Classical and Bayesian Analysis of a Probit Panel Data Model with Unobserved Individual Heterogeneity and Autocorrelated Errors

Martin Burda  
Department of Economics, University of Pittsburgh,  
4521 Posvar Hall, Pittsburgh, PA 15213, USA

Roman Liesenfeld  
Department of Economics, Christian-Albrecht-Universität,  
Ohlshausenstr. 40-60, 24118 Kiel, Germany

Jean-François Richard  
Department of Economics, University of Pittsburgh,  
4917 Posvar Hall, Pittsburgh, PA 15213, USA

March 15, 2007

Abstract

In this paper, we perform both classical and Bayesian analysis of a panel probit model with unobserved individual heterogeneity and serially correlated errors. In the classical part, we use Efficient Importance Sampling (EIS) in evaluating a sequentially factorized simulated maximum likelihood function. In the Bayesian part, we utilize the Markov Chain Monte Carlo (MCMC) sampling scheme, augmenting the data with latent variables. We sample the unobserved individual heterogeneity component as one Gibbs block drawing from a piece-wise linear approximation to the marginal posterior density constructed with a nonparametric form of EIS. The time effects are simulated as another Gibbs block with a parametric EIS proposal density for an Acceptance-Rejection Metropolis-Hastings step. We apply our methods to the product innovation activity of a panel of German manufacturing firms in response to imports, foreign direct investment and other control variables. This dataset was analyzed by Bertschek and Lechner (1998) and Greene (2004) under more restrictive assumptions that we use as a benchmark for our analysis. Compared to these authors, our coefficient estimates of the key variables were somewhat smaller, which can be explained by the exclusion of three far outliers from our estimation and also by our flexible model assumptions. Nonetheless, our results confirm the positive effect of imports and FDI on firms’ innovation activity found in the previous literature. Moreover, unobserved firm heterogeneity is shown to play a far more significant role in the application than time effects.

Keywords: Dynamic latent variables, Markov Chain Monte Carlo, panel probit, simulated likelihood, importance sampling.

JEL Classification: C11, C13, C15, C23, C25.
1 Introduction

It has long been recognized that maximum likelihood analysis of limited dependent variable (LDV) models with panel data is feasible only under relatively restrictive assumptions (Butler and Moffitt, 1982). The difficulty that such models pose in general lies in the likelihood function containing multivariate integrals that are often analytically intractable. This obstacle is typically overcome with the use of simulation methods (see e.g. Geweke and Keane, 2001, and references therein) that replace integrals with computationally inexpensive Monte Carlo (MC) estimates. By the law of large numbers, such integral estimates can be made arbitrarily accurate by increasing the size of the simulated data.

Simulation-based estimation methods for LDV models generally take one of two approaches (Hyslop, 1999). The first approach, often called the Simulated Maximum Likelihood\(^1\) (SML), involves obtaining an unbiased simulator\(^2\) for the likelihood function and maximizing the resulting log simulated likelihood function instead of the actual likelihood function. The second approach utilizes simulation of an expression for the score of the likelihood. Two leading examples are the Method of Simulated Moments (MSM) estimator (McFadden, 1989) and the Method of Simulated Scores (MSS) estimator (Hajivassiliou and McFadden, 1998). Under the MSM estimator, the score of the likelihood is first expressed as a moment condition, the moment condition is then simulated and the estimator solves for the root of the simulated condition. The MSS estimator solves for the root of the simulated score directly. Based on available MC evidence, Geweke and Keane (2001, p. 3505) report that "in most contexts the choice between SML and MSM is not important."\(^3\) On the other hand, Hyslop (1999, p. 1268-1269) expresses preference for SML based on ease of implementation, numerical stability and computational burden. Notably, while SML is comparatively simple to implement, "MSM and MSS often require significant manipulation of the score function."

For a successful implementation of any of these estimators, it is essential to use a highly accurate smooth probability simulator. Among the currently available methods, the GHK simulator\(^4\) (developed by Geweke, 1991; Börsch-Supan and Hajivassiliou, 1993; Keane, 1994) is the most popular one and it has been reported to perform very well in MC studies for simulating the multivariate normal choice probabilities (see Geweke and Keane, 2001, and references therein). However, the

---

1. Gourieroux and Monfort (1996) provide the essential statistical background for the SML estimator.
2. Here we refer to a method of drawing random numbers involving an appropriate density for the random draws.
3. These authors note that one known exception is the case of panel data models with serially correlated errors - the type of models considered in this paper. This conclusion is based on a study by Lee (1997) that compared the performance of SML and MSM based on the GHK simulator and found GHK-SML serial correlation parameters severely biased relative to GHK-MSM. However, in this paper we use a different simulator, the EIS, which has been found to improve on the GHK simulator in terms of bias (Zhang and Lee, 2004).
4. It is sometimes also called the Smooth Recursive Conditioning (SRC) simulator (Börsch-Supan and Hajivassiliou, 1993).
recently developed Efficient Importance Sampling technique (Richard and Zhang, 2000, 2006) has been found highly competitive with the GHK sampler. Zhang and Lee (2004) show in an MC study that while the performance of GHK-SML and EIS-SML is comparable for short panels \( (T = 8) \), for longer panels \( (T > 50) \) the GHK-SML estimates of the lagged dependent variable coefficient and the serial correlation coefficient are biased (upward and downward, respectively), while the EIS-SML estimates avoid this bias. The appealing theoretical justification for EIS is one of minimizing the MC sampling variance in construction of the SML whereas the GHK simulator lacks this property. Moreover, the EIS sequential implementation (Danielsson and Richard, 1993; Richard and Zhang, 2006) is well suited for evaluation of likelihood functions expressed as integrals with very high dimensions (>1,000).

In this paper, we perform EIS-SML classical and Bayesian analysis of a panel probit model with unobserved individual heterogeneity and autocorrelated errors. We do not impose any orthogonality condition on the unobserved individual effects with respect to the observed regressors. Our model thus falls outside of the class of what is called in the traditional econometric parlance "random effects" models (Wooldridge, 2001, p. 252). In the LDV context\(^5\), the EIS-SML approach has been implemented in Richard and Zhang (2006) as a binary logit model in a Monte Carlo simulation pilot study, and in Liesenfeld and Richard (2006b) analyzing the union/non-union decision of young men with the data set of Vella and Verbeek (1998). Here we adopt the EIS-SML procedure to the panel probit case. Two other studies that used the SML method for the panel probit model with the same assumptions as ours are Falcetti and Tudela (2006), and Hyslop (1999). However, these authors utilized the competing GHK simulator which is tantamount to using a different estimation technique in the construction of the simulated log likelihood function.

In the Bayesian part, we embed EIS within the Markov Chain Monte Carlo (MCMC) simulation method to perform posterior analysis. Specifically, we implement the Gibbs sampling scheme where we augment the data with latent variables. We sample the unobserved individual heterogeneity component as one Gibbs block drawing from a piece-wise linear approximation to the marginal posterior density constructed with a nonparametric form of EIS. The time effects are simulated as another Gibbs block with a parametric EIS proposal density for an Acceptance-Rejection Metropolis-Hastings step. The general approach to augmented Gibbs sampling has been implemented in Liesenfeld and Richard (2003, 2006a) in models of stochastic volatility for sampling the autocorrelated error component. However, Bayesian analysis of an LDV model with unobserved heterogeneity and autocorrelated errors has, to our knowledge, thus far not been performed and represents a methodological

\(^5\) The EIS technique has been successfully implemented in other models, specifically stochastic volatility models (Liesenfeld and Richard, 2003, 2006a), dynamic parameter models involving counts (Jung and Liesenfeld, 2001), and stochastic autoregressive duration models (Bauwens and Hautsch, 0003).
contribution of this paper. The use of nonparametric EIS represents another novel feature.

We apply our method to the product innovation activity of a panel of German manufacturing firms in response to imports, foreign direct investment and other control variables. The same dataset was analyzed by Bertschek and Lechner (1998) and Greene (2004) for different types of estimators under more restrictive assumptions providing a useful benchmark for comparison with our results. Specifically, Bertschek and Lechner (1998) proposed several variants of a GMM estimator based on the period specific regression functions. Greene (2004) performed maximum likelihood analysis with GHK-SML and the Butler and Moffitt (1982) Hermite quadrature method. None of these authors considered a model with unobserved individual heterogeneity and autocorrelated errors as analyzed in this paper.

The remainder of the paper is organized as follows. Section 2 outlines the empirical example and the GMM and ML estimators of the dynamic panel probit models considered by Bertschek and Lechner (1998) and Greene (2004). In Section we elaborate our classical EIS-SML and Bayesian estimation techniques. Empirical results are reported and discussed in Section 4. Section 5 concludes.

2 Empirical Example and Estimation Methods

The goal of our empirical application is to investigate firms’ innovative activity as a response to imports and foreign direct investment (FDI). This problem was originally considered in Bertschek (1995) who suggested that imports and inward FDI had a positive effect on the innovative activity of domestic firms. The rationale behind this argument is that imports and FDI represent a competitive threat to domestic firms. Competition on the domestic market is enhanced and the profitability of the domestic firms might be reduced. Consequently, these firms have to produce more efficiently. One possibility to react to this competitive threat is to increase innovative activity.

