University of Toronto
Department of Economics

Working Paper 321

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

By Martin Burda, Roman Liesenfeld and Jean-Francois Richard

June 16, 2008

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

Martin Burda† Roman Liesenfeld‡ Jean-François Richard§

First draft: August 22, 2006
This version: May 16, 2008

Abstract

In this paper, we perform Bayesian analysis of a panel probit model with unobserved individual heterogeneity and serially correlated errors. We augment the data with latent variables and sample the unobserved heterogeneity component as one Gibbs block per individual using a flexible piecewise linear approximation to the marginal posterior density. The latent time effects are simulated as another Gibbs block. For this purpose we develop a new user-friendly form of the Efficient Importance Sampling proposal density for an Acceptance-Rejection Metropolis-Hastings step. We apply our method to the analysis of product innovation activity of a panel of German manufacturing firms in response to imports, foreign direct investment and other control variables. The dataset used here was analyzed under more restrictive assumptions by Bertschek and Lechner (1998) and Greene (2004). Although our results differ to a certain degree from these benchmark studies, we confirm the positive effect of imports and FDI on firms' innovation activity. Moreover, unobserved firm heterogeneity is shown to play a far more significant role in the application than the latent time effects.

Keywords: Dynamic latent variables, Markov Chain Monte Carlo, importance sampling.
JEL Classification: C11, C13, C15, C23, C25.

---

*The authors would like to thank session participants of the 14^{th} International Conference on Panel Data at WISE, Xiamen University, China, the 2008 Conference of the Society for Nonlinear Dynamics & Econometrics, San Francisco, CA, and seminar participants at Pittsburgh and HEC, Montreal, for helpful comments and suggestions.

†Department of Economics, University of Toronto, 5025A Sidney Smith Hall, Toronto, ON, M5S 3G7, Canada. Phone: (416) 978-4479; Email: martin.burda@utoronto.ca

‡Institute for Statistics and Econometrics, Christian-Albrecht-Universität, Ohlshausenstr. 40-60, 24118 Kiel, Germany. Phone: +49-431-880-2166; E-mail: liesenfeld@stat-econ.uni-kiel.de

§Department of Economics, University of Pittsburgh, 4917 Posvar Hall, Pittsburgh, PA 15213, USA. Phone: 412-648-1750; E-mail: fantin@pitt.edu

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.

Fuelled by advances in computation, the last two decades witnessed an explosion of interest in Bayesian models using data augmentation (Tanner and Wong, 1987) that had previously been regarded as numerically unfeasible. Under this framework, the latent variables within these multivariate integrals are treated as model parameters and hence are sampled along with the key model parameters of economic interest. The Bayesian Gibbs sampling scheme is naturally suited for such purpose: an often high-dimensional multivariate integral forming the likelihood function is factorized into a sequence of low-dimensional conditional densities each of which is sampled individually. Proceeding iteratively along a Markov Chain of these low-dimensional subproblems yields draws from the joint posterior which are the used directly for inference.

Due to their flexibility and conceptual simplicity, Bayesian methods successfully compete against simulation-based frequentist techniques, such as Simulated Maximum Likelihood (SML)[1]. The advantages of the former become more pronounced with increased dimensionality of the underlying problem. For example, in our setup the SML approach would require a large number of latent variable draws for each parameter value in order to approximate numerically the integral likelihood function to an acceptable degree of accuracy. In contrast, Gibbs sampling takes one latent variable draw for each parameter value. In many cases, this implies that Bayesian parameter estimation is faster than SML. Multiple local modes in the SML objective function for a given dataset are another concern which is completely bypassed using the Bayesian setup. Numerous other technical details regarding the advantages of Bayesian inference in latent variable models are discussed in Paap (2002). Moreover, Bayesian hierarchical models can be readily extended to incorporate inference on latent classes of similar individuals or mixtures of distributions for various objects of interest (see e.g. Rossi et al., 2005). Implementation of such features in high-dimensional latent variable models using SML would arguably pose a far more taxing problem.

In this paper, we perform 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 context of dynamic Bayesian models, previous literature has considered either a latent time dimension or latent individid-

[1] Gourieroux and Monfort (1996) provide the essential statistical background for the SML estimator.

ual heterogeneity (Paap, 2002; Franses, 2006; Liesenfeld and Richard, 2006). Our approach, based on proposal densities constructed with the Efficient Imporance Sampling (EIS) procedure (Richard and Zhang, 2007), combines both features and thus allows for inference within a rich economic model environment. Specifically, we augment the data with latent variables and sample the unobserved individual heterogeneity component as one Gibbs block per individual, drawing from a piecewise 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 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 frequentist estimators under more restrictive assumptions providing a useful benchmark for comparison with our results.[2] 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 3 we elaborate on our estimation technique. The results of our empirical application are discussed in Section 4. Section 5 concludes. A simulation study with artificial data is reported in the Appendix.

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

[2] Similar data set was used in an interesting paper by Inkmann (2000) but with some regressors different from ours.

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 $\underline{z}_{it}$ and the corresponding vector of parameters to be estimated by $\underline{\beta}$. 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. All 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.[3] 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 τ_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 economy-wide factors influencing all companies alike 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 λ_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

[3] An illuminating discussion is provided in Falcetti and Tudela (2006, p. 454), drawing on Heckman (1981) and Börsch-Supan et al. (1992).

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

$$y_{it}^* = \underline{\beta}_0' \underline{z}_{it} + \epsilon_{it} \tag{1}$$

with the observation rule

$$y_{it} = \mathbf{1}\left(y_{it}^* \geq 0\right), \qquad i : 1, ..., N \ ; \qquad t : 1, ...T \tag{2}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. The error terms ϵ_{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 $\underline{\epsilon}_i = (\epsilon_{i1}, ..., \epsilon_{iT})'$ are modeled as jointly normally distributed with mean zero and covariance matrix Σ. Also, $\underline{\epsilon}_i$ 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 Σ are set to unity to facilitate identification of $\underline{\beta}$ 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

$$\begin{aligned} E[W(Z, \beta_0)|X] &= 0 \\ W(z, \beta) &= [w_1(Z_1, \beta), ..., w_T(Z_T, \beta)]' \\ w_t(Z_t, \beta) &= Y_t - \Phi(\underline{\beta}' \underline{z}_{it}) \end{aligned} \tag{3}$$

where Φ 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 Σ. 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 β_0 is then defined as

$$\widehat{\beta}_N = \arg\min_{\beta} g'_N(\beta)\Omega^{-1}g_N(\beta) \tag{4}$$

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 Ω 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

$$y^*_{it} = \underline{\beta}'_0 \underline{z}_{it} + \epsilon_{it}$$

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

$$\underline{\beta}_i = \mu + \Delta z_i + \Gamma w_i$$

where μ is $K \times 1$ vector of location parameters, Δ is $K \times L$ matrix of unknown location parameters, Γ 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 $E[\underline{\beta}_i|X_i, z_i] = \mu + \Delta z_i$ and $Var[\underline{\beta}_i|X_i, z_i] = \Gamma V \Gamma'$. Conditional on w_i, observations of y_{it} are independent across time; timewise correlation would arise through correlation of elements of $\underline{\beta}_i$. The joint conditional density on y_{it} is

$$f(y_i|X_i,\beta) = \prod_{t=1}^{T} \Phi[(2y_{it}-1)\underline{\beta}'\underline{z}_{it}] \tag{5}$$

