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

Martin Burda† Roman Liesenfeld‡ Jean-François Richard§

September 8, 2008

Abstract

In this paper, we perform a 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 with one Gibbs block per individual using a flexible piecewise log-linear approximation to the marginal posterior density. The latent time effects are simulated with one Gibbs block per time period. 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 in our application 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 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.

†Corresponding author. 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 evaluation of a likelihood function containing multivariate integrals that are often analytically intractable.

Fuelled by advances in computation such as data augmentation (Tanner and Wong, 1987), the last two decades witnessed an explosion of interest in Bayesian models that had previously been regarded as numerically unfeasible. Under this framework, the latent variables within multivariate integrals are treated as model parameters and are sampled along with them. The Bayesian Gibbs sampling scheme is naturally suited for such purpose: high-dimensional multivariate integrals forming the likelihood function are factorized into sequences of low-dimensional conditional densities each of which is sampled individually. Embedding these low-dimensional subproblems within a Markov chain 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 in order to accurately approximate the integral likelihood function for each parameter value embedded within an optimization algorithm. In contrast, Gibbs sampling takes one latent variable draw for each parameter value until convergence. In many cases, this implies that Bayesian parameter estimation is substantially faster than SML. Multiple local modes in the SML objective function for a given dataset are another potential concern which is alleviated using the Bayesian setup. Other advantages of Bayesian inference in latent variable models are discussed e.g. 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).

In this paper, we perform a Bayesian analysis of a panel probit model with unobserved individual heterogeneity and autocorrelated errors with unobserved random heterogeneity along both dimensions. The unobserved time random effect is assume to be serially correlated and common to all individuals. Our approach, based on proposal densities partially constructed with the Efficient Importance Sampling (EIS) procedure (Richard and Zhang, 2007), allows for inference within a rich high-dimensional economic model environment. We augment the data with latent variables and sample the unobserved individual heterogeneity

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

component with one Gibbs block per individual, drawing from a flexible piecewise log-linear approximation to the marginal posterior density. The time effects are simulated in a series of Gibbs blocks 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 Hermite quadrature method (see Butler and Moffitt, 1982). 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 present our Bayesian analysis. The results of our empirical application are 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 zero otherwise. The $K-$vector of control variables is denoted by $\underline{x}_{it}$ 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

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

(*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).

Companies may differ in their propensity to innovate. Two components can be distinguished with regard to this phenomenon. The first one relates to the existence of company-specific attributes that are time-invariant. This component is typically called *unobserved heterogeneity* and we account for it by including a time-invariant company-specific error term τ_i. It may reflect unobserved institutional factors such as managerial entrepreneurship that cannot be directly included among the regressors. The second component takes into account that economy-wide factors influencing all companies alike may be correlated over time.

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 analyzed by these authors.

2.1.1 Model 1: Pooled Probit

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

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

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, $\underline{\epsilon}_i$ is 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 across firms. The diagonal elements of Σ are set to unity for 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:

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

where Φ denotes the CDF of a univariate normal distribution, $Z = (Z_1', \ldots, Z_T')'$, $Z_t = (Y_t, X_t')$. The main advantage of using these moments is that their evaluation does not require multidimensional integration and that they do not depend on the $T(T-1)/2$ off-diagonal elements of Σ. In line with the GMM literature, (3) implies

$$E\left[A(X)W(Z, \beta_0)\right] = 0$$

where $A(X)$ is a $P \times T$ matrix of instrumental variables. An 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

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

and Ω is a consistent estimator of the covariance matrix of $A(X)W(Z, \beta_0)$ – see Hansen (1982). 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 by Bertschek and Lechner (1998) presents 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, it is likely to contain a significant degree of unobserved individual heterogeneity which the model presented above ignores. 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 ('Hierarchical' or 'Multilevel' Model) is based on the latent variable probit model specification

$$y_{it}^* = \underline{\beta}_i' x_{it} + \epsilon_{it}$$

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 with $E[w_i|X_i, z_i] = 0$ 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 of y_{it} is

