Bayesian Inference for Duration Data with Unobserved and Unknown Heterogeneity: Monte Carlo Evidence and an Application*

M. Daniele Paserman
Department of Economics
Hebrew University, CEPR and IZA
dpaserma@shum.huji.ac.il

September 2004

Abstract

This paper describes a semiparametric Bayesian method for analyzing duration data. The proposed estimator specifies a complete functional form for duration spells, but allows flexibility by introducing an individual heterogeneity term, which follows a Dirichlet mixture distribution. I show how to obtain predictive distributions for duration data that correctly account for the uncertainty present in the model. I also directly compare the performance of the proposed estimator with Heckman and Singer's (1984) Non Parametric Maximum Likelihood Estimator (NPMLE).

The methodology is applied to the analysis of youth unemployment spells. Compared to the NPMLE, the proposed estimator reflects more accurately the uncertainty surrounding the heterogeneity distribution.

JEL codes: C11, C41
Keywords: Duration data; Dirichlet process; Bayesian inference; Markov chain Monte Carlo simulation;

*I wish to thank Gary Chamberlain and Keisuke Hirano for many lengthy discussions. Caroline Hoxby, Guido Imbens, Larry Katz, Jack Porter, Tiemen Woutersen, and participants in seminars at Harvard University, Hebrew University, and at the Second Haifa Winter Workshop on Computer Science and Statistics (CSStat 2003) also provided helpful suggestions. All errors are my own.

1. Introduction

This paper develops a semiparametric Bayesian methodology for analyzing duration data. The methodology specifies a hazard model belonging to a parametric family, and allows a flexible distribution for a residual heterogeneity term, by modeling it as a mixture of Dirichlet processes (Ferguson, 1973; Antoniak, 1974). Markov chain Monte Carlo methods are then used to simulate posterior quantities for parameters of interest, and to generate predictive distributions.

It is common to model duration data as a combination of a baseline hazard and a mixing distribution, and to interpret the baseline hazard as representing structural duration dependence, and the mixing distribution as unobserved heterogeneity. It is well known that parameter estimates in this model are sensitive to the assumptions made about the mixing distribution. In an important contribution, Heckman and Singer (1984) propose a Non Parametric Maximum Likelihood Estimator (NPMLE) that overcomes the excessive sensitivity of parameter estimates to assumptions about the distribution of residual heterogeneity. The NPMLE specifies a hazard function up to a finite number of unknown parameters, and then lets the heterogeneity term follow a discrete mixture structure. This estimator has been used frequently in the literature to model unobserved heterogeneity in a variety of settings: unemployment duration for Canadian men (Ham and Rea, 1987); the effects of training on the length of unemployment and employment spells in an experimental study (Ham and Lalonde, 1996); welfare spells (Blank, 1989); and transitions in and out of poverty (Stevens, 1999). The NPMLE performs well in estimating the structural parameters of the duration model, but proves to be an unreliable guide to the shape of the true mixing distribution of unobservables. In the case where the number of mixture points is unknown, a distribution theory for the proposed estimator has not yet, to the best of my knowledge, been developed.[1]

The estimator proposed here addresses the shortcomings of the NPMLE. The Bayesian approach enables one to obtain, conditional on the prior distribution, exact finite sample posterior probability intervals for the parameters of interest, that correctly account for the uncertainty present in the model.[2] The Dirichlet process is a prior on the space of distribution functions, and allows flexibility in the heterogeneity distribution: this can be multimodal, skewed, or fat-tailed. The posterior distribution is a mixture of a continuous density and a discrete density. Importantly, the algorithm used to obtain posterior distributions of the parameters of interest generates also a posterior distribution for the number of mass points in the heterogeneity distribution. Therefore, the marginal distribution of the parameters reflects the uncertainty surrounding the number of mixture points. This enables one to directly compare the performance of my estimator to Heckman and Singer's NPMLE.

In most applications of the Dirichlet process, the data is modeled as a normal density, mixed with respect to the distribution of the parameters. If the common prior distribution follows a Dirichlet process, then the data will come from a Dirichlet mixture of normals (Ferguson, 1983; Escobar, 1994; Escobar and West, 1995). An interesting economic application is in Hirano (2002), who uses this

[1] Van der Vaart (1996) proves asymptotic normality for the NPMLE in certain special cases, but does not provide a general proof.

[2] In a parametric model, the 95 percent posterior probability interval does have the frequentist property that, in repeated samples, and for large sample sizes, it contains the true parameter 95 percent of the time. In semiparametric applications such as the one studied here, it is no longer clear that the posterior probability interval has the desired frequentist property. Nevertheless, it can still be a useful summary measure of uncertainty.

methodology to study the structure of earning dynamics in a longitudinal data set.

Non parametric analysis of duration data presents some peculiarities, because of the nonlinearity of the problem and because the residual heterogeneity term usually enters the model multiplicatively. The normal model is not convenient in this case. I overcome these difficulties by specifying a Weibull hazard function, and letting the heterogeneity term follow a Dirichlet mixture of Gamma distributions. The posterior distribution for the mixture density in this case has not been previously derived.

Semiparametric Bayesian analysis for proportional hazard models has been described in Kalbfleisch and Prentice (1980). Hjort (1990) proposes a nonparametric Bayes estimator based on Beta processes. In economic applications, Ruggiero (1994) proposes a fully Bayesian estimator for the regression parameters in a proportional hazards model, by specifying a Dirichlet prior distribution for the baseline hazard, treated as a nuisance parameter. He then computes the posterior distribution of the parameter of interest, conditional on the data and integrated with respect to the nuisance parameter, and applies this methodology to an analysis of survival times of job vacancies. My approach differs in that I specify the complete distribution of duration times, up to a finite dimensional parameter vector, and allow a flexible mixture for the distribution of the individual heterogeneity term. This allows one to generate predictive distributions for duration spells, possibly at the cost of additional functional form assumptions. Moreover, the Dirichlet process mixes over one of the parameters of the Weibull baseline hazard, implying that the resulting model is a mixture of proportional hazards models, which is not necessarily a proportional hazard model in itself.

My approach is similar to that developed independently by Campolieti (2001): the difference lies in the fact that Campolieti models the hazard in discrete time using a multiperiod probit model and a normal prior for the Dirichlet process. The Weibull-Gamma combination used in this paper adheres more closely to the types of models commonly analyzed in duration studies.[3] In addition, I present results from a small Monte Carlo study showing that proposed estimator has the desired frequentist properties of unbiasedness (i.e., the posterior mean approximates the true parameter value) and correct coverage rates of the posterior interval.

The rest of the paper is structured as follows: in Section 2 I present first a brief description of the Dirichlet process and discuss of some of its properties; then I describe its application to the Bayesian estimation of duration data. In Section 3 I present some suggestive Monte Carlo evidence on the performance of the estimation technique on simulated data sets. Section 4 applies this methodology to an analysis of unemployment spells of young men. It also compares the performance of the proposed estimator to Heckman and Singer's NPMLE: parameter estimates and standard errors based on the Dirichlet model reflect substantially more accurately the uncertainty surrounding the distribution of unobserved heterogeneity. Section 5 concludes.

2. Dirichlet Mixture Models for Duration Data

2.1. The Dirichlet Process

The following definitions and properties of a Dirichlet process are due to Antoniak (1974).

[3] This specification has been used in a non-economic apllucation by Merrick, Soyer and Mazzuchi (2003).

Definition 1. *Let Θ be a set, and $\mathcal{A}$ a σ-field of subsets of Θ. Let ν be a finite, non-null, non-negative, finitely additive measure on $(\Theta, \mathcal{A})$. We say a random probability measure P on $(\Theta, \mathcal{A})$ is a* Dirichlet *process on $(\Theta, \mathcal{A})$ with base measure ν, denoted $P \in \mathcal{D}(\nu)$, if for every $k = 1, 2, ...$ and measurable partition $B_1, B_2, ..., B_k$ of Θ, the joint distribution of the random probabilities $(P(B_1), ..., P(B_k))$ is Dirichlet with parameters $(\nu(B_1), ..., \nu(B_k))$. (Based on Antoniak, 1974, Definition 1).*

Following are some useful properties of the Dirichlet process:

1. If $P \in \mathcal{D}(\gamma)$ and $A \in \mathcal{A}$,then $E(P(A)) = \gamma(A)/\gamma(\Theta)$.

2. If $P \in \mathcal{D}(\gamma)$ and conditional given P, θ_1, $\theta_2, ...,$ θ_N are i.i.d. P, then $P|\theta_1, \theta_2, ..., \theta_N \in \mathcal{D}(\gamma + \sum_{i=1}^{N} \delta_{\theta_i})$ where δ_x denotes the probability measure giving mass one to the point x.