The analyzed dataset contains $N = 1270$ cross-section units observed over $T = 5$ time periods. The dependent variable $y_{it}$ in the data takes the value one if a product innovation occurred within the last year and the value zero otherwise. The $K$-vector of control variables is denoted by $\bar{z}_{it}$ and the corresponding vector of parameters to be estimated by $\beta$. The independent variables refer to the market structure, in particular the market size of the industry ($\ln(sales)$), the shares of imports and FDI in the supply on the domestic market ($import\ share$ and $FDI\ share$), the productivity as a measure of the competitiveness of the industry as well as two variables indicating whether a firm belongs to the raw materials or to the investment goods industry. Also, including the relative firm size accounts for the innovation – firm size relation often discussed in the literature.

---

6 Similar data set was used in an interesting paper by Inkmann (2000) but with some regressors different from ours.
variables with exception of the firm size are measured at the industry level. Descriptive statistics and further discussion appear in Bertschek and Lechner (1998) and Greene (2004).

Two distinct sources of time dependence have been identified in the literature. In the context of our empirical application, the first arises from the possibility that innovation occurring in the present period may alter the conditions for the occurrence of innovation in the next period. In this case past experience has a behavioral effect in the sense that otherwise identical company that did not experience the event would behave differently from the company that experienced the event. This phenomenon is known as true state dependence and is typically captured by including a lagged dependent variable among the regressors.

The second source of time dependence derives from the fact that companies may differ in their propensity to innovate. Two components are distinguished in this case. The first one relates to the existence of company-specific attributes that are time-invariant. This component is typically called unobserved heterogeneity and we allow for it by including a time-invariant company-specific error term $\tau_i$. It may reflect institutional factors that are difficult to control for by direct inclusion among the regressors. The second component takes into account that companies’ differences may be correlated over time. Improper treatment of the error structure may result in a conditional relationship between future and past experience that is termed spurious state dependence (Hyslop, 1999). We avoid this problem by assuming an $AR(1)$ structure for the latent error term $\lambda_t$.

### 2.1 Alternative Panel Probit Model Specifications

The panel probit model has been analyzed extensively under various assumptions in the literature. In this Section, in addition to the basic probit model, we briefly review two studies, Bertschek and Lechner (1998) and Greene (2004), which used the same dataset as in this paper and are therefore of particular relevance as benchmarks for discussion of our results. In doing so, we present only the least restrictive models of the ones analyzed by these authors.

#### 2.1.1 Model 1: Pooled Probit

This is the simplest probit estimator that treats the entire sample as if it were a large cross-section. Specifically, it postulates the latent variable probit model specification

$$y_{it}^* = \beta' z_{it} + \epsilon_{it}$$

(1)
with the observation rule

\[ y_{it} = \mathbf{1}(y_{it}^* \geq 0), \quad i : 1, \ldots, N ; \quad t : 1, \ldots, T \]  

(2)

where \( \mathbf{1}(\cdot) \) denotes the indicator function. The error terms \( \epsilon_{it} \) are normally distributed with zero mean and unit variance.

### 2.1.2 Model 2: Panel Probit with Autocorrelated Errors

Bertschek and Lechner (1998) assume the latent variable probit model specification (1) with the observation rule (2). However, their error terms \( \xi_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})' \) are modeled as jointly normally distributed with mean zero and covariance matrix \( \Sigma \). Also, \( \xi_i \) are independent of the explanatory variables which implies strict exogeneity of the latter. The error terms may be correlated over time for a given firm, but uncorrelated over firms. The diagonal elements of \( \Sigma \) are set to unity to facilitate identification of \( \beta \) and the off-diagonal elements are considered nuisance parameters. On the basis of the model (1) Bertschek and Lechner (1998) formulated the following set of moment conditions

\[
E[W(Z, \beta_0)|X] = 0 \\
W(z, \beta) = [w_1(Z_1, \beta), \ldots, w_T(Z_T, \beta)]' \\
w_t(Z_t, \beta) = Y_t - \Phi(\beta'Z_{it}) 
\]

(3)

where \( \Phi \) denotes the CDF of a univariate normal distribution. The main advantage of using these moments is that their evaluation does not require multidimensional integration and they do not depend on the \( T(T - 1)/2 \) off-diagonal elements of \( \Sigma \). In line with the GMM literature, (3) implies

\[
E\{A(X)W(Z, \beta_0)\} = 0 
\]

where \( A(X) \) is a \( P \times T \) matrix of instrumental variables. The efficient GMM estimator of \( \beta_0 \) is then defined as

\[
\hat{\beta}_N = \arg\min_{\beta} g_N'(\beta)\Omega^{-1}g_N(\beta) 
\]

(4)

where

\[
g_N(\beta) = \frac{1}{N} \sum_{i=1}^{N} A(x_i)W(Z_i, \beta) 
\]

Bertschek and Lechner (1998) obtained a nonparametric estimate of the optimal weighting matrix \( \Omega \) using a \( k \)-nearest neighbor (\( k \)-NN) approach.
2.1.3 Model 3: Random Parameters Model

Greene (2004) noted that the dataset used contains a considerable amount of between group variation (97.6% of the FDI variation and 92.9% of the imports share variation is accounted for by differences in the group means). Thus, the dataset was likely to contain significant degree of unobserved individual heterogeneity, while none of the models above accounted for it. Greene (2004) suggested two alternative formulations of the panel probit model: the Random Parameters Model and the Latent Class Model (discussed further below). The Random Parameters Model (or 'Hierarchical' or 'Multilevel' Model) is based on the latent variable probit model specification

\[ y_{it} = \beta'_i z_{it} + \epsilon_{it} \]

with the observation rule (2), \( \epsilon_{it} \sim NID[0, 1] \), and

\[ \beta_i = \mu + \Delta z_i + \Gamma w_i \]

where \( \mu \) is \( K \times 1 \) vector of location parameters, \( \Delta \) is \( K \times L \) matrix of unknown location parameters, \( \Gamma \) is \( K \times K \) lower triangular matrix of unknown variance parameters, \( z_i \) is \( L \times 1 \) vector of individual characteristics, \( w_i \) is \( K \times 1 \) vector of random latent individual effects. It holds that \( E[w_i|X_i, z_i] = 0 \) and \( Var[w_i|X_i, z_i] = V \), a \( K \times K \) diagonal matrix of known constants. Hence \( E[\beta_i|X_i, z_i] = \mu + \Delta z_i \) and \( Var[\beta_i|X_i, z_i] = \Gamma \Gamma' \). Conditional on \( w_i \), observations of \( y_{it} \) are independent across time; timewise correlation would arise through correlation of elements of \( \beta_i \). The joint conditional density on \( y_{it} \) is

\[ f(y_i|X_i, \beta) = \prod_{t=1}^{T} \Phi[(2y_{it} - 1)\beta'_i z_{it}] \]  

(5)

The contribution of this observation to the log-likelihood function for the observed data is obtained by integrating the latent heterogeneity out of the distribution. Thus

\[ \log L = \sum_{i=1}^{N} \log L_i = \sum_{i=1}^{N} \log \int_{\beta_i} \prod_{t=1}^{T} \Phi[(2y_{it} - 1)\beta'_i z_{it}] g(\beta_i|\mu, \Delta, \Gamma, z_i) d\beta_i \]  

(6)

Estimates of \( \mu, \Delta \) and \( \Gamma \) are obtained by maximizing the SML version of (6).

2.1.4 Model 4: Latent Class, Finite Mixture Model

This model arises if we assume a discrete distribution for \( \beta_i \) instead of the continuous one postulated in the previous Random Parameters Model. Alternatively, the Latent Class model can be viewed as arising from a discrete, unobserved sorting of firms into groups, each of which has its own set
of characteristics. If the distribution of \( \beta_i \) has finite, discrete support over \( J \) points (classes) with probabilities \( p(\beta_j | \mu, \Delta, \Gamma, z_i) \), \( j = 1, \ldots, J \), then the resulting formulation of the analog of \( L_i \) from (6) is

\[
L_i = \sum_{j=1}^{J} p(\beta_j | \mu, \Delta, \Gamma, z_i) f(y_i | X_i, \beta_j)
\]

The model can then be estimated using the EM algorithm (see Greene, 2004, for details).

