Simulation Based Inference for Dynamic Multinomial Choice Models

Geweke, John and Houser, Dan and Keane, Michael

July 1999

Simulation Based Inference for Dynamic Multinomial Choice Models

By

John Geweke
University of Minnesota
and
University of Iowa

Daniel Houser
University of Arizona

Michael Keane
University of Minnesota
and
New York University

March, 1999
Revised July, 1999

1. Introduction

Over the last decade econometric inference based on simulation techniques has become increasingly common, particularly for latent variable models. The reason is that such models often generate econometric objective functions that embed high-order integrals, and which, consequently, can be most easily evaluated using simulation techniques.[1] There are several well known classical techniques for inference by simulation. Perhaps most common are the Method of Simulated Moments (McFadden (1989)) and Simulated Maximum Likelihood or SML (Lerman and Manski, (1981)). In practice, both methods require that reasonably accurate simulators be used to evaluate the integrals that enter the objective function (see Geweke, et. al., (1994)). Bayesian techniques are also becoming quite popular. These techniques typically entail Markov Chain – Monte Carlo (MCMC) simulation to evaluate the integrals that define the posterior densities of a model's parameters (see Geweke and Keane (1999b) for an overview of MCMC methods).

Our goal in this chapter is to explain concretely how to implement simulation methods in a very general class of models that are extremely useful in applied work: dynamic discrete choice models where one has available a panel of multinomial choice histories and partially observed payoffs. Many general surveys of simulation methods are now available (see Geweke (1996), Monfort, et. al., (1995), and Gilks, et. al., (1996)), so in our view a detailed illustration of how to implement such methods in a specific case has greater marginal value than an additional broad survey. Moreover, the techniques we describe are directly applicable to a general class of models that includes static discrete choice models, the Heckman (1976) selection model, and all of the Heckman (1981) models (such as static and dynamic Bernoulli models, Markov models, and renewal processes.) The particular procedure that we describe derives from a suggestion by Geweke and Keane (1999a), and has the advantages that it does not require the econometrician to solve the agents' dynamic optimization problem, or to make strong assumptions about the way individuals form expectations.

This chapter focuses on Bayesian inference for dynamic multinomial choice models via the MCMC method. Originally, we also hoped to discuss classical estimation of such models, so that readers could compare the two approaches. But, when we attempted to estimate the model developed below using SML it proved infeasible. The high dimension of the parameter vector caused iterative search for the maximum of the simulated likelihood function via standard gradient based methods to fail rather dismally. In fact, unless the initial parameter values were set very close to the true values, the search algorithm would quickly stall. In contrast, the MCMC procedure was computationally feasible and robust to initial conditions. We concluded that Bayesian inference via MCMC has an important advantage over SML for high dimensional problems because it does not require search for the optimum of the likelihood.

[1] Currently, approaches to numerical integration such as quadrature and series expansion are not useful if the dimension of the integration is greater than four.

We consider dynamic, stochastic, parametric models with intertemporally additively separable preferences and a finite time horizon. Suppose that in each period $t = 1,...,T$ $(T < \infty)$ each agent chooses among a finite set A_t of mutually exclusive alternatives. Let $\Re^{k_t}$ be the date-t state space, where k_t is a positive integer. Choosing alternative $a_t \in A_t$ in state $I_t \in \Re^{k_t}$ leads to period payoff $R(I_t, a_t; \theta)$, where θ is a finite-vector denoting the model's structural parameters.

The value to choosing alternative a_t in state I_t, denoted by $V_t(I_t, a_t)$, depends on the period payoff and on the way agents expect that choice to affect future payoffs. For instance, in the familiar case when agents have rational expectations, alternative specific values can be expressed:

$$V_t(I_t, a_t) = R(I_t, a_t; \theta) + \delta E_t \max_{a_{t+1} \in A_{t+1}} V_{t+1}(I_{t+1}, a_{t+1} \mid I_t, a_t) \quad (t = 1,...,T) \tag{1.1}$$

$$V_{T+1}(\cdot) \equiv 0 \tag{1.2}$$

$$I_{t+1} = H(I_t, a_t; \theta) \tag{1.3}$$

where δ is the constant rate of time preference, $H(I_t, a_t; \theta)$ is a stochastic law of motion that provides an intertemporal link between choices and states, and E_t is the date-t mathematical expectations operator so that expectations are taken with respect to the true distribution of the state-variables $P_H(I_{t+1} \mid I_t, a_t; \theta)$ as generated by $H(\cdot)$. Individuals choose alternative a_t^* if and only if $V_t(I_t, a_t^*) > V_t(I_t, a_t)\ \forall a_t \in A_t, a_t \neq a_t^*$. See Eckstein and Wolpin (1989) for a description of many alternative structural models that fit into this framework.

The econometrician is interested in drawing inferences about θ, the vector of structural parameters. One econometric procedure to accomplish this (see Rust (1987) or Wolpin (1984)) requires using dynamic programming to solve system (1.1)-(1.3) at many trial parameter vectors. At each parameter vector, the solution to the system is used as input to evaluate a prespecified econometric objective function. The parameter space is systematically searched until a vector that "optimizes" the objective function is found. A potential drawback of this procedure is that, in general, solving system (1.1)-(1.3) with dynamic programming is extremely computationally burdensome. The reason is that the mathematical expectations that appear on the right-hand side of (1.1) are often impossible to compute analytically, and very time-consuming to approximate well numerically. Hence, as a practical matter, this estimation procedure is useful only under very special circumstances (for instance, when there is a small number of state-variables.) Consequently, a literature has arisen that suggests alternative approaches to inference in dynamic multinomial choice models.

Some recently developed techniques for estimation of the system (1.1) – (1.3) focus on circumventing the need for dynamic programming. Several good surveys of this literature already exist, and we will not attempt one here (see Rust (1994)). Instead, we simply note that the idea underlying the more well-known of these approaches, i.e., Hotz and Miller (1993) and Manski (1993), is to use choice and payoff data to draw inferences about the values of the expectations on the right-hand side of (1.1). A key limitation of these procedures is that, in order to learn about expectations, each requires the data to satisfy a strict form of stationarity in order to rule out cohort effects.

The technique proposed by Geweke and Keane (1999a) for structural inference in dynamic multinomial choice models also circumvents the need for dynamic programming. A unique advantage of their method is that it does not require the econometrician to make strong assumptions about the way people form expectations. Moreover, their procedure is not hampered by strong data requirements. It can be implemented when the data includes only partially observed payoffs from a single cohort of agents observed over only part of their life cycle.

To develop the Geweke-Keane approach, it is useful to express the value function (1.1) as:

$$V_t(I_t, a_t) = R(I_t, a_t; \theta) + F^H(I_t, a_t) \tag{1.4}$$

where $F^H(I_t, a_t) \equiv \delta E_t \max_{a_{t+1} \in A_{t+1}} V_{t+1}(a_{t+1}, H(I_t, a_t))$. Geweke and Keane (1999a) observed that the definition of $F^H(\cdot)$, henceforth referred to as the 'future component', makes sense independent of the meaning of E_t. If, as assumed above, E_t is the mathematical expectations operator then $F^H(\cdot)$ is the rational expectations future component. On the other hand, if E_t is the zero operator, then future payoffs do not enter the individuals' decision rules, and $F^H(\cdot)$ is identically zero. In general, the functional form of the future component $F^H(\cdot)$ will vary with the way people form expectations. Unfortunately, in most circumstances the way people form expectations is unknown. Accordingly, the correct specification of the future component $F^H(\cdot)$ is also unknown.

There are, therefore, two important reasons an econometrician may prefer not to impose strong assumptions about the way people form expectations, or, equivalently, on the admissible forms of the future component. First, such assumptions may lead to an intractable econometric model. Second, the econometrician may see some advantage to taking a less dogmatic stance with respect to behaviors about which very little, if any, a-priori information is available.

When the econometrician is either unwilling or unable to make strong assumptions about the way people form expectations, Geweke and Keane (1999a) suggest that the future component $F^H(\cdot)$ be represented by a parameterized flexible functional form such as a high-order polynomial. The resulting value function can be written

$$V_t(I_t, a_t) = R(I_t, a_t; \theta) + F^H(I_t, a_t \mid \pi) \tag{1.5}$$

where π is a vector of polynomial coefficients that characterize expectation formation. Given functional forms for the contemporaneous payoff functions, and under the condition that θ and π are jointly identified, it is possible to draw inferences both about the parameters of the payoff functions and the structure of expectations.

This chapter focuses on an important case in which key structural and expectations parameters are jointly identified. We consider a model where an alternative's payoff is partially observed if and only if that alternative is chosen. In this case, after substitution of a flexible polynomial function for the future component as in (1.5), the model takes on a form similar to a static Roy (1951) model augmented to include influences on choice other than the current payoffs, as in Heckman and Sedlacek (1986). The key difference is that $F^H(\cdot)$ incorporates overidentifying restrictions on the non-payoff component of the value function that are implied by (1.1)-(1.3) and that are not typically invoked in the estimation of static selection models. Specifically, the parameters of the non-payoff component of the value function are constant across alternatives, and the arguments of the non-payoff component vary in a systematic way across alternatives that is determined by the law of motion $H(\cdot)$ for the state variables.

The structural model (1.1)-(1.3) also implies restrictions on the nature of the future component's arguments. For instance, if $H(\cdot)$ and $R(\cdot)$ jointly imply that the model's payoffs are path-independent, then the future component should be specified so that path-dependent expectation formation is ruled out.[2] Similarly, contemporaneous realizations of serially independent stochastic variables contain no information relevant for forecasting future outcomes, so they should not enter the arguments of the flexible functional form. Without such coherency conditions one might obtain results inconsistent with the logic of the model's specification.

A finite order polynomial will in general provide only an approximation to the true future component. Hence, it is important to investigate the extent to which misspecification of the future component may affect inference for the model's structural parameters. Below we report the outcome of some Monte Carlo experiments that shed light on this issue. The experiments are conducted under both correctly and incorrectly specified future components. We find that the Geweke-Keane approach performs extremely well when $F^H(\cdot)$ is correctly specified, and still very well under a misspecified future component. In particular, we find that assuming the future component is a polynomial when it is actually generated by rational expectations leads to only "second order" difficulties in two senses. First, it has a small effect on inferences with regard to the structural parameters of the payoff functions.[3] Second, the decision rules inferred from the data in the misspecified model are very close to the optimal rule in the sense that agents using the suboptimal rule incur 'small' lifetime payoff losses.