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

$$\log L = \sum_{i=1}^{N} \log L_i = \sum_{i=1}^{N} \log \int_{\beta_i} \prod\nolimits_{t=1}^{T} \Phi[(2y_{it} - 1)\underline{\beta}' \underline{z}_{it}] g(\beta_i | \mu, \Delta, \Gamma, z_i) d\beta_i \tag{6}$$

Estimates of μ, Δ and Γ 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 β_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 β_i has finite, discrete support over J points (classes) with probabilities $p(\beta_j | \mu, \Delta, \Gamma, z_i)$, $j = 1, ..., J$, then the resulting formulation of the analog of L_i from (6) is

$$L_i = \sum\nolimits_{j=1}^{J} p(\beta_j | \mu, \Delta, \Gamma, z_i) f\left(y_i | X_i, \beta_j\right)$$

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

$$y_{it}^* = \underline{\beta}' \underline{z}_{it} + \tau_i + \lambda_t + \epsilon_{it}, \qquad i : 1, ..., N \; ; \qquad t : 1, ...T \tag{7}$$

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

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

where $\eta_t \sim N(0, \sigma_\eta^2)$ such that the mean of λ_t is zero and the variance σ_λ^2 is stationary. It is assumed that ϵ_{ti}, τ_i and η_t are mutually independent. The vector of parameters to be estimated is $\underline{\theta} = (\underline{\beta}', \sigma_\tau, \rho_1, ..., \rho_k, \sigma_\eta)'$. Denote $\underline{\lambda} = (\lambda_1, ..., \lambda_T)'$ and $\underline{\tau} = (\tau_1, ..., \tau_N)'$.

The likelihood function associated with $\underline{y} = (y_{11}, ..., y_{TN})'$ can be written as

$$L(\underline{\theta}; \underline{y}) = \int g(\underline{\tau}, \underline{\lambda}; \underline{\theta}, \underline{y}) p(\underline{\tau}, \underline{\lambda}; \underline{\theta}) d\underline{\tau} d\underline{\lambda} \tag{8}$$

with

$$g(\underline{\tau}, \underline{\lambda}; \underline{\theta}, \underline{y}) = \prod_{i=1}^{N} \prod_{t=1}^{T} [\Phi(v_{it})]^{y_{it}} [1 - \Phi(v_{it})]^{1-y_{it}}$$

where

$$\Phi(v_{ti}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{v_{ti}} \exp\left(-\frac{1}{2}t^2\right) dt$$

$$v_{it} = \underline{\beta}' \underline{z}_{it} + \tau_i + \lambda_t$$

$$p(\underline{\tau}, \underline{\lambda}; \underline{\theta}) = \sigma_\tau^{-N} (2\pi)^{-N/2} \exp\left[-\frac{1}{2\sigma_\tau^2} \sum_{i=1}^{N} \tau_i^2\right] (2\pi)^{-T/2} |\Sigma_\lambda|^{-1/2} \exp\left[-\frac{1}{2} \underline{\lambda}' \Sigma_\lambda^{-1} \underline{\lambda}\right] \tag{9}$$

and Σ_λ denotes the stationary variance-covariance matrix of $\underline{\lambda}$.

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 $\underline{\tau}$ and $\underline{\lambda}$. The complete joint posterior $f(\underline{\theta}, \underline{\tau}, \underline{\lambda}|Z)$ can then be drawn from using Gibbs sampling. The main difficulty with such an MCMC approach is that of efficiently sampling from τ_i and $\underline{\lambda}$ since the corresponding multivariate posterior distributions are high-dimensional and have no closed-form solution.

To overcome this problem, Liesenfeld and Richard (2006) 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 along the time dimension. In this paper, we also take the general approach of combining EIS with AR-MH but introduce a new user-friendly parametrization of the EIS proposal density for the time dimension $\underline{\lambda}|\underline{\theta}, Z$. Specifically, we approximate with a first-stage EIS kernel only the part of the likelihood that arises from the LDV model specification, and then recombine this approximation analytically with the known AR(p) likelihood assumed for the latent time process to form the desired second-stage EIS proposal density. Thus we avoid the need to pass integrating constants across periods which results in significant simplification of constructing the proposal density. The unobserved individual heterogeneity component $\tau_i|\underline{\theta}, Z$ as N individual Gibbs blocks drawing from a piecewise linear approximation to the marginal posterior density constructed with a nonparametric form of EIS. The basis of these procedures is that the EIS proposal densities for $\tau_i|\underline{\theta}, Z$ and $\underline{\lambda}|\underline{\theta}, Z$ provide very close approximations to $f(\tau_i|\underline{\theta}, Z)$ and $f(\underline{\lambda}|\underline{\theta}, Z)$, respectively. The piece-wise linear approximation to $f(\tau_i|\underline{\theta}, Z)$ freely adapts to the shape of the posterior can be made arbitrarily precise by increasing the size of the simulated grid. For $f(\underline{\lambda}|\underline{\theta}, Z)$ given the model assumptions, one can expect that the EIS parametric density provides an efficient proposal density for the target posterior $f(\underline{\lambda}|\underline{\theta}, Z)$ in the AR-MH step. This conjecture has been validated by AR-MH acceptance rates close to 100% in our empirical application.

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

$$f(\theta_j|\underline{\theta}_{/\theta_j}, \Upsilon, Z) \propto L(\underline{\theta}, \underline{\tau}, \underline{\lambda}; Z)\pi(\theta_j) \tag{10}$$

The individual Gibbs blocks used are $\underline{\beta}$, σ_τ, σ_η, $\underline{\rho}$, $\underline{\lambda}$, and $\underline{\tau}$, given data and the remaining augmented parameters. Throughout the analysis we make use of diffuse 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 Ω 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 Bayesian posterior means and medians of parameters in the model (7). Posterior marginal densities of the Bayesian analysis, MCMC chains and autocorrelation functions of the parameter draws are presented in Figures 2 - 6. The latter two results indicate very good mixing properties of the Markov chains.

Table 1: Models 1-3

	Pooled Probit[a]		Model 1[b]		Model 2[c]	
Variable	Estimate	Std.Err.	Estimate	Std.Err.	Estimate	Std.Err.
Constant	−1.960**	0.230	−1.74**	0.37	−3.134	0.191
log sales	0.177**	0.022	0.15**	0.034	0.306	–
Rel size	1.072**	0.142	0.95**	0.20	3.735	0.184
Imports	1.133**	0.151	1.14**	0.24	1.582	0.126
FDI	2.853**	0.402	2.59**	0.59	3.111	0.320
Prod.	−2.341**	0.715	−1.91**	0.82	−5.786	0.755
Raw Mtl	−0.279**	0.081	−0.28**	0.12	−0.346	0.077
Inv good	0.188**	0.039	0.21**	0.063	0.238	0.453

[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]

	Class 1		Class 2		Class 3	
Variable	Estimate	Std.Err.	Estimate	Std.Err.	Estimate	Std.Err.
Constant	−2.32**	0.768	−2.71**	0.766	−8.97**	2.50
log sales	0.323**	0.075	0.233**	0.0675	0.571**	0.197
Rel size	4.38**	0.882	0.720**	0.253	1.42*	0.616
Imports	0.936**	0.491	2.26**	0.503	3.12*	1.35
FDI	2.20	2.54	2.80**	0.926	8.37**	2.27
Prod.	−5.86**	1.69	−7.70**	1.16	−0.910	1.26
Raw Mtl	−0.110	0.172	−0.599**	0.295	−0.856*	0.424
Inv good	0.131	0.143	0.413**	0.132	0.469*	0.225

[d] Greene (2004), Table 7
* Indicates significant at the 95% level
** Indicates significant at the 99% level

We excluded from estimation three distant outliers with relative firm size larger than 0.1 and productivity larger than 0.8 (see Figure 1) as these observations may potentially induce numerical instabilities. The three excluded observations with large relative size have also disproportionately large values of *import share* and *FDI* - our two key variables of interest. 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 82^{nd} percentile and 98^{th} percentile, respectively, of the remaining observations of these variables. The exclusion reduced our sample size to $N = 1267$ and $T = 5$.