3. If $P \in \mathcal{D}(\gamma)$, then P is almost surely discrete.

The almost sure discreteness of the Dirichlet process is a key feature for model analysis. Suppose that $P \sim \mathcal{D}(\gamma P_0)$ is a Dirichlet process defined by γ, a positive scalar, and P_0, a probability measure. The probability measure P_0 can be thought of as the prior expectation of P. The scalar γ is a precision parameter which determines the prior concentration of P around P_0. In other words, γ represents the weight of the belief that P is centered around the distribution P_0.

Briefly, in any sample θ of size N from P, there is positive probability of coincident values. For any $i = 1, 2, ..., N$, let $\theta^{(i)}$ denote the vector θ without element i: $\theta^{(i)} = \{\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_N\}$. Then the conditional prior for $(\theta_i|\theta^{(i)})$

is

$$(\theta_i|\theta^{(i)}) \sim \frac{\gamma}{\gamma+N-1}P_0 + \frac{1}{\gamma+N-1}\sum_{j=1,\, j\neq i}^{N} \delta_{\theta_j}. \tag{1}$$

Similarly, the distribution of a new draw $(\theta_{N+1}|\,\theta)$ is given by:

$$(\theta_{N+1}|\theta) \sim \frac{\gamma}{\gamma+N}P_0 + \frac{1}{\gamma+N}\sum_{j=1}^{N} \delta_{\theta_j}. \tag{2}$$

Thus, given θ, a sample of size N from P, the next case θ_{N+1} represents a new, distinct value with probability $\frac{\gamma}{\gamma+N}$ and is otherwise drawn uniformly from among the first N values. These first N values themselves behave as described by (1) and so with positive probability reduce to $k < N$ distinct values. If we write the k distinct values among the N elements of θ as $\theta_j^*,\ j = 1, ..., k$, and let N_j be the number of occurrences of θ_j^*, then we can rewrite equation (2) as

$$(\theta_{N+1}|\theta) \sim \frac{\gamma}{\gamma+N}P_0 + \frac{1}{\gamma+N}\sum_{j=1}^{k} N_j \delta_{\theta_j^*}. \tag{3}$$

Antoniak (1974) summarizes the prior distribution for k induced by this process, and shows that it depends critically on γ. A value of $\gamma = 1$ indicates that we are giving the prior P_0 the same weight as every other observation. For instance, for N relatively large, $E(k|\gamma, N) \approx \gamma \ln(1 + \frac{N}{\gamma})$; for N between 50 and 250, the prior for k heavily favors single digit values.

Now assume the data $t = (t_1, ..., t_N)$ are conditionally independent and follow a distribution with density $f(t_i|\theta_i)$. It then follows, from simple application of Bayes' Theorem that the posterior distribution of θ_i given $\theta^{(i)}$ and t is

$$(\theta_i|\theta^{(i)}, t_i) \sim q_0 P_{1i} + \sum_{j=1,\, j\neq i}^{N} q_j \delta_{\theta_j}, \tag{4}$$

where

$$q_0 \quad \propto \quad \gamma \cdot \int f(t_i \,|\theta) dP_0; \tag{5}$$

$$q_j \quad \propto \quad f(t_i \,|\theta_j), \tag{6}$$

and P_{1i} is the marginal posterior distribution of θ_i given the data t and the prior P_0. This posterior distribution has an analogous meaning as above: with probability proportional to q_0 we draw a new value of θ from the posterior distribution P_{1i}, and with probability proportional to q_j we draw from one of the already existing values, θ_j. The proportionality factor is easily obtained by noting that $q_0 + q_1 + ... + q_N$ always sum up to one. The conditional distribution of $(\theta_i \mid \theta^{(i)}, t)$ is easily sampled from, given a convenient choice of the prior distribution P_0. Given some starting value for θ, (possibly drawn from the P_1 distribution), one can sample new elements of θ sequentially, by drawing from the distribution of $(\theta_1 \mid \theta^{(1)}, t)$, $(\theta_2 \mid \theta^{(2)}, t)$, and so on up to $(\theta_N \mid \theta^{(N)}, t)$, with the relevant elements of the most recently sampled $\theta^{(i)}$ values inserted in the conditioning vectors at each step, and repeat this procedure until convergence (Escobar and West, 1995).

2.2. Dirichlet Mixture Models for Duration Data

I now apply the methodology described above to a model of duration data. I first describe a simple model with exponential distribution of duration spells and no covariates, where the constant hazard rate is allowed to differ between individuals. I then extend the model to include covariates; finally, duration dependence can be introduced by letting spell duration follow a Weibull distribution.

An Exponential Model. For a sample of size N, let $t_1, ..., t_N$ be a sample of independent duration times, distributed exponentially:

$$f(t_i \mid \theta_i) = \theta_i \exp(-\theta_i t_i),$$

with

$$\theta_i \ \sim \ P, \qquad \text{and} \qquad P \ \sim \ \mathcal{D}(\gamma P_0). \tag{7}$$

Note that, conditional on θ_i, the model belongs to the proportioanl hazards class. However, by introducing the Dirichlet process based mixture over θ_i, we obtain a a mixture of proportional hazards models, which is not necessarily a proportional hazard model in itself. For the moment let γ be a fixed scalar. The prior distribution P_0 is a Gamma distribution with mean a/b and variance a/b^2, denoted $\mathcal{G}(a,b)$. Its density is

$$p(\theta) = \frac{b^a \theta^{a-1} \exp(-b\theta)}{\Gamma(a)}. \tag{8}$$

It then follows from the analysis in equations (4)-(6) that the posterior distribution of θ_i given $\theta^{(i)}$ and t is given by:

$$\theta_i \mid \theta^{(i)}, t \ \sim \ q_0 P_{1i} + \sum_{j \neq i,\, j=1}^{N} q_j \delta_{\theta_j}, \tag{9}$$

with

$$\begin{aligned} q_0 \ &\propto \ \gamma \cdot \int f(t_i|\theta) dP_0(\theta) = \gamma \cdot \int \theta e^{-\theta t_i} \frac{b^a \theta^{a-1} e^{-b\theta}}{\Gamma(a)} d\theta \\ &= \ \gamma \cdot \frac{a \cdot b^a}{(b+t_i)^{a+1}}; \qquad (10) \\ q_j \ &\propto \ f\left(t_i|\theta_j\right) \qquad (11) \end{aligned}$$

and

$$P_{1i} \sim \mathcal{G}(a+1, b+t_i). \tag{12}$$

Given these distributions, one can then implement a Gibbs sampling algorithm as described at the end of Section 2.1 above to draw from the posterior distribution of θ. It is worth noting that q_0 represents the marginal density of t_i unconditional on θ, and that this density is equivalent to the density of $b \cdot V$, where V is distributed as the ratio of two independent Gamma random variables:

$$V \sim \frac{\mathcal{G}(1, \frac{1}{2})}{\mathcal{G}(a, \frac{1}{2})}. \tag{13}$$

It is possible to also place prior distribution on the parameters that characterize P_0, and the Gibbs sampling algorithm can be augmented by steps that draw from the posterior distribution of these parameters. In this paper, however, the hyperprior parameters a and b are held fixed.

Predictive Distributions. It is also easy to draw from the posterior predictive distribution of a new observation t_{N+1} given the existing draw of the parameters $\theta_1, ..., \theta_N$. This is important if we wish to generate simulated data sets that can be used to assess goodness of fit of the model. There are two ways of generating simulated data for this model. The first method involves drawing a new sample of size N, $\tilde{t}_1, ..., \tilde{t}_N$, given the drawn values $\theta_1, ..., \theta_N$ from an exponential distribution with parameter θ_i, for $i = 1, ..., N$. This is equivalent to generating a predictive distribution for exactly the same N individuals in the sample. These draws tell us about the appropriateness of the underlying exponential model conditional on the unobserved heterogeneity.

The alternative is to take advantage of the fact that the distribution of a new value θ_{N+1} conditional on $\theta_1, ..., \theta_N$ follows the distribution described in equation (2); one can easily draw a new value θ_{N+1} given $\theta_1, ..., \theta_N$, and then draw a duration spell $\tilde{t}_{N+1}$ from an exponential distribution with density $\theta_{N+1}e^{-\theta_{N+1}t}$.