---

[2] Restrictions of this type can be tested easily by estimating versions of the model with different but nested future components.

[3] These findings are related to those of Lancaster (1996), who considered Bayesian inference in the stationary job search model. He found that if the future component is treated as a free parameter (rather than being set "optimally" as dictated by the offer wage function, offer arrival rate, unemployment benefit and discount rate) there is little loss of information about the structural parameters of the offer wage functions. (As in our example, however, identification of the discount factor is lost.) The stationary job search model considered by Lancaster (1996) has the feature that the future component is a constant (i.e. it is not a function of state variables). Our procedure of treating the future component as a polynomial in state variables can be viewed as extending Lancaster's approach to a much more general class of models.

The remainder of this chapter is organized as follows. Section two describes the application, and section three details the Gibbs sampling algorithm. Section four reviews our experimental design and results, and section five concludes.

2. The Dynamic Multinomial Choice Model

In this section we present an example of Bayesian inference for dynamic discrete choice models using the Geweke-Keane method of replacing the future component of the value function with a flexible polynomial function. The discussion is based on a model that is very similar to ones analyzed by Keane and Wolpin (1994, 1997).

In the model we consider, $i = 1, ..., N$ agents choose among $j = 1, ..., 4$ mutually exclusive alternatives in each of $t = 1, ..., 40$ periods. One can think of the first two alternatives as work in one of two occupations, the third as attending school and the fourth alternative as remaining home. One component of the current period payoff in each of the two occupational alternatives is the associated wage, w_{ijt} $(j = 1, 2)$. The log-wage equation is:

$$\begin{aligned} \ln w_{ijt} &= \beta_{0j} + \beta_{1j} X_{i1t} + \beta_{2j} X_{i2t} + \beta_{3j} S_{it} + \beta_{4j} X_{ijt}^2 + \varepsilon_{ijt} \quad (j = 1, 2) \\ &= Y_{ijt}' \beta_j + \varepsilon_{ijt} \quad (j = 1, 2) \end{aligned} \tag{2.1}$$

where Y_{ijt} is the obvious vector, $\beta_j = \left(\beta_{0j}, ..., \beta_{4j}\right)'$, S_{it} is the periods of school completed, $(X_{ijt})_{j=1,2}$ is the periods of experience in each occupation *j*, and the ε_{ijt} are serially independent productivity shocks, with $(\varepsilon_{i1t}, \varepsilon_{i2t})' \sim N(0, \Sigma_\varepsilon)$. Each occupational alternative also has a stochastic nonpecuniary payoff, v_{ijt}, so the complete current period payoffs are

$$u_{ijt} = w_{ijt} + v_{ijt}, \; (j = 1, 2). \tag{2.2}$$

The schooling payoffs include tuition costs. Agents begin with a tenth-grade education, and may complete two additional grades without cost. We assume there is a fixed undergraduate tuition rate α_1 for attending grades 13 through 16, and a fixed graduate tuition rate α_2 for each year of schooling beyond 16. We assume a "return to school" cost α_3 that agents face if they did not choose school the previous period. Finally, school has a nonstochastic, nonpecuniary benefit α_0 and a mean zero stochastic nonpecuniary payoff v_{i3t}. Thus we have

$$u_{i3t} = \alpha_0 + \alpha_1 \chi(12 \le S_{it} \le 15) + \alpha_2 \chi(S_{it} \ge 16) + \alpha_3 \chi(d_{i,t-1} \ne 3) + v_{i3t} \equiv \Lambda_{it}' \alpha + v_{i3t} \tag{2.3}$$

where χ is an indicator function that takes value one if the stated condition is true and is zero otherwise, Λ_{it} is a vector of zeros and ones corresponding to the values of the indicator functions, $\alpha = \left(\alpha_0, ..., \alpha_3\right)'$, $d_{it} \in \{1, 2, 3, 4\}$ denotes the choice of i at *t*. Lastly, we assume that option four, home, has both a nonstochastic nonpecuniary payoff ϕ and a stochastic nonpecuniary payoff v_{ijt}, so

$$u_{i4t} = \phi + v_{i4t}. \tag{2.4}$$

We will set $u_{ijt} = \bar{u}_{ijt} + v_{ijt}$, $(j = 1,...,4)$. The nonpecuniary payoffs $\left(v_{ijt}\right)_{j=1,4}$ are assumed serially independent.

The state of the agent at the time of each decision is

$$I_{it} = \{(X_{ijt})_{j=1,2}, S_{it}, t, d_{i,t-1}, (\varepsilon_{ijt})_{j=1,2}, (v_{ijt})_{j=1,..4}\}. \tag{2.5}$$

We assume $d_{i0} = 3$. The laws of motion for experience in the occupational alternatives and school are: $X_{ij,t+1} = X_{ijt} + \chi(d_{it} = j),\ j = 1, 2,\ S_{i,t+1} = S_{it} + \chi(d_{it} = 3).$ The number of 'home' choices is excluded from the state-space as it is linearly dependent on the level of education, the period, and experience in the two occupations.

It is convenient to have notation for the elements of the state vector whose value in period *t+1* depends nontrivially on their value in period *t* or on the current choice. The reason, as we note below, is that these elements are the natural arguments of the future component. We define

$$I^*_{it} = \{(X_{ijt})_{j=1,2}, S_{it}, t, d_{i,t-1}\}.$$

The value of each alternative is the sum of its current period payoff, the stochastic non-pecuniary payoff and the future component:

$$\begin{aligned} V_{ijt}(I_{it}) &= \bar{u}_{ijt}(I_{it}) + v_{ijt} \\ &\quad + F(X_{i1t} + \chi(j=1), X_{i2t} + \chi(j=2), S_{it} + \chi(j=3), \mathrm{t}+1, \chi(j=3)) \quad (j=1,\ldots4)\ (t=1,\ldots,40) \\ &\equiv \bar{u}_{ijt}(I_{it}) + v_{ijt} + F(I^*_{it}, j) \end{aligned} \tag{2.6}$$

The function F represents agents' forecasts about the effects of their current state and choice on their future payoff stream. The function is fixed across alternatives, implying that forecasts vary across alternatives only because different choices lead to different future states, and it depends only on the choice and the state variables in I^*_{t+1}.[4]

Since choices depend only on relative alternative values, rather than their levels, we define for $j \in \{1, 2, 3\}$:

$$\begin{aligned} Z_{ijt} &\equiv V_{ijt} - V_{i4t} \\ &= \bar{u}_{ijt} + v_{ijt} + F(I^*_{it}, j) - \bar{u}_{i4t} - v_{i4t} - F(I^*_{it}, 4) \\ &= \tilde{u}_{ijt} + f(I^*_{it}, j) + \eta_{ijt} \end{aligned} \tag{2.7}$$

where $\tilde{u}_{ijt} \equiv \bar{u}_{ijt} - \bar{u}_{i4t}$, $\{\eta_{ijt}\}_{j=1,2,3} \equiv (v_{ijt} - v_{i4t})_{j=1,2,3} \sim N(0, \Sigma_\eta)$ and $f(I^*_{it}, j) = F(I^*_{it}, j) - F(I^*_{it}, 4)$. Importantly, after differencing, the value ϕ of the home payoff is subsumed in f the relative future component. Clearly, if an alternative's future component has an intercept (as each of ours does) then it and the period return to home cannot be separately identified.

[4] As noted earlier, the future component's arguments reflect restrictions implied by the model. For instance, because the productivity and preference shocks are serially independent, they contain no information useful for forecasting future payoffs and do not appear in the future component's arguments. Also, given total experience in each occupation, the order in which occupations one and two were chosen in the past does not bear on current or future payoffs. Accordingly, only total experience in each occupation enters the future component.

The value function differences Z_{it} are latent variables unobserved by the econometrician. The econometrician only observes the agents' choices $\{d_{it}\}$ for t=1,…,40, and, in the periods when the agent works, the wage for the chosen alternative. Thus, payoffs are never completely observed, both because wages are censored and because the nonpecuniary components of the payoffs (v_{ijt}) are never observed. Nevertheless, given observed choices and partially observed wages, along with the functional form assumptions about the payoff functions, it is possible to learn both about the future component $F(\cdot)$ and the structural parameters of the payoff functions without making strong assumptions about how agents form expectations. Rather, we simply assume that the future component lies along a fourth-order polynomial in the state variables. After differencing to obtain $\{f(I_{it}^{*}, j)\}_{j=1,2,3}$, the polynomial contained 53 terms of order three and lower (see Appendix A.) We express the future component as

$$f(I_{it}^{*}, j) = \psi_{ijt}^{'}\pi \qquad (j = 1,2,3) \tag{2.8}$$

where ψ_{ijt} is a vector of functions of state-variables that appear in the equation for $f(I_{it}^{*}, j)$ and π is a vector of coefficients common to each choice. Cross-equation restrictions of this type are a consequence of using the same future component function *F* for each alternative and reflect the consistency restrictions discussed earlier.