Table3: EIS-MCMC

Variable	Posterior mean	Posterior median	Std.Dev.
Constant	-2.399^{**}	-2.393^{**}	0.548
log sales	0.2428^{**}	0.2428^{**}	0.053
Rel size	1.3443^{**}	1.3408^{**}	0.324
Imports	1.5598^{**}	1.5634^{**}	0.362
FDI	3.5451^{**}	3.5250^{**}	0.906
Prod.	-5.2336^{*}	-5.2621^{*}	2.886
Raw Mtl	-0.2203	-0.2231	0.245
Inv good	0.2985^{**}	0.2991^{**}	0.093
σ_τ	1.1625^{**}	1.1610^{**}	0.046
σ_η	0.4729^{**}	0.4373^{**}	0.172
ρ	0.0025	-0.0099	0.498

Posterior moments are based on $20,000$ Gibbs cycles discarding the first $5,000$ cycles and keeping every fifth draw thereafter resulting in 3000 MC draws for each parameter. One Gibbs iteration took approximately 3.5 seconds on a 2.2 GHz unix machine. The nonparametric EIS sampler was perfomed over a grid of size 200. 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 τ_i and λ. The AR and MH acceptance rates for λ were 99.00% and 99.85%, respectively.

All of our coefficient estimates fit into a convex combination of the results found in the previous literature reported in Tables 1 and 2. In most cases, our finding are close to the mean values of the previous findings. The pattern of parameter significance matches closely previous results with the exception of sector dummy variables; these were previously found either both significantly different from zero or the converse. In our case, the *raw materials* dummy turned out not significant while the *investment good* dummy was estimated as significant. The estimates of the two key parameters of *FDI* and *import share* are positive, further validating the original economic hypothesis that imports and inward FDI had a positive effect on the innovative activity of domestic firms.

The posterior mean of the unobserved heterogeneity parameter σ_τ was estimated at 1.1625 which is closely matches the value 1.1707 of an analogous parameter reported by Greene (2004, p.35) for the random effects model. In addition, the posterior mean standard deviation of the latent time effects σ_η was estimated at 0.4729 which is roughly half the magnitude of its cross-sectional counterpart. The posterior mean of the autoregressive parameter ρ is not statistically different from zero. Unobserved individual heterogeneity thus appears to play a more important role in this application than latent time effects.

5 Conclusion

In this paper, we performed Bayesian analysis of a panel probit model with unobserved individual heterogeneity and autocorrelated errors. We embedded EIS within a Gibbs sampling method to perform posterior analysis augmented with both the time and cross-section latent variables. The posterior for the unobserved individual heterogeneity was sampled from as one Gibbs block per individual, using a nonparametric version of EIS to form a piecewise 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 new form of 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 fit into a convex combination of the ones reported by Bertschek and Lechner (1998) and Greene (2004) who analyzed the same dataset under more restrictive model assumptions. The difference can be explained by the exclusion of three far outliers from our estimation and also by our weak model assumptions relative to these authors.

6 Appendix 1: Empirical Results

Figure 1: Descriptive Histograms for the Data

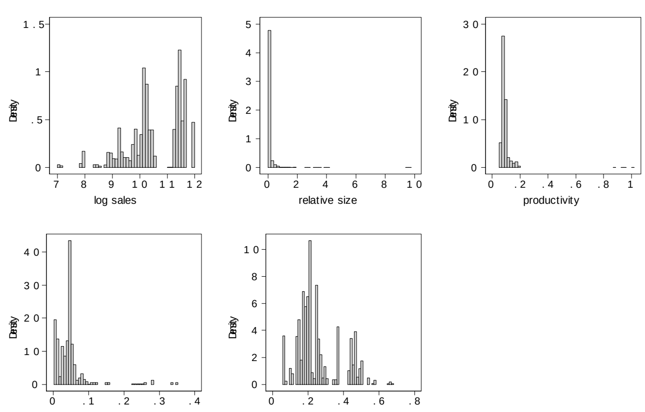


Figure 2: Marginal Posterior Densities of $\underline{\beta}$

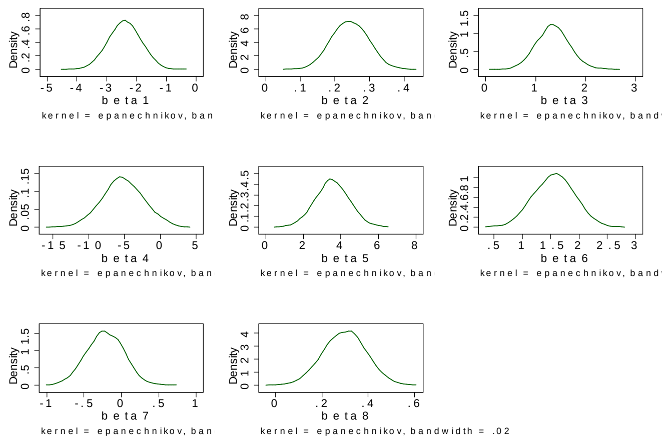


Figure 3: Left: MCMC chain for draws of $\underline{\beta}$. Right: Autocorrelations of draws of $\underline{\beta}$.

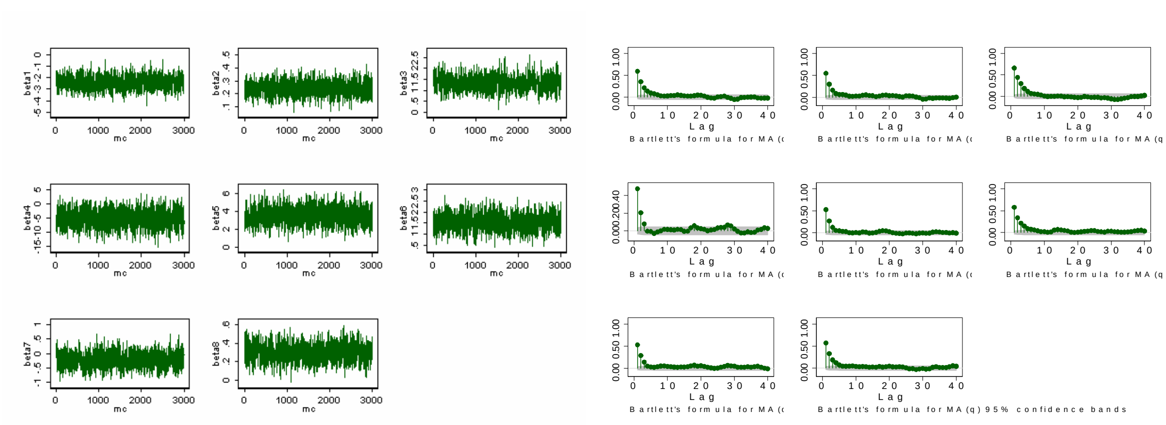


Figure 4: Left: Posterior density of σ_τ. Middle: MCMC chain for draws of σ_τ. Right: Autocorrelations of draws of σ_τ.