Then, given $\theta_1, ..., \theta_N, \theta_{N+1}$, one can draw a new value θ_{N+2} and a new duration spell $\tilde{t}_{N+2}$, and continue similarly until a new sample of size N has been generated. This is equivalent to generating a predictive distribution of duration spells for a new sample of size N with identical demographic characteristics as the original sample.[4]

Duration Dependence and Covariates. The model described above can be easily extended to allow for duration dependence and the presence of covariates. Suppose that the distribution of completed duration spells is given by:

$$f(t_i \mid \theta_i) = \theta_i e^{X_i\beta} \alpha t_i^{\alpha-1} \exp(-\theta_i e^{X_i\beta} t_i^{\alpha}), \tag{14}$$

and θ follows a Dirichlet process as described in equations (7)-(8). Assume for the moment that the duration dependence and regression parameters are known. The posterior distribution of θ_i conditional on $\theta^{(i)}, t, X, \alpha$, and β has then a form similar to equation (9) with

$$q_0 \propto \gamma \frac{\Gamma(a+1)}{\Gamma(a)} \frac{e^{X_i\beta}}{b} \alpha t_i^{\alpha-1} \left(1 + \frac{e^{X_i\beta}}{b} t_i^{\alpha}\right)^{-(a+1)}; \tag{15}$$

$$q_j \propto \theta_j e^{X_i\beta} \alpha t_i^{\alpha-1} \exp\left(-\exp\left(X_i\beta\right) t_i^{\alpha}\right) \tag{16}$$

and

$$P_{1i} = \mathcal{G}(a+1, b + e^{X_i\beta} t_i^{\alpha}). \tag{17}$$

The form of equation (15) can also be used to deduce the marginal distribution of duration, unconditional on θ_i. Let V be distributed as the ratio of two gamma random variables as in equation (13). It can then be shown that the distribution of t_i is equivalent to the distribution of $\left(\frac{b}{e^{X_i\beta}} V\right)^{\frac{1}{\alpha}}$.

[4] Note that we cannot draw independently N times from the distribution of $\theta_{N+1}|P, \theta_1, ..., \theta_N$, since $\theta_{N+2}|P, \theta_1, ..., \theta_N \nsim \theta_{N+2}|P, \theta_1, ..., \theta_N, \theta_{N+1}$. Thus the sequential nature of this sampling scheme. An alternative would be to obtain approximate exact draws from the Dirichlet process using the Sethuraman and Tiwari construction (Sethuraman and Tiwari, 1982).

Prior distribution for γ and k. As noted above, k, the number of distinct elements of θ induced by the posterior distribution, depends critically on γ. Therefore, one can specify different values for γ, and analyze the sensitivity of the results to different assumptions. Alternatively, as in Escobar and West (1995), and Hirano (2002), one can specify a prior Gamma distribution for the measure of the Dirichlet process γ, and add a step to the Gibbs sampling algorithm to draw from the posterior distribution of γ given k. I follow this latter procedure.

Summarizing the Gibbs Sampler. We can now summarize all the steps in the Gibbs sampling algorithm:[5]

1. Pick some initial values $\alpha^{(1)}, \beta^{(1)}, a^{(1)}$ and $b^{(1)}$. (Details on the choice of initial values is given in the appendix). Then draw $\theta_{1,1}, ..., \theta_{N,1}$ from the posterior distribution P_1.

   Then, for $m = 1, 2, ...$:

2. Given $k^{(m)}$, the number of distinct elements of θ_m, draw $\gamma^{(m+1)}$ following the procedure described in Escobar and West (1995).

3. Draw new values $\theta_{i,m+1}|\theta_m^{(i)}$, for $i = 1, ..., N$ following the steps described in Section 2.1.

4. Given θ, draw from a, b, α and β from their posterior distributions, using the Metropolis-Hastings algorithm when necessary.

5. Once a whole Gibbs iteration has been completed, generate simulated data that will be used for sensitivity analysis following the description in Section 2.2, adapted to the Weibull model with covariates.

[5] Details on the choice of initial values, on the prior distributions for a, b, α, and β, and on the Metropolis-Hastings algorithm are in the Appendix.

6. Repeat steps (2)-(5) until convergence.

To monitor convergence, I follow the methods described in Gelman and Rubin (1992). I run several parallel Gibbs sampling simulations. If the process has reached convergence after a burn-in period, we expect the variation within runs to be roughly the same as the variation between runs. For each scalar estimand ψ, we denote draws from R parallel Gibbs runs of length M as ψ_{mr} ($m = 1, 2, ..., M$; $r = 1, 2, ...R$). The between and within-sequence variances are

$$\begin{aligned} B &= \frac{M}{R-1} \sum_{r=1}^{R} \left(\bar{\psi}_{.R} - \bar{\psi}_{..} \right)^2 ; \\ W &= \frac{1}{R} \sum_{r=1}^{R} s_r^2 \text{ where } \quad s_r^2 = \frac{1}{M-1} \sum_{m=1}^{M} \left(\psi_{mr} - \bar{\psi}_{.r} \right)^2 , \end{aligned}$$

in obvious notation. We then calculate the Gelman-Rubin scale reduction statistic

$$GR = \sqrt{\frac{\frac{M-1}{M} W + \frac{1}{M} B}{W}}.$$

We then continue the simulation runs until the scale reduction statistics for all scalar estimands of interest are "near" 1.[6]

3. Monte Carlo Evidence

In this Section I present some limited Monte Carlo evidence on the performance, in the frequentist sense, of the proposed estimator. I generate repeated samples of duration spells, and ask whether a) the mean over repeated samples of the posterior mean of the parameters of interest approximates the true parameter; and b) the 95 percent posterior interval for the parameters of interest does indeed

[6] Gelman et al. (1996, page 332) suggest that values of $\sqrt{\hat{R}}$ below 1.2 are acceptable.

contain the true parameter 95 percent of the time. The data generating process is

$$\begin{aligned}
f\left(t_i|\alpha,\beta,x_i,\theta_i\right) &= \theta_i e^{x_i\beta}\alpha t_i^{\alpha-1}\exp\left(-\theta_i e^{x_i\beta}\right)\ i=1,...,N; \\
\alpha &= 1;\ \beta=1;\ N=500; \\
x_i &\sim N\left(0,1\right).
\end{aligned}$$

I study the performance of the Dirichlet model under three alternative assumptions for the heterogeneity distribution.

1. $p\left(\theta_i\right)\sim\mathcal{G}\left(1/2,1\right).$

2. $p\left(\theta_i\right)\sim$ Multinomial:

$$\begin{aligned}
\theta &= \left(0.25,0.75,1,1.25,1.75\right); \\
P &= \left(0.3,0.2,0.1,0.15,0.25\right).
\end{aligned}$$

3. $p\left(\theta_i\right)\sim$ Mixture of normals truncated at zero:

$$0.5N\left(0.5,0.25^2\right)+0.5N\left(3,0.25^2\right).$$

The choice of distributions is dictated by our interest in the question of whether the Dirichlet model does a good job at recovering the underlying heterogeneity distribution. Since the prior and the posterior expectations of the Dirichlet process are themselves Gamma distributions, we would expect that the Gibbs sampler should do a particularly good job in generating posterior distributions centered around the true parameters when the true heterogeneity is Gamma. Moreover, we expect in this case that a relatively large weight should be given to the posterior expectation P_1: the posterior distribution for γ should be centered around large values, and the number of distinct elements in θ should also be large – the non-parametric model should approximate the baseline parametric model.

The multinomial and bimodal distributions, on the other hand, depart substantially from the baseline Gamma distribution. The behavior of the estimator is studied under these two alternatives to assess the flexibility of our semiparametric estimator. In these alternative specifications, we expect small values for the posterior distribution of γ and a small number of distinct elements in θ.

For each simulated data set, I ran 8 parallel Gibbs sequences, and constructed posterior distributions by discarding the first 25% of draws in each sequence. Initial values for the parameters were drawn randomly from an overdispersed distribution centered at the maximum likelihood estimates of a parametric model with Gamma heterogeneity. The hyperprior parameters a and b were both set to 0.01 in all three models, indicating a very non-informative prior expectation for the Dirichlet process.

The Monte Carlo analysis is based on only 50 replications. This is far from being a comprehensive Monte Carlo exercise, but it should give us at least suggestive evidence on the properties of posterior distributions based on the Dirichlet process prior.

The basic results of the Monte Carlo simulations are presented in Table 1. Altogether the results are encouraging. The average of the posterior means for α and β are in the neighborhood of the true parameter values, and the coverage rates of the 95% posterior interval appear appropriate given the small number of replications. The behavior of the parameters governing the heterogeneity distribution also conforms to expectation. With Gamma unobserved heterogeneity, the posterior means of γ and k are rather high, indicating that the data supports giving relatively high weight to the baseline parametric distribution. In the two alternative models, the posterior means of γ and k are relatively low: the data supports a substantial departure from the baseline parametric model.

4. An Application to Unemployment Spells in the NLSY

4.1. A Model With Parametric Heterogeneity

I now apply the proposed estimator to data on the duration of unemployment spells for young men. My sample includes 1000 spells of unemployment (in weeks) for males in the National Longitudinal Survey of Youth (NLSY).[7] I first estimate the parameters of a benchmark Weibull duration model with parametric Gamma heterogeneity by maximum likelihood. The density for a completed duration spell is

$$\begin{aligned} f\left(t_i|X_i\right) &= \int f\left(t_i|X_i,\theta\right) dP\left(\theta\right) \\ &= \int \theta e^{X_i\beta}\alpha t_i^{\alpha-1} \exp\left(-\theta e^{X_i\beta} t_i^{\alpha}\right) \frac{b^a\theta^{a-1}e^{-b\theta}}{\Gamma\left(a\right)} d\theta \\ &= \frac{e^{X_i\beta}\alpha t_i^{\alpha-1} b^a}{\Gamma\left(a\right)} \frac{\Gamma\left(a+1\right)}{\left(b+e^{X_i\beta}t_i^{\alpha}\right)^{a+1}}. \end{aligned}$$

I include as explanatory variables a constant, age, education, two race dummies and a dummy for having received training. The maximum likelihood estimates are presented in Table 2. As we can see, age and education have a positive and significant effect on the hazard rate, the exit rate for hispanics, but not for blacks, is significantly lower than that of whites, and training raises the exit by approximately 39 percent.[8] There is slight evidence of positive duration dependence, but

[7] I thank Guido Imbens for making this data available to me.