3. Implementing the Gibbs Sampling Algorithm

Bayesian analysis of this model entails deriving the joint posterior distribution of the model's parameters and unobserved variables. Recall that the value function differences $\mathbf{Z} = \left\{(Z_{ijt})_{j=1,2,3;i=1,N;t=1,40}\right\}$ are never observed, and that wages $\mathbf{W} = \left\{(w_{ijt})_{j=1,2,3;i=1,N;t=1,40}\right\}$ are only partially observed. Let $\mathbf{W}_1$ and $\mathbf{W}_2$ denote the set of observed and unobserved wages, respectively, and let $Y = \left\{y_{ijt}\right\}_{i=1,N;\,j=1,2;t=1,40}$ denote the log wage equation regressors. Then the joint posterior density is $p(\mathbf{W}_2, \mathbf{Z}, \beta_1, \beta_2, \alpha, \pi, \Sigma_{\varepsilon}^{-1}, \Sigma_{\eta}^{-1} \mid \mathbf{W}_1, Y, \Lambda)$. By Bayes' law, this density is proportional to

$$p(\mathbf{W}, \mathbf{Z} \mid Y, \Lambda, \beta_1, \beta_2, \alpha, \pi, \Sigma_{\varepsilon}^{-1}, \Sigma_{\eta}^{-1}) \cdot p(\beta_1, \beta_2, \alpha, \pi, \Sigma_{\varepsilon}^{-1}, \Sigma_{\eta}^{-1}) \tag{3.1}$$

The first term in (3.1) is the so-called "complete data" likelihood function. It is the likelihood function that could be formed in the hypothetical case that we had data on N individuals observed over 40 periods each, and we observed all of the value function differences $\mathbf{Z}$ and the complete set of wages $\mathbf{W}$ for all alternatives. This is:

$$\begin{aligned} p(\mathbf{W}, \mathbf{Z} \mid Y, \Lambda, \beta_1, \beta_2, \alpha, \pi, \Sigma_{\varepsilon}^{-1}, \Sigma_{\eta}^{-1}) \propto \prod_{i,t} |\Sigma_{\varepsilon}^{-1}|^{1/2} (w_{i1t} w_{i2t})^{-1} \exp\left\{-\frac{1}{2}\begin{pmatrix} \ln w_{i1t} - Y_{i1t}^{'}\beta_1 \\ \ln w_{i2t} - Y_{i2t}^{'}\beta_2 \end{pmatrix}^{'} \Sigma_{\varepsilon}^{-1} \begin{pmatrix} \ln w_{i1t} - Y_{i1t}^{'}\beta_1 \\ \ln w_{i2t} - Y_{i2t}^{'}\beta_2 \end{pmatrix}\right\} \\ \cdot |\Sigma_{\eta}^{-1}|^{1/2} \exp\left\{-\frac{1}{2}\begin{pmatrix} Z_{i1t} - w_{i1t} - \Psi_{i1t}^{'}\pi \\ Z_{i2t} - w_{i2t} - \Psi_{i2t}^{'}\pi \\ Z_{i3t} - \Lambda_{it}^{'}\alpha - \Psi_{i3t}^{'}\pi \end{pmatrix}^{'} \Sigma_{\eta}^{-1} \begin{pmatrix} Z_{i1t} - w_{i1t} - \Psi_{i1t}^{'}\pi \\ Z_{i2t} - w_{i2t} - \Psi_{i2t}^{'}\pi \\ Z_{i3t} - \Lambda_{it}^{'}\alpha - \Psi_{i3t}^{'}\pi \end{pmatrix}\right\} \\ \cdot \chi(Z_{ijt} > 0, Z_{ikt}(k \neq j) \text{ if } d_{it} = j \text{ and } j \in \{1,2,3\}, \{Z_{ijt}\}_{j=1,2,3} < 0 \text{ otherwise}) \end{aligned} \tag{3.2}$$

The second term in (3.1) is the joint prior distribution. We assume flat priors on all parameters except the two precision matrices, for which we assume the standard noninformative priors (see Zellner, (1971), section 8.1):

$$p(\Sigma_{\varepsilon}^{-1}) \propto |\Sigma_{\varepsilon}^{-1}|^{-3/2},\ p(\Sigma_{\eta}^{-1}) \propto |\Sigma_{\eta}^{-1}|^{-2} \tag{3.3}$$

The Gibbs sampler draws from a density that is proportional to the product of (3.2) and the two densities in (3.3).

The Gibbs sampling algorithm is used to form numerical approximations of the parameters' marginal posterior distributions. It is not feasible to construct these marginal posteriors analytically, since doing so requires high dimensional integrations over unobserved wages and value function differences. Implementing the Gibbs sampling algorithm requires us to factor the joint posterior defined by (3.1)-(3.3) into a set of conditional posterior densities, in such a way that each can be drawn from easily. Then, we cycle through these conditionals, drawing a block of parameters from each in turn. As the number of cycles grows large, the parameter draws so obtained converge in distribution to their respective marginal posteriors, given certain mild regularity conditions (see Tierney (1994) for a discussion of these conditions). An important condition is that the posterior distribution be finitely integrable, which we verify for this model in Appendix B. Given the posterior distribution of the parameters, conditional on the data, the investigator can draw exact finite sample inferences.

Our Gibbs sampling-data augmentation algorithm consists of six steps or 'blocks.' These steps, which we now briefly describe, are cycled through repeatedly until convergence is achieved.

Step 1. Draw value function differences $\{Z_{ijt}, i=1,N; j=1,3; t=1,40\}$

Step 2. Draw unobserved wages $\{w_{ijt} \text{ when } d_{it} \neq j,\ (j=1,2)\}$

Step 3. Draw the log-wage equation coefficients β_j.

Step 4. Draw the log-wage equation error-covariance matrix Σ_{ε}.

Step 5. Draw the parameters of the future component π and school payoff parameters α.

Step 6. Draw the nonpecuniary payoff covariance matrix Σ_{η}.

<u>Step 1:</u> We chose to draw the $\{Z_{ijt}, i=1,N; j=1,3; t=1,40\}$ one-by-one. Taking everything else in the model as given, it is evident from (3.1)-(3.3) that the conditional distribution of a single Z_{ijt} is truncated Gaussian. Dealing with the truncation is straightforward. There are three ways in which the Gaussian distribution might be truncated.

Case 1: Z_{ijt} is the value function difference for the chosen alternative. Thus, we draw $Z_{ijt} > \max\left\{0, (Z_{ikt})_{\substack{k\in\{1,2,3\}\\ k\neq j}}\right\}$.

Case 2: Z_{ijt} is not associated with the chosen alternative, and 'home' was not chosen. Thus, we draw $Z_{ijt} < Z_{id_{it}t}$.

Case 3: 'Home' was chosen. In this case, we draw $Z_{ijt} < 0$.

We draw from the appropriate univariate, truncated Gaussian distributions using standard inverse CDF methods.

Step 2: We chose to draw the unobserved wages $\{w_{ijt} \text{ when } d_{it} \neq j,\ (j = 1,2,3)\}$ one-by-one. Suppose w_{i1t} is unobserved. Its density, conditional on every other wage, future component difference and parameter being known, is from (3.1), (3.2) and (3.3) evidently given by:

$$g(w_{i1t} \mid \cdot) \propto \frac{1}{\tilde{w}_{i1t}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} \ln w_{i1t} - Y_{i1t}'\beta_1 \\ \ln w_{i2t} - Y_{i2t}'\beta_2 \end{pmatrix}' \Sigma_\varepsilon^{-1} \begin{pmatrix} \ln w_{i1t} - Y_{i1t}'\beta_1 \\ \ln w_{i2t} - Y_{i2t}'\beta_2 \end{pmatrix} \right\} \exp\left\{ -\frac{1}{2} \begin{pmatrix} Z_{i1t} - w_{i1t} - \Psi_{i1t}'\pi \\ Z_{i2t} - w_{i2t} - \Psi_{i2t}'\pi \\ Z_{i3t} - \Lambda_{it}'\alpha - \Psi_{i3t}'\pi \end{pmatrix}' \Sigma_\eta^{-1} \begin{pmatrix} Z_{i1t} - w_{i1t} - \Psi_{i1t}'\pi \\ Z_{i2t} - w_{i2t} - \Psi_{i2t}'\pi \\ Z_{i3t} - \Lambda_{it}'\alpha - \Psi_{i3t}'\pi \end{pmatrix} \right\} \quad (3.4)$$

This distribution is nonstandard as wages enter in both logs and levels. Nevertheless, it is straightforward to sample from this distribution using rejection methods (see Geweke (1995) for a discussion of efficient rejection sampling). In brief, we first drew a candidate wage w^c from the distribution implied by the first exponential of (3.4), so that $\ln w^c \sim N(Y_{i1t}'\beta_1 + \lambda_{it}, \sigma_*^2)$, where $\lambda_{it} \equiv \Sigma_\varepsilon(1,2)\varepsilon_{i2t} / \Sigma(2,2)$ and $\sigma_*^2 \equiv \Sigma_\varepsilon(1,1)(1-(\Sigma_\varepsilon(1,2)^2/(\Sigma_\varepsilon(1,1)\Sigma_\varepsilon(2,2))$. This draw is easily accomplished, and w^c is found by exponentiating. The probability with which this draw is accepted is found by dividing the second exponential in (3.4) by its conditional maximum over w_{i1t} and evaluating the resulting expression at $w_{i1t} = w^c$. If the draw is accepted then the unobserved w_{i1t} is set to w^c. Otherwise, the process is repeated until a draw is accepted.

Step 3: Given all wages, value function differences, and other parameters, the density of (β_1, β_2) is:

$$g(\beta_1, \beta_2) \propto \exp\left\{ -\frac{1}{2} \begin{pmatrix} \ln w_{i1t} - Y_{i1t}'\beta_1 \\ \ln w_{i2t} - Y_{i2t}'\beta_2 \end{pmatrix}' \Sigma_\varepsilon^{-1} \begin{pmatrix} \ln w_{i1t} - Y_{i1t}'\beta_1 \\ \ln w_{i2t} - Y_{i2t}'\beta_2 \end{pmatrix} \right\} \quad (3.5)$$

So that (β_1, β_2) is distributed according to a multivariate normal. In particular, it is easy to show that

$$\beta \sim N[(\mathbf{Y}'\Sigma^{-1}\mathbf{Y})^{-1}\mathbf{Y}'\Sigma^{-1}\ln \mathbf{W}, (\mathbf{Y}'\Sigma^{-1}\mathbf{Y})^{-1}]$$

where $\beta \equiv (\beta_1', \beta_2')'$, $\Sigma = \Sigma_\varepsilon \otimes I_{NT}$, $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 & 0 \\ 0 & \mathbf{Y}_2 \end{bmatrix}$ and $\ln \mathbf{W} = [\ln \mathbf{W}_1', \ln \mathbf{W}_2']'$, where $\mathbf{Y}_1$ is the regressor matrix for the first log-wage equation naturally ordered through all individuals and periods, and similarly for $\mathbf{Y}_2, \mathbf{W}_1$ and $\mathbf{W}_2$. It is straightforward to draw β from this multivariate normal density.