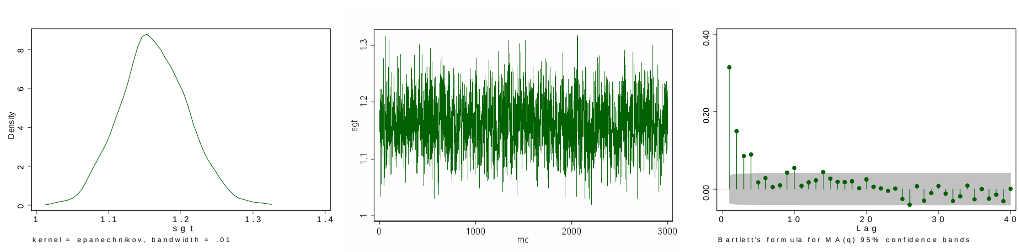


Figure 5: Left: Posterior density of σ_η. Middle: MCMC chain for draws of σ_η. Right: Autocorrelations of draws of σ_η.

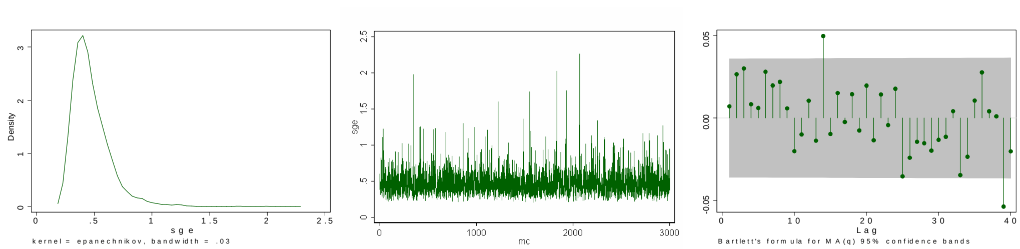
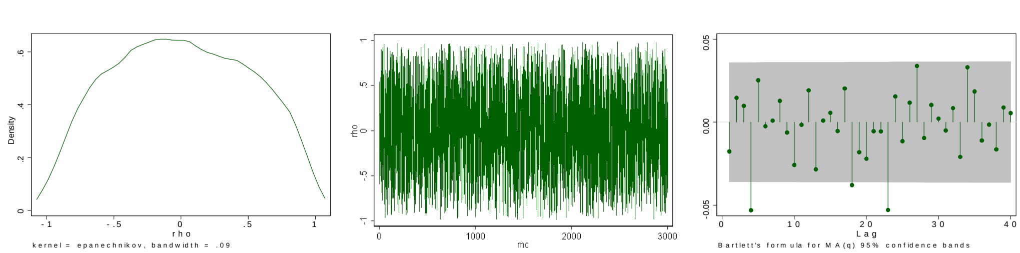


Figure 6: Left: Posterior density of ρ. Middle: MCMC chain for draws of ρ. Right: Autocorrelations of draws of ρ.



Appendix 2: Sampling from Posterior Densities

Sampling from $f(\underline{\beta}|\underline{\theta}_{/\underline{\beta}}, \Upsilon, Z)$

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

$$\begin{aligned}
Y_i^* &= Z_i\underline{\beta} + \underline{\lambda} + \tau_i\underline{\iota} + \underline{\varepsilon}_i \\
Y_{/\Upsilon,i}^* &= Y_i^* - \underline{\lambda} - \tau_i\underline{\iota} + \underline{\varepsilon}_i \\
Y_{/\Upsilon,i}^* &= Z_i\underline{\beta} + \underline{\varepsilon}_i
\end{aligned}$$

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

$$f(\underline{\beta}|\underline{\theta}_{/\underline{\beta}}, \Upsilon, Z) = N(\widehat{\underline{\beta}}, \widehat{\Sigma}_{\underline{\beta}}) \tag{11}$$

where $\widehat{\underline{\beta}} = \left(Z'Z\right)^{-1} Z'Y_{/\Upsilon}^*$, the dependent variable is a $(NT \times k)$ matrix $Y_{/\Upsilon}^* = (Y_{/\Upsilon,1}^{*\prime}, ..., Y_{/\Upsilon,N}^{*\prime})'$ and $\widehat{\Sigma}_{\underline{\beta}} = \left(Z'Z\right)^{-1}$. The random variables Y_{it}^* are independent with

$$\begin{aligned}
f(Y_{it}^*|\underline{\theta}, \Upsilon, Z) &= N(\mu_{it}^*, 1) \\
\mu_{it}^* &= Z_{it}\underline{\beta} + \lambda_t + \tau_i
\end{aligned} \tag{12}$$

truncated at the left by 0 if $Y_{it} = 1$ and truncated at the right by 0 if $Y_{it} = 0$. Given a previous value of $\underline{\beta}$, τ_i and λ_t, one cycle the Gibbs algorithm would produce Y_{it}^* and $\underline{\beta}$ from the distributions (12) and (11); see Train (2003, p. 210) for simulation algorithm. The starting value $\underline{\beta}^{(0)}$ may be taken to be the ML estimate.

Sampling from $f(\tau_i|\underline{\lambda}, \underline{\theta}, Z)$

From (8),

$$f(\tau_i|\underline{\lambda}, \underline{\theta}, Z) \propto \sigma_\tau^{-1} \exp\left[-\frac{1}{2\sigma_\tau^2}\tau_i^2\right] \prod_{t=1}^{T} \left[\Phi(v_{it})\right]^{y_{it}} \left[1 - \Phi(v_{it})\right]^{1-y_{it}} \tag{13}$$

The posterior $f(\tau_i|\underline{\lambda}, \underline{\theta}, Z)$ 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 $\underline{y}_i$. Therefore, we use a piece-wise linear approximation to $f(\tau_i|\underline{\lambda}, \underline{\theta}, Z)$ which is a form of nonparametric EIS capable of accurately sampling from any univariate distribution irrespective of its shape. The procedure works as follows. First, we obtain an empirical distribution function of $f(\tau_i|\underline{\lambda}, \underline{\theta}, Z)$ evaluated over an equispaced grid of τ_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 τ_i via the final edf to obtain the new value of the τ_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 $f(\underline{\lambda}|\underline{\tau}, \underline{\theta}, Z)$

From (9),

$$
\begin{aligned}
f(\underline{\lambda}|\underline{\tau}, \underline{\theta}, Z) &= p(\underline{\lambda}|\underline{\theta}) \prod_{t=1}^{T} \prod_{i=1}^{N} [\Phi(v_{it})]^{y_{it}} [1 - \Phi(v_{it})]^{1-y_{it}} \\
&= p(\underline{\lambda}|\underline{\theta}) \prod_{t=1}^{T} g(\lambda_t)
\end{aligned}
$$

where

$$
\begin{aligned}
g(\lambda_t) &\equiv \prod_{i=1}^{N} [\Phi(v_{it})]^{y_{it}} [1 - \Phi(v_{it})]^{1-y_{it}} \\
v_{it} &= \underline{\beta}' \underline{z}_{it} + \tau_i + \lambda_t
\end{aligned}
$$

The procedure is built on the fact that serial dependence in $\underline{\lambda}$ occurs only in $p(\underline{\lambda}|\underline{\theta})$ but not in $g(\lambda_t)$. We approximate $g(\lambda_t)$ with a first-stage EIS kernel for each period individually and then match these approximations analytically with the known $p(\underline{\lambda}|\underline{\theta})$ to form a second-stage EIS proposal density $\widetilde{m}(\underline{\lambda}|\underline{\tau}, \underline{\theta}, Z, \underline{\gamma})$ for the AR-MH step. Moreover, this procedure yields accurate scale factors $c(\underline{\lambda})$ for the AR-MH step that are functions of the proposed value of $\underline{\lambda}$ resulting in a very close overlap between the proposal and the posterior.