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

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

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

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

The Latent Class model arises if we assume a discrete distribution for $\underline{\beta}_i$ instead of the continuous one postulated above. Alternatively, it 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 $\underline{\beta}_i$ has finite, discrete support over J points (classes) with probabilities $p(\underline{\beta}_j|\mu, \Delta, \Gamma, z_i)$, $j = 1, ..., J$, then the resulting formulation of the analog of L_i from (6) is

$$L_i = \sum_{j=1}^{J} p(\underline{\beta}_j|\mu, \Delta, \Gamma, z_i) f\left(y_i|X_i, \underline{\beta}_j\right)$$

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 latent variables for both individual heterogeneity and time effects. Specifically, our standardized probit model specification assumes a latent variable regression for individual i and time period t

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

under the observation rule (2), where x_{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. λ_t captures latent time effects and is assumed to follow a stationary autoregressive process

$$\lambda_t = \sum_{j=1}^{k} \rho_j \lambda_{t-j} + \eta_t$$

where $\eta_t \sim N(0, \sigma_\eta^2)$. 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

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

and

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

and Σ_λ denotes the stationary variance-covariance matrix of $\underline{\lambda}$. See Richard (1977) for an analytical expression for Σ_λ^{-1}.

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 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 along the time dimension of $\underline{\lambda}|\underline{\theta}, Z$. Specifically, we approximate with a first-stage EIS kernel only the part of the likelihood that arises from the probit model specification, and then recombine this approximation analytically with the known AR likelihood assumed for the latent time process in order to form the desired second-stage EIS proposal density. This results in significant simplification in constructing the proposal density. The unobserved individual heterogeneity components $\tau_i|\underline{\theta}, Z$ are sampled as N Gibbs blocks drawing from a piecewise log-linear approximation to the marginal posterior density constructed as described in DeJong et al. (2007). The justification for these procedures is that the resulting 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 piecewise log-linear approximation to $f(\tau_i|\underline{\theta}, Z)$ freely adapts to the shape of the posterior and can be made arbitrarily precise by increasing the size of the simulation grid. For $f(\underline{\lambda}|\underline{\theta}, Z)$, given the model assumptions, one finds 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 with acceptance rates close to 100% in our empirical application.

Actually, we shall apply Gibbs sampling to an augmented vector which includes not only the parameters $\underline{\theta}$ but also the latent variables $\underline{\lambda}$, $\underline{\tau}$ and y^*_{it}. Therefore, let $\xi = (Y^*, \underline{\beta}, \underline{\lambda}, \underline{\tau}, \sigma^2_\eta, \sigma^2_\tau, \rho)$ and let ξ without a generic subvector ξ_j be denoted as $\xi_{/\xi_j}$. For a given ξ the augmented likelihood $L(\xi; \underline{y})$ is defined as the integrand in equation (8). For each Gibbs block of ξ_j the Bayesian optimal updating of prior beliefs, $\pi(\xi_j)$, with new information (data Z) takes the form

$$f(\xi_j|\xi_{/\xi_j}, Z) \propto L(\xi; \underline{y})\pi(\xi_j) \tag{10}$$

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* (pooled probit) in (1) estimated in Stata using the command 'probit'. Table 1 also reports the Bertschek and Lechner (1998) GMM parameter estimates of *Model 2* (autocorrelated errors) with a k-NN estimate of Ω in (4) and the Greene (2004) random parameter model prior means estimates of *Model 3* (random parameters). 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* (finite mixture). 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 of our latent panel probit model as defined by equation (7) with first order AR. Posterior marginal densities, MCMC chains and autocorrelation functions of the parameter draws are presented in Figures 2 to 6. The latter two results highlight the excellent mixing properties of our Markov chains.

We excluded from estimation three distant outliers with relative firm size larger than 0.1 and productivity larger than 0.8 (compare with the horizontal scales in 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 exclusion reduced our sample size to $N = 1267$ and $T = 5$.