3 Panel Probit with Unobserved Individual Heterogeneity and Autocorrelated Errors

Our panel probit model differs from the ones described above by an explicit inclusion of variables for both individual unobserved heterogeneity and time effects accounting for spurious state dependence. Specifically, our standardized probit model specification assumes a latent variable regression for individual \( i \) and time period \( t \)

\[
y_{it} = \beta' z_{it} + \tau_i + \lambda_t + \epsilon_{it}, \quad i : 1, \ldots, N ; \quad t : 1, \ldots, T
\]

under the observation rule (2), where \( z_{it} \) is a vector of explanatory variables and \( \epsilon_{it} \sim N(0,1) \) is a stochastic error component uncorrelated with any other regressor. \( \tau_i \sim N(0, \sigma^2_{\tau}) \) represents individual unobserved heterogeneity that can be arbitrarily correlated with other regressors. \( \lambda_t \) captures latent time effects and is assumed to follow a stationary autoregressive process

\[
\lambda_t = \rho \lambda_{t-1} + \eta_t
\]

where \( \eta_t \sim N(0, \sigma^2_{\eta}) \) such that the mean of \( \lambda_t \) is zero and the variance \( \sigma^2_{\lambda} \) is stationary. It is assumed that \( \epsilon_{it}, \tau_i \) and \( \eta_t \) are mutually independent. The vector of parameters to be estimated is \( \theta = (\beta', \sigma_{\tau}, \rho_1, \ldots, \rho_k, \sigma_{\eta})' \). Denote \( \lambda = (\lambda_1, \ldots, \lambda_T)' \) and \( \tau = (\tau_1, \ldots, \tau_N)' \).

The likelihood function associated with \( y = (y_{i1}, \ldots, y_{iN})' \) can be written as

\[
L(\theta; y) = \int g(\tau, \lambda; \theta, y) p(\tau, \lambda; \theta) d\tau d\lambda
\]

with

\[
g(\tau, \lambda; \theta, y) = \prod_{i=1}^{N} \prod_{t=1}^{T} \left[ \Phi(v_{it}) \right]^{y_{it}} \left[ 1 - \Phi(v_{it}) \right]^{1-y_{it}}
\]
where
\[
\Phi(v_{it}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v_{it}} \exp\left(-\frac{1}{2}t^2\right) dt
\]
\[
v_{it} = \beta \xi_{it} + \tau_i + \lambda_t
\]
\[
p(\tau, \Delta; \theta) = \sigma^{-N}_\tau (2\pi)^{-N/2} \exp \left[ - \frac{1}{2\sigma_\tau^2} \sum_{i=1}^{N} \tau_i^2 \right] (2\pi)^{-T/2} |\Sigma_\lambda|^{-1/2} \exp \left[ - \frac{1}{2} \Delta \lambda^{-1} \lambda \right]
\] (9)

and \(\Sigma_\lambda\) denotes the stationary variance-covariance matrix of \(\lambda\).

### 3.1 EIS Evaluation of the Likelihood

We factorize the global high-dimensional Efficient Importance Sampling (EIS) optimization problem associated with (8) into a sequence of low-dimensional subproblems according to an appropriate factorization of the integrand \(g(\tau, \Delta; \theta, y)p(\tau, \Delta; \theta)\). Thus (8) becomes

\[
L(\theta; y) = \int \left[ \prod_{i=1}^{N} \prod_{t=1}^{T} \left[ \Phi(v_{it}) \right]^{y_{it}} \left[ 1 - \Phi(v_{it}) \right]^{1-y_{it}} \right] \times \sigma^{-N}_\tau (2\pi)^{-N/2} \exp \left[ - \frac{1}{2\sigma_\tau^2} \sum_{i=1}^{N} \tau_i^2 \right] (2\pi)^{-T/2} |\Sigma_\lambda|^{-1/2} \exp \left[ - \frac{1}{2} \Delta \lambda^{-1} \lambda \right] d\tau d\lambda
\]
\[
= \int (2\pi)^{-T/2} |\Sigma_\lambda|^{-1/2} \exp \left[ - \frac{1}{2} \Delta \lambda^{-1} \lambda \right] \sigma^{-N}_\tau (2\pi)^{-N/2} \times \prod_{i=1}^{N} \left\{ \exp \left[ - \frac{1}{2\sigma_\tau^2} \tau_i^2 \right] \prod_{t=1}^{T} \left[ \Phi(v_{it}) \right]^{y_{it}} \left[ 1 - \Phi(v_{it}) \right]^{1-y_{it}} \right\} d\tau d\lambda
\]
\[
= \int \phi_0(\Delta; \theta) \prod_{i=1}^{N} \phi_i(\tau_i, \Delta; \theta, y) d\tau d\lambda
\] (10)

where
\[
\phi_0(\Delta; \theta) = (2\pi)^{-T/2} |\Sigma_\lambda|^{-1/2} \exp \left[ - \frac{1}{2} \Delta \lambda^{-1} \lambda \right] \sigma^{-N}_\tau (2\pi)^{-N/2}
\]
\[
\phi_i(\tau_i, \Delta; \theta, y) = \exp \left[ - \frac{1}{2\sigma_\tau^2} \tau_i^2 \right] \prod_{t=1}^{T} \left[ \Phi(v_{it}) \right]^{y_{it}} \left[ 1 - \Phi(v_{it}) \right]^{1-y_{it}}
\] (11)

Since \(\phi_i\) introduces interdependencies between \(\tau_i\) and \(\lambda_i\), the efficient sampler can be constructed as a sequence of sampling densities with an unconditional density \(m_0(\lambda; \alpha_0)\) for \(\lambda\) and a sequence of conditional densities \(m_i(\tau_i | \lambda; \alpha_i)\) for \(\tau_i | \lambda\). The resulting factorization is given by

\[
m(\tau, \lambda | \alpha) = m_0(\lambda; \alpha_0) \prod_{i=1}^{N} m_i(\tau_i | \lambda; \alpha_i)
\]
For any given value of \( \alpha \), the likelihood (10) can be rewritten as

\[
L(\theta; y) = \int \frac{\phi_0(\lambda; \theta)}{m_0(\lambda; \alpha_0)} \prod_{i=1}^{N} \frac{\phi_i(\tau_i; \lambda; \theta, y)}{m_i(\tau_i|\lambda; \alpha_i)} \frac{m(\tau, \lambda|\alpha)}{d\tau d\lambda}
\]

(12)

The corresponding EIS estimate is given by

\[
\tilde{L}_{S,m}(\theta; y) = \frac{1}{S} \sum_{r=1}^{S} \frac{\phi_0(\lambda_r; \alpha_0)}{m_0(\lambda_r; \alpha_0)} \prod_{i=1}^{N} \frac{\phi_i(\tau_{ir}; \lambda_r; \alpha_0)}{m_i(\tau_{ir}|\lambda_r; \alpha_i)} \frac{m(\tau, \lambda|\alpha)}{d\tau d\lambda}
\]

(13)

where \( \{\tau_{1r}(\alpha_1), \ldots, \tau_{Nr}(\alpha_N), \lambda_r(\alpha_0)\; ; \; r = 1, \ldots, S\} \) are iid draws from the auxiliary importance sampling density \( m(\tau, \lambda|\alpha) \).

A density kernel \( k_i(\tau_i; \lambda; \alpha_i) \) for \( m_i(\tau_i|\lambda; \alpha_i) \) is given by

\[
m_i(\tau_i|\lambda; \alpha_i) = \frac{k_i(\tau_i; \lambda; \alpha_i)}{\chi_i(\lambda; \alpha_i)}
\]

with

\[
\chi_i(\lambda; \alpha_i) = \int k_i(\tau_i; \lambda, \alpha_i) d\tau_i
\]

The likelihood (12) can now be rewritten as

\[
L(\theta; y) = \int \frac{\phi_0(\lambda; \theta)}{m_0(\lambda; \alpha_0)} \prod_{i=1}^{N} \frac{\phi_i(\tau_i; \lambda; \theta, y)}{k_i(\tau_i; \lambda; \alpha_i)} \frac{m(\tau, \lambda|\alpha)}{d\tau d\lambda}
\]

where \( \gamma_i \) is a proportionality constant.

The EIS optimization problem requires solving a sequence of \( N + 1 \) weighted LS problems of the form

\[
\hat{\alpha}_i = \arg \min_{\alpha_i} \sum_{r=1}^{S} \left\{ \ln \phi_i(\tau_{ir}; \lambda_r; \theta, y) - q_i - \ln k_i(\tau_{ir}; \lambda_r; \alpha_i) \right\}^2 g_i(\tau_{ir}, \lambda_r; \theta, y)
\]

(14)

for \( i = 1, \ldots, N \) and

\[
\hat{\alpha}_0 = \arg \min_{\alpha_0} \sum_{r=1}^{S} \left\{ \ln \left[ \frac{\phi_0(\lambda_r; \theta)}{m_0(\lambda_r; \alpha_0)} \prod_{i=1}^{N} \chi_i(\lambda_r; \alpha_i) - q_0 - \ln m_0(\lambda_r; \alpha_0) \right] \right\}^2
\]

where \( \lambda_r, \lambda \) are draws from \( m(\tau, \lambda|\alpha) \). Based on these draws, the EIS estimate of the likelihood (13) is calculated as

\[
\tilde{L}_{r,m}(\theta; y) = \frac{\phi_0(\lambda_r; \theta)}{m_0(\lambda_r; \alpha_0)} \prod_{i=1}^{N} \frac{\phi_i(\tau_{ir}; \lambda_r; \theta, y)}{k_i(\tau_{ir}|\lambda_r; \alpha_i)}
\]