The joint EIS sampler takes the form

$$
\begin{aligned}
\widetilde{m}(\underline{\lambda}|\underline{\tau}, \underline{\theta}, Z, \underline{\gamma}) &= p(\lambda_1|\underline{\theta}) h\left(\lambda_1; \underline{\gamma}\right) \left[\prod_{t=2}^{T} p(\lambda_t|\lambda_{t-1}, \underline{\theta}) h\left(\lambda_t; \underline{\gamma}\right)\right] \\
&= \widetilde{m}(\lambda_1|\underline{\tau}, \underline{\theta}, Z, \underline{\gamma}) \prod_{t=2}^{T} \widetilde{m}(\lambda_t|\underline{\tau}, \underline{\theta}, Z, \underline{\gamma})
\end{aligned}
$$

where

$$
\begin{aligned}
p(\underline{\lambda}|\underline{\theta}) &= (2\pi)^{-T/2} |\Sigma_\lambda|^{-1/2} \exp\left[-\frac{1}{2} \underline{\lambda}' \Sigma_\lambda^{-1} \underline{\lambda}\right] \\
p(\lambda_1|\underline{\theta}) &= \frac{1}{\sigma_{\lambda_1}^2 \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_{\lambda_1}^2} \lambda_1^2\right] \\
\sigma_{\lambda_1}^2 &= \frac{\sigma_\eta^2}{1 - \rho^2} \\
\lambda_1 &\sim N(0, \sigma_{\lambda_1}^2) \\
p(\lambda_t|\lambda_{t-1}, \underline{\theta}) &= \frac{1}{\sigma_\eta \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_\eta^2} \left(\lambda_t - \rho\lambda_{t-1}\right)^2\right] \\
\lambda_t &\sim N(\mu_{\lambda_t}, \sigma_{\lambda_t}^2) \quad \text{for } t > 1 \\
\mu_{\lambda_t} &= \rho\lambda_{t-1} \\
\sigma_{\lambda_t}^2 &= \sigma_\eta^2
\end{aligned}
$$

and

$$
h\left(\lambda_t; \underline{\gamma}\right) = N\left(\mu_{h,\lambda_t}, \sigma_{h,\lambda_t}^2\right)
$$

The EIS regression for $t = 1, ..., T$ is

$$
\ln g(\lambda_t) = \gamma_{t,0} + \gamma_{t,1}\lambda_t + \gamma_{t,2}\lambda_t^2
$$

resulting in

$$\begin{aligned}\mu_{h,\lambda_t} &= -\frac{1}{2}\frac{\gamma_{t,1}}{\gamma_{t,2}}\\ \sigma^2_{h,\lambda_t} &= -\frac{1}{2}\frac{1}{\gamma_{t,2}}\end{aligned}$$

Estimates of γ are obtained using the iterative EIS procedure (Richard and Zhang, 2007). Then, for AR-MH step proposals are generated from $\widetilde{m}(\underline{\lambda}|\underline{\tau},\underline{\theta},Z,\underline{\gamma})$ that accounts for temporal dependence of $\underline{\lambda}$ by inclusion of $p(\underline{\lambda}|\underline{\theta})$. To construct $\widetilde{m}(\underline{\lambda}|\underline{\tau},\underline{\theta},Z,\underline{\gamma})$, we match Gaussian kernels of $h\left(\lambda_t;\underline{\gamma}\right)$ and $p(\underline{\lambda}|\underline{\theta})$ as follows:

For $t=1$ with $\widetilde{m}(\lambda_1|\underline{\tau},\underline{\theta},Z,\underline{\gamma}) = N\left(\widetilde{\mu}_{\lambda_1},\widetilde{\sigma}^2_{\lambda_1}\right)$,

$$\begin{aligned}\ln f(\lambda_1|\underline{\tau},\underline{\theta},Z) &= \ln p(\lambda_1|\underline{\theta}) + \gamma_{t,0} + \gamma_{t,1}\lambda_1 + \gamma_{t,2}\lambda_1^2 + \varepsilon_1\\ &= \ln\left(\frac{1}{\sigma^2_{\lambda_1}\sqrt{2\pi}}\right) + \gamma_{t,0} - \frac{1}{2}\left(\frac{1}{\sigma^2_{\lambda_1}} + \frac{1}{\sigma^2_{h,\lambda_t}}\right)\lambda_1^2 + \frac{\mu_{h,\lambda_t}}{\sigma^2_{h,\lambda_t}}\lambda_1 + \varepsilon_1\end{aligned}$$

and hence

$$\begin{aligned}\widetilde{\sigma}^2_{\lambda_1} &= \left(\frac{1}{\sigma^2_{\lambda_1}} + \frac{1}{\sigma^2_{h,\lambda_t}}\right)^{-1}\\ \widetilde{\mu}_{\lambda_1} &= \widetilde{\sigma}^2_{\lambda_1}\frac{\mu_{h,\lambda_t}}{\sigma^2_{h,\lambda_t}}\end{aligned}$$

For the scale factor,

$$\begin{aligned}\ln f(\lambda_1|\underline{\tau},\underline{\theta},Z) &= -\frac{1}{2}\ln\left(2\pi\sigma^2_{\lambda_1}\right) + \gamma_{t,0} - \frac{1}{2\widetilde{\sigma}^2_{\lambda_1}}\lambda_1^2 + \frac{\widetilde{\mu}_{\lambda_1}}{\widetilde{\sigma}^2_{\lambda_1}}\lambda_1 + \varepsilon_1\\ &= -\frac{1}{2}\ln\left(2\pi\sigma^2_{\lambda_1}\right) + \gamma_{t,0} + \frac{1}{2}\ln\left(2\pi\widetilde{\sigma}^2_{\lambda_1}\right) - \frac{1}{2}\ln\left(2\pi\widetilde{\sigma}^2_{\lambda_1}\right)\\ &\quad - \frac{1}{2\widetilde{\sigma}^2_{\lambda_1}}\lambda_1^2 + \frac{\widetilde{\mu}_{\lambda_1}}{\widetilde{\sigma}^2_{\lambda_1}}\lambda_1 - \frac{\widetilde{\mu}^2_{\lambda_1}}{2\widetilde{\sigma}^2_{\lambda_1}} + \frac{\widetilde{\mu}^2_{\lambda_1}}{2\widetilde{\sigma}^2_{\lambda_1}} + \varepsilon_1\\ &= \ln c_1 + \ln\widetilde{m}(\lambda_1) + \varepsilon_1\end{aligned}$$

resulting in

$$\ln c_1 = -\frac{1}{2}\ln\left(2\pi\sigma^2_{\lambda_1}\right) + \gamma_{t,0} + \frac{1}{2}\ln\left(2\pi\widetilde{\sigma}^2_{\lambda_1}\right) + \frac{\widetilde{\mu}^2_{\lambda_1}}{2\widetilde{\sigma}^2_{\lambda_1}}$$