Step 4: With everything else known Σ_ε^{-1} has a Wishart distribution. Specifically, it is immediate from the joint posterior that $p(\Sigma_\varepsilon^{-1}) \propto |\Sigma_\varepsilon^{-1}|^{\frac{NT-3}{2}} \exp\left\{-\frac{1}{2} tr(S(\beta)\Sigma_\varepsilon^{-1})\right\}$, so that

$$\Sigma_\varepsilon^{-1} \sim W(S(\beta), NT), \text{ where } S(\beta) = (\ln \mathbf{W}_1 - \mathbf{Y}_1\beta_1, \ln \mathbf{W}_2 - \mathbf{Y}_2\beta_2)'(\ln \mathbf{W}_1 - \mathbf{Y}_1\beta_1, \ln \mathbf{W}_2 - \mathbf{Y}_2\beta_2). \quad (3.6)$$

It is easy to draw from the Wishart and then invert the 2 X 2 matrix to obtain Σ_ε.

Step 5: It is convenient to draw both the future component π parameters and the parameters α of the school payoff jointly. Since the future component for school contains an intercept, it and the constant in Λ cannot be separately identified. Hence, we omit α_0 as well as the first row from each Λ_{it}. Define the vector $\pi^* \equiv [\pi', \alpha']'$, where $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$. and define $\Psi^*_{ijt} \equiv [\Psi'_{ijt}, 0'_3]'$ $(j = 1, 2)$, and $\Psi^*_{i3t} = [\Psi'_{i3t}, \Lambda'_{it}]'$. Note that π^* and the Ψ^*_{ijt} are 56-vectors. Then define $\Psi_k = [\Psi^*_{1k1}\, \Psi^*_{1k2} \ldots \Psi^*_{Nk,T-1}\, \Psi^*_{NkT}]$, and set $\Psi = [\Psi_1\, \Psi_2\, \Psi_3]'$, so that Ψ is a ($3 \cdot$ NT x 56) stacked-regressor matrix. Similarly, define the corresponding $3 \cdot$ NT-vector Γ by $\Gamma = (\{Z_{i1t} - w_{i1t}\}'_{i,t}, \{Z_{i2t} - w_{i2t}\}'_{i,t}, \{Z_{i3t}\}'_{i,t})'$. It is immediate from (3.2), in which π^* enters only through the second exponential expressions that, conditional on everything else known, π^* has a multivariate normal density given by:

$$\pi^* \sim N[(\Psi'\Omega^{-1}\Psi)^{-1}\Psi'\Omega^{-1}\Gamma, (\Psi'\Omega^{-1}\Psi)^{-1}] \tag{3.7}$$

where $\Omega = \Sigma_\eta \otimes I_{NT}$. We draw from this using a standard, multivariate normal random number generator.

Step 6: With everything else known the distribution of Σ_η^{-1} is Wishart; $\Sigma_\eta^{-1} \sim W(SST_\eta, NT)$, where $SST_\eta = \sum_{i,t} (\eta_{i1t}\, \eta_{i2t}\, \eta_{i3t})'(\eta_{i1t}\, \eta_{i2t}\, \eta_{i3t})$, and with the η_{ijt} defined by (2.7). It is easy to draw from this distribution and then invert the 3 X 3 matrix to obtain Σ_η.

4. Experimental Design and Results

This section details the design and results of a Monte Carlo experiment that we conducted to shed light on the performance of the Gibbs sampling algorithm discussed in section two. We generated data according to equations (2.1) – (2.7), using the true parameter values that are listed in column two of Table 3. The table does not list the discount rate and the intercepts in the school and home payoff functions, which were set to 0.95, 11000 and 17000 respectively, since these are not identified. In all of our experiments we set the number of people, N, to 2000.

Data from this model were generated using two different assumptions about the way people formed expectations. First, we assumed that people had rational expectations. This required us to solve the resulting dynamic optimization problem once to generate the optimal decision rules. Since the choice set includes only four discrete alternatives it is feasible to do this. Then, to simulate choice and wage paths requires only that we generate realizations of the appropriate stochastic variables. It is important to note that the polynomial future component used in the estimation procedure does not provide a perfect fit to the rational expectations future component. Hence, analysis of this data sheds light on the effect that misspecification of the future component may have on inference.

Next, we assumed that agents used a future component that was actually a polynomial in the state variables to form decisions. Analysis of this data set sheds light on how the algorithm performs when the model is correctly specified. To ensure close comparability with the rational expectations case, we constructed this polynomial by regressing the rational expectations future components on a fourth-order polynomial in the state-variables, constructed as described in the discussion preceding (2.8). We used the point estimates from this regression as the coefficients of our polynomial future component (see Appendix A for the specific form of the polynomial).

We found that a fourth order polynomial provided a good approximation to the future component in the sense that if agents used the approximate instead of optimal decision rule they suffered rather small lifetime earnings losses. Evidence of this is given in Table 1, where we report the results of simulations under optimal and suboptimal decision rules. The simulations were conducted as follows. First, for N=2000 people we drew five sets of lifetime (T=40) realizations of the model's stochastic components $\{\varepsilon_{i1t}, \varepsilon_{i2t}, (\eta_{ijt})_{j=1,4}\}$. In Table 1 these are referred to as error sets one to five. For each of the five error sets we simulated lifetime choice histories for each of the 2000 people under the optimal and approximate decision rules. We refer to the 10 data sets constructed in this way as 1-EMAX through 5-EMAX and 1-POLY through 5-POLY, respectively. We then calculated the mean of the present value of lifetime payoffs (pecuniary plus nonpecuniary) for each of the 2000 people under the optimal and approximate decision rules, respectively, for each of the five error sets. These are reported in the second and third rows of Table 1. Holding the error set fixed, the source of any difference in the mean present value of lifetime payoffs lies in the use of different decision rules. The mean present values of dollar equivalent losses from using the suboptimal polynomial rules are small, ranging from 287 to 491. The percentage loss ranges from 8 hundredths of one percent to 14 hundredths of one percent. These findings are similar to those reported by Geweke and Keane (1999a) and Krusell and Smith (1995).

Table 2 reports the mean accepted wages and choice frequencies for the data generated from error-set two. The first set of columns report statistics for data generated according to the polynomial approximation (data set 2-POLY) while the second set of columns report results from the optimal decision rule (data set 2-EMAX). Under our parameterization, occupation one can be thought of "unskilled" labor, while occupation two can be understood as "skilled" labor. The reason is the mean of the wage offer distribution is lower in occupation two early in life, but it rises more quickly with experience. The choice patterns and mean accepted wages are similar under the two decision rules. School is chosen somewhat more often under the optimal decision rule, which helps to generate slightly higher lifetime earnings. Finally, note that selection effects leave the mean accepted wage in occupation two higher than that in occupation one throughout the life-cycle under both decision rules.

Next, we ran the Gibbs algorithm described in section two for 40,000 cycles on each data set. We achieved about three cycles per minute on a Sun ultra-2 workstation.[5] Thus, while time requirements were substantial, they

[5] To begin the Gibbs algorithm we needed an initial guess for the model's parameters (although the asymptotic behavior of the Gibbs sampler as the number of cycles grows large is independent of starting values). We chose to set the log-wage equation β's equal to the value from an OLS regression on observed wages. The diagonal elements of Σ_ε were set to the variance of observed log-wages, while the off-diagonal elements were set to zero. The school payoff parameters were all initialized at zero. All of the future component's π values were also started at zero, with the exception of the alternative-specific intercepts. The intercepts for alternatives one, two and three

were minor compared to what estimation of such a model using a full solution of the dynamic programming problem would entail. Visual inspection of graphs of the draw sequences, as well as application of the split sequence diagnostic suggested by Gelman (1996)-which compares variability of the draws across subsequences-suggests that the algorithm converged for all ten artificial data sets. In all cases, the final 15,000 draws from each run were used to simulate the parameters' marginal posterior distributions.

Table 3 reports the results of the Gibbs sampling algorithm when applied to the data generated with a polynomial future component. In this case, the econometric model is correctly specified. The first column of Table 3 is the parameter label, the second column is the true value, and the remaining columns report the structural parameters' posterior means and standard deviations for each of the five data sets.[6] The results are extremely encouraging. Across all runs, there was only one instance in which the posterior mean of a parameter for the first wage equation was more than two posterior standard deviations away from its true value: the intercept in data set one. In data sets four and five, all of the structural parameters' posterior means are within two posterior standard deviations of their true values. In the second data set, only the second wage equation's own experience term is slightly more than two posterior standard deviations from its true value. In the third data set the mean of the wage equation's error correlation is slightly more than two posterior standard deviations from the true value, as are a few of the second wage equation's parameters.

Careful examination of Table 3 reveals that the standard deviation of the nonpecuniary payoff was the most difficult parameter to pin down. In particular, the first two moments of the marginal posteriors of these parameters vary considerably across experiments, in relation to the variability of the other structural parameters' marginal posteriors. This result reflects earlier findings reported by Geweke and Keane (1999a). In the earlier work they found that relatively large changes in the value of the nonpecuniary component's standard deviation had only a small effect on choices. It appears that this is the case in the current experiment as well.

It is interesting to note that an OLS regression of accepted (observed) log-wages on the log-wage equation's regressors yields point estimates that differ sharply from the results of the Gibbs sampling algorithm. Table 4 contains point estimates and standard errors from such an accepted wage regression. Selection bias is apparent in the estimates of the log-wage equation's parameters in all data sets. This highlights that the Bayesian simulation algorithm is doing an impressive job of implementing the appropriate dynamic selection correction.