[8] Given the absence of randomization, these should not be interpreted as causal effects.

the null hypothesis of $\alpha = 1$ can not be rejected based on the ML estimates.

4.2. Results Based on The Dirichlet Model

In this section I describe the results obtained from application of the Dirichlet model to the NLSY data. I assume that the unobserved heterogeneity term θ has a Dirichlet process prior distribution and I follow the methodology described in Section 2.2. The Gibbs sampling algorithm described in section 2.2 is augmented with a step for drawing completed duration spells for censored observations.[9] Altogether I ran 8 parallel Gibbs sequences of length 8,000, and retained only the last 6,000 draws in each sequence, for a total of 48,000 draws.

The prior distributions for α and β were chosen to be diffuse but proper. The prior for α was a $\mathcal{G}(3,2)$ distribution; the prior for β was $N(0, 5\mathbf{I})$. After some experimentation, it was decided to use information from the parametric model to determine the values of the hyperparameters a and b. Let $\hat{a}_{ML}$ and $\hat{b}_{ML}$ be the maximum likelihod estimates of a and b in the parametric model above: I set $a = \hat{a}_{ML}/5$ and $b = \hat{b}_{ML}/5$, meaning that the prior expectation of the Dirichlet process has the same mean as the heterogeneity distribution estimated by maximum likelihood, but has five times larger variance. I have found that if the prior is less informative, it is difficult to achieve convergence, whereas a more informative prior will dominate the data, and the posterior distributions will not depart meaningfully from that obtained in a fully parametric model. Finally, the Dirichlet precision parameter γ has a $\mathcal{G}(3,2)$ prior distribution. This distribution favors heavily low values of γ and consequently a small number of mixture points in the heterogeneity distribution.

[9] Alternatively, one could directly use the survival function as the likelihood contribution for the censored observations.

Table 3 shows the posterior distribution of the model parameters. First of all, note that for all the elements of β the Gelman-Rubin statistic is very close to 1, indicating that the sequences have reached approximate convergence. The Gelman-Rubin statistic for α is slightly above 1.2, a value considered borderline acceptable.[10] The posterior distributions for selected parameters are depicted graphically in Figures 1 to 3. The continuous lines in the figures show the prior distribution for the chosen parameters: it is not surprising that the flat prior is quickly dominated for the parametric part of the model. The distributions of the elements of β are roughly symmetric, while the distribution of α has a long left tail. This result seems to be due to one Gibbs sequence that wandered off testing low values for α. A longer Gibbs chain would probably be required to verify whether this left tail is a true feature of the posterior distribution.

Compared to the parametric model, we find that the posterior distribution for α is located much to the right of the maximum likelihood estimate, indicating now a substantial degree of positive duration dependence. The distributions of the β coefficients are also shifted relative to their maximum likelihood estimates. However, it may make more sense to look at the distribution of $-\beta/\alpha$, which tells us the percentage effect of the explanatory variables on expected duration in the Weibull model (conditional on the heterogeneity parameter θ).[11] This distribution is shown in Table 4. The posterior means and standard deviation for this quantity

[10] Inspection of the Gibbs sequences reveals that the apparent non-convergence of α is due to one of the eight sequences, which wandered off to low values.

[11] In the Weibull model, conditional expected duration is

$$E\left(T|X,\alpha,\beta,\theta\right) = \Gamma\left(1+\frac{1}{\alpha}\right)\theta^{-1/\alpha}e^{-X\beta/\alpha}.$$

Therefore $-\beta/\alpha = \partial \ln E\left(T|X,\alpha,\beta,\theta\right)/\partial X$.

are relatively close to the maximum likelihood estimates and their standard errors respectively for most variables. This is also shown in Figure 4, which show the posterior distribution of the percentage effects on expected duration together with the normal distribution implied by the maximum likelihood estimates. For all explanatory variables but age, there is substantial agreement between the two distributions.

One would also like to learn about the underlying mixing distribution. Table 5 shows the posterior distribution of the Dirichlet measure parameter γ and of the number of mixture points k. For comparison, I also report the theoretical prior distribution of γ and of the expected number of mixture points, $E(k|\gamma, N) \approx \gamma \ln(1 + N/\gamma)$.[12] The posterior and prior distributions of γ and k are also plotted in Figure 5. The striking finding here is that the prior distribution for γ and k is completely dominated by the data. Despite our prior preference for low values of γ and a moderate number of distinct clusters in the heterogeneity distribution, the data strongly supports a posterior distribution highly concentrated around the baseline distribution P_1 and with a large number of distinct clusters.

Finally, we would like to learn about features of the heterogeneity distribution θ. The posterior distribution of the mean, 25th, 50th, and 75th percentiles of the θ distribution are presented in Table 6. Compared to the gamma heterogeneity distribution implied by the parametric model, the estimated distribution of θ has a substantial portion of its mass at very low values, even though the mean is only slightly lower. It appears that the heterogeneity distribution departs considerably from the distribution implied from the parametric model. Moreover, the posterior

[12] The prior for γ is a $\mathcal{G}(3, 2)$ distribution. The values of $E(k|\gamma, N)$ are simulated using draws from a $\mathcal{G}(3, 2)$ distribution.

distribution for the distribution of θ seems to be dominated by the data, not the prior. This can be seen in Figure 6, where we graph the prior and posterior distributions for the mean of θ: as can be clearly seen, the posterior distribution is considerably more concentrated.

4.3. Comparison to the NPMLE

It is worthwhile to compare the results the methodology described here to more traditional procedures used to model non-parametrically unobserved heterogeneity in duration data. In particular, I compare my results to those obtained using Heckman and Singer's NPMLE. The Heckman and Singer estimator specifies the conditional density of duration to be the same as in (14), but now the distribution of the heterogeneity term is given by

$$\theta_i = \begin{cases} \eta_1 & \text{with probability } p_1 \\ \eta_2 & \text{with probability } p_2 \\ & \cdots \\ \eta_K & \text{with probability } 1 - \sum_{k=1}^{K-1} p_k \end{cases}.$$

I estimate this model by maximum likelihood using the EM algorithm,[13] and analyze the sensitivity of the results to the number of mixture points K in the heterogeneity distribution. As in most of the applied literature, I limit the analysis to models with low values of K.

The Heckman-Singer estimates are presented in Table 7. One notes immediately that the parameter estimates are quite sensitive to the number of mixture points used. This is true not only of the duration dependence parameter α, but also of the coefficients on the explanatory variables. We conjecture that a model that ignores the uncertainty surrounding the number of mixture points in the

[13] See Dempster, Laird and Rubin (1977).

heterogeneity distribution may lead to biased inference and excessively tight confidence intervals. In this respect, the estimator based on the Dirichlet process prior represents a substantial improvement: by integrating over the posterior distribution of the *distribution* of the unobserved heterogeneity term, it may generate more reliable inference on the extent of duration dependence and on the effect of the explanatory variables on the length of unemployment spells. This statement should be qualified by two observations. First, one should ensure that the Gibbs sampling algorithm has indeed converged to the true posterior distribution. Second, the posterior intervals based on the Dirichlet model are exact probability intervals conditional on a given prior. It remains to be seen whether these intervals have adequate coverage rates in the frequentist sense and can be interpreted as confidence intervals.

5. Conclusion

In this paper I have presented and described a methodology for drawing semiparametric Bayesian inference for duration data and for generating predictive distributions. Modelling unobserved heterogeneity as a Dirichlet process, I describe how to draw from the posterior distribution of the parameters of interest using the methodology described in Escobar and West (1995). I then apply the methodology to data on the duration of unemployment spells of young men in the NLSY, and compare it to results from a simple parametric model. The marginal effects of explanatory variables on expected duration differ significantly between the two models. These differences can lead a decision maker to implement different courses of action, depending on whether he is using the parameteric or the non-parametric model. Consider for example the worker who must decide whether to enroll in a training program: the predicted effect of training on unemployment duration is a full two weeks larger in the Dirichlet model than in the parametric one.[14]

I also compare the performance of my estimator with conventional methods used in modelling duration data with unobserved heterogeneity. The Heckman-Singer estimator is quite sensitive to the number of mixture points in the distribution of unobserved heterogeneity. By integrating over the uncertainty in the distribution of unobserved heterogeneity, my estimator overcomes this problem and generates more reliable results.