We note that with few exceptions the point estimates reported in Tables 1 and 2 lie within two standard deviations of our EIS-MCMC posterior means. 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 posterior means of the two key parameters of *FDI* and *import share* are positive and generally higher than the point esti-

Table 1: Models 1-3

	Model 1[a]		Model 2[b]		Model 3[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**	0.03	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**	0.06	0.238	0.453

[a] Pooled probit, estimated in Stata by the simple command 'probit'.
[b] Probit with autocorrelated errors, Bertschek and Lechner (1998), WNP-joint uniform estimates with $k = 880$, Table 9, standard errors from Table 10
[c] Random parameters, 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**	0.76	-2.71**	0.77	-8.97**	2.50
log sales	0.32**	0.07	0.23**	0.07	0.57**	0.20
Rel size	4.38**	0.88	0.72**	0.25	1.42*	0.62
Imports	0.93**	0.49	2.26**	0.50	3.12*	1.35
FDI	2.20	2.54	2.80**	0.93	8.37**	2.27
Prod.	-5.86**	1.69	-7.70**	1.16	-0.91	1.26
Raw Mtl	-0.11	0.17	-0.60**	0.30	-0.86*	0.42
Inv good	0.13	0.14	0.41**	0.13	0.50*	0.23

[d] Finite mixture (3 classes), Greene (2004), Table 7

Table 3: EIS-MCMC

Variable	Posterior mean	Posterior median	Std.Err.
Constant	-2.484**	-2.479**	0.433
log sales	0.248**	0.247**	0.044
Rel size	1.308**	1.312**	0.231
Imports	1.434**	1.437**	0.286
FDI	3.701**	3.698**	0.722
Prod.	-6.127**	-6.086**	2.459
Raw Mtl	-0.144	-0.141	0.200
Inv good	0.266**	0.267**	0.073
σ_τ	0.868**	0.868**	0.025
σ_η	0.477**	0.443**	0.167
ρ	0.006	0.008	0.499

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.

mates in Tables 1 and 2, 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 0.868 which somewhat smaller than 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 0.477 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 than latent time effects in this application.

5 Conclusion

In this paper, we performed a Bayesian analysis of a panel probit model with unobserved individual heterogeneity and autocorrelated errors. We embedded EIS within a Gibbs sampling method augmented with both the time and cross-sectional latent variables. The posterior for the unobserved individual heterogeneity was sampled from univariate individual Gibbs blocks, using a piecewise log-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 strengthen the positive effect of imports and FDI on firms' innovation activity found in previous literature. Our posterior means for the key coefficients lie on the high end of the point estimates reported by Bertschek and Lechner (1998) and Greene (2004) who analyzed the same dataset under more restrictive model assumptions.

