potentially allow for a huge amount of heterogeneity in productivity.

An MCMC algorithm can be derived by sequentially drawing from (A.1) along with  $p(\pi|W, \gamma, \mu_\pi, \sigma_\pi^2)$ ,  $p(\mu_\pi|W, \gamma, \pi, \sigma_\pi^2)$  and  $p(\sigma_\pi^{-2}|W, \gamma, \mu_\pi)$  where  $\pi$  now equals  $(\pi_1, \dots, \pi_N)'$ . The latter two posterior conditionals simplify to:

$$p(\mu_\pi|\pi, \sigma_\pi^2) = f_N(p_1, P_1) \tag{A.3}$$

and

$$p(\sigma_\pi^{-2}|\pi, \mu_\pi) = f_G(\mu_1, \nu_1), \tag{A.4}$$

where  $P_1^{-1} = P_0^{-1} + N\sigma_\pi^{-2}$ ,  $p_1 = P_1(P_0^{-1}\mu_0 + \sigma_\pi^{-2}\sum \pi_i)$ ,  $\nu_1 = \nu_0 + N$  and  $\mu_1 = \frac{\nu_1}{\nu_0\mu_0 + \sum(\pi_i - \mu_\pi)^2}$ .

$p(\pi|W, \gamma, \mu_\pi, \sigma_\pi^2) = \prod_{i=1}^N p(\pi_i|w_i, \gamma, \mu_\pi, \sigma_\pi^2)$  and draws from  $p(\pi_i|w_i, \gamma, \mu_\pi, \sigma_\pi^2)$  for  $i = 1, \dots, N$  can be taken using rejection methods analogous to those used for the basic model. To be precise, we use  $p(\pi_i)$  truncated to ensure that  $p_i \geq h_i$  as the source density to generate draws. An acceptance probability can be obtained by noting that  $p(\pi_i|w_i, \gamma, \mu_\pi, \sigma_\pi^2)$  is decreasing in  $\pi_i$  and has a finite value at  $\exp(\pi_i) = h_i$ .

The marginal likelihood is calculated using the method outlined in Geweke (1999, section 4.3). However, it is slightly complicated by the fact that the parameter vector includes  $\pi$  which is N-dimensional. To simplify computation, we only include  $\gamma, \mu_\pi, \sigma_\pi^2$  as the para-

meters used to construct the function Geweke (1999) labels  $f(\theta)$ . This choice is simulation consistent and experimentation with artificial data indicates it to be a good one.

### 8.2.2 Measurement Error in Wages

The likelihood function is given by (2.4) and (3.5). The prior for  $\gamma$  and  $\pi$  is as in the basic model. The first stage in the hierarchical prior is given by (3.4). The second stage in the hierarchy is given by  $p(\mu_\varepsilon) = f_N(e_0, E_0)$  and  $p(\sigma_\varepsilon^{-2}) = f_G(\mu_{\varepsilon 0}, \nu_{\varepsilon 0})$ . We set  $e_0 = 0, E_0 = 10.0, \mu_{\varepsilon 0} = 10.0$  and  $\nu_{\varepsilon 0} = 0.0001$ . These values indicate that we expect measurement error to have mean zero and be relatively small (i.e. typically less than \$20), however we are very uncertain about this prior expectation and hence make very noninformative choices for the other hyperparameters.

Our MCMC algorithm involves sequentially drawing from (A.1) and (A.2) (with  $B$  defined by  $r < w_i + \varepsilon_i < h$ ) and  $p(\varepsilon|W, \gamma, \mu_\varepsilon, \sigma_\varepsilon^2)$ ,  $p(\mu_\varepsilon|W, \gamma, \pi, \sigma_\varepsilon^2)$  and  $p(\sigma_\varepsilon^{-2}|W, \gamma, \pi, \mu_\varepsilon)$  where  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)'$ . The latter two posterior conditionals simplify to:

$$p(\mu_\varepsilon|\varepsilon, \sigma_\varepsilon^2) = f_N(e_1, E_1) \tag{A.6}$$

and

$$p(\sigma_\varepsilon^{-2}|\varepsilon, \mu_\varepsilon) = f_G(\mu_{\varepsilon 1}, \nu_{\varepsilon 1}), \tag{A.7}$$

where  $E_1^{-1} = E_0^{-1} + N\sigma_\varepsilon^{-2}$ ,  $e_1 = E_1(E_0^{-1}e_0 + \sigma_\varepsilon^{-2}\sum \varepsilon_i)$ ,  $\nu_{\varepsilon 1} = \nu_{\varepsilon 0} + N$  and  $\mu_{\varepsilon 1} = \frac{\nu_{\varepsilon 1}}{\nu_{\varepsilon 0}\mu_{\varepsilon 0} + \sum(\varepsilon_i - \mu_\varepsilon)^2}$ .