Perhaps more interesting is the performance of the algorithm when taken to data that was generated using optimal decision rules. Table 5 reports the results of this analysis on data sets 1-EMAX to 5-EMAX. Again, the

---

were initialized at –5,000, 10,000, and 20,000, respectively. These values were chosen with an eye towards matching aggregate choice frequencies in each alternative. We initialized the Σ_η covariance matrix by setting all off-diagonal elements to zero, and each diagonal element to $5X10^8$. We used large initial variances because doing so increases the size of the initial Gibbs steps, and seems to improve the rate of convergence of the algorithm.

first column labels the parameter, and the second contains its data generating value. The performance of the algorithm is quite impressive. In almost all cases, the posterior means of the wage function parameters deviate only slightly from the true values in percentage terms. Also, the posterior standard deviations are in most cases quite small, suggesting that the data contain a great deal of information about these structural parameters – even without imposing the assumption that agents form the future component "optimally." Finally, despite the fact that the posterior standard deviations are quite small, the posterior means are rarely more than two posterior deviations away from the true values.[7] As with the polynomial data, the standard deviation of the nonpecuniary component seems difficult to pin down. Unlike the polynomial data, the school payoff parameters are not pinned down as well as the wage equation parameters. This is perhaps not surprising since school payoffs are never observed.

Figure 1 contains the simulated posterior densities for a subset of the structural parameters based on data set 3-EMAX. Each figure includes three triangles on its horizontal axis. The middle triangle defines the posterior mean, and the two flanking triangles mark the points two posterior standard deviations above and below the mean. A vertical line is positioned at the parameters' data generating (true) value. These distributions emphasize the quality of the algorithm's performance in that the true parameter values are typically close to the posterior means. The figures also make clear that not all the parameters have approximately normal distributions. For instance, the posterior density of the wage equations' error correlation is multi-modal.

The results of Table 5 indicate that in a case where agents form the future component optimally, we can still obtain reliable and precise inferences about structural parameters of the current payoff functions using a simplified and misspecified model that says the future component is a simple fourth-order polynomial in the state variables. But we are also interested in how well our method approximates the decision rule used by the agents. In Table 6 we consider an experiment in which we use the posterior means for the parameters π that characterize how agents form expectations to form an estimate of agents' decision rules. We then simulate five new artificial data sets, using the exact same draws for the current period payoffs as were used to generate the original five artificial data sets. The only difference is that the estimated future component is substituted for the true future component in forming the decision rule. The results in Table 6 indicate that the mean wealth losses from using the estimated decision rule range from five-hundredths to three-tenths of one percent. The percentage of choices that agree between agents who use the optimal versus the approximate rules ranges from 89.8% to 93.5%. These results suggest that our estimated polynomial approximations to the optimal decision rules are reasonably accurate.

---

[6] Space considerations prevent us from reporting results for individual expectations parameters. Instead, below we will graphically compare the form of the estimated future component to that which was used to generate the data.

[7] We also ran OLS accepted log-wage regressions for the 1-EMAX through 5-EMAX data sets. The results are very similar to those in Table 4, so we do not report them here. The estimates again show substantial biases for all the wage equation parameters. Thus, the Gibbs sampling algorithm continues to do an impressive job of implementing a dynamic selection correction despite the fact that the agents' decision rules are misspecified.

Figure 2 provides an alternative way to examine the quality of the polynomial approximation to the future component. This figure plots the value of the approximate and the true EMAX future components when evaluated at the mean of each period's state-vector.[8] Each vertical axis corresponds to the value of the future component, and the horizontal axis is the period. Clearly, the approximation reflects the true EMAX future component's main features. The fit of the polynomial seems relatively strong for each occupational alternative throughout the life-cycle. The fit is good for school in early periods, but begins to deteriorate later. One reason is that school is chosen very infrequently after the first five periods, so there is increasingly less information about its future component. A second reason is that the contemporaneous return begins to dominate the future component in alternative valuations. Consequently, each data point in latter periods contains relatively less information about the future component's value. Overall, however, these figures furnish additional evidence that the polynomial approximation does a reasonable job of capturing the key characteristics of the true future component.

5. Conclusion

This chapter described how to implement a simulation based method for inference that is applicable to a wide class of dynamic multinomial choice models. The results of a Monte Carlo analysis demonstrated that the method works very well in relatively large state-space models with only partially-observed payoffs where very high dimensional integrations are required. Although our discussion focused on models with discrete choices and independent and identically distributed stochastic terms, the method can also be applied to models with mixed continuous/discrete choice sets and serially correlated shocks (see Houser, (1999)).

Appendix A. The future component

The future component we used was a fourth-order polynomial in the state variables. Below, in the interest of space and clarity, we will develop that polynomial only up to its third order terms. The extension to the higher order terms is obvious. From equation (2.6), the future component is the flexible functional form

$$F(X_{i1t} + \chi(j=1),\ X_{i2t} + \chi(j=2),\ S_{it} + \chi(j=3), \mathrm{t}+1, \chi(j=3))$$

Define $\iota_k \equiv \chi(j=k)$. Then, to third order terms, we used the following polynomial to represent this function.

$$\begin{aligned}
&F(X_1+\iota_1,\ X_2+\iota_2,\ S+\iota_3,\ \mathrm{t}+1,\ \iota_3) = P_1 + P_2(X_1+\iota_1) + P_3(X_2+\iota_2) + P_4(S+\iota_3) + P_5(t+1) \\
&+P_6(X_1+\iota_1)^2 + P_7(X_2+\iota_2)^2 + P_8(S+\iota_3)^2 + P_9(t+1)^2 + P_{10}(X_1+\iota_1)^3 + P_{11}(X_2+\iota_2)^3 + P_{12}(S+\iota_3)^3 + P_{13}(t+1)^3 \\
&+P_{14}(X_1+\iota_1)^2(X_2+\iota_2) + P_{15}(X_1+\iota_1)^2(S+\iota_3) + P_{16}(X_1+\iota_1)^2(t+1) + P_{17}(X_2+\iota_2)^2(X_1+\iota_1) \\
&+P_{18}(X_2+\iota_2)^2(S+\iota_3) + P_{19}(X_2+\iota_2)^2(t+1) + P_{20}(S+\iota_3)^2(X_1+\iota_1) + P_{21}(S+\iota_3)^2(X_2+\iota_2) + P_{22}(S+\iota_3)^2(t+1) \\
&+P_{23}(t+1)^2(X_1+\iota_1) + P_{24}(t+1)^2(X_2+\iota_2) + P_{25}(t+1)^2(S+\iota_3) + P_{26}\iota_3 + P_{27}\iota_3(X_1+\iota_1) + P_{28}\iota_3(X_2+\iota_2) \\
&+P_{29}\iota_3(S+\iota_3) + P_{30}\iota_3(t+1) + P_{31}\iota_3(X_1+\iota_1)^2 + P_{32}\iota_3(X_2+\iota_2)^2 + P_{33}\iota_3(S+\iota_3)^2 + P_{34}\iota_3(t+1)^2
\end{aligned}$$

[8] The mean state vectors were derived from the choices in data set 5-EMAX, and the coefficients of the polynomial were valued at the posterior means derived from the analysis of data set 5-EMAX.

The differenced future components used above are defined by $f(I_{it}^*, j) = F(I_{it}^*, j) - F(I_{it}^*, 4)$. Several of the parameters of the level future component drop out due to differencing. For instance, the intercept P_1 and all coefficients of terms involving only *(t+1)* vanish. Simple algebra reveals the differenced future components have the following forms.

$$f(I_{it}^*,1) = \pi_1 + \pi_2 g(X_1) + \pi_3 h(X_1) + \pi_4 X_2 g(X_1) + \pi_5 Sg(X_1) + \pi_6 (t+1)g(X_1) + \pi_7 X_2^2 + \pi_8 S_2^2 + \pi_9 (t+1)^2.$$
$$f(I_{it}^*,2) = \pi_4 X_1^2 + \pi_7 X_1 g(X_2) + \pi_{10} + \pi_{11} g(X_2) + \pi_{12} h(X_2) + \pi_{13} Sg(X_2) + \pi_{14} (t+1) g(X_2) + \pi_{15} S^2 + \pi_{16} (t+1)^2.$$
$$f(I_{it}^*,3) = \pi_5 X_1^2 + \pi_8 X_1 g(S) + \pi_{13} X_2^2 + \pi_{15} X_2 g(S) + \pi_{17} + \pi_{18} g(S) + \pi_{19} h(S) + \pi_{20} X_1^2 + \pi_{21} X_2^2 + \pi_{22} (t+1) g(S)$$
$$+ \pi_{23} (t+1)^2 + \pi_{24} X_1 + \pi_{25} X_2 + \pi_{26} (t+1).$$

where $g(x) = 2x+1$, and $h(x) = 3x^2 + 3x + 1$. Several of the parameters appear in multiple equations. Such cross equation restrictions reflect the specification's logical consistency. The future components' asymmetry arises since choosing school both augments school experience and removes the cost of returning to school that one would otherwise face. In contrast, choosing alternative one or two only augments experience within that alternative.

Appendix B. Existence of joint posterior distribution

Let ω denote the number of missing wage observations, and let $\Omega \equiv [A_\Omega, B_\Omega]^\omega \in \Re_{++}^\omega$ be the domain of unobserved wages, where $0 < A_\Omega < B_\Omega < \infty$. Also, let $\Delta \equiv [A_\Delta, B_\Delta]^{3NT} \in \Re^{3NT}$ be the domain of latent relative utilities, where $-\infty < A_\Delta < B_\Delta < \infty$. We want to show that:

$$\int_{\Omega,\Delta,\beta,\pi,\Sigma_\varepsilon^{-1},\Sigma_\eta^{-1}} \left(\prod_{i,t} (w_{i1t} w_{i2t})^{-1} \right) g(V, \beta, \Sigma_\varepsilon^{-1}) h(Z, W, \pi, \Sigma_\eta^{-1}) < \infty. \tag{B1}$$

Here, we have subsumed α and Λ into π and Ψ, respectively (as we did in step 5 of section 3) and defined

$$g(V, \beta, \Sigma_\varepsilon^{-1}) = |\Sigma_\varepsilon^{-1}|^{\frac{NT-3}{2}} \exp\left\{ -\frac{1}{2} (V - Y\beta)' (\Sigma_\varepsilon^{-1} \otimes I_{NT}) (V - Y\beta) \right\} \tag{B2}$$

where

$$V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}, V_i = \begin{pmatrix} \ln w_{1i1} \\ \vdots \\ \ln w_{NiT} \end{pmatrix}, Y = \begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}. \tag{B3}$$