6 Appendix 1: Details of Empirical Results

Figure 1: Descriptive Histograms for the Data

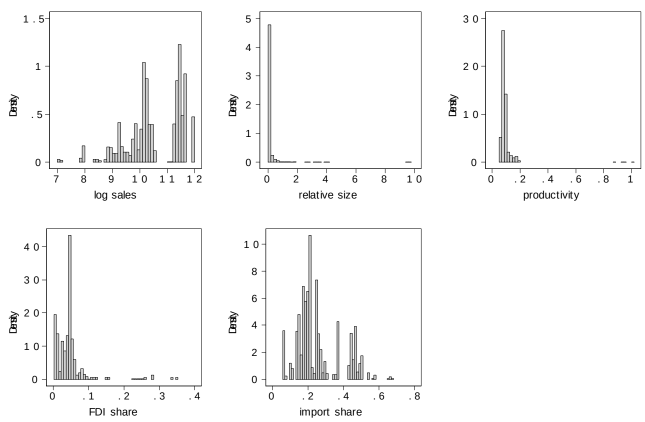


Figure 2: Marginal Posterior Densities of $\underline{\beta}$

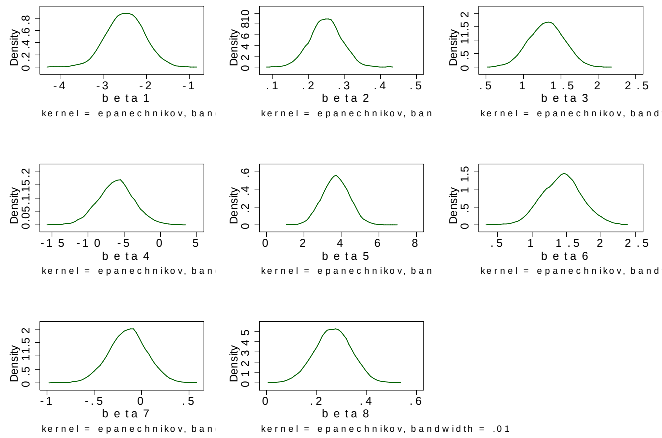


Figure 3: Left: MCMC chain for draws of $\underline{\beta}$. Right: Autocorrelations of draws of $\underline{\beta}$.

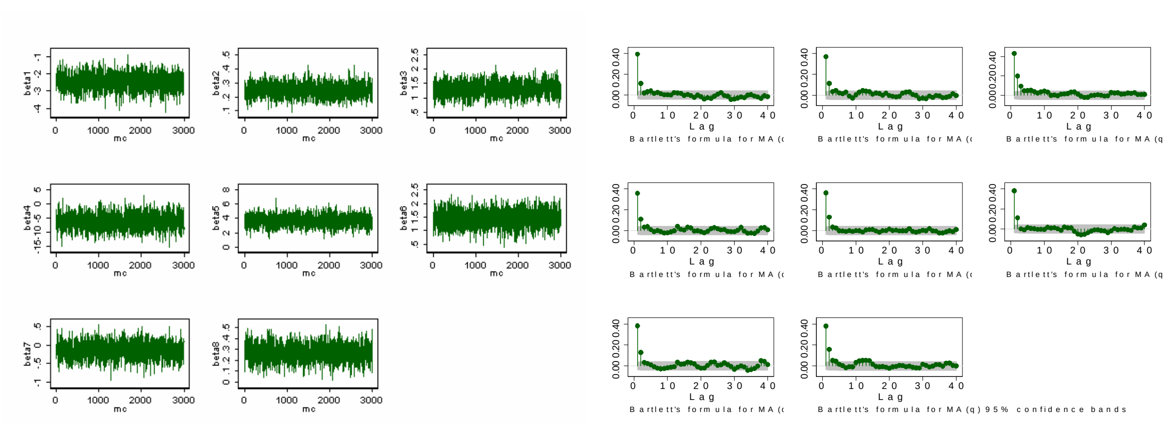


Figure 4: Left: Posterior density of σ_τ. Middle: MCMC chain for draws of σ_τ. Right: Autocorrelations of draws of σ_τ.

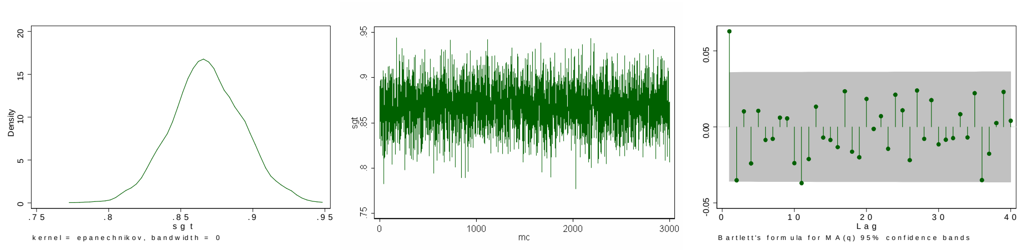


Figure 5: Left: Posterior density of σ_η. Middle: MCMC chain for draws of σ_η. Right: Autocorrelations of draws of σ_η.