$p(\varepsilon|W, \gamma, \mu_\varepsilon, \sigma_\varepsilon^2) = \prod_{i=1}^N p(\varepsilon_i|w_i, \gamma, \mu_\varepsilon, \sigma_\varepsilon^2)$  and each of  $p(\varepsilon_i|w_i, \gamma, \mu_\varepsilon, \sigma_\varepsilon^2)$  can be drawn from using rejection methods analogous to those used for the previous models. We use  $p(\varepsilon_i)$  truncated to ensure that  $r < w_i + \varepsilon_i < h$  as the source density to generate draws. An acceptance probability can be obtained by noting that  $p(\varepsilon_i|w_i, \gamma, \mu_\varepsilon, \sigma_\varepsilon^2)$  is increasing in  $\varepsilon_i$  and has a finite value at  $\varepsilon_i = h - w_i$ .



The marginal likelihood is calculated as in the previous model with heterogeneity in productivity.

## 8.3 Nonparametric Extensions of the Equilibrium Search Model

### 8.3.1 Heterogeneity in Productivity

The setup for this model is as the parametric case except for the treatment of  $\pi$ . Thus, the MCMC algorithm and priors for  $\gamma, \mu_\pi, \sigma_\pi^2$  conditional on  $\pi$  are exactly as described above. Similarly the marginal likelihood calculation is (apart from the changes due to the use of a different prior for  $\pi$ ) identical to that outlined in the previous sub-section. Hence, we need only describe how to draw from  $p(\pi|W, \gamma, \mu_\pi, \sigma_\pi^2)$  in order to fully specify our computational methods. We do not do so in any detail since, once methods for calculating (4.4), (4.5) and (4.6) are specified, the remainder of the algorithm is identical to that given in West, Muller and Escobar (1994) pages 367-369 (or Campolieti, 1997) and the reader is referred there for precise details. Equations (4.4) and (4.6) can be calculated directly. However, it is not possible to evaluate the integral in (4.5) analytically. Hence, it is evaluated using Monte Carlo integration, In particular, we take 1,000 draws from  $g_0(\pi_i)$  (i.e.  $f_N(\mu_\pi, \sigma_\pi^2)$  evaluated at the MCMC draws of  $\mu_\pi, \sigma_\pi^2$ ), and average the  $f(w_i|\pi_i)$ s evaluated at each draw.

The nonparametric model has the additional parameter,  $\alpha$ , for which we use the prior  $p(\alpha) = f_G(\mu_\alpha, \nu_\alpha)$ . By setting  $\mu_\alpha = N$  we allocate roughly half the weight to the parametric part of the model (see equation 4.2), by setting  $\nu_\alpha = 3$  we are making the prior relatively noninformative. Note that West, Muller and Escobar (1994) also describes how to draw from the conditional posterior of  $\alpha$ .

### 8.3.2 Measurement Error

The setup for this model is as the parametric model with measurement error except for the treatment of  $\varepsilon$ . Thus, the MCMC algorithm and priors for  $\gamma, \pi, \mu_\varepsilon, \sigma_\varepsilon^2$  conditional on  $\varepsilon$  are

exactly as described above. The prior for  $\alpha$  is the same as for the previous nonparametric model. To draw from  $p(\varepsilon|W, \pi, \gamma, \mu_\varepsilon, \sigma_\varepsilon^2)$ , we use the same method as for the nonparametric model with the material relating to heterogeneity in productivity altered in the obvious way. That is, (4.4)-(4.6) are calculated using densities from the model with measurement error (rather than the model with heterogeneity in productivity).

[ natheight=10.9918in, natwidth=8.4838in, height=8.0307in, width=5.0194in]  
C:/localapp/swp50/temp/graphics/fig1<sub>1</sub>.pdf

[ natheight=10.9918in, natwidth=8.4838in, height=8.0307in, width=5.0194in]  
C:/localapp/swp50/temp/graphics/fig2<sub>2</sub>.pdf