Despite implying rather different conclusions about the nature of unemployment, both a simple parametric model and the model based on the Dirichlet process prior seem to fit the data rather well. To be precise: neither model seems to be strikingly at odds with some of the important features of the data examined. The final judgment about which model to adopt depends on our own economic knowledge. If we interpret the semiparametric model as a model for duration data with unobserved heterogeneity, and if we believe that substantial heterogeneity persists even after controlling for the observed covariates, then we should base our inference on a model that allows for unobserved heterogeneity. It should be noted though, that it may be preferable to have flexibility in the baseline hazard rather than in the heterogeneity distribution, to capture for example deadline effects associated with exhaustion of Unemployment Insurance benefits (Meyer, 1990).

The distinction between unobserved heterogeneity and structural duration dependence may be purely academic: in the end it is impossible know whether the

[14]I abstract here from the issue of whether results can be given a causal interpretation.

population is heterogeneous or whether duration spells are drawn from a complicated mixing distribution. Nevertheless, the methodology presented here can still be useful because of its ability to generate predictive distributions for duration data. Predictive distributions incorporate parameter uncertainty present in the model, so that it is relevant for decision making under uncertainty in the expected utility framework of microeconomics.[15] For example, consider the case of an agent who has just become unemployed, and must now decide how to reallocate consumption of durable and non-durable goods over the course of the unemployment spell. If the agent believes that he will soon return to work, he probably will not alter substantially his consumption when unemployed. On the other hand, if he believes that the unemployment spell will last long, he may immediately adjust his consumption path.[16] Clearly, the agent must have knowledge of the entire distribution of the length of unemployment spells to maximize expected utility. The Bayesian methodology provides a powerful and relatively simple tool for generating predictive distributions for spells, which can be used to inform the agent's optimal consumption decision.

[15] See Chamberlain and Imbens (2003) for some examples of this approach.

[16] For example, Dynarski and Sheffrin (1987) find that consumption changes following unemployment spells are smaller for workers with high recall probabilities.

6. Appendix

The duration model is

$$p(t_i|\theta_i, X_i, \alpha, \beta) = \theta_i e^{X_i\beta} \alpha t_i^{\alpha-1} e^{-\theta_i e^{X_i\beta} t_i^{\alpha}}$$

The heterogeneity term θ_i is assumed to have a Dirichlet process prior $\mathcal{D}(\gamma P_0)$ with $P_0 = \mathcal{G}(a, b)$. The parameters of the Weibull distribution, α and β, are given a diffuse bu proper prior:

$$\begin{aligned} \alpha &\sim \mathcal{G}(3, 2), \\ \beta &\sim N(0, 5\mathbf{I}). \end{aligned}$$

After some experimentation, it was decided to use information from a parametric model to determine the values of the hyperparameters a and b. I first estimated a parametric model with Gamma heterogeneity by maximum likelihood, yielding values $\hat{a}_{ML}$ and $\hat{b}_{ML}$. I then set $a = \hat{a}_{ML}/10$ and $b = \hat{b}_{ML}/10$, meaning that the prior expectation of the Dirichlet process has the same mean as the heterogeneity distribution estimated by maximum likelihood, but has ten times larger variance.

The Metropolis-Hastings Algorithm for α and β. To draw from the posterior distribution of α and β, I use a Metropolis-Hastings algorithm. The conditional posterior distribution is proportional to the likelihood function times the prior:

$$\begin{aligned} p(\phi|t^*, X, \theta) &= p(\alpha, \beta|t^*, X, \theta) \\ &\propto \prod_{i=1}^{N} \theta_i e^{X_i\beta} \alpha t_i^{*\alpha-1} e^{-\theta_i e^{X_i\beta} t_i^{*\alpha}} \times p(\alpha, \beta), \end{aligned}$$

where t^* is the vector of completed durations.

Given an initial value $\phi^{(m)} = \left(\alpha^{(m)}, \beta^{(m)}\right)$, I draw a candidate value ϕ^* from a normal distribution with mean $\phi^{(m)}$, and variance $\Sigma^{(m)} = -\left(\frac{\partial^2 \log p(\phi^{(m)}|t^*, X)}{\partial\phi\partial\phi'}\right)^{-1}$, i.e., the inverse of the information matrix evaluated at the current parameter draw. I then calculate the importance ratio

$$r_\phi = \frac{p\left(\phi^*|t^*, X\right)}{p\left(\phi^{(m)}|t^*, X\right)} \frac{N\left(\phi^{(m)}|\phi^*, \Sigma^*\right)}{N(\phi^*|\phi^{(m)}, \Sigma^{(m)})},$$

where $N(\cdot|\mu, \Sigma)$ is the multivariate normal density evaluated at $\cdot$, and Σ^* is the inverse of the information matrix evaluated at the candidate value ϕ^*. I then accept the new value ϕ^* with probability $\min(r_\phi, 1)$.

Drawing from the posterior distribution of γ and k. The prior distribution of γ is $\mathcal{G}(a_1, a_2)$, with $a_1 = 3$, and $a_2 = 2$. At the beginning of each Gibbs iteration I draw a random variable η from a $Beta(\gamma^{(m-1)}, N)$ distribution. The posterior distribution of $\gamma^{(m)}$ is then given by:

$$\gamma^{(m)} \sim \pi_\eta \mathcal{G}(a_1 + k^{(m-1)}, a_2 - \log\eta) + (1 - \pi_\eta)\mathcal{G}(a_1 + k^{(m-1)} - 1, a_2 - \log\eta)$$

where $\frac{\pi_\eta}{1-\pi_\eta} = \frac{a_1 + k^{(m-1)} - 1}{N(a_2 - \log\eta)}$.

References

[1] **Antoniak, Charles E.** "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems." *The Annals of Statistics*, November 1974, *2*(6), pp. 1152–1174.

[2] **Blank, Rebecca M.** "Analyzing the Length of Welfare Spells." *Journal of Public Economics*, August 1989, *39*(3), pp. 245–273.

[3] **Campolieti**, **Michael**. "Bayesian Semiparametric Estimation of Discrete Duration Models: An Application of the Dirichlet Process Prior." *Journal of Applied Econometrics,* January 2001, *16*(1), pp. 1-22.

[4] **Chamberlain, Gary and Imbens, Guido W.** "Nonparametric Applications of Bayesian Inference." *Journal of Business and Economic Statistics*, January 2003, *21*(1), pp. 12-18.

[5] **Dempster, Arthur P.; Laird, Nan M. and Rubin, Donald B.** "Maximum Likelihood from Incomplete Data via the *EM* Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological),* 1977, *39*(1), pp.1-38.

[6] **Dynarski, Mark and Sheffrin, Steven M.** "Consumption and Unemployment." *The Quarterly Journal of Economics*, May 1987, *102*(2), pp. 411-428.

[7] **Escobar**, **Michael D.** "Estimating Normal Means with a Dirichlet Process Prior." *Journal of the American Statistical Association,* March 1994, *89*(425), pp. 268–277.

[8] **Escobar, Michael D. and West, Michael.** "Bayesian Density Estimation and Inference Using Mixtures." *Journal of the American Statistical Association,* June 1995, *90*(430), pp. 577–588.

[9] **Ferguson, Thomas S.** "A Bayesian Analysis of Some Nonparametric Problems." *The Annals of Statistics*, March 1973, *1*(2), pp. 209–230.

[10] **Ferguson, Thomas S.** "Bayesian Density Estimation by Mixtures of Normal Distributions," in M. Haseeb Rizvi; Jagdish S. Rustagi and David Siegmund, eds., *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday.* New York: Academic Press, 1983, pp. 287–302.

[11] **Gelman, Andrew; Carlin, John B.; Stern, Hal S. and Rubin, Donald B.** *Bayesian Data Analysis.* London: Chapman & Hall, 1996.

[12] **Gelman, Andrew and Rubin, Donald B.** "Inference from Iterative Simulation Using Multiple Sequences (with discussion)." *Statistical Science*, November 1992, *7*(4), pp. 457–511.

[13] **Geman, Stuart, and Geman, Donald**. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence.* November 1984, *6*(6), pp. 721–741.

[14] **Ham, John C. and Rea, Samuel Jr. A.** "Unemployment Insurance and Male Unemployment Duration in Canada." *Journal of Labor Economics,* July 1987, *5*(3), pp. 325–353.

[15] **Ham, John C. and Lalonde, Robert J.** "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training." *Econometrica*, January 1996, *64*(1), pp. 175–205.

[16] **Heckman, James J. and Singer, Burton.** "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica,* March 1984, *52*(2), pp. 271-320.

[17] **Hirano, Keisuke.** "Semiparametric Bayesian Inference in Autoregressive Panel Data Models." *Econometrica*, March 2002, *70*(2), pp.781-799.

