Panel Probit with Flexible Correlated Effects: Quantifying Technology Spillovers in the Presence of Latent Heterogeneity*

Martin Burda†   Matthew Harding‡

October 18, 2011

Abstract

In this paper, we introduce a Bayesian panel probit model with two flexible latent effects: first, unobserved individual heterogeneity that is allowed to vary in the population according to a nonparametric distribution, and second, a latent serially correlated common error component. In doing so, we extend the approach developed in Albert and Chib (1993, 1996), and Chib and Carlin (1999) by releasing restrictive parametric assumptions on the latent individual effect and eliminating potential spurious state dependence with latent time effects. The model is found to outperform more traditional approaches in an extensive series of Monte Carlo simulations. We then apply the model to the estimation of a patent equation using firm level data on research and development (R&D). We find a strong effect of technology spillovers on R&D but little evidence of product market spillovers, consistent with economic theory. The distribution of latent firm effects is found to have a multimodal structure featuring within-industry firm clustering.

JEL: C11, C13, C15, C23, C25

Keywords: Dynamic latent variables, Markov Chain Monte Carlo, Dirichlet Process prior, R&D, Spillover effects

*We are grateful to Nick Bloom for sharing the data for the empirical application with us. We also thank Siddhartha Chib, David Dahl, Christian Gourieroux, Jerry Hausman, Ivan Jeliazkov, Andriy Norets, seminar participants at UC Berkeley and UC Irvine, and conference audiences at the North American Summer Meetings of the Econometric Society in Boston 2009, and the Canadian Econometics Study Group 2009 Ottawa meetings for useful comments. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca).

†Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada; Phone: (416) 978-4479; Email: `martin.burda@utoronto.ca`

‡Corresponding author. Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305; Phone: (650) 723-4116; Fax: (650) 725-5702; Email: `mch@stanford.edu`

1. Introduction

There is broad agreement that individual heterogeneity plays a crucial role in many economic models. In linear models, panel data can be used to identify the effects of interest while at the same time controlling for unobserved individual heterogeneity (Hausman and Taylor, 1981). Nonlinear models with unobserved heterogeneity pose substantial theoretical and computational challenges (Arellano and Hahn, 2006). In particular, in the case of nonlinear panel data models it is in general not possible to remove the unobserved effects by differencing as is commonly done in linear models. Convenient solutions can be obtained in some cases when a specific parametric form is assumed for the distribution of heterogeneity, such as in the negative binomial regression. Nonetheless, relaxing parametric assumptions on the distribution of unobserved heterogeneity in nonlinear models is important as often such restrictions cannot be justified by economic theory.

One possibility is to treat the unobserved effects as nuisance parameters to be estimated along with the parameters of interest. This approach requires large amounts of data though, as consistency is guaranteed only in the large N and large T limit. In most microeconomic applications, the econometrician only has a small number of repeated cross-sections to work with and the estimation of the individual fixed effects as incidental parameters induces bias. In the logit case Abrevaya (1999) shows that the model with fixed effects and only two time periods leads to severe bias and the estimated coefficients can reach twice their true value. In the parametric setting it is possible in some cases to circumnavigate this problem by redefining the quantity of interest. Fernandez-Val (2007) shows that under certain assumptions the inclusion of fixed effects does not affect the consistency of the marginal effects. More recently, Arellano and Bonhomme (2009) show that well chosen weights in average or integrated likelihood settings can produce estimators that are first-order unbiased.

Removing the parametric assumptions on the distribution of unobserved heterogeneity is also beneficial since economic models are usually silent on how to formally describe individual heterogeneity. At the same time, recent attempts at estimating nonlinear models nonparametrically are often rather difficult to implement (Berry and Haile, 2009).

Fuelled by advances in computation, as well as their flexibility and conceptual simplicity, Bayesian methods provide a powerful alternative to the more traditional approaches to solving these problems. In particular, 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. This makes Bayesian modeling an extremely flexible tool and a promising avenue to explore relaxing the assumptions discussed.

In some special cases such as the probit model, Bayesian data augmentation completely avoids the need to specify the likelihood in the form of a multivariate integral. This feature was introduced for the probit model in a seminal paper by Albert and Chib (1993). Instead of formulating the likelihood by integrating out the latent utility, the estimation problem is re-cast in the form of an iterative scheme of linear regressions where the latent utility is explicitly sampled along with other model parameters. Thus, the estimation is free from the curse of dimensionality that plagues inference with integral-based likelihoods. The approach was further developed for limited dependent variable (LDV) models to include parametric random latent effects in Albert and Chib (1996), Chib and Carlin (1999), and Gu, Fiebig, Cripps, and Kohn (2009).