We have also defined

$$h(Z, W, \pi, \Sigma_\eta^{-1}) = |\Sigma_\eta^{-1}|^{\frac{NT-4}{2}} \exp\left\{ -\frac{1}{2} (Z - W - \Psi\pi)' (\Sigma_\eta^{-1} \otimes I_{NT}) (Z - W - \Psi\pi) \right\}$$
$$\cdot I(Z_{ijt} > 0, Z_{ikt}(k \neq j) \text{ if } d_{it} = j \text{ and } j \in \{1,2,3\}, \{Z_{ijt}\}_{j=1,2,3} < 0 \text{ otherwise}) \tag{B4}$$

Where

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}, Z_i = \begin{pmatrix} Z_{1i1} \\ \vdots \\ Z_{NiT} \end{pmatrix}, W = \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix}, W_i = \begin{pmatrix} w_{1i1} \\ \vdots \\ w_{NiT} \end{pmatrix} (i = 1,2), \text{ and } W_3 = (0).$$

Since the only arguments common to both *g* and *h* are those that include functions of wages, we can express (B1) as

$$\int_\Omega \left\{ \left(\prod_{i,t} (w_{i1t} w_{i2t})^{-1} \right) \int_{\beta,\Sigma_\varepsilon^{-1}} g(V, \beta, \Sigma_\varepsilon^{-1}) \int_{\Delta,\pi,\Sigma_\eta^{-1}} h(Z, W, \pi, \Sigma_\eta^{-1}) \right\}. \tag{B5}$$

We first observe that for any configuration of unobserved wages in Ω,

$$g(V) \equiv \int_{\beta,\Sigma_\varepsilon^{-1}} g(V, \beta, \Sigma_\varepsilon^{-1}) < \infty. \tag{B6}$$

To see this, note that we can express $g(V, \beta, \Sigma_\varepsilon^{-1})$ as

$$g(V,\beta,\Sigma_\varepsilon)=|\Sigma_\varepsilon^{-1}|^{\frac{NT-3}{2}}\exp\left\{-\frac{1}{2}tr(S(\hat\beta)\Sigma_\varepsilon^{-1})-\frac{1}{2}(\beta-\hat\beta)'Y'(\Sigma_\varepsilon^{-1}\otimes I_{NT})Y\left(\beta-\hat\beta\right)\right\} \tag{B7}$$

where

$$\begin{aligned}\hat\beta&=(Y'(\Sigma_\varepsilon^{-1}\otimes I_{NT})Y)^{-1}Y'(\Sigma_\varepsilon^{-1}\otimes I_{NT})V\\ S(\hat\beta)&=(V_1-Y_1\hat\beta_1,\ V_2-Y_2\hat\beta_2)'(V_1-Y_1\hat\beta_1,\ V_2-Y_2\hat\beta_2)\end{aligned} \tag{B8}$$

Hence, $g(V,\beta,\Sigma_\varepsilon)$ is proportional to a normal-Wishart density (Bernardo and Smith, (1994), p. 140), hence finitely integrable to a value $g(V)$ that, in general, will depend on the configuration of unobserved wages. Since $g(V)$ is finite over the compact set Ω, it follows that $g(V)$ is bounded over Ω.

Turn next to $h(Z,W,\pi,\Sigma_\eta^{-1})$. Since (B4) has the same form as (B2), just as (B7) we can write:

$$\begin{aligned}h(Z,W,\pi,\Sigma_\eta^{-1})&=|\Sigma_\eta^{-1}|^{\frac{NT-4}{2}}\exp\left\{-\frac{1}{2}tr(S(\hat\pi)\Sigma_\eta^{-1})-\frac{1}{2}(\pi-\hat\pi)'\Psi'(\Sigma_\eta^{-1}\otimes I_{NT})\Psi\left(\pi-\hat\pi\right)\right\}\\ &\cdot I(Z_{ijt}>0,Z_{ikt}(k\neq j)\ \text{if}\ d_{it}=j\ \text{and}\ j\in\{1,2,3\},\ \{Z_{ijt}\}_{j=1,2,3}<0\ \text{otherwise})\end{aligned} \tag{B9}$$

where $\hat\pi$ and $S(\hat\pi)$ are defined in a way that is exactly analogous to (B8).

Hence,

$$h(Z,W)\equiv\int_{\Delta,\pi,\Sigma_\eta^{-1}}h(Z,W,\pi,\Sigma_\eta^{-1})<\infty \tag{B10}$$

for any configuration of unobserved wages in Ω and latent utilities in Δ. It follows that $h(Z,W)$ is bounded over $\Omega\times\Delta$. Thus, the integral (B1) reduces to

$$\int_{\Omega,\Delta}\left\{\left(\prod_{i,t}(w_{i1t}w_{i2t})^{-1}\right)g(V)h(Z,W)\right\} \tag{B11}$$

which is finite since each element of the integrand is bounded over the compact domain of integration. //

References

Bernardo, Jose and Smith, Adrian F. M. 1994. *Bayesian Theory*. John Wiley and Sons, Ltd., New York.

Eckstein, Z. and K. Wolpin. (1989): The specification and estimation of dynamic stochastic discrete choice models. Journal of Human Resources, 24, 562-598.

Gelman, A. 1996: Inference and monitoring convergence. In W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London. 131-143

Geweke, J. 1995: Priors for macroeconomic time series and their application. Working paper, Federal Reserve Bank of Minneapolis.

Geweke, J. 1996: Monte Carlo simulation and numerical integration. In H. M. Amman, D. A. Kendrick and J. Rust (eds), *Handbook of Computational Economics*, Vol. 1, 731-800.

Geweke, J. and M. Keane. 1999a: Bayesian inference for dynamic discrete choice models without the need for dynamic programming. In Mariano, Schuermann and Weeks, (eds), *Simulation Based Inference and Econometrics: Methods and Applications*. Cambridge University Press. (Also available as Federal Reserve Bank of Minneapolis working paper #564, January, 1996.)

Geweke, J. and M. Keane. 1999b: Computationally intensive methods for integration in econometrics. In J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 5. North Holland.

Geweke, J., M. Keane and D. Runkle. 1994: Alternative computational approaches to inference in the multinomial probit model. *The Review of Economics and Statistics*, 76, 609-32.

Gilks, W. R., S. Richardson and D. J. Spiegelhalter. 1996: Introducing Markov chain Monte Carlo. In W. R.

Gilks, S. Richardson and D. J. Spiegelhalter (eds), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London. 1-20.

Heckman, J. 1976: The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *The Annals of Economic and Social Measurement*, 5, 475-92.

Heckman, J. 1981: Statistical models for discrete panel data. In Manski, C. and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA. 115-178

Heckman, J. and G. Sedlacek. 1985: Heterogeneity, aggregation and market wage functions: an empirical model of self-selection in the labor market. *Journal of Political Economy*, 93, 1077-1125.

Hotz, V.J. and R.A. Miller 1993: Conditional choice probabilities and the estimation of dynamic programming models. *Review of Economic Studies*, 60, 497-530.

Houser, D. 1999: Bayesian analysis of a dynamic, stochastic model of labor supply and saving. Manuscript, University of Arizona.

Keane, M. and K. Wolpin 1994: Solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo Evidence. *Review of Economics and Statistics*, 76, 648-672.

Keane, M. and K. Wolpin 1997: The career decisions of young men. *Journal of Political Economy*, 105, 473-522.

Krusell, P. and A. A. Smith. 1995: Rules of thumb in macroeconomic equilibrium: a quantitative analysis. *Journal of Economic Dynamics and Control*, 20, 527-58.

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

Lerman, S. and C. Manski. 1981: On the use of simulated frequencies to approximate choice probabilities. In C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge, MA. 305-319.

Manski, C. 1993: Dynamic choice in social settings. *Journal of Econometrics*, 58, 121-136.

McFadden, D. 1989: A method of simulated moments for estimation of discrete response models without numerical integration', *Econometrica*, 57, 995-1026.

Monfort, A., Van Dijk, H. and Brown, B. (eds) 1995: *Econometric Inference Using Simulation Techniques*. New York: John Wiley and Sons, Ltd.

Roy, A.D. 1951: Some thoughts on the distribution of earnings. *Oxford Economics Papers*, 3, 135-146.

Rust, J. 1987: Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher. *Econometrica*, 55, 999-1033.

Rust, J. 1994: Estimation of dynamic structural models, problems and prospects: discrete decision processes. In C. Sims (ed.*), Advances in Econometrics*, *Sixth World Congress Vol. II*. Cambridge University Press, 119-170.

Tierney, L. 1994: Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22, 1701-32.

Wolpin, K. 1984: An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy*, 92, 852-74.

Zellner, A. 1971. *An introduction to Bayesian inference in econometrics*. New York. John Wiley and Sons.

Table 1

Quality of the Polynomial Approximation to the True Future Component

Error Set	1	2	3	4	5
Mean present value of payoffs with true future component*	356796	356327	355797	355803	355661
Mean present value of payoffs with polynomial approximation*	356306	355978	355337	355515	355263
Mean dollar equivalent loss*	491	349	460	287	398
Mean percent loss*	0.14%	0.10%	0.13%	0.08%	0.11%
Percent choice agreement					
Aggregate	91.81%	91.71%	91.66%	92.30%	91.80%
By Period					
1	95.80%	95.55%	96.30%	96.10%	96.15%
2	95.35%	94.90%	95.85%	95.95%	95.30%
3	91.30%	90.45%	90.25%	91.15%	89.90%
4	88.00%	87.75%	89.00%	88.90%	88.45%
5	87.00%	88.30%	87.00%	89.20%	87.60%
10	92.70%	92.60%	92.70%	92.30%	92.30%
20	92.20%	93.00%	92.70%	93.05%	93.10%
30	91.55%	90.90%	90.55%	91.85%	90.85%
40	92.80%	92.15%	91.70%	92.75%	92.10%

*The mean present value of payoffs is the equally-weighted average discounted sum of ex-post lifetime payoffs over 2000, 40 period lived agents. The values are dollar equivalents.