(15)

For further details on implementation, see Appendix 1.
3.2 Bayesian MCMC Approach Based on EIS

Bayesian MCMC simulation methods such as Gibbs sampling rely upon sampling from conditional posterior distributions in order to construct a Markov chain whose equilibrium distribution is the joint posterior of the parameters given the data. For the panel probit model, the joint posterior distribution of parameters can be augmented with the vectors of latent variables $\tau$ and $\lambda$. The complete joint posterior $f(\tilde{\theta}, \tau, \lambda; Z)$ can then be drawn from using Gibbs sampling. The main difficulty with such an MCMC approach is that of efficiently sampling from $\tau$ and $\lambda$ since the corresponding multivariate posterior distributions are high-dimensional and have no closed-form solution. To overcome this problem, Liesenfeld and Richard (2006a) proposed combining the EIS sampler with the Acceptance-Rejection Metropolis-Hastings (AR-MH) algorithm of Tierney (1994) in simulating the autocorrelated error component in stochastic volatility models. We adopt the approach to the panel probit model by simulating $\tau(\tilde{\theta}, Z)$ and $\lambda(\tilde{\theta}, Z)$ as Gibbs blocks: We sample the unobserved individual heterogeneity component $\tau(\tilde{\theta}, Z)$ as one Gibbs block drawing from a piece-wise linear approximation to the marginal posterior density constructed with a nonparametric form of EIS. The time effects $\lambda(\tilde{\theta}, Z)$ are simulated as another Gibbs block with a parametric EIS proposal density for an AR-MH step. The basis of this procedure is that the EIS densities for $\tau(\tilde{\theta}, Z)$ and $\lambda(\tilde{\theta}, Z)$ provide very close approximations to $f(\tau; \lambda; Z)$ and $f(\lambda; \tau; Z)$, respectively. The piece-wise linear approximation to $f(\tau; \lambda; Z)$ is dominated by $f(\tau; \lambda; Z)$ everywhere and can be made arbitrarily precise by increasing the size of the simulated grid. For $f(\lambda; \tau; Z)$ given the model assumptions, one can expect that the EIS parametric density provides an efficient proposal density for the target posterior $f(\lambda; \tau; Z)$ in the AR-MH step. This conjecture has been validated by AR-MH acceptance rates close to 100% in our empirical application.

Liesenfeld and Richard (2006a) list three attractive features that hold for the EIS-AR-MH approach in general: 1) only minor modifications of the code for the classical SML analysis are necessary in order to obtain a corresponding code for the Bayesian analysis (and vice versa), 2) it allows for a direct comparison between Bayesian and classical estimation results and for a corresponding analysis of the impact of the prior density on the inference process, and 3) its basic structure does not depend upon a specific model.

For a given vector of parameters $(\tilde{\theta}, \Upsilon)$ the augmented likelihood $L(\tilde{\theta}, \Upsilon; Z)$ is defined in (8). Let $\tilde{\theta}$ without the subvector $\theta_j$ be denoted by $\tilde{\theta}_{j\setminus j}$. For each Gibbs block of a generic parameter $\theta_j$ the Bayesian optimal updating of prior beliefs, $\pi(\theta_j)$, with new information (data $Z$) takes the form

$$f(\theta_j|\tilde{\theta}_{j\setminus j}, \Upsilon, Z) \propto L(\tilde{\theta}, \Upsilon; Z)\pi(\theta_j)$$

(16)
The individual Gibbs blocks used are $\beta, \sigma, \sigma', \rho$, and $X$, given data and the remaining augmented parameters. Throughout the analysis we make use of non-informative priors. Details of sampling from the posterior distributions are described in Appendix 2.

4 Empirical Results

In this section, we first reproduce the pooled probit estimates and the results obtained by Bertschek and Lechner (1998) and Greene (2004) as a benchmark for comparison with our results. Although these authors also report estimates of models other than shown below, we only select the ones with the least restrictive assumptions on the underlying probit models.

Table 1 presents the basic case of Pooled Estimator of Model 1 in (1) estimated in Stata using the command 'probit'. Table 1 also reports the Bertschek and Lechner (1998) GMM parameter estimates of Model 2 with a $k$-NN estimate of $\Omega$ in (4) and the Greene (2004) random parameter model prior means estimates of Model 3. As discussed in Greene (2004), there are some substantial differences compared to the other two models. Especially noteworthy are the greater impacts of the two central parameters of imports and FDI share on innovations as implied by the random parameters model. Nonetheless, these effects are positive in all cases as predicted.

Table 2 lists the Greene (2004) latent class estimates of Model 4. According to Greene (2004), working down from the number of classes $J = 5$ the estimates stabilized at the reported $J = 3$. Despite a large amount of variation across the three classes, the original conclusion that FDI and imports positively affect the probability of product innovation continued to be supported.

Table 3 presents our classical EIS-SML estimates and Bayesian posterior means of parameters in the model (7) with unobserved heterogeneity and autocorrelated errors. Posterior marginal densities of the Bayesian analysis are given in Figure 1 and autocorrelation functions of the parameter draws in Figure 3.
We excluded from estimation three distant outliers with relative firm size larger than 0.1 and productivity larger than 0.8 (see Figure 2) as these observations were inducing numerical instabilities into our EIS-SML estimator. Thus our sample size was $N = 1267$ and $T = 5$. The EIS-SML asymptotic (statistical) standard errors were obtained as the square root of the diagonal of the negative of the inverse of the Hessian evaluated at the estimated parameter values. The EIS-SML estimates are all within one standard deviation from the EIS-MCMC posterior means. One exception is the unobserved heterogeneity parameter $\sigma_r$. Its EIS-MCMC value $\hat{\sigma}_r = 1.021$ lies close to the value 1.1707 of an analogous parameter reported by Greene (2004, p.35) for the random effects model, but its EIS-SML value is about half that size. This can be explained by a potentially high skewness of its sampling density for companies whose response variable was constant (1 or 0) throughout

---

### Table 1: Models 1-3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Pooled Probit$^a$</th>
<th>Model 1$^b$</th>
<th>Model 2$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>$-1.960^{**}$</td>
<td>0.230</td>
<td>$-1.74^{**}$</td>
</tr>
<tr>
<td>log sales</td>
<td>0.177$^{**}$</td>
<td>0.022</td>
<td>0.15$^{**}$</td>
</tr>
<tr>
<td>Rel size</td>
<td>1.072$^{**}$</td>
<td>0.142</td>
<td>0.95$^{**}$</td>
</tr>
<tr>
<td>Imports</td>
<td>1.133$^{**}$</td>
<td>0.151</td>
<td>1.14$^{**}$</td>
</tr>
<tr>
<td>FDI</td>
<td>2.853$^{**}$</td>
<td>0.402</td>
<td>2.59$^{**}$</td>
</tr>
<tr>
<td>Prod.</td>
<td>$-2.341^{**}$</td>
<td>0.715</td>
<td>$-1.91^{**}$</td>
</tr>
<tr>
<td>Raw Mtl</td>
<td>$-0.279^{**}$</td>
<td>0.081</td>
<td>$-0.28^{**}$</td>
</tr>
<tr>
<td>Inv good</td>
<td>0.188$^{**}$</td>
<td>0.039</td>
<td>0.21$^{**}$</td>
</tr>
</tbody>
</table>

$^a$ Estimated in Stata by the simple command 'probit'.

$^b$ Bertschek and Lechner (1998), WNP-joint uniform estimates with $k = 880$, Table 9, standard errors from Table 10

$^c$ Greene (2004), $\hat{\mu}$ in Table 5

* Indicates significant at the 95% level

** Indicates significant at the 99% level

### Table 2: Model 4$^d$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>$-2.32^{**}$</td>
<td>0.768</td>
<td>$-2.71^{**}$</td>
</tr>
<tr>
<td>log sales</td>
<td>0.323$^{**}$</td>
<td>0.075</td>
<td>0.233$^{**}$</td>
</tr>
<tr>
<td>Rel size</td>
<td>4.38$^{**}$</td>
<td>0.882</td>
<td>0.720$^{**}$</td>
</tr>
<tr>
<td>Imports</td>
<td>0.936$^{**}$</td>
<td>0.491</td>
<td>2.26$^{**}$</td>
</tr>
<tr>
<td>FDI</td>
<td>2.20</td>
<td>2.54</td>
<td>2.80$^{**}$</td>
</tr>
<tr>
<td>Prod.</td>
<td>$-5.86^{**}$</td>
<td>1.69</td>
<td>$-7.70^{**}$</td>
</tr>
<tr>
<td>Raw Mtl</td>
<td>$-0.110$</td>
<td>0.172</td>
<td>$-0.599^{**}$</td>
</tr>
<tr>
<td>Inv good</td>
<td>0.131</td>
<td>0.143</td>
<td>0.413$^{**}$</td>
</tr>
</tbody>
</table>

$^d$ Greene (2004), Table 7

* Indicates significant at the 95% level

** Indicates significant at the 99% level
the sample period. Both estimates of $\tilde{\sigma}_\eta$ indicate that the role of time effects in this dataset is very small relative to individual unobserved effects. Large standard errors on $\hat{\rho}$ and its posterior distribution imply that this parameter could not be empirically identified, which further confirms the small significance of the time effects.