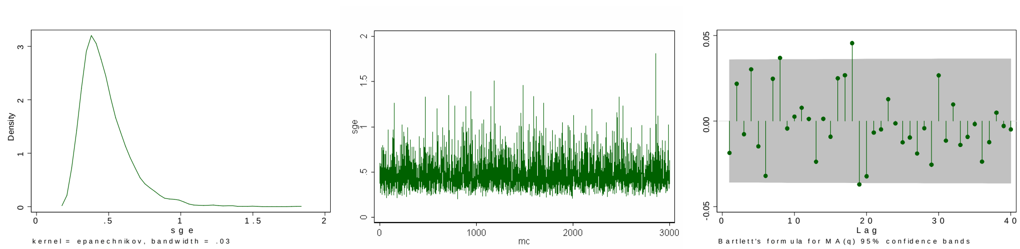
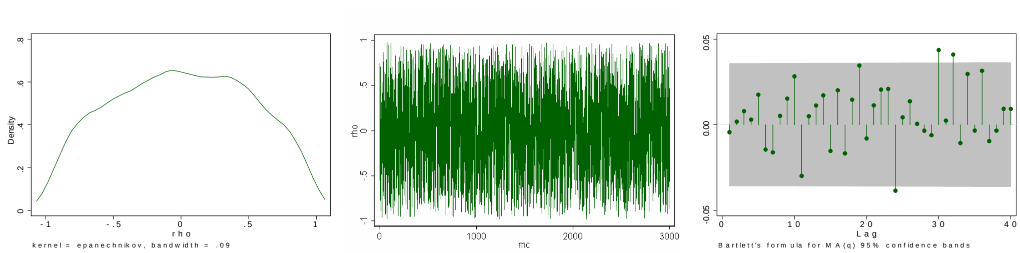


Figure 6: Left: Posterior density of ρ. Middle: MCMC chain for draws of ρ. Right: Autocorrelations of draws of ρ.



Appendix 2: Sampling from Posterior Densities

Note that the variable y_{it}^* in equation (7) is not observed, only its sign. Whence, in order to facilitate the Gibbs step for β, it proves convenient to include the $T \times N$ matrix $Y^* = (Y_1^*, \ldots, Y_N^*)$ with $Y_i^* = (y_{i1}, \ldots, y_{iT})'$ in the draw. Thus our Gibbs cycle operates on the variables $\xi = (Y^*, \underline{\beta}, \underline{\lambda}, \underline{\tau}, \sigma_\eta^2, \sigma_\tau^2, \rho)$.

Sampling Y^* given $(\xi_{/Y^*}, Z)$

It follows from equations (2) and (7) that the y_{it}^*'s given $\xi_{/Y^*}$, Z are independently distributed from one another with the truncated Normal densities

$$f\left(y_{it}^*|\xi_{/Y^*}, Z\right) \propto \mathbf{1}\left(y_{it}^*\left(2y_{it}-1\right) \geq 0\right) f_N\left(y_{it}^*|\mu_{it}^*, 1\right)$$

where $\mu_{it}^* = \underline{\beta}'x_{it} + \lambda_t + \tau_i$ (see e.g. Devroye (1986) for sampling from a truncated Normal density).

Sampling $\underline{\beta}$ given $(\xi_{/\underline{\beta}}, Z)$

Let $Y_{i/\underline{\lambda},\tau_i}^* = Y_i^* - \underline{\lambda} - \tau_i \underline{\iota}$ and $Y_{/\underline{\lambda},\tau_i}^* = (Y_{1/\underline{\lambda},\tau_1}^{*\prime}, ..., Y_{N/\underline{\lambda},\tau_N}^{*\prime})'$. It follows from equation (7) that under a uniform prior for $\underline{\beta}$ its posterior density given $(\xi_{/\underline{\beta}}, Z)$ is the joint Normal density

$$f\left(\underline{\beta}|\xi_{/\underline{\beta}}, Z\right) \propto f_N(\underline{\beta}|\widehat{\underline{\beta}}, \widehat{\Sigma}_{\underline{\beta}})$$

with $\widehat{\underline{\beta}} = \left(Z'Z\right)^{-1} Z'Y_{/\underline{\lambda},\tau_i}^*$ and $\widehat{\Sigma}_{\underline{\beta}} = \left(Z'Z\right)^{-1}$.