[18] **Hjort, Nils Lid**. "Nonparametric Bayes Estimators Based on Beta Processes in Models for Life History Data." *The Annals of Statistics*, September 1990, *18*(3), pp. 1259–1294.

[19] **Kalbfleisch, John D. and Prentice, Ross L.** *The Statistical Analysis of Failure Time Data*. New York: Wiley and Sons, 1980.

[20] **Merrick, Jason R. W.; Soyer, Refik and Mazzuchi, Thomas A.** "A Bayesian Semiparametric Analysis of the Reliability and Maintenance of Machine Tools." *Technometrics,* February 2003, *45*(1), pp. 58-69.

[21] **Meyer, Bruce D.** "Unemployment Insurance and Unemployment Spells." *Econometrica,* July 1990, *58*(4), pp. 757–782.

[22] **Ruggiero, Michael.** "Bayesian Semiparametric Estimation of Proportional Hazards Models." *Journal of Econometrics*, June 1994, *62*(2), pp. 277–300.

[23] **Sethuraman, Jayaram and Tiwari, Ram C.** "Convergence of Dirichlet Measures and the Interpretation of their Parameter," in Shanti S. Gupta and

James O. Berger, eds., *Statistical Decision Theory and Related Topics III.* New York: Academic Press, 1982, Volume 2, pp. 305–315.

[24] **Stevens, Ann Huff.** “Climbing out of Poverty, Falling Back In.” *Journal of Human Resources*, Summer 1999, *34*(3), pp. 557–588.

[25] **Van der Vaart, Aad.** “Efficient Maximum Likelihood Estimation in Semiparametric Mixture Models.” *The Annals of Statistics,* April 1996, *24*(2), pp. 862–878.

Table 1 - Monte Carlo Evidence

	Heterogeneity Distribution		
	Gamma	**Multinomial**	**Bimodal**
α - Posterior Mean	1.001	1.0481	1.0668
α - Coverage Rate of 95% Posterior Interval	0.96	0.98	0.92
β - Posterior Mean	1.000	1.0707	1.0736
β - Coverage Rate of 95% Posterior Interval	0.96	0.98	0.92
γ (Dirichlet parameter) - Posterior Mean	18.013	6.666	9.222
k (number of distinct elements of θ) - Posterior Mean	15.98	4.65	7.21

Table 2 - Weibull Model with Parametric Gamma Heterogeneity Maximum Likelihood Estimates

	Coefficient	Std. Error
α	1.1333	0.0555
Age	0.0575	0.0229
Education	0.0849	0.0253
Black	0.0072	0.1184
Hispanic	-0.5029	0.1024
Training	0.3947	0.1588
a	2.6971	0.7138
b	546.196	224.452
$\log L$	-3885.2	

Table 3: Posterior Distributions, Dirichlet Model						
	α	Age	Education	Black	Hispanic	Training
min	1.2919	-0.0782	-0.0327	-0.9257	-1.4469	-0.5976
5%	1.5043	-0.0183	0.0554	-0.3110	-1.0193	0.2845
25%	1.6559	0.0046	0.0974	-0.1203	-0.8481	0.5206
median	1.7121	0.0022	0.1252	-0.0025	-0.7392	0.6778
75%	1.7670	0.0398	0.1534	0.1075	-0.6274	0.8334
95%	1.8438	0.0669	0.1888	0.2807	-0.4563	1.0602
max	1.9865	0.1398	0.2918	0.6640	0.0327	1.5777
mean	**1.7021**	**-0.0017**	**0.1246**	**-0.0081**	**-0.7375**	**0.6741**
std. error	**0.1010**	**0.0333**	**0.0409**	**0.1815**	**0.1718**	**0.2402**
Gelman-Rubin Statistic	1.2760	1.0410	1.0172	1.0079	1.0162	1.0159

Table 4: Posterior Distributions of the Marginal Effect of the Explanatory Variables on Expected Duration					
	Age	Education	Black	Hispanic	Training
min	-0.1030	-0.1767	-0.5061	-0.0200	-1.0616
5%	-0.0417	-0.1104	-0.1658	0.2729	-0.6294
25%	-0.0238	-0.0894	-0.0634	0.3696	-0.4883
median	-0.0130	-0.0736	0.0014	0.4315	-0.3965
75%	-0.0027	-0.0577	0.0706	0.4975	-0.3069
95%	0.0104	-0.0329	0.1824	0.5960	-0.1684
max	0.0419	0.0241	0.6442	0.8904	0.4442
mean	-0.0139	-0.0731	0.0049	0.4334	-0.3965
std. deviation	0.0162	0.0236	0.1091	0.0995	0.1440
ML estimate	-0.0507	-0.0749	-0.0063	0.4439	-0.3484
Std. error	0.0197	0.0226	0.1045	0.0894	0.1385

Table 5: Distribution of the mixing distribution parameters

	$\gamma-$posterior	k-posterior	γ - prior	$E(k\|\gamma,N)$
min	95.64	206	0	0.3731
5%	121.07	249	0.4089	3.1486
25%	133.24	269	0.8637	6.1875
median	141.92	282	1.3370	8.8466
75%	150.89	296	1.9602	12.2103
95%	164.11	315	3.1479	18.1804
max	196.04	359	∞	39.3952
mean	142.17	282.12	1.5	9.5426
std. error	13.07	19.99	0. 8660	4.6476
Gelman-Rubin Statistic	1.0040	1.0025	-	-

Table 6: Features of the θ Distribution

	Mean	25%	Median	75%
min	2.5490	0.1316	0.4219	2.5148
5%	3.4844	0.2273	0.7509	3.7674
25%	3.9233	0.2987	0.9462	4.4318
median	4.2880	0.3628	1.1131	4.9543
75%	4.6994	0.4465	1.3183	5.5380
95%	5.4027	0.6495	1.6792	6.5122
max	8.2578	1.2150	2.6662	9.8376
mean	4.3458	0.3881	1.1510	5.0218
std. dev.	0.5937	0.1304	0.2844	0.8379

Features of the θ Distribution in Parametric Model $[\mathcal{G}\,(2.7, 546)]$

Mean	25%	Median	75%
4.9380	2.7273	4.3430	6.5062

Note: All entries in the table are blown up by a factor of 10^3

Table 7: NPMLE Estimates				
	2 mass points	3 mass points	4 mass points	5 mass points
α	1.2188	1.4992	1.8726	1.9437
	(0.0666)	(0.0944)	(0.1376)	(0.1634)
age	0.0406	0.0422	0.0750	0.0735
	(0.0199)	(0.0270)	(0.0328)	(0.0345)
education	0.1171	0.1420	0.1364	0.1489
	(0.0292)	(0.0330)	(0.0399)	(0.0426)
black	0.0154	0.0086	-0.0248	-0.0023
	(0.1071)	(0.1396)	(0.1708)	(0.1842)
hispanic	-0.6249	-0.6015	-0.8613	-0.8168
	(0.0940)	(0.1213)	(0.1499)	(0.1608)
training	0.4296	0.4439	0.5111	0.6349
	(0.1317)	(0.1685)	(0.2077)	(0.2227)

Notes: standard errors in parentheses.