Most of our coefficient estimates fit into a convex combination of Greene’s (2004) Class 1 – Class 3. The estimates of the two key parameters of FDI and import share are positive, further validating the original hypothesis. However, both our estimates of the FDI coefficient are smaller relative to previous results. The import share coefficient estimates are also very close to the lower bound of Greene’s (2004) Class1 – Class 3 estimates. We attribute this finding to our flexible model assumptions whereby the influence the unobserved effects on product innovation was previously unaccounted for and channeled through FDI and import share in the model. Also, the exclusion of far outliers on this variables from our estimates may have played a role in this respect. The three excluded observations with large relative size have also disproportionately large values of import share and FDI; the means of the three outliers are 0.402 and 0.208 contrasting with means of the rest of the sample of 0.252 and 0.045 for import share and FDI, respectively. The outliers’ means thus correspond to approximately to the 82nd percentile and 98th percentile, respectively, of the remaining observations of these variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>EIS-SML</th>
<th>EIS-MCMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>$-1.612^{**}$</td>
<td>0.215</td>
</tr>
<tr>
<td>log sales</td>
<td>0.155**</td>
<td>0.022</td>
</tr>
<tr>
<td>Rel size</td>
<td>0.613**</td>
<td>0.134</td>
</tr>
<tr>
<td>Imports</td>
<td>0.947**</td>
<td>0.176</td>
</tr>
<tr>
<td>FDI</td>
<td>2.057**</td>
<td>0.465</td>
</tr>
<tr>
<td>Prod.</td>
<td>$-3.035^{**}$</td>
<td>1.592</td>
</tr>
<tr>
<td>Raw Mtl</td>
<td>$-0.108$</td>
<td>0.308</td>
</tr>
<tr>
<td>Inv good</td>
<td>0.141**</td>
<td>0.046</td>
</tr>
<tr>
<td>$\sigma_\tau$</td>
<td>0.471**</td>
<td>0.015</td>
</tr>
<tr>
<td>$\sigma_\eta$</td>
<td>0.036*</td>
<td>0.010</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.002</td>
<td>0.567</td>
</tr>
</tbody>
</table>

$^e$ EIS-SML estimates are the averages of 10 estimation rounds starting with different CRNs. Each round is based on an MC sample size $S = 600$. An average of 6-7 EIS iterations were needed for full parameter convergence. A grid search optimization procedure in Fortran 90 took approximately 9 hours, with relative function tolerance of $10^{-4}$ on a 1.7 GHz opteron unix machine.

$^f$ Posterior moments are based on 12,000 Gibbs iterations (discarding the first 2,000 draws). One Gibbs iteration took approximately 28 seconds on a 1.7 GHz opteron unix machine. The EIS simulation smoother is based on an MC sample of 400. On average, it took less than 6 EIS iterations for full convergence of the EIS parameters in sampling from the posteriors of the latent variables $\mu$ and $\lambda$. The AR and MH acceptance rates for $\lambda$ were 99.00% and 99.85%, respectively.

$^g$ Due to the skewness of the marginal posterior distribution [see Figure 1], the median is reported. The mean is 0.07842, interquartile range [0.022, 0.072], and the 95% confidence interval is [0.006, 0.255].

$^*$ Indicates significant at the 95% level

$^{**}$ Indicates significant at the 99% level
5 Conclusion

In this paper, we performed classical simulated maximum likelihood (SML) and Bayesian analysis of a panel probit model with unobserved individual heterogeneity and autocorrelated errors. The SML analysis was facilitated with the Efficient Importance Sampling (EIS) method that was found competitive with the GHK simulator in previous studies and was newly adopted to the panel probit case in this paper. In the Bayesian part, we embedded EIS within an Markov Chain Monte Carlo (MCMC) simulation method to perform posterior analysis augmented with both the time and cross-section latent variables. Thus, the posterior for the unobserved individual heterogeneity was sampled from as one Gibbs block, using a nonparametric version of EIS to form a piece-wise linear approximation to the posterior as a proposal density. The posterior for the vector of latent time effects was treated as another Gibbs block, using a parametric EIS approximation as the proposal density for an AR-MH step. This approach represents a methodological contribution to the limited dependent variable panel literature.

We applied our method to the product innovation activity of a panel of German manufacturing firms in response to imports, foreign direct investment and other control variables. Our findings confirm the positive effect of imports and FDI on firms’ innovation activity found in previous literature. However, our coefficient estimates of these variables were smaller than the ones reported by Bertschek and Lechner (1998) and Greene (2004) who analyzed the same dataset under more restrictive model assumptions. This discrepancy can be explained by the exclusion of three far outliers from our estimation and also by our weak model assumptions relative to these authors.

The work presented in this paper can be extended in several directions. First, the parametric EIS used in the classical evaluation of the likelihood function can be replaced by the nonparametric EIS version used in sampling from the posterior of $\tau_i$. Implementation of the nonparametric EIS for approximating the density kernels of the unobserved firm heterogeneity component is currently subject to our research. We anticipate further efficiency improvements in the SML evaluation relative to the present parametric EIS. Second, despite the theoretical appeal of EIS, the current Monte Carlo evidence comparing its performance to other samplers is rather sparse. An MC study comparing both the parametric and nonparametric EIS to, for example, GHK in the SML, MSM, and MSS environments would undoubtedly be of interest to applied researchers using simulation estimators. Furthermore, EIS as a procedure for fast and accurate numerical evaluation of multivariate integrals can be imbedded in more complicated structural models beyond its current reduced form use.
In line with Bertschek and Lechner (1998) and Greene (2004) we have normalized the relative size by the factor of 30.
The autocorrelation functions are based on 12,000 parameter draws.
Appendix 1: Implementation of EIS Likelihood Evaluation

We consider the density kernel $k_i$ for $\tau_i|\tilde{\Lambda}$ as given by

$$k_i(\tau_i; \tilde{\Lambda}, \underline{\alpha}_i) = \exp \left\{ -\frac{1}{2} \left( \underline{b}_i^T \underline{v}_i + \underline{v}_i^T C_i \underline{v}_i \right) - \frac{\tau_i^2}{2\sigma_i^2} \right\} \quad (17)$$

where

$$\underline{b}_i = (b_{1i}, ..., b_{Ti})'$$
$$C_i = \text{diag}(c_i)$$
$$\underline{v}_i = (c_{1i}, ..., c_{Ti})'$$
$$\underline{L}_i = \tilde{\Lambda} + \tau_i \underline{L} + Z_i \underline{L}_i$$
$$\underline{L} = (1, ..., 1)'$$
$$Z_i = (\tilde{z}_{1i}, ..., \tilde{z}_{Ti})'$$

and the auxiliary parameters are $\underline{\alpha}_i = (b_i^0, c_i^0)'$.

Let $\underline{L}_i = \tilde{\Lambda} + Z_i \underline{L}$ which implies $\underline{v}_i = \underline{L}_i + \tau_i \underline{L}$. Then

$$k_i(\tau_i; \tilde{\Lambda}, \underline{\alpha}_i) = \exp \left\{ -\frac{1}{2} \left( \frac{1}{\sigma_i^2} + \underline{L}_i^T C_i \underline{L}_i \right) \tau_i^2 + (\underline{b}_i^0 L_i + 2 \underline{L}_i^T C_i \underline{L}_i \tau_i + \underline{b}_i^0 L_i + \underline{L}_i^T C_i \underline{L}_i) \right\} \quad (18)$$

Matching (18) with a Gaussian kernel we obtain the conditional mean of $\tau_i|\tilde{\Lambda}$ as

$$\mu_i(\underline{\alpha}_i) = -\sigma_i^2 \left( \frac{1}{2} \underline{b}_i^0 L_i + \underline{L}_i^T C_i \underline{L}_i \right) \quad (19)$$

and variance of $\tau_i|\tilde{\Lambda}$ as

$$\sigma_i^2(\underline{\alpha}_i) = \left( \frac{1}{\sigma_i^2} + \underline{L}_i^T C_i \underline{L}_i \right)^{-1}$$
$$= \frac{\sigma_i^2}{1 - \sigma_i^2 \underline{L}_i^T C_i \underline{L}_i} \quad (20)$$

In what follows we will suppress dependence of $\mu_i$ and $\sigma_i^2$ on $\underline{\alpha}_i$ for notational convenience. Integrating $k_i$ with respect to $\tau_i$ leads to the following form of the integrating constant

$$\chi_i(\tilde{\Lambda}; \underline{\alpha}_i) = \sqrt{2\pi\sigma_i} \exp \left\{ -\frac{1}{2} \left( \underline{b}_i^0 L_i + \underline{L}_i^T C_i \underline{L}_i \right) + \frac{1}{2} \mu_i^2 \right\} \quad (21)$$

which itself is a Gaussian density kernel for $\tilde{\Lambda}$.