For $t>1$

$$\begin{aligned}\ln f(\lambda_t|\underline{\tau},\underline{\theta},Z) &= \ln p(\lambda_t|\lambda_{t-1},\underline{\theta}) + \gamma_{t,0} + \gamma_{t,1}\lambda_t + \gamma_{t,2}\lambda_t^2 + \varepsilon_t\\ &= -\frac{1}{2}\ln\left(2\pi\sigma^2_\eta\right) - \frac{1}{2\sigma^2_\eta}\left(\lambda_t - \rho\lambda_{t-1}\right)^2 + \gamma_{t,0} + \frac{\mu_{h,\lambda_t}}{\sigma^2_{h,\lambda_t}}\lambda_t - \frac{1}{2\sigma^2_{h,\lambda_t}}\lambda_t^2 + \varepsilon_t\\ &= -\frac{1}{2}\ln\left(2\pi\sigma^2_\eta\right) - \frac{1}{2\sigma^2_\eta}\left(\lambda_t^2 - 2\rho\lambda_t\lambda_{t-1} + \rho^2\lambda_{t-1}^2\right) + \gamma_{t,0} + \frac{\mu_{h,\lambda_t}}{\sigma^2_{h,\lambda_t}}\lambda_t - \frac{1}{2\sigma^2_{h,\lambda_t}}\lambda_t^2 + \varepsilon_t\\ &= -\frac{1}{2}\ln\left(2\pi\sigma^2_\eta\right) - \frac{1}{2\sigma^2_\eta}\rho^2\lambda_{t-1}^2 + \gamma_{t,0} - \frac{1}{2}\left(\frac{1}{\sigma^2_\eta} + \frac{1}{\sigma^2_{h,\lambda_t}}\right)\lambda_t^2 + \left(\frac{\lambda_{t-1}\rho}{\sigma^2_\eta} + \frac{\mu_{h,\lambda_t}}{\sigma^2_{h,\lambda_t}}\right)\lambda_t + \varepsilon_t\end{aligned}$$

and hence

$$\begin{aligned}
\widetilde{\sigma}^2_{\lambda_t} &= \left(\frac{1}{\sigma^2_\eta} + \frac{1}{\sigma^2_{h,\lambda_t}}\right)^{-1} \\
\widetilde{\mu}_{\lambda_t} &= \widetilde{\sigma}^2_{\lambda_t}\left(\frac{\lambda_{t-1}\rho}{\sigma^2_\eta} + \frac{\mu_{h,\lambda_t}}{\sigma^2_{h,\lambda_t}}\right)
\end{aligned}$$

For the scale factors,

$$\begin{aligned}
\ln f(\lambda_t|\underline{\tau},\underline{\theta},Z) &= -\frac{1}{2}\ln\left(2\pi\sigma^2_\eta\right) - \frac{1}{2\sigma^2_\eta}\rho^2\lambda^2_{t-1} + \gamma_{t,0} - \frac{1}{2}\frac{1}{\widetilde{\sigma}^2_{\lambda_t}}\lambda^2_t + \frac{\widetilde{\mu}_{\lambda_t}}{\widetilde{\sigma}^2_{\lambda_t}}\lambda_t + \varepsilon_t \\
&= -\frac{1}{2}\ln\left(2\pi\sigma^2_\eta\right) - \frac{1}{2\sigma^2_\eta}\rho^2\lambda^2_{t-1} + \gamma_{t,0} + \frac{1}{2}\ln\left(2\pi\widetilde{\sigma}^2_{\lambda_t}\right) \\
&\quad -\frac{1}{2}\ln\left(2\pi\widetilde{\sigma}^2_{\lambda_t}\right) - \frac{1}{2}\frac{1}{\widetilde{\sigma}^2_{\lambda_t}}\lambda^2_t + \frac{\widetilde{\mu}_{\lambda_t}}{\widetilde{\sigma}^2_{\lambda_t}}\lambda_t - \frac{\widetilde{\mu}^2_{\lambda_t}}{2\widetilde{\sigma}^2_{\lambda_t}} + \frac{\widetilde{\mu}^2_{\lambda_t}}{2\widetilde{\sigma}^2_{\lambda_t}} + \varepsilon_t \\
&= \ln c_t(\lambda_{t-1}) + \ln \widetilde{m}(\lambda_t) + \varepsilon_t
\end{aligned}$$

where

$$\begin{aligned}
\ln c_t(\lambda_t,) &= -\frac{1}{2}\ln\left(2\pi\sigma^2_\eta\right) - \frac{1}{2\sigma^2_\eta}\rho^2\lambda^2_{t-1} + \gamma_{t,0} + \frac{1}{2}\ln\left(2\pi\widetilde{\sigma}^2_{\lambda_t}\right) + \frac{\widetilde{\mu}^2_{\lambda_t}}{2\widetilde{\sigma}^2_{\lambda_t}} \\
\ln m(\lambda_t) &= -\frac{1}{2}\ln\left(2\pi\widetilde{\sigma}^2_{\lambda_t}\right) - \frac{1}{2}\frac{1}{\widetilde{\sigma}^2_{\lambda_t}}\lambda^2_t + \frac{\widetilde{\mu}_{\lambda_t}}{\widetilde{\sigma}^2_{\lambda_t}}\lambda_t - \frac{\widetilde{\mu}^2_{\lambda_t}}{2\widetilde{\sigma}^2_{\lambda_t}}
\end{aligned}$$

Subsequently

$$\begin{aligned}
\widetilde{m}(\underline{\lambda}|\underline{\tau},\underline{\theta},Z,\underline{\gamma}) &= \widetilde{m}(\lambda_1|\underline{\tau},\underline{\theta},Z,\underline{\gamma})\prod_{t=2}^{T}\widetilde{m}(\lambda_t|\lambda_{t-1},\underline{\tau},\underline{\theta},Z,\underline{\gamma}) \\
\ln c(\underline{\lambda}) &= \ln c_1 + \sum_{t=2}^{T} c_t(\lambda_{t-1})
\end{aligned}$$

The result regarding $\ln c_t(\lambda_{t-1})$ as a function of λ_{t-1} follows from the intrinsic adaptability of the kernel of the EIS sampler $\widetilde{m}(\underline{\lambda})$. The kernel changes shape depending on the path $\underline{\lambda}$ which then in turn changes $\ln c_t(\lambda_{t-1})$. A potential absence of such dependency would imply a global sampler unable to overlap closely with $f(\underline{\lambda}|\underline{\tau},\underline{\theta},Z)$ for a given $\underline{\lambda}$, resulting in deterioration of acceptance probabilities.