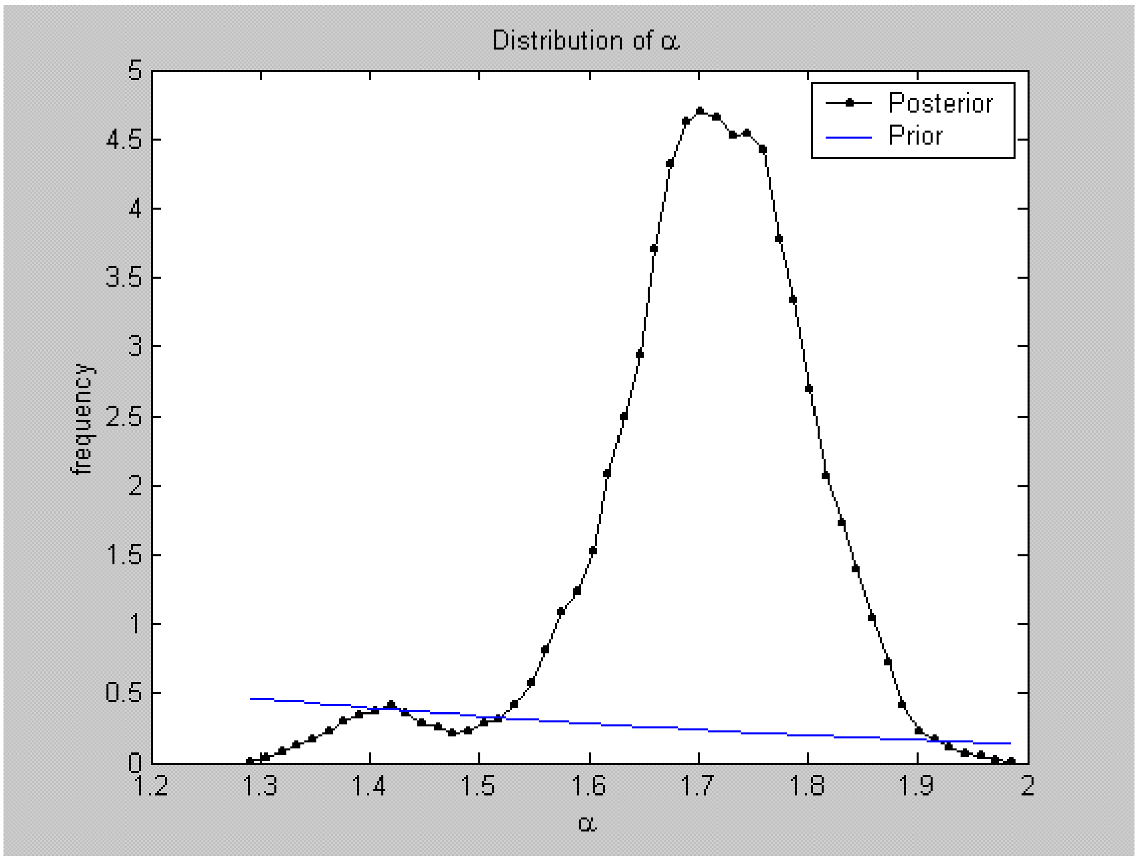

Figure 1: Distribution of α

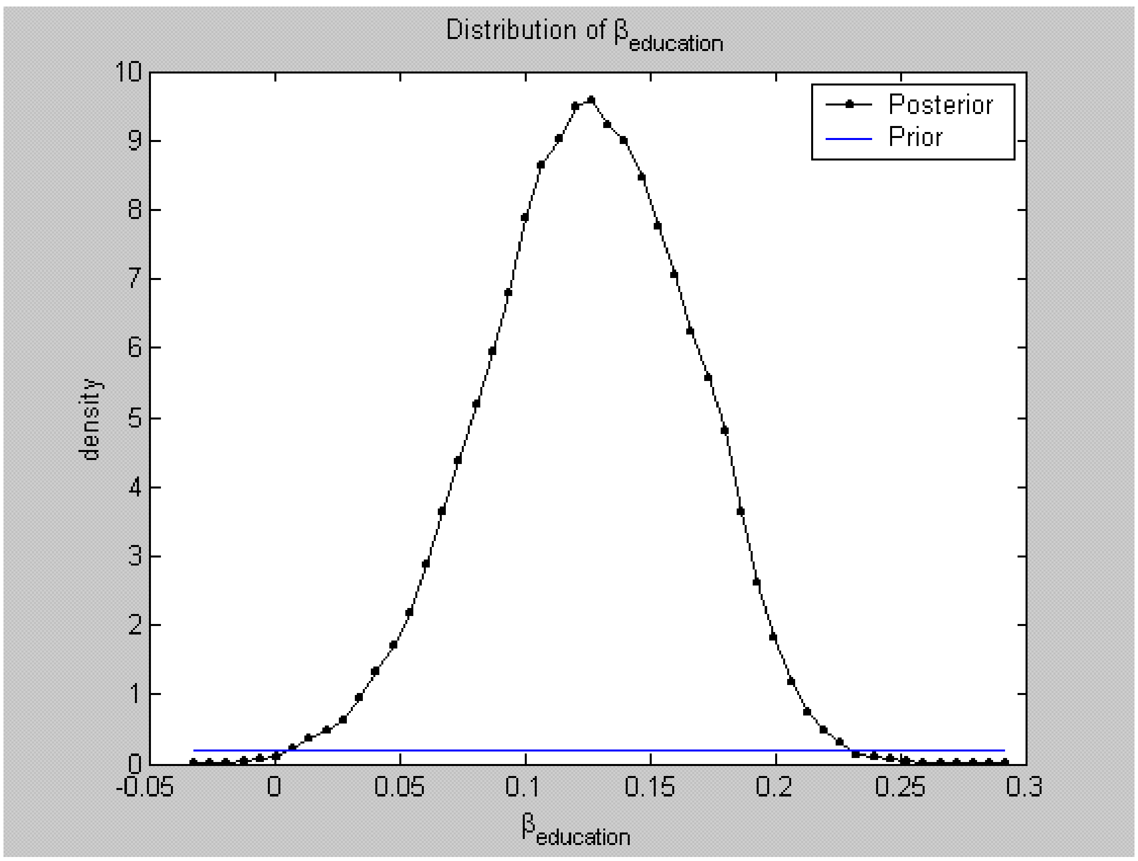

Figure 2: Distribution of $\beta_{education}$

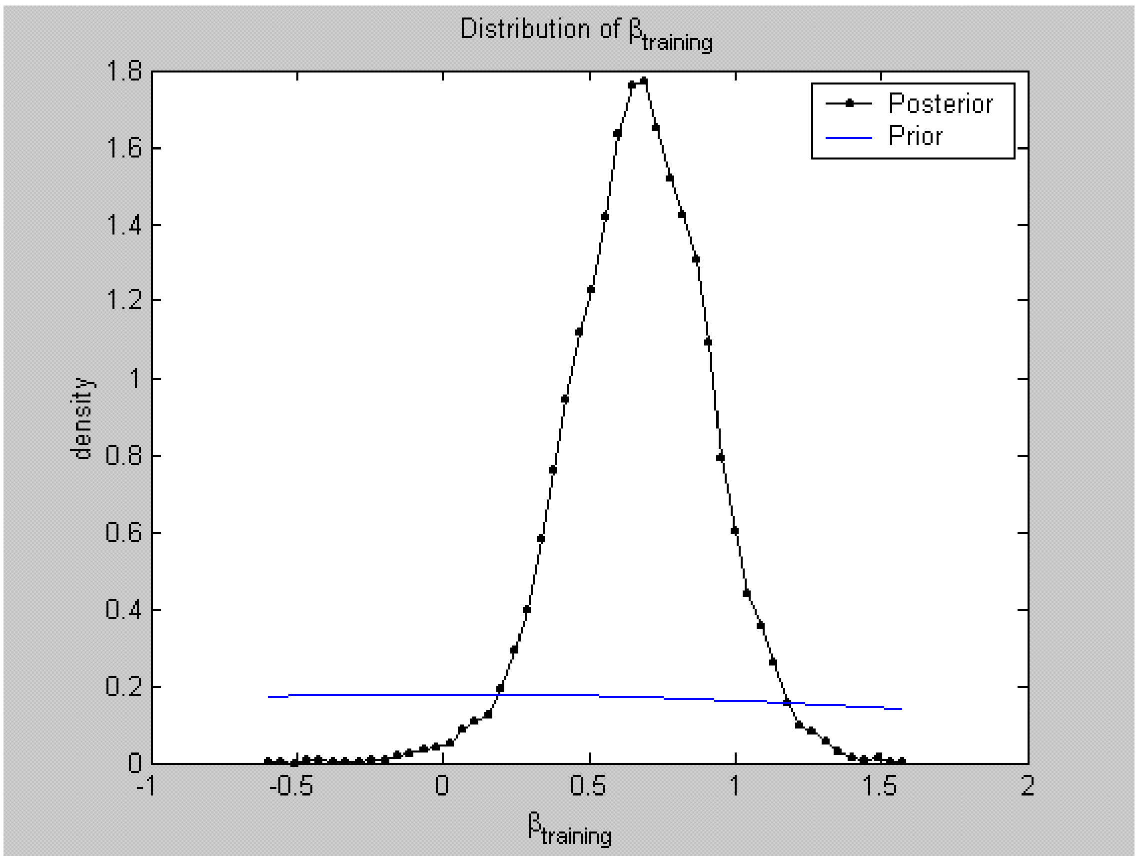

Figure 3: Distribution of $\beta_{training}$

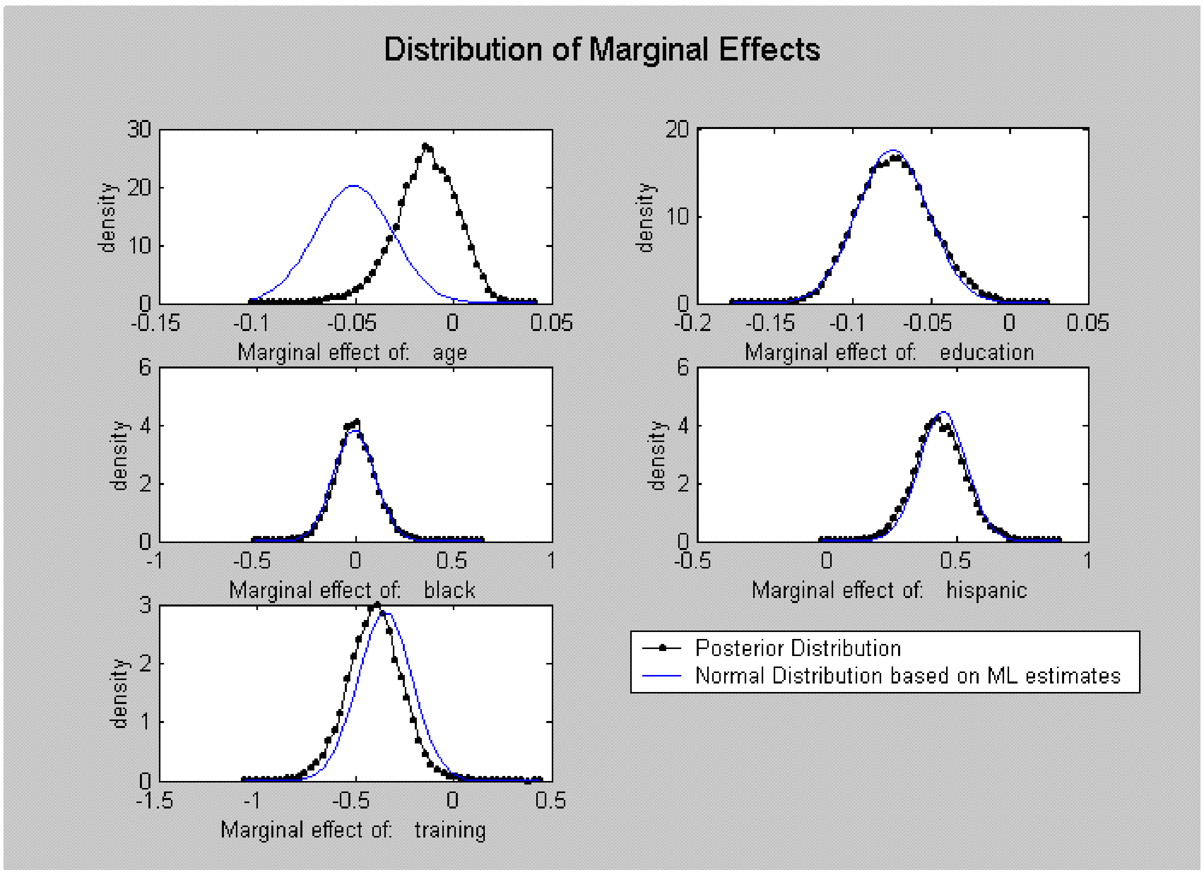
Distribution of Marginal Effects
density
Marginal effect of: age