Let

$$m_i(\tau_i|\tilde{\Lambda}; \underline{\alpha}_i) = \frac{k_i(\tau_i; \tilde{\Lambda}, \underline{\alpha}_i)}{\chi_i(\tilde{\Lambda}; \underline{\alpha}_i)}$$

The EIS regression (without weights) introduced in (14) is derived for each $i$ from (11) and (17) as

$$\ln \phi_i(\tau_i; \tilde{\Lambda}, \underline{\theta}, y) = q_i + \ln k_i(\tau_i; \tilde{\Lambda}, \underline{\alpha}_i) + \xi_i$$

$$-\frac{\tau_i^2}{2\sigma_i^2} + \frac{1}{T} \sum_{t=1}^{T} \left[ (1 - y_{ti}) \ln (1 - \Phi(\tilde{v}_{itr})) + y_{ti} \ln \Phi(\tilde{v}_{itr}) \right]$$

$$= q_i - \frac{1}{2} \left( \underline{b}_i^0 \tilde{v}_{itr} + \underline{v}_i^T C_i \tilde{v}_{itr} \right) + \frac{\tau_i^2}{2\sigma_i^2} + \xi_i$$
\[
\sum_{t=1}^{T} [(1 - y_{it}) \ln [1 - \Phi(\bar{v}_{it})] + y_{it} \ln \Phi(\bar{v}_{it})] = q_i - \frac{1}{2} \left( b_i^2 + \bar{v}_{it}^2 \right) + \xi_{ir} \\
= q_i + (-b_{1t}/2) \bar{v}_{1ir} + \ldots + (-b_{Tt}/2) \bar{v}_{Tir} + (-c_{1t}/2) \bar{v}_{1ir}^2 + \ldots + (-c_{Tt}/2) \bar{v}_{Tir}^2 + \xi_{ir} 
\] 

(22)

with weights

\[
g_i(\bar{v}_{ir}, \lambda, \beta, \varphi) = \exp \left[ -\frac{\bar{v}_{ir}^2}{2\sigma_i^2} \right] \prod_{t=1}^{T} \Phi(\bar{v}_{it})^{y_{it}} [1 - \Phi(\bar{v}_{it})]^{1-y_{it}}
\]

where \( \xi_{ir} \) denotes the regression error term and \( \{ \bar{v}_{ir} : r = 1, \ldots, S \} \) are the simulated draws \( v_{ir} \).

Using (21), the function to be approximated by the Gaussian sampler \( m_0 \) is given by

\[
\phi_0(\lambda; \varphi) = (2\pi)^{-T/2} |\Sigma_\lambda|^{-1/2} \exp \left[ -\frac{1}{2} \lambda' \Sigma_\lambda^{-1} \lambda \right] \sigma_{e}^{-N} (2\pi)^{-N/2} \times \prod_{i=1}^{N} \sqrt{2\pi} \sigma_i \exp \left\{ -\frac{1}{2} \left[ (\beta_i - C_i \beta)^2 + (b_i - C_i \beta)^2 + b_i^2 \right] \right\}
\]

(23)

Consider for the moment the very last term \( \frac{\mu_i^2}{\sigma_i^2} \) of (23)

\[
\frac{\mu_i^2}{\sigma_i^2} = 2\sigma_i^2 \lambda_i \sigma_i \left( -\beta_i + 2\lambda_i \lambda_i + \sigma_i^2 \lambda_i \sigma_i \lambda_i + |\lambda_i| \right)
\]

(24)

Substituting (24) into (23) yields

\[
\phi_0(\lambda; \varphi) \prod_{i=1}^{N} \chi_i(\lambda; \sigma_i) = (2\pi)^{-(T+N)/2} |\Sigma_\lambda|^{-1/2} \sigma_{e}^{-N} (2\pi)^{-N/2} (2\pi)^{1/2} \times \prod_{i=1}^{N} \sigma_i \exp \left\{ -\frac{1}{2} \left[ \lambda_i \lambda_i \sigma_i + \sum_{i=1}^{N} \left( \lambda_i \sigma_i \lambda_i + \lambda_i \lambda_i \right) \right] \right\}
\]

(25)

where

\[
\psi = (2\pi)^{-(T+N+1)/2} |\Sigma_\lambda|^{-1/2} \sigma_{e}^{-N} \prod_{i=1}^{N} \sigma_i \\
\tau = \sum_{i=1}^{N} \left[ \beta_i \beta_i + \lambda_i \lambda_i + \sigma_i^2 \lambda_i \sigma_i \lambda_i + |\lambda_i| \right]
\]

Matching a multivariate Gaussian kernel with (25) yields the variance-covariance matrix of \( \lambda \) on \( m_0 \)

\[
\Sigma_0(\varphi_0) = \left[ \Sigma_\lambda + \sum_{i=1}^{N} [C_i - \sigma_i^2 \lambda_i \lambda_i] \right]^{-1}
\]

(26)

and the mean of \( \lambda \) on \( m_0 \)

\[
\mu_0(\varphi_0) = \Sigma_0 \sum_{i=1}^{N} \left( \sigma_i^2 \lambda_i \left( \beta_i + \frac{1}{2} \lambda_i \lambda_i \right) - \frac{1}{2} \lambda_i - C_i \lambda_i \beta \right)
\]

(27)
Matching the last term of (25) with a multivariate Gaussian kernel we obtain the integrating constant of $\Lambda$ on $m_0$

$$
\chi_0 = (2\pi)^{T/2} |\Sigma_0 (\mathbf{a}_0)|^{1/2} \psi \exp \left\{ \frac{1}{2} \left( r - \mu_0' \Sigma_0^{-1} \mu_0 \right) \right\}
$$

The EIS estimate of the likelihood (13) is calculated from (15) as

$$
\tilde{L}_{r,m}(\theta; y) = (2\pi)^{-(N-1)/2} |\Sigma_0 (\mathbf{a}_0)|^{1/2} |\Sigma_\lambda|^{-1/2} \exp \left\{ -\frac{1}{2} \left( r - \mu_0' \Sigma_0^{-1} \mu_0 \right) \right\} \sigma_r^{-N} \prod_{i=1}^N \sigma_i \prod_{i=1}^N \phi_i(\tilde{\tau}_{ir}, \tilde{\lambda}_r; \theta, y) k_i(\tilde{\tau}_{ir}|\tilde{\lambda}_r; \mathbf{a}_i)
$$

and the log-likelihood as

$$
\ln \tilde{L}_{S,m}(\theta; y) = \frac{1}{S} \sum_{r=1}^S \ln \tilde{L}_{r,m}(\theta; y) = - \frac{N-1}{2} \ln (2\pi) + \frac{1}{2} \left( \ln |\Sigma_0 (\mathbf{a}_0)| - \ln |\Sigma_\lambda| \right) - \frac{1}{2} \left( r - \mu_0' \Sigma_0^{-1} \mu_0 \right) - N \ln \sigma_r
$$

$$
+ \sum_{i=1}^N \ln \sigma_i + \ln \left[ \frac{1}{S} \sum_{r=1}^S \exp \left\{ \sum_{i=1}^N \left( \ln \phi_i(\tilde{\tau}_{ir}, \tilde{\lambda}_r; \theta, y) - \ln k_i(\tilde{\tau}_{ir}|\tilde{\lambda}_r; \mathbf{a}_i) \right) \right\} \right]
$$

**Algorithm**

Based on these derivations, the computation of an efficient MC estimate of the likelihood for the panel probit model requires the following steps:

**Step (1):** Use the natural sampling density $p$ in (9) to draw $S$ independent realizations of the latent process $(\tilde{\tau}, \tilde{\lambda})$.

**Step (2):** Use these random draws to solve the sequence of $N$ weighted (unweighted for the first iteration of the importance sampling construction) LS problems defined in equation (22).

**Step (3):** Use the sampling density from $m_0$ with moments given in (26) and (27) to draw $S$ trajectories $\left\{ \tilde{\lambda}_r(\mathbf{a}_0) : r = 1, ..., S \right\}$. Conditional on these trajectories, draw from the conditional densities $\{m_i\}$ characterized by the moments (19) and (20) the vectors $\left\{ \tilde{\tau}_{r}, \tilde{\lambda}_1(\mathbf{a}_1), ..., \tilde{\lambda}_N(\mathbf{a}_N) : r = 1, ..., S \right\}$. Throughout the text, these draws are denoted by a shorthand notation $\tilde{\tau}_{r}$ and $\tilde{\lambda}_r$.