Table 2

Choice Distributions and Mean Accepted Wages in the Data Generated with True and OLS Polynomial Future Components

	Data Set 2 - POLY						Data Set 2 - EMAX					
					Mean Accepted Wage						Mean Accepted Wage	
Period	Percent in Occ. 1	Percent in Occ. 2	Percent in School	Percent at Home	Occ. 1	Occ. 2	Percent in Occ. 1	Percent in Occ. 2	Percent in School	Percent at Home	Occ. 1	Occ. 2
1	0.10	0.00	0.75	0.15	13762.19	17971.88	0.09	0.00	0.79	0.12	13837.93	19955.02
2	0.23	0.01	0.58	0.17	11822.16	19032.16	0.23	0.01	0.60	0.15	11941.31	18502.51
3	0.41	0.04	0.34	0.21	11249.67	16521.61	0.39	0.03	0.39	0.18	11275.37	16786.46
4	0.52	0.06	0.20	0.22	11167.03	16209.88	0.50	0.04	0.27	0.19	11208.15	17778.61
5	0.60	0.06	0.14	0.20	11417.94	16141.39	0.57	0.05	0.20	0.18	11598.75	16965.70
6	0.63	0.08	0.10	0.19	11802.61	16427.58	0.63	0.06	0.14	0.16	11897.28	17100.50
7	0.65	0.09	0.07	0.18	12257.30	16987.46	0.68	0.08	0.08	0.16	12286.26	17634.68
8	0.69	0.10	0.05	0.16	12701.01	17067.03	0.72	0.09	0.05	0.13	12751.92	17300.99
9	0.69	0.10	0.05	0.16	13167.06	18442.74	0.72	0.10	0.05	0.13	13159.23	19498.18
10	0.70	0.11	0.05	0.14	13709.21	18274.23	0.74	0.09	0.05	0.12	13790.83	19125.16
11	0.72	0.12	0.05	0.11	14409.81	19391.23	0.75	0.11	0.04	0.10	14546.63	19867.95
12	0.71	0.14	0.04	0.12	14511.54	19730.21	0.74	0.12	0.04	0.10	14650.45	20320.53
13	0.72	0.14	0.05	0.10	15216.89	21641.41	0.75	0.14	0.03	0.08	15439.04	21723.49
14	0.74	0.13	0.03	0.09	15943.12	21866.44	0.76	0.14	0.02	0.08	16150.59	22096.07
15	0.73	0.16	0.03	0.07	16507.05	22177.61	0.75	0.16	0.03	0.06	16773.15	22764.69
16	0.74	0.16	0.03	0.07	17129.96	22624.51	0.75	0.16	0.03	0.06	17437.26	22786.18
17	0.75	0.17	0.02	0.06	17886.20	24194.23	0.75	0.18	0.02	0.05	18276.74	24804.59
18	0.73	0.17	0.02	0.07	18408.75	24318.34	0.74	0.18	0.01	0.06	18786.43	24476.18
19	0.72	0.19	0.02	0.06	19590.88	25385.99	0.73	0.20	0.01	0.05	19961.17	25719.67
20	0.74	0.19	0.02	0.05	20186.07	25161.39	0.75	0.20	0.01	0.04	20571.89	25422.50
21	0.71	0.23	0.01	0.05	21113.74	26409.20	0.71	0.24	0.01	0.04	21613.91	26613.20
22	0.70	0.25	0.01	0.04	22002.82	26935.39	0.70	0.25	0.00	0.04	22488.94	27566.90
23	0.67	0.29	0.01	0.03	23259.72	28191.41	0.67	0.29	0.01	0.03	23655.61	28952.46
24	0.66	0.30	0.00	0.03	23119.46	28634.21	0.66	0.31	0.00	0.03	23706.23	29491.67
25	0.66	0.30	0.00	0.04	24085.78	30826.10	0.66	0.31	0.00	0.03	24535.54	31403.33
26	0.62	0.34	0.01	0.04	25399.34	30707.99	0.63	0.34	0.00	0.03	26003.31	31157.72
27	0.62	0.34	0.00	0.04	26971.71	32251.61	0.62	0.35	0.00	0.03	27482.83	33112.86
28	0.60	0.37	0.00	0.03	27074.62	32024.07	0.60	0.37	0.00	0.02	27805.46	32743.79
29	0.57	0.40	0.00	0.03	29049.11	32411.14	0.58	0.39	0.00	0.03	29596.82	33872.97
30	0.55	0.42	0.00	0.04	30492.25	34513.76	0.56	0.41	0.00	0.03	31216.48	35462.35
31	0.52	0.45	0.00	0.03	30745.54	35672.21	0.52	0.45	0.00	0.03	31744.63	36763.93
32	0.50	0.48	0.00	0.03	32078.16	36076.17	0.51	0.47	0.00	0.03	33016.52	37028.17
33	0.46	0.50	0.00	0.04	34202.82	37460.57	0.47	0.51	0.00	0.03	34905.34	38435.23
34	0.43	0.54	0.00	0.02	34578.60	38293.38	0.44	0.54	0.00	0.02	35656.54	39212.14
35	0.42	0.55	0.00	0.04	37084.91	39690.50	0.43	0.54	0.00	0.03	38195.57	40767.33
36	0.39	0.58	0.00	0.03	37580.47	40970.75	0.40	0.57	0.00	0.03	39119.51	41740.41
37	0.36	0.60	0.00	0.04	40129.34	41885.28	0.37	0.60	0.00	0.03	41228.89	42901.62
38	0.33	0.64	0.00	0.03	40101.57	43929.61	0.34	0.63	0.00	0.03	41477.02	45076.14
39	0.28	0.67	0.00	0.05	43282.44	44724.22	0.30	0.66	0.00	0.04	44266.27	46039.50
40	0.26	0.70	0.00	0.03	44462.69	45703.45	0.28	0.69	0.00	0.03	45668.36	46847.68

Table 3

Descriptive Statistics for Posterior Distributions of the Model's Structural Parameters for Several Different Data Sets Generated Using Polynomial Future Component

		Data Set 1-POLY		Data Set 2-POLY		Data Set 3-POLY		Data Set 4-POLY		Data Set 5-POLY	
Parameter	**True**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Occ. 1 Intercept	9.00000	9.01300	0.00643	9.00125	0.00797	9.00845	0.00600	9.00256	0.00661	9.00309	0.00598
Occ. 1 Own Experience	0.05500	0.05440	0.00080	0.05540	0.00086	0.05429	0.00076	0.05462	0.00080	0.05496	0.00060
Occ. 2 Experience	0.00000	0.00093	0.00095	-0.00121	0.00103	-0.00084	0.00092	0.00114	0.00112	-0.00129	0.00115
Education	0.05000	0.04806	0.00130	0.04881	0.00132	0.04850	0.00137	0.04938	0.00140	0.04924	0.00139
Occ. 1 Exp. Squared	-0.00025	-0.00024	0.00003	-0.00026	0.00003	-0.00025	0.00003	-0.00024	0.00003	-0.00026	0.00002
Occ. 1 Error SD	0.40000	0.398	0.002	0.402	0.002	0.401	0.002	0.400	0.002	0.400	0.002
Occ. 2 Intercept	8.95000	8.91712	0.01625	8.97693	0.01582	8.91699	0.01583	8.96316	0.01551	8.92501	0.01590
Occ. 2 Own Experience	0.04000	0.04049	0.00039	0.03918	0.00034	0.03946	0.00037	0.03968	0.00037	0.04055	0.00042
Occ. 2 Experience	0.06000	0.06103	0.00175	0.06016	0.00178	0.06461	0.00173	0.05946	0.00180	0.06009	0.00178
Education	0.07500	0.07730	0.00187	0.07245	0.00161	0.07782	0.00178	0.07619	0.00167	0.07650	0.00171
Occ. 2 Exp. Squared	-0.00090	-0.00088	0.00008	-0.00093	0.00008	-0.00109	0.00008	-0.00092	0.00008	-0.00087	0.00008
Occ. 2 Error SD	0.40000	0.409	0.003	0.399	0.003	0.407	0.003	0.397	0.003	0.399	0.003
Error Correlation	0.50000	0.481	0.031	0.512	0.025	0.420	0.033	0.528	0.037	0.438	0.042
Undergraduate Tuition	-5000	-4629	363	-5212	464	-5514	404	-4908	464	-4512	399
Graduate Tuition	-15000	-18006	2085	-16711	1829	-16973	1610	-16817	1692	-15091	1972
Return Cost	-15000	-14063	531	-16235	894	-15809	554	-14895	822	-15448	679
Preference Shock SD											
Occ. 1	9082.95	10121.05	255.49	9397.38	577.26	10578.31	537.01	9253.22	615.27	9494.74	354.07
Occ. 2	9082.95	8686.12	456.77	11613.38	346.03	10807.31	519.13	9610.23	457.45	9158.09	228.32
Occ. 3	11821.59	11569.93	281.18	12683.67	803.09	13418.79	358.98	12019.38	417.75	12247.80	401.15
Preference Shock Corr.											
Occ. 1 with Occ. 2	0.89	0.90	0.01	0.96	0.01	0.93	0.01	0.91	0.01	0.90	0.02
Occ. 1 with Occ. 3	0.88	0.89	0.01	0.88	0.01	0.90	0.01	0.88	0.01	0.88	0.01
Occ. 2 with Occ. 3	0.88	0.89	0.01	0.89	0.01	0.89	0.01	0.89	0.01	0.89	0.01

Table 4

Log Wage Equation Estimates from OLS on Observed Wages Generated Under the Polynomial Future Component*

	Occupation One					Occupation Two					Wage Error SDs	
Data Set	Intercept	Occ. 1 Experience	Occ. 2 Experience	Education	Occ. 1 Exp. Squared	Intercept	Occ. 1 Experience	Occ. 2 Experience	Education	Occ. 2 Exp. Squared	Occ. 1	Occ. 2
TRUE	9.00000	0.05500	0.00000	0.05000	-0.00025	8.95000	0.04000	0.06000	0.07500	-0.00090	0.40000	0.40000
1-POLY	9.15261	0.04236	0.01708	0.04247	0.00012	9.46735	0.03516	0.01953	0.05589	0.00038	0.38845	0.36574
	0.00520	*0.00076*	*0.00074*	*0.00133*	*0.00003*	0.00906	0.00037	0.00146	0.00179	0.00008		
2-POLY	9.14715	0.04320	0.01586	0.04309	0.00010	9.47924	0.03446	0.02261	0.05311	0.00017	0.38940	0.36300
	0.00528	*0.00076*	*0.00073*	*0.00134*	*0.00003*	*0.00888*	*0.00036*	*0.00136*	*0.00170*	*0.00007*		
3-POLY	9.14895	0.04230	0.01732	0.04420	0.00011	9.45851	0.03482	0.02335	0.05665	0.00016	0.38935	0.36495
	0.00528	*0.00076*	*0.00072*	*0.00135*	*0.00003*	*0.00914*	*0.00036*	*0.00140*	*0.00177*	*0.00007*		
4-POLY	9.15157	0.04220	0.01734	0.04261	0.00012	9.46413	0.03480	0.02346	0.05469	0.00015	0.38900	0.36217
	0.00527	*0.00076*	*0.00074*	*0.00136*	*0.00003*	*0.00896*	*0.00037*	*0.00138*	*0.00174*	*0.00007*		
5-POLY	9.14838	0.04274	0.01695	0.04408	0.00011	9.45131	0.03570	0.02021	0.05671	0.00035	0.38772	0.35781
	0.00521	*0.00076*	*0.00073*	*0.00135*	*0.00003*	*0.00880*	*0.00036*	*0.00139*	*0.00173*	*0.00007*		

*Standard errors in italics.