Following the construction of $\widetilde{m}(\underline{\lambda}|\underline{\tau},\underline{\theta},Z,\underline{\gamma})$, the AR-MH step then completes the sampling procedure. Given K draws $\{\underline{\lambda}_1,...,\underline{\lambda}_K\}$ from the EIS-MCMC algorithm, potential new candidate draws are sampled from $\widetilde{m}(\underline{\lambda}|\underline{\tau},\underline{\theta},Z,\underline{\widehat{\gamma}})$ until acceptance of a candidate $\underline{\widetilde{\lambda}}$ in the AR step with probability

$$P(\underline{\lambda}) = \min\left(\frac{f(\underline{\lambda}|\underline{\tau},\underline{\theta},Z)}{c(\underline{\lambda})\widetilde{m}(\underline{\lambda}|\underline{\tau},\underline{\theta},Z,\underline{\widehat{\gamma}})},1\right)$$

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

$$\alpha(\underline{\lambda}_K,\underline{\widetilde{\lambda}}) = \min\left(\frac{f(\underline{\widetilde{\lambda}}|\underline{\tau},\underline{\theta},Z)\min\left[f(\underline{\lambda}|\underline{\tau},\underline{\theta},Z),c(\underline{\widetilde{\lambda}})\widetilde{m}(\underline{\lambda}|\underline{\tau},\underline{\theta},Z,\underline{\widehat{\gamma}})\right]}{f(\underline{\lambda}|\underline{\tau},\underline{\theta},Z)\min\left[f(\underline{\widetilde{\lambda}}|\underline{\tau},\underline{\theta},Z),c(\underline{\widetilde{\lambda}})\widetilde{m}(\underline{\widetilde{\lambda}}|\underline{\tau},\underline{\theta},Z,\underline{\widehat{\gamma}})\right]},1\right)$$

Note that we could technically use a random walk chain or an independence chain, tweaking the proposal dispersion for low-dimensional integrals. However, with increased dimensionality of integration, such methods would result in acceptance probability approaching zero as $f(\underline{\lambda}|\underline{\tau}, \underline{\theta}, Z)$ takes strictly positive values on a very small region in the domain of $\underline{\lambda}$.

Sampling from $f(\sigma_\tau^2|\underline{\theta}_{/\sigma_\tau}, \Upsilon, Z)$

Since for all τ_i it holds that $\tau_i \sim N(0, \sigma_\tau^2)$ we can apply result B (p. 300) of Train (2003): for a $IG(s_0, v_0)$ prior, the posterior of σ_τ^2 is given by $IG(s_1, v_1)$ with $v_1 = v_0 + N$ and $s_1 = (v_0 s_0 + N\bar{s})/(v_0 + N)$ where $\bar{s} = N^{-1} \sum_{i=1}^{N_{c_i}} \tau_i^2$. We utilize a diffuse prior $s_0 = 1$ and $v_0 \to 1$.

Sampling from $f(\sigma_\eta^2|\underline{\theta}_{/\sigma_\eta}, \Upsilon, Z)$

Derivation of the posterior proceeds along the lines of result B (p. 300) of Train (2003) but care needs to be taken with regard to the serial dependence of $\underline{\lambda}$. Conditional on $\underline{\lambda}$ and ρ, the likelihood function of σ_η^2 takes the form

$$L(\sigma_\eta^2|\underline{\lambda}, \underline{\theta}_{/\sigma_\eta^2}) \propto \frac{\sqrt{1-\rho^2}}{\sigma_\eta\sqrt{2\pi}} \exp\left[-\frac{1-\rho^2}{2\sigma_\eta^2}\lambda_1^2\right] \prod_{t=2}^{T} \frac{1}{\sigma_\eta\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_\eta^2}\left(\lambda_t - \rho\lambda_{t-1}\right)^2\right]$$

An $IG(v_0, s_0)$ prior has density

$$k(\sigma_\eta^2) = \frac{1}{m_0 \sigma_\eta^{(v_0+1)/2}} \exp\left[-\frac{v_0 s_0}{2\sigma_\eta}\right]$$

where m_0 is a normalizing constant. The posterior is then

$$\begin{aligned} L(\sigma_\eta^2|\underline{\lambda}, \underline{\theta}_{/\sigma_\eta^2}) &\propto L(\sigma_\eta^2|\underline{\lambda}, \underline{\theta}_{/\sigma_\eta^2})k(\sigma_\eta^2) \\ &\propto \frac{1}{\sigma_\eta^{(T+v_0+1)/2}} \exp\left[-\frac{\left(1-\rho^2\right)\lambda_1^2 + \sum_{t=2}^{T}\left(\lambda_t - \rho\lambda_{t-1}\right)^2 + v_0 s_0}{2\sigma_\eta^2}\right] \\ &= IG(v_1, s_1) \end{aligned}$$

where

$$\begin{aligned} v_1 &= v_0 + T \\ s_1 &= \frac{v_0 s_0 + \left(1-\rho^2\right)\lambda_1^2 + \sum_{t=2}^{T}\left(\lambda_t - \rho\lambda_{t-1}\right)^2}{v_0 + T} \end{aligned}$$

Sampling from $f(\rho|\underline{\theta}_{/\rho}, \Upsilon, Z)$

Under the assumption of AR(1) process for λ_t, the posterior density is given by

$$f(\rho|\underline{\theta}_{/\rho}, \underline{\lambda}, \underline{\tau}, Z) \propto \frac{1}{\sqrt{2\pi\frac{\sigma_\eta^2}{(1-\rho^2)}}} \exp\left(-\frac{(1-\rho^2)}{2\sigma_\eta^2}\lambda_1^2\right) \prod_{t=1}^{T} \left\{ \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \exp\left(-\frac{1}{2\sigma_\eta^2}\left(\lambda_t - \rho\lambda_{t-1}\right)^2\right) \right\} \quad (14)$$

Since ρ is univariate in this application, we utilize the same procedure as for sampling $\tau_i|\underline{\lambda}, \underline{\theta}, Z$ but with the posterior (14).

Appendix 3: Simulation Study with Artificial Data

We performed a small-scale pilot simulation study in order to assess the performance of our method on artifical data of the same sample size $N = 1267$ and $T = 5$. In order to distinguish the properties of the technique from the generic small-sample properties that concern any model, we selected a series of simulated true values of $\underline{\lambda}_0 = \{-0.21738, -0.63577, -0.45545, 0.67616, 0.63244\}$ that result in classical maximum likelihood estimates $\widehat{\rho} = 0.4$ and $\widehat{\sigma}_\eta = 0.5$ which we set as the true values for these parameters. Then we generated a sample of true $\underline{\tau}_0$ from $N(0, \sigma_\tau^2)$ with true $\sigma_\tau = 0.5$. Finally, we drew sample data z_{it} from a uniform density on the interval $(-0.5, 0.5)$. We generated the latent utility according to (7) and observed responses according to (2) for a vector of true values $\underline{\beta}_0 = \{0, -0.4, -0.4, -0.4, 0.3, 0.3, 0.3, 0.3\}$. These true values are marked by solid lines in Figure 7. Maximum likelihood estimates conditional on true values of the latent variables $\underline{\lambda}_0$ and $\underline{\tau}_0$ are shown with dashed lines in Figure 7. The ML estimates differ in some cases substantially from the true values which serves as a benchmark for assessing the performance of the posterior estimation of our model.

As in the empirical application, the posterior draws are based on 20,000 Gibbs cycles discarding the first 5,000 cycles and keeping every fifth draw thereafter resulting in 3000 MC draws for each parameter. One Gibbs iteration takes approximately 3.5 seconds on a 2.2 GHz unix machine. The nonparametric EIS sampler was perfomed over a grid of size 200. 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 $\underline{\lambda}$ and $\underline{\tau}$. The AR and MH acceptance rates for $\underline{\lambda}$ and $\underline{\tau}$ were close to 99%. Posterior means of all parameters are very close to the true values. Most means of $\underline{\beta}$ draw are in absolute value closer to the true values than the ML conditional estimates. Aided by a 5 cycle thinning, all autocorrelation functions indicate superior mixing properties of the Markov chains on which the posterior analysis is based. Most autocorrelation coefficients falter to a statistical zero after a handful of cycles, with the exception of the σ_τ chain that takes 6 cycles due to the incremental nature of the sampler for τ_i.