Sampling τ_i given $(\xi_{/\tau_i}, Z)$

It follows from equation (8) that the conditional posterior density of τ_i given $(\xi_{/\tau_i}, Z)$ is proportional to

$$\varphi_i(\tau_i|\xi_{/\tau_i}, Z) = \sigma_\tau^{-1} \exp\left[-\frac{1}{2\sigma_\tau^2}\tau_i^2\right] \prod_{t=1}^{T} \left[\Phi(v_{it})\right]^{y_{it}} \left[1-\Phi(v_{it})\right]^{1-y_{it}}$$

This density can be quite skewed depending on the observation vector Y_i. Whence φ_i is approximated by a piecewise log-linear density k_i :

$$\ln k_i(\tau) \propto \widehat{\alpha}_j + \widehat{\beta}_j\tau \quad \text{for} \quad \tau \in \left]\widehat{\tau}_{i,j-1}, \widehat{\tau}_{i,j}\right],\ j = 1 \to J$$

where $\{\widehat{\tau}_{i,0}, \ldots, \widehat{\tau}_{i,J}\}$ denotes an auxiliary grid (iteratively) constructed in such a way that its intervals are equiprobable according to the corresponding empirical c.d.f. K_i, i.e. that

$$K_i(\widehat{\tau}_{i,j}) = \frac{j}{J}, \quad j : 0 \to J$$

within a selected fixed point tolerance level. See DeJong et al. (2007) for details.

Sampling λ_t given $(\xi_{/\lambda_t}, Z)$

Under a stationary AR(1) for $\underline{\lambda}$, the conditional posterior density of λ_t given $(\xi_{/\lambda_t}, Z)$ obtains from equation (9) and is given by

$$f(\lambda_t|\xi_{/\lambda_t}, Z) \propto p(\lambda_{t+1}|\lambda_t, \underline{\theta})p(\lambda_t|\lambda_{t-1}, \underline{\theta})g(\lambda_t)$$

where

$$\begin{aligned}
g(\lambda_t) &\equiv \prod_{i=1}^{N} [\Phi(v_{it})]^{y_{it}} [1-\Phi(v_{it})]^{1-y_{it}} \\
p(\lambda_1|\lambda_0,\underline{\theta}) &\equiv p(\lambda_1|\underline{\theta}) = f_N(0|\sigma_\lambda^2), \quad \sigma_\lambda^2 = \sigma_\eta^2/(1-\rho^2) \\
p(\lambda_t|\lambda_{t-1},\underline{\theta}) &= f_N(\rho\lambda_{t-1}|\sigma_\eta^2) \\
p(\lambda_{T+1}|\lambda_T,\underline{\theta}) &= 1
\end{aligned}$$

The corresponding EIS sampler for λ_t is defined as the product

$$\widetilde{m}(\lambda_t|\xi_{/\lambda_t}, Z) = \widetilde{c}_t p(\lambda_{t+1}|\lambda_t,\underline{\theta}) p(\lambda_t|\lambda_{t-1},\underline{\theta}) h\left(\lambda_t;\widehat{\gamma}_t\right)$$

where $h_t(\lambda_t)$ denotes an auxiliary density kernel of the form

$$\ln h\left(\lambda_t;\widehat{\gamma}_t\right) = \widehat{\gamma}_{t,0} + \widehat{\gamma}_{t,1}\lambda_t + \widehat{\gamma}_{t,2}\lambda_t^2$$

which obtains as a (fixed point) EIS-OLS approximation of $\ln g_t(\lambda_t)$. Also, $\widetilde{c_t}^{-1}$ denotes the integrating factor for $\widetilde{m}_t$. See Richard and Zhang (2007) for details of the EIS algorithm. The moments and integrating factor for $\widetilde{m}_t$ obtain from standard Gaussian algebra and are given by