**Step (4):** Maximize the simulated log-likelihood (28), evaluated at $\tilde{\tau}_{r}$ and $\tilde{\lambda}_r$, in each step, with respect to the parameters $\theta$. 
Appendix 2: Sampling from Posterior Densities

Sampling from $f(\beta|\theta, Y, Z)$

Here we adopt the methodology elaborated in (Albert and Chib, 1993). In our panel application,

$$Y_i^* = Z_i\beta + \Lambda + \tau_i4 + \epsilon_i$$
$$Y_{i,T,i}^* = Y_i^* - \Lambda - \tau_i4 + \epsilon_i$$
$$Y_{i,T,i}^* = Z_i\beta + \epsilon_i$$

Assigning a noninformative prior $\pi(\beta)$ to $\beta$ results in

$$f(\beta|\theta, Y, Z) = N(\tilde{\beta}, \tilde{\Sigma}_{\beta})$$ (29)

where $\tilde{\beta} = (Z'Z)^{-1} Z'Y_{i,T}^*$, the dependent variable is a $(NT \times k)$ matrix $Y_{i,T}^* = (Y_{i,T,1}^*, ..., Y_{i,T,N}^*)'$ and $\tilde{\Sigma}_{\beta} = (Z'Z)^{-1}$. The random variables $Y_{i,t}^*$ are independent with

$$f(Y_{i,t}^*|\theta, Y, Z) = N(\mu_{i,t}^*, 1)$$
$$\mu_{i,t}^* = Z_{it}\beta + \lambda_t + \tau_i$$ (30)

truncated at the left by 0 if $Y_{i,t} = 1$ and truncated at the right by 0 if $Y_{i,t} = 0$. Given a previous value of $\beta$, $\tau_i$ and $\lambda_t$, one cycle the Gibbs algorithm would produce $Y_{i,t}^*$ and $\beta$ from the distributions (30) and (29); see Train (2003, p. 210) for simulation algorithm. The starting value $\beta^{(0)}$ may be taken to be the ML estimate.

Sampling from $f(\beta|\theta, Y, Z)$

From (10),

$$f(\beta|\theta, Y, Z) \propto \sigma^{-1}_\tau \exp \left[ -\frac{1}{2\sigma^2_\tau} \tau^2 \right] \prod_{t=1}^{T} \Phi(v_{i,t})^{y_{i,t}} [1 - \Phi(v_{i,t})]^{1-y_{i,t}}$$ (31)

The posterior $f(\beta|\Lambda, \theta, Z)$ is a convolution of a gaussian density and a product of standard normal cdfs. As such, it can be asymmetric with the direction of skewness depending on the particular realization of the vector of dependent variables $y$. Therefore, for our simulator we use a piece-wise linear approximation to $f(\beta|\Lambda, \theta, Z)$ which is a form of nonparametric EIS capable of accurately sampling from any distribution irrespective of its shape. The procedure works as follows. First, we obtain an empirical distribution function of $f(\beta|\Lambda, \theta, Z)$ evaluated over an equispaced grid of $\tau_i$ around the importance region and then we invert $S$ draws from $U[0,1]$ through this edf to obtain a new grid whose values are concentrated in the importance region. We update the edf over this grid and iterate this process until the change of the maxima of the edf parameters (intercept and slope of individual segments) converges within a tolerance level around zero. Then we invert one draw from $U[0,1]$ for the given $\tau_i$ via the final edf to obtain the new value of the $\tau_i$ in the Gibbs block. Aside from shape adaptability, another advantage of this nonparametric form of EIS is that the degree of accuracy of this procedure can be made arbitrarily precise by increasing the size of the mesh, at the expense of computational cost.

Sampling from $f(\Lambda|\beta, \theta, Z)$

From (10),

$$f(\Lambda|\beta, \theta, Z) \propto p(\Lambda|\theta) \prod_{t=1}^{T} \prod_{i=1}^{N} \Phi(v_{i,t})^{y_{i,t}} [1 - \Phi(v_{i,t})]^{1-y_{i,t}}$$

$$= p(\Lambda|\theta) \prod_{t=1}^{T} g(\lambda_t)$$
where
\[ g(\lambda_t) = \prod_{i=1}^{N} [\Phi(v_{it})]^{y_{it}} [1 - \Phi(v_{it})]^{1 - y_{it}} \]

The joint EIS sampler takes the form
\[
m(\Delta|\theta, Z, \gamma) = p(\Delta|\theta) \prod_{t=1}^{T} h(\lambda_t; \gamma) = p(\lambda_1|\theta) \prod_{t=2}^{T} p(\lambda_t|\lambda_{t-1}, \theta) \prod_{t=1}^{T} h(\lambda_t; \gamma)
\]

where
\[
p(\Delta|\theta) = (2\pi)^{-T/2} |\Sigma_{\lambda}|^{-1/2} \exp \left[ -\frac{1}{2} \lambda^T \Sigma_{\lambda}^{-1} \lambda \right]
p(\lambda_1|\theta) = \frac{\sqrt{1 - \rho^2}}{\sigma_\eta \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma_\eta^2} (\lambda_1 - \rho \lambda_{t-1})^2 \right]np(\lambda_t|\lambda_{t-1}, \theta) = \frac{1}{\sigma_\eta \sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma_\eta^2} (\lambda_t - \rho \lambda_{t-1})^2 \right]h(\lambda_t; \gamma) = N(\mu_\lambda, \Omega_\lambda)\]
\[
\prod_{t=1}^{T} h(\lambda_t; \gamma) = NID(\mu_\lambda, \Omega_\lambda)\]
\[
\Omega_\lambda = \text{diag}(\sigma^2_\lambda)
\]
with \( \mu_\lambda = (\mu_{\lambda_1}, \ldots, \mu_{\lambda_T})' \) and \( \sigma^2_\lambda = (\sigma^2_{\lambda_1}, \ldots, \sigma^2_{\lambda_T})' \).

Let
\[
\ln k_t(\lambda_t|\pi, \theta, Z, \gamma) = -\ln \lambda_{\pi_t} + \ln h(\lambda_t; \gamma_t) = \gamma_{t,1} \lambda_t + \gamma_{t,2} \lambda_t^2
\]

The EIS regression for \( t = 1, \ldots, T \) is
\[
\ln g(\lambda_t) = \gamma_{t,0} + \ln k_t(\lambda_t|\pi, \theta, Z, \gamma) = \gamma_{t,0} + \gamma_{t,1} \lambda_t + \gamma_{t,2} \lambda_t^2
\]

Then
\[
\mu_{\lambda_t} = -\frac{1}{2} \frac{\gamma_{t,1}}{\gamma_{t,2}}
\]
\[
\sigma^2_{\lambda_t} = -\frac{1}{2} \frac{1}{\gamma_{t,2}}
\]

and
\[
m(\Delta|\pi, \theta, Z, \gamma) = p(\Delta|\theta) \prod_{t=1}^{T} h(\lambda_t; \gamma) = p(\Delta|\theta) NID(\mu_\lambda, \Omega_\lambda)
\]
\[
= \exp \left\{ -\frac{1}{2} \left( \lambda^T \Sigma_\lambda^{-1} + \Omega_\lambda^{-1} \right) \Delta - 2 \mu_\lambda^T \Omega_\lambda^{-1} \Delta + \mu_\lambda^T \Omega_\lambda^{-1} \mu_\lambda \right\}
\]
Matching (32) with a multivariate Gaussian kernel yields

\[ \Sigma_m = \left[ \Sigma_\lambda^{-1} + \Omega_\lambda^{-1} \right]^{-1} \]

and

\[ \mu'_{m} = \frac{\mu'_{t} \Omega^{-1}_{m}}{\mu'_{t} \Omega^{-1}_{m} + \lambda^{-1} \left[ \Sigma^{-1}_{\lambda} + \Omega^{-1}_{\lambda} \right]^{-1}} \]

In the AR-MH algorithm, we will utilize

\[ g(\lambda_t) \approx \exp(\gamma_{t,0})h(\lambda_t, \gamma) \]

\[ = \exp(\gamma_{t,0})k_t(\lambda_t | \theta, Z, \gamma) \]

and

\[ p(\lambda | \theta) \prod_{t=1}^{T} g(\lambda_t) \approx M(\lambda | \theta, Z, \gamma) \]

\[ = p(\lambda | \theta) \prod_{t=1}^{T} \exp(\gamma_{t,0})k_t(\lambda_t | \theta, Z, \gamma) \]

\[ = (2\pi)^{-T/2} |\Sigma_{\lambda}|^{-1/2} \exp \left[ -\frac{1}{2} \sum_{t=0}^{2} \gamma_{t,0} + \gamma_{t,1} \lambda_1 + \gamma_{t,2} \lambda_2 \right] \]

**AR-MH Algorithm**

Given \( K \) draws \( \{\lambda_1, \ldots, \lambda_K\} \) from the EIS-MCMC algorithm, potential new candidate draws are sampled from \( m(\lambda | \tau, \theta, Z, \gamma) \) until acceptance of a candidate \( \hat{\lambda} \) in the AR step with probability

\[ P(\hat{\lambda}) = \min \left( \frac{f(\hat{\lambda} | \tau, \theta, Z)}{M(\hat{\lambda} | \tau, \theta, Z, \gamma)}, 1 \right) \]

In the MH-step \( \hat{\lambda} \) is accepted as the \( K + 1 \)-th draw \( \lambda_{K+1} \) from the EIS-MCMC algorithm with probability \( \alpha(\lambda_{K}, \lambda_{K+1}) \), otherwise \( \lambda_{K+1} \) is set to equal \( \lambda_K \). It holds that

\[ \alpha(\lambda_{K}, \lambda) = \min \left( \frac{f(\lambda | \tau, \theta, Z) \min \left[ f(\lambda | \tau, \theta, Z), M(\lambda | \tau, \theta, Z, \gamma) \right]}{f(\lambda | \tau, \theta, Z) \min \left[ f(\lambda | \tau, \theta, Z), M(\lambda | \tau, \theta, Z, \gamma) \right]}, 1 \right) \]

The AR-MH step for \( \tau_t \) is repeated 10 times before the parameters are updated in the Gibbs sequence.