Marginal effect of: education
Marginal effect of: black
Marginal effect of: hispanic
Marginal effect of: training
Posterior Distribution
Normal Distribution based on ML estimates

Figure 4:

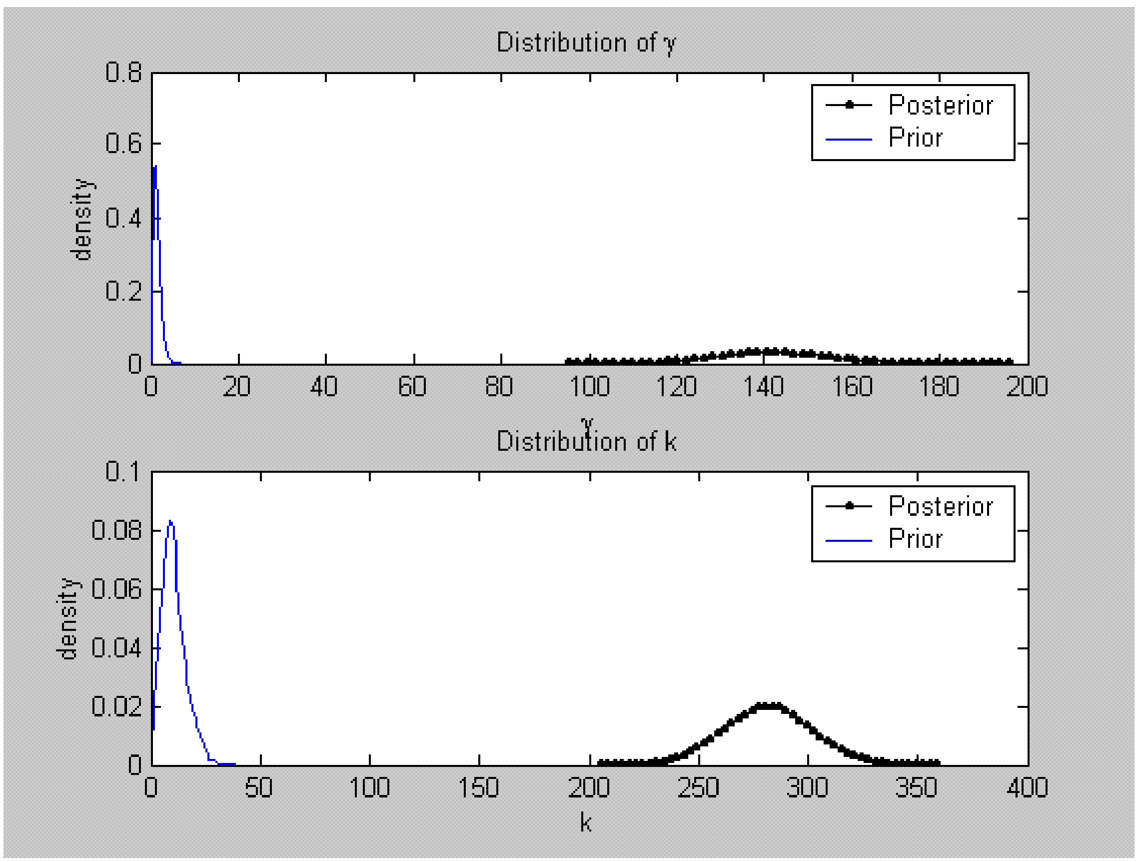

Figure 5: Distribution of γ and k

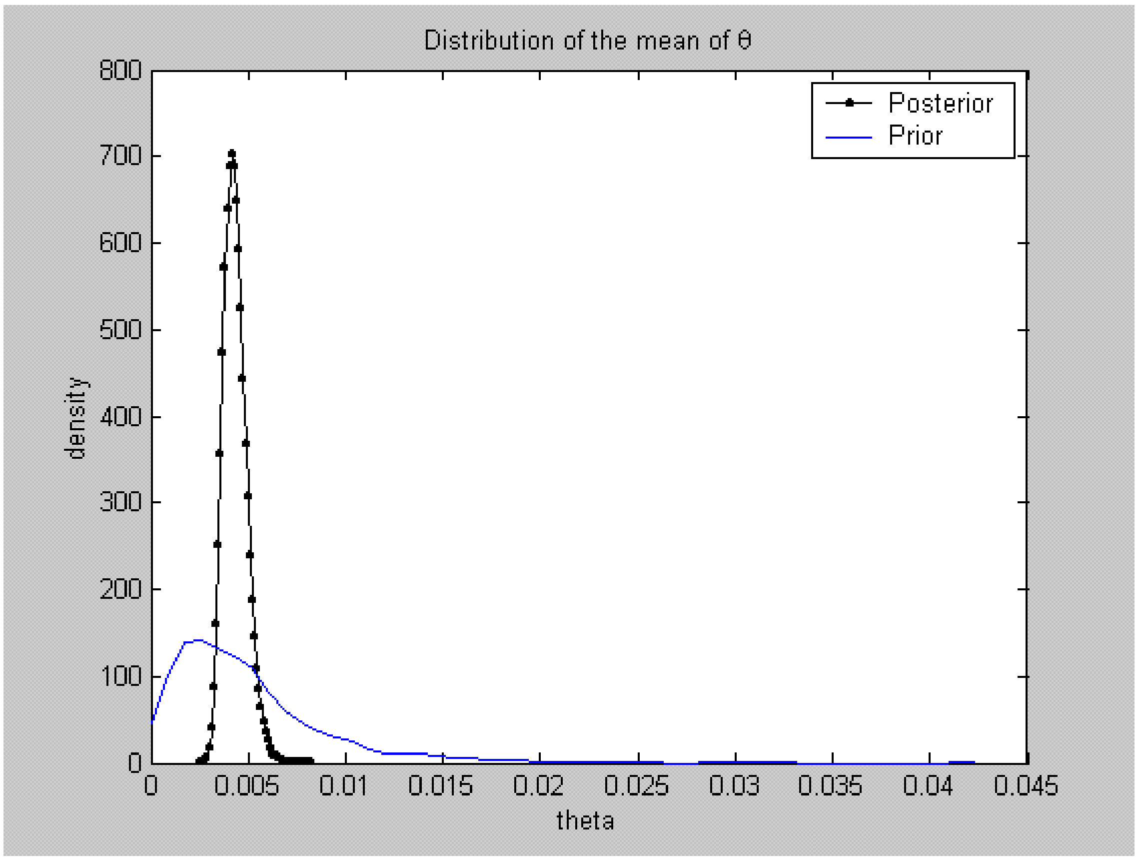

Figure 6: Distribution of the mean of θ