$$\begin{aligned}
\widetilde{\sigma}^2_{\lambda_1} &= \left(\frac{1}{\sigma_\lambda^2} + \frac{\rho^2}{\sigma_\eta^2} - 2\widehat{\gamma}_{1,2}\right) \\
\widetilde{\mu}_{\lambda_1} &= \widetilde{\sigma}^2_{\lambda_1}\left(\frac{\rho\lambda_2}{\sigma_\eta^2} + \widehat{\gamma}_{1,1}\right) \\
\widetilde{c}_1 &= \frac{1}{\sqrt{2\pi}}\frac{\widetilde{\sigma}_{\lambda_1}}{\sigma_\lambda\sigma_\eta}\exp\left\{-\frac{1}{2}\left(\frac{\lambda_2^2}{\sigma_\eta^2} - 2\widehat{\gamma}_{1,0} - \frac{\widetilde{\mu}^2_{\lambda_1}}{\widetilde{\sigma}^2_{\lambda_1}}\right)\right\} \\
\widetilde{\sigma}^2_{\lambda_t} &= \left(\frac{1+\rho^2}{\sigma_\eta^2} - 2\widehat{\gamma}_{t,2}\right)^{-1} \\
\widetilde{\mu}_{\lambda_t} &= \widetilde{\sigma}^2_{\lambda_t}\left(\frac{\rho\left(\lambda_{t-1}+\lambda_{t+1}\right)}{\sigma_\eta^2} + \widehat{\gamma}_{t,1}\right) \\
\widetilde{c}_t &= \frac{1}{\sqrt{2\pi}}\frac{\widetilde{\sigma}_{\lambda_t}}{\sigma_\eta^2}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_\eta^2}(\lambda_{t+1}^2 + \rho^2\lambda_{t-1}^2) - 2\widehat{\gamma}_{t,0} - \frac{\widetilde{\mu}_{\lambda_t}}{\widetilde{\sigma}^2_{\lambda_t}}\right)\right\} \\
\widetilde{\sigma}^2_{\lambda_T} &= \left(\frac{1+\rho^2}{\sigma_\eta^2} - 2\widehat{\gamma}_{T,2}\right)^{-1} \\
\widetilde{\mu}_{\lambda_T} &= \widetilde{\sigma}^2_{\lambda_T}\left(\frac{\rho\lambda_{T-1}}{\sigma_\eta^2} + \widehat{\gamma}_{T,1}\right) \\
\widetilde{c}_T &= \frac{\widetilde{\sigma}_{\lambda_t}}{\sigma_\eta}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma_\eta^2}\rho^2\lambda_{T-1}^2 - 2\widehat{\gamma}_{T,0} - \frac{\widetilde{\mu}_{\lambda_T}}{\widetilde{\sigma}^2_{\lambda_T}}\right)\right\}
\end{aligned}$$

The EIS samplers $\widetilde{m}_t$ are then used to construct AR-MH draws of $\underline{\lambda}$. Let $\underline{\widetilde{\lambda}}_K = (\widetilde{\lambda}_{1,K},\ldots,\widetilde{\lambda}_{T,K})$ denote the $K-$th Gibbs (AR-MH) draw of $\underline{\lambda}$ (which is included in the conditioning set $\underline{\widetilde{\xi}}_K$ for draw $K+1$). Draw $K+1$ of λ_t given $\left(\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right)$ obtains from the following two steps.

- An AR step which consists of drawing candidates l_t from $\widetilde{m}_t$ until one is accepted with acceptance probability

$$P(l_t) = \min\left[\frac{f\left(l_t|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right)}{\widetilde{c}_t\widetilde{m}\left(l_t|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right)}, 1\right]$$

Let $\widetilde{l}_t$ denote the accepted candidate;