Figure 7: Marginal Posterior Densities of $\underline{\beta}$

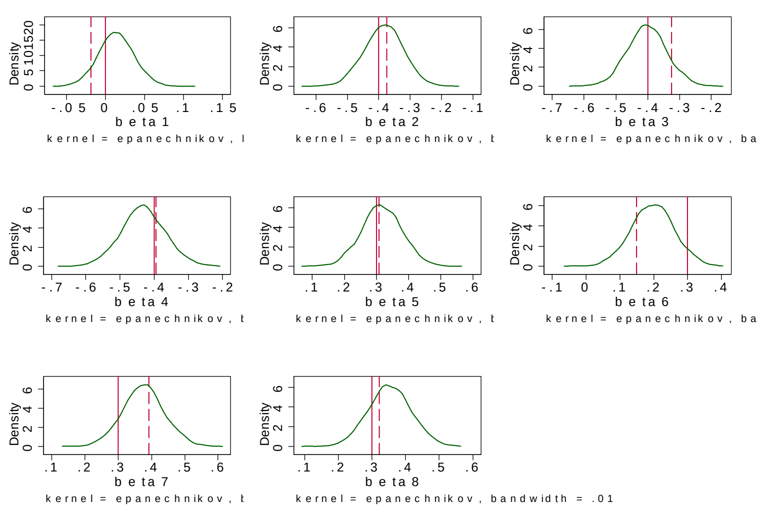


Figure 8: Left: MCMC chain for draws of $\underline{\beta}$. Right: Autocorrelations of draws of $\underline{\beta}$.

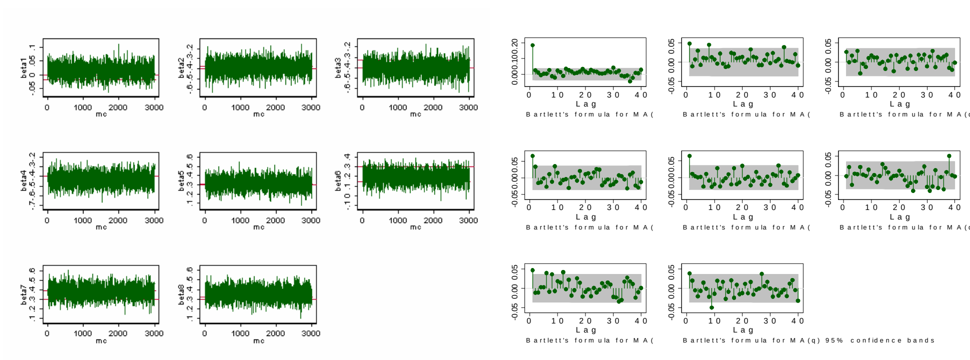


Figure 9: Left: Posterior density of σ_τ. Middle: MCMC chain for draws of σ_τ. Right: Autocorrelations of draws of σ_τ.

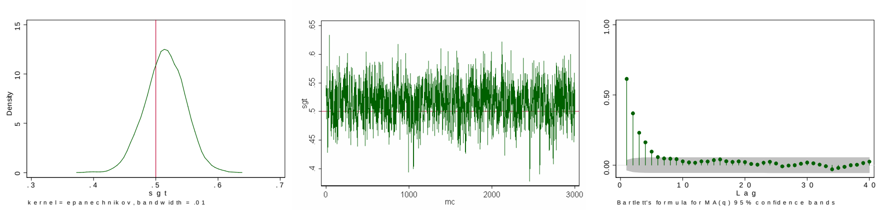


Figure 10: Left: Posterior density of σ_η. Middle: MCMC chain for draws of σ_η. Right: Autocorrelations of draws of σ_η.

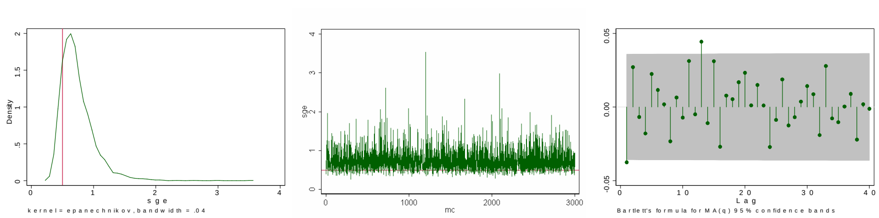
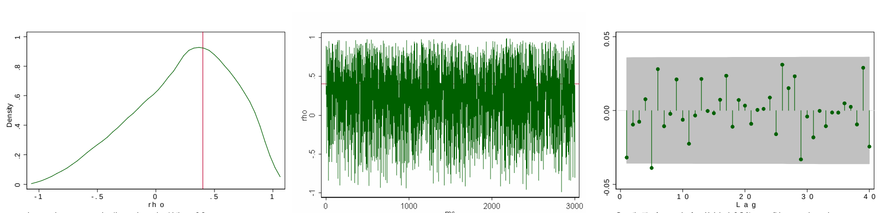


Figure 11: Left: Posterior density of ρ. Middle: MCMC chain for draws of ρ. Right: Autocorrelations of draws of ρ.



References

Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*(422), 669–679.

Bertschek, I. (1995). Product and process innovation as a response to increasing import and foreign direct investment. *Journal of Industrial Economics 43*(4), 341–57.

Bertschek, I. and M. Lechner (1998). Convenient estimators for the panel probit model. *Journal of Econometrics 87*(2), 329–371.

Börsch-Supan, A., V. A. Hajivassiliou, L. J. Kotlikoff, and J. N. Morris (1992). Health, children, and elderly living arrangements: a multiperiod-multinomial probit model with unobserved heterogeneity and autocorrelated errors. In W. D. A. (Ed.), *Topics in the Economics of Aging*. The University of Chicago Press.

Butler, J. S. and R. Moffitt (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica 50*(3), 761–764.

Falcetti, E. and M. Tudela (2006). modelling currency crises in emerging markets: a dynamic probit model with unobserved heterogeneity and autocorrelated errors. *Oxford Bulleting of Economics and Statistics 68*(4), 445–471.

Franses, P. H. (2006). On modeling panels of time series. *Statistica Neerlandica 60*(4), 438Ű456.

Gourieroux, C. and A. Monfort (1996). *Simulation-based econometric methods*. Oxford University Press.

Greene, W. (2004). Convenient estimators for the panel probit model: Further results. *Empirical Economics 29*(1), 21–47.

Heckman, J. J. (1981). Statistical models for discrete panel data. In M. C. F. and M. D. (Eds.), *Structural Analysis of Discrete Data with Econometrics Applications*. The MIT Press, Cambridge, Masschusetts; London, England.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica 67*(6), 1255–1294.

Inkmann, J. (2000). Misspecified heteroskedasticity in the panel probit model: A small sample comparison of gmm and sml estimators. *Journal of Econometrics 97*(2), 227–259.

Liesenfeld, R. and J.-F. Richard (2006). Classical and bayesian analysis of univariate and multivariate stochastic volatility models. *Econometric Reviews 25*(2), 1–26.

Paap, R. (2002). What are the advantages of mcmc based inference in latent variable models? *Statistica Neerlandica 56*(1), 2–22.

Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*(2), 1385–1411.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing*. Wiley series in Probability and Statistics.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association 82*, 528–550.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics 22*, 1701–1762.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.