Table 5

Descriptive Statistics for Final 15,000 Gibbs Sampler Parameter Draws for Several Different Data Sets Generated Using True Future Component

		Data Set 1-EMAX		Data Set 2-EMAX		Data Set 3-EMAX		Data Set 4-EMAX		Data Set 5-EMAX	
Parameter	True	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Occ. 1 Intercept	9.00000	9.01342	0.00602	9.00471	0.00527	9.01436	0.00584	9.01028	0.00593	9.00929	0.00550
Occ. 1 Own Experience	0.05500	0.05427	0.00073	0.05489	0.00071	0.05384	0.00072	0.05394	0.00072	0.05410	0.00071
Occ. 2 Experience	0.00000	0.00111	0.00093	0.00092	0.00114	0.00078	0.00126	0.00107	0.00100	0.00051	0.00093
Education	0.05000	0.04881	0.00118	0.05173	0.00126	0.04869	0.00129	0.04961	0.00123	0.05067	0.00124
Occ. 1 Exp. Squared	-0.00025	-0.00023	0.00002	-0.00025	0.00002	-0.00023	0.00002	-0.00022	0.00002	-0.00023	0.00002
Occ. 1 Error SD	0.40000	0.397	0.002	0.399	0.002	0.399	0.002	0.397	0.002	0.397	0.002
Occ. 2 Intercept	8.95000	8.90720	0.01704	8.98989	0.01970	8.93943	0.01850	8.93174	0.01649	8.94097	0.01410
Occ. 2 Own Experience	0.04000	0.04093	0.00037	0.03967	0.00037	0.03955	0.00038	0.04001	0.00037	0.04060	0.00039
Occ. 2 Experience	0.06000	0.06087	0.00178	0.05716	0.00190	0.06200	0.00201	0.06211	0.00179	0.05880	0.00157
Education	0.07500	0.07822	0.00166	0.07338	0.00171	0.07579	0.00165	0.07743	0.00167	0.07613	0.00159
Occ. 2 Exp. Squared	-0.00090	-0.00087	0.00008	-0.00081	0.00008	-0.00098	0.00008	-0.00101	0.00008	-0.00084	0.00007
Occ. 2 Error SD	0.40000	0.409	0.003	0.397	0.003	0.404	0.003	0.402	0.003	0.397	0.003
Error Correlation	0.50000	0.517	0.023	0.607	0.029	0.484	0.044	0.521	0.035	0.488	0.028
Undergraduate Tuition	-5000	-2261	313	-2937	358	-3407	371	-3851	426	-3286	448
Graduate Tuition	-15000	-10092	1046	-10788	1141	-11983	1188	-10119	1380	-11958	1823
Return Cost	-15000	-14032	482	-16014	431	-16577	500	-16168	662	-18863	1065
Preference Shock SD											
Occ. 1	9082.95	10634.90	423.85	10177.24	165.11	11438.63	438.72	9973.32	371.64	9071.29	509.80
Occ. 2	9082.95	9436.10	372.86	12741.02	405.25	11432.19	287.69	9310.37	718.15	7770.66	555.39
Occ. 3	11821.59	11450.65	338.28	12470.12	259.81	13999.95	351.33	13183.33	471.47	13897.62	533.67
Preference Shock Corr.											
Occ. 1 with Occ. 2	0.89	0.93	0.01	0.98	0.00	0.94	0.01	0.91	0.02	0.86	0.03
Occ. 1 with Occ. 3	0.88	0.89	0.01	0.88	0.01	0.90	0.01	0.88	0.01	0.88	0.01
Occ. 2 with Occ. 3	0.88	0.87	0.01	0.90	0.01	0.90	0.01	0.89	0.02	0.89	0.02

Table 6

Wealth Loss when Posterior Polynomial Approximation is used in Place of True Future Component*

	Using True EMAX**	Using Posterior EMAX**						
Data Set	Mean Present Value of Payoffs***	Mean Present Value of Payoffs***	Mean Dollar Equivalent Loss	Mean Percent Loss	Aggregate Choice Agreement	Percent with 0-35 Agreements	Percent with 36-39 Agreeements	Percent Choosing Same Path
1-EMAX	356796	356134	663	0.19%	90.80%	34.25%	42.65%	23.10%
2-EMAX	356327	355836	491	0.14%	91.34%	33.00%	44.00%	23.00%
3-EMAX	355797	354746	1051	0.30%	89.79%	39.00%	38.95%	22.05%
4-EMAX	355803	355450	353	0.10%	93.48%	24.60%	38.45%	36.95%
5-EMAX	355661	355485	176	0.05%	93.18%	24.95%	30.50%	44.55%

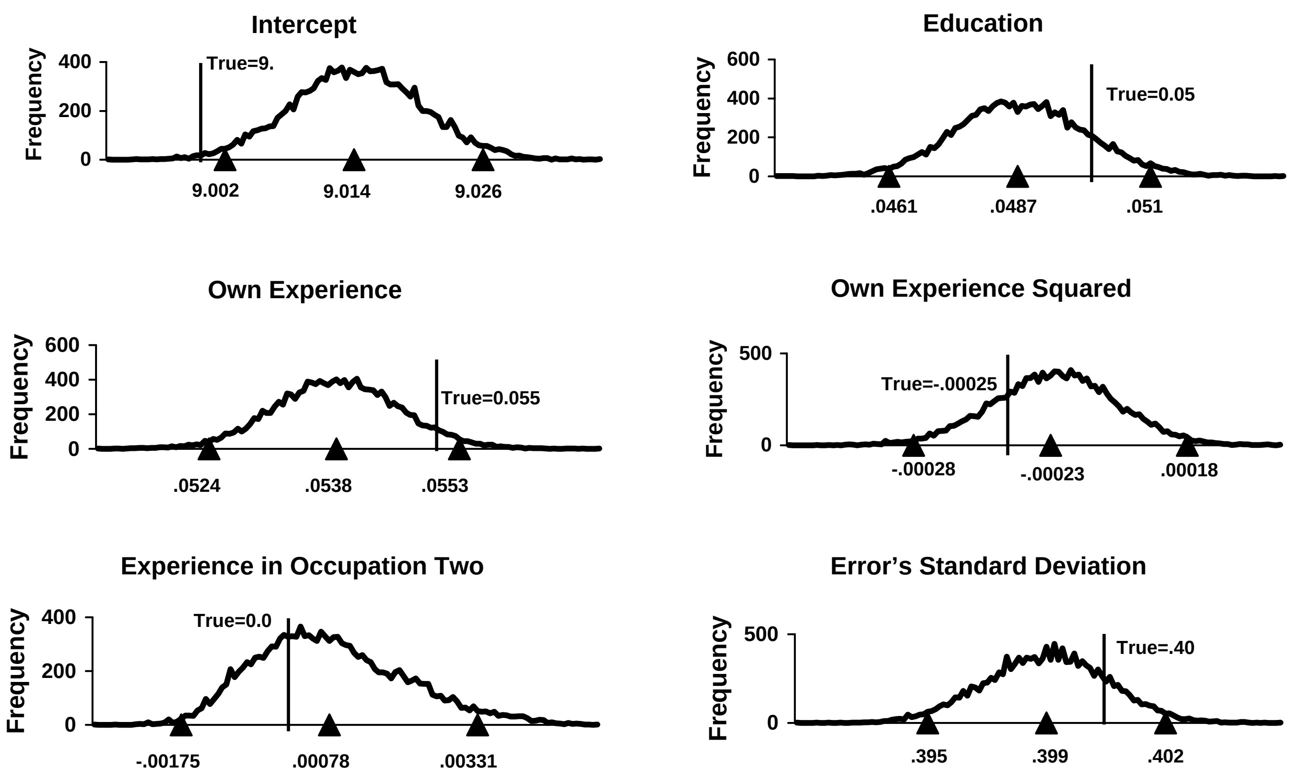
* Polynomial parameter values are set to the mean of their respective empirical posterior distributions.

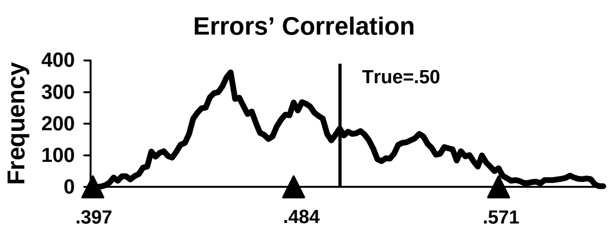
** Each simulation includes 2000 agents that live for exactly 40 periods.

*** The mean present value of payoffs is the equal-weight sample average of the discounted streams of ex-post lifetime payoffs.

Figure 1

Marginal Posterior Densities of First Log-Wage Equation's Parameters from Data Set 3-Emax*





***On each graph, the vertical line indicates the data generating parameter value. The middle triangle indicates the empirical mean, and the two flanking triangles are located two standard deviations from the mean.**

Figure 2

EMAX and Polynomial Future Components Evaluated at Mean Values of State Variables at Each Period