- An MH step, whereby

$$\begin{aligned}\widetilde{\lambda}_{t,K+1} &= \widetilde{l}_t, \text{ with probability } \alpha\left(\widetilde{\lambda}_{t,K}, \widetilde{l}_t\right),\\ \widetilde{\lambda}_{t,K+1} &= \widetilde{\lambda}_{t,K} \text{ otherwise,}\end{aligned}$$

where

$$\alpha\left(\widetilde{\lambda}_{t,K}, \widetilde{l}_t\right) = \min\left[\frac{f\left(l_t|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right)}{f\left(\widetilde{\lambda}_{t,K}|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right)}\frac{\min\left[f\left(\widetilde{\lambda}_{t,K}|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right), \widetilde{c}_t\widetilde{m}_t\left(\widetilde{\lambda}_{t,K}|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right)\right]}{\min\left[f\left(l_t|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right), \widetilde{c}_t\widetilde{m}_t\left(l_t|\underline{\widetilde{\xi}}_{K/\widetilde{\lambda}_{t,K}}, Z\right)\right]}, 1\right]$$

See also Liesenfeld and Richard (2008) for a block version of this algorithm for high-dimensional $\underline{\lambda}$ vectors. However, with the present application with small T and generally low autocorrelation, we have found that individual AR-MH draws of $\{\lambda_1\}$ – i.e. blocks of size 1 – produce overall higher acceptance rates for $\widetilde{\lambda}$.

Sampling σ_τ^2 given $(\xi_{/\sigma_\tau^2}, Z)$

Under a non-informative prior density for σ and following e.g. Zellner (1971), the posterior density σ_τ is an Inverted Gamma with degrees of freedom N and a scale parameter $s = \left(\frac{1}{N}\sum_{i=1}^N \tau_i^2\right)^{1/2}$. We use N instead of $N-1$ degrees of freedom because this posterior density is conditional on the mean of τ_i being set equal to zero.

Sampling σ_η^2 given $(\xi_{/\sigma_\eta^2}, Z)$ and ρ given $(\xi_{/\rho}, Z)$

We follow Zellner (1971) (Section 4.1) with minor adjustments for the assumptions that λ_1 follows the stationary distribution of the AR(1) process and that our prior is $p(\sigma_\eta, \rho) \propto \sigma^{-1}$. Whence,

$$f\left(\sigma_\eta, \rho|\xi_{/(\sigma_\eta,\rho)}, Z\right) \propto \sqrt{1-\rho^2}\sigma_\eta^{-(T+1)} \exp\left\{-\frac{1}{2}\left(\frac{Ts_\eta^2}{\sigma_\eta^2}\right)\right\}$$

with

$$s_\eta^2 = \frac{1}{T}\left[\left(1-\rho^2\right)\lambda_1^2 + \sum_{t=2}^T(\lambda_t - \rho\lambda_{t-1})^2\right]$$

It immediately follows that:

- σ_τ^2 given $(\xi_{/\sigma_\tau^2}, Z)$ has an Inverted Gamma density with degrees of freedom T and scale parameter s_η;
- $f(\rho|\xi_{/\rho}, Z) \propto \sqrt{1-\rho^2}\exp\left\{-\frac{1}{2}\left(\frac{Ts_\eta^2}{\sigma_\eta^2}\right)\right\}$. Since this is a univariate density we follow the same procedure as for the individual τ_i and construct a piecewise log-linear approximation to $f(\rho|\xi_{/\rho}, Z)$.

References

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.

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

DeJong, D., H. Dharmarajan, R. Liesenfeld, and J. Richard (2007). An efficient approach to analyzing state-space representations. Mimeo.

Devroye, L. (1986). *Non-uniform Random Variate Generation*. New York: Springer-Verlag.

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.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), 1029–54.

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.

Liesenfeld, R. and J.-F. Richard (2008). Improving MCMC using efficient importance sampling. *Computational Statistics and Data Analysis*. forthcoming.

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

Richard, J.-F. (1977). Bayesian analysis of the regression model when the disturbances are generated by an autoregressive process. In A. Aykac and C. Brumat (Eds.), *New Developments in the Applications of Bayesian Methods*. North-Holland, Amsterdam.

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.

Zellner, A. (1971). *Introduction to Bayesian Inference in Econometrics*. New York: John Wiley and Sons.