**Sampling from** \( f(\sigma_t^2 | \theta_t, \gamma_t, \tau, Z) \)

We follow the same philosophy of simulated data augmentation as applied in (Albert and Chib, 1993) to draws from \( f(\beta | \theta, \gamma, \tau, Z) \). Since

\[ \tau_t \sim N(0, \sigma_t^2) \]

the likelihood of the sample \( \tau_t \) treated as a function of \( \sigma_t^2 \), is

\[ L(\tau | \sigma_t^2) = (2\pi \sigma_t^2)^{-N/2} \exp \left( -\frac{1}{2} \frac{S_t}{\sigma_t^2} \right) \]
where
\[ S_r = \sum_{i=1}^{N} \tau_i^2 \]

A commonly used prior for the variance of Gaussian random variables is the inverted gamma-2 density \( IG(s_0, v_0) \) with kernel
\[ k_\tau(\sigma_\tau^2) = \sigma_\tau^{-(v_0+2)} \exp \left( -\frac{s_0}{2\sigma_\tau^2} \right) \]
(see Train, 2003, ch. 12). The posterior is then
\[
f(\sigma_\tau^2|\theta, \tau, Y, Z) \propto \sigma_\tau^{-N} \exp \left( -\frac{1}{2} \frac{S_r}{\sigma_\tau^2} \right) \sigma_\tau^{-(v_0+2)} \exp \left( -\frac{s_0}{2\sigma_\tau^2} \right) = \sigma_\tau^{-(v_0+2+N)} \exp \left( -\frac{1}{2} \left( \frac{S_r + s_0}{\sigma_\tau^2} \right) \right) = \sigma_\tau^{-(v_3+2)} \exp \left( -\frac{1}{2} \frac{s_1}{\sigma_\tau^2} \right)
\]
which is a kernel of \( IG(s_1, v_1) \) with \( s_1 = s_0 + S_r \) and \( v_1 = v_0 + N \).

Following Bauwens et al. (1999, p. 114) we specify a non-informative prior \( \pi(\sigma_\tau^2) \) as the limit of the \( IG(s_0, v_0) \) kernel
\[ k_\tau(\sigma_\tau^2) = \sigma_\tau^{-(v_0+2)} \exp \left( -\frac{s_0}{2\sigma_\tau^2} \right) \]
where \( s_0 \to 0 \) and \( v_0 = 0 \). Thus
\[
f(\sigma_\tau^2|\theta, \tau, Y, Z) = IG(s_1, v_1) \quad s_1 = S_r \quad v_1 = N
\]

To draw from this posterior, draw \( z \sim U[0, 1] \), compute \( y_1 = Ga^{-1} \left( \frac{s_1}{2\tau}, 1, z \right) \), \( y_2 = \frac{2}{s_1} y_1 \) and \( \sigma_\tau^2 = y_2^{-1} \).

**Sampling from** \( f(\rho|\theta, \rho, Y, Z) \)

The time random effects \( \lambda_t \) are assumed to follow a stationary autoregressive process of order \( p \)
\[
A(L) \lambda_t = \sum_{i=0}^{p} (a_i L^i) \lambda_t = \eta_t
\]
with \( \eta_t \sim N(0, \sigma^2_\eta) \). For AR(1) process, the likelihood function is given by
\[
L(\theta; y) \propto f(\lambda_1, ..., \lambda_T | \sigma^2_\eta, \rho) = \prod_{t=1}^{T} p(\lambda_t | \lambda_{t-1}, \cdot)
\]
where \( f(\lambda_1, ..., \lambda_T | \sigma^2_\eta, \rho) \) is the joint density of \( \{\lambda_t\}_{t=1}^{T} \), and
\[
p(\lambda_t | \lambda_{t-1}, \cdot) \propto \begin{cases} \exp \left( -\frac{(1-\rho^2)}{2\sigma^2_\eta} \lambda_t^2 \right), & t = 1 \\ \exp \left( -\frac{1}{2\sigma^2_\eta} (\lambda_t - \rho \lambda_{t-1})^2 \right), & t = 2, ..., T \end{cases}
\]
The joint density is given by
\[
f(\lambda_1, ..., \lambda_t | \sigma^2, \rho) \propto \frac{1}{\sqrt{2\pi \sigma^2(1-\rho^2)}} \exp \left( -\frac{(1-\rho^2)}{2\sigma^2} \lambda^2 \right) \times \prod_{t=1}^{T} \left\{ \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{1}{2\sigma^2}(\lambda_t - \rho\lambda_{t-1})^2 \right) \right\} \times \alpha_{\rho} \exp \left( -\frac{1}{2} \left[ \rho \frac{1}{\sigma^2} \left( \sum_{t=2}^{T} \lambda_{t-1}^2 - \lambda_1^2 \right) - \frac{2}{\sigma^2} \sum_{t=2}^{T} \lambda_t \lambda_{t-1} + \frac{1}{\sigma^2} \sum_{t=1}^{T} \lambda_t^2 \right] \right) \] (33)

where
\[
\alpha_{\rho} = \sqrt{(1-\rho^2)}
\]

Matching (33) with a Gaussian kernel yields
\[
\begin{align*}
\sigma^2_{\rho} &= \sigma^2 \left( \sum_{t=2}^{T} \lambda_t^2 \right)^{-1} \\
\mu_{\rho} &= \frac{\sigma^2}{\sigma^2_{\eta}} \sum_{t=2}^{T} \lambda_t \lambda_{t-1} \\
\gamma_{\rho} &= -\frac{1}{2\sigma^2} \sum_{t=1}^{T} \lambda_t^2 - \frac{\mu_{\rho}^2}{\sigma^2_{\rho}}
\end{align*}
\]

Hence, draw \( \rho \) from \( \frac{1}{\exp(\gamma_{\rho})} N(\mu_{\rho}, \sigma^2_{\rho}) \) truncated to \( |\rho| < 1 \).

**Sampling from \( f(\sigma^2_{\eta}|\theta/\sigma, \gamma, Z) \)**

For a given \( \lambda \) and \( \rho \) the likelihood function can be formulated as
\[
L(\sigma_{\eta}; \alpha, y) \propto \sigma_{\eta}^{-T} \left[ \exp \left( \frac{S_{\lambda}}{2\sigma^2_{\eta}} \right) \right] \]

where, for AR(1),
\[
S_{\lambda} = (1-\rho^2)\lambda_1^2 \sum_{t=2}^{T} (\lambda_t - \rho\lambda_{t-1})^2
\]

Similarly to the case of \( f(\sigma^2_{\eta}|\theta/\sigma, \gamma, Z) \), we postulate the prior on \( \sigma^2_{\eta} \) as the inverted gamma-2 density \( IG(s_0, v_0) \) with kernel
\[
k_{\lambda}(\sigma^2_{\lambda}) = \sigma_{\lambda}^{-(v_0+2)} \exp \left( -\frac{s_0}{2\sigma^2_{\lambda}} \right)
\]

The posterior becomes
\[
f(\sigma^2_{\eta}|\theta/\sigma, \gamma, Z) \propto \sigma_{\eta}^{-T} \left[ \exp \left( \frac{-S_{\lambda}}{2\sigma^2_{\eta}} \right) \right] \sigma_{\eta}^{-(v_0+2)} \exp \left( -\frac{s_0}{2\sigma^2_{\lambda}} \right)
\]
\[
= \sigma_{\eta}^{-(2+T+v_0)} \exp \left( -\frac{1}{2} \frac{S_{\lambda} + s_0}{\sigma^2_{\eta}} \right)
\]
\[
= \sigma_{\eta}^{-(v_1+2)} \exp \left( -\frac{1}{2} \frac{S_{\lambda}}{\sigma^2_{\eta}} \right)
\]

24
which is a kernel of $IG(s_1, v_1)$ with $s_1 = s_0 + S$ and $v_1 = v_0 + T$. We again specify a non-informative prior $\pi(\sigma^2)$ as the limit of the $IG(s_0, v_0)$ kernel with $s_0 \to 0$ and $v_0 = 0$. Thus

$$f(\sigma^2 | g_1, \sigma_2, Y, Z) = IG(s_1, v_1)$$
$$s_1 = S$$
$$v_1 = T$$

To draw from this posterior, draw $z \sim U[0, 1]$, compute $y_1 = Ga^{-1} \left( \frac{1}{2}, 1, z \right)$, $y_2 = \frac{2}{\sigma^2} y_1$ and $\sigma^2_2 = y_2^{-1}$. 

References