In this paper, we further extend this line of research by introducing a model with two latent variables: first, we introduce unobserved individual heterogeneity that is allowed to vary in the population according to a nonparametric distribution, and second a latent error component that is serially correlated over time. The unobserved individual effects are allowed to be correlated with the observed regressors, in the spirit of Chamberlain (1982, 1984). Our model thus extends beyond the class of traditional random effects models (for a discussion on this issue, see e.g. Wooldridge, 2001). We model the distribution of the unobserved heterogeneity component with a nonparametric Dirichlet Process (DP) mixture model. The prior for the latent time component is specified as a parametric autoregressive process but its influence decreases linearly with the amount of data available. Due to its structure we label the proposed model as the "flexible latent effects probit" (FLEP). We note that individual building blocks of our model have been used in separate settings, such as modeling autoregressive processes in discrete choice models (Allenby and Lenk, 1994) and implementing the Dirichlet Process prior for studying heterogeneity in choice models (Li and Zheng, 2008; Rossi, 2010). However, the combined model with panel latent effects considered here has not yet been applied in the literature. Our aim in this paper is to show how to account for both flexible forms of unobserved heterogeneity and common latent time effects within the same framework.

We conduct an extensive empirical analysis of the decision to innovate where we suspect that unobserved heterogeneity plays an important role at the firm level. At the same time patenting activity may also be driven by a common time trend reflecting the macroeconomic environment or the overall stock of scientific knowledge. Without properly accounting for these latent effects it is not possible to correctly identify effects of interest or test hypotheses based on economic theory. We use data from a recent study of firm level research and development (R&D) by Bloom, Schankerman, and Van Reenen (2010) (henceforth BSV) to estimate a patent equation and test theoretical predictions on R&D spillovers. The dataset captures the majority of the patents granted between 1980 and 2001 in the US.

We explore the possibility that R&D leads to two major externalities. One the one hand, R&D may increase the productivity of firms using similar technology whereby a firm can benefit from the R&D conducted by another firm in the same technology area. On the other hand, R&D can have a product market rivalry effect with a number of firms striving to develop essentially the same product, which is detrimental to social welfare. Economic theory predicts that the marginal effect of technology spillovers on patenting activity is positive while the marginal effect of product market spillovers on patenting activity is zero. Our econometric approach allows us to additionally account for the two important types of latent effects in the analysis of R&D spillovers mentioned above: firm level heterogeneity and common time factors. We document the presence of both statistically significant technology spillover effects and firm level heterogeneity. The estimated distribution of firm level heterogeneity shows many interesting features and its multimodality suggests the clustering of heterogeneity across different firms. One important advantage of our approach is that it estimates firm level clustering without having to rely on a priori guesses of the form of heterogeneity. As we shall see industry classifications, a common proxy for heterogeneity, does a poor job at capturing the measured variation in latent firm level heterogeneity.

Our paper also introduces a series of computational innovations for the Bayesian estimation of this class of models. An core component in our implementation strategy is the efficient computation of the posteriors using a recent Sequentially Allocated Merge-Split (SAMS) algorithm (Dahl, 2005) that is substantially more efficient than samplers used previously in similar contexts. The SAMS sampler can update in one move large blocks of elements involved in implementation of the Dirichlet Process sampling scheme. It thus avoids a shortcoming of sequential samplers, such as the Polya urn scheme, that can get stuck in particular clustering configurations due to the one-at-a-time nature of their updates. Moreover, the SAMS algorithm is applicable to both conjugate and non-conjugate DP mixture models.

Our approach builds on the stream of literature aiming to relax restrictive assumptions of existing limited dependent variable models. A recent state-of-the-art Bayesian nonparametric analysis was introduced in Chib and Jeliazkov (2006) who study a binary dependent variable model with AR(p) errors and normally distributed unobserved individual heterogeneity. These authors focus on a non-parametric estimation of an unknown function of the model covariates, while we model nonparametrically the distribution of the unobserved individual heterogeneity. Burda, Harding, and Hausman (2008) analyze a flexible model for multinomial discrete choice with a flexible distribution of several parameters on the observable regressors. Their unobserved error component was fully parametric with an extreme-value type 1 distribution. As a result, their model was based on the logit closed-form solution facilitated by such assumption. Moreover, their model did not incorporate any dynamic element. In contrast, the error component in our model contains both flexible unobserved individual heterogeneity and common latent time effects which makes our estimation method suitable for panel data with a dynamic latent factor structure. The Normal distribution of the transitory idiosyncratic component stipulates a probit structure here, precluding the closed-form logit likelihood derivation utilized in Burda, Harding, and Hausman (2008). Instead, here we rely on data augmentation due to Albert and Chib (1993) using an iterative scheme of linear regressions in sampling the latent utility along with other model parameters.

Random error components that induce correlation over alternatives and time can also be accommodated by frequentist procedures. Such approach would assume a model for the distribution of the latent components and then specify the model likelihood in the form of an integral whose dimensions are formed by the individual unobserved components. Typically, such integral is analytically intractable and hence is estimated by numerical simulation methods, such as the GHK simulator developed by Geweke (1991), Hajivassiliou (1990), and Keane (1990). The resulting simulated likelihood (SML) is then maximized with respect to the model parameters. The GHK procedure thus numerically approximates the likelihood integral until convergence at every iteration of the model parameters within the optimization procedure. In contrast, Bayesian Gibbs sampling factorizes the high-dimensional multivariate integral into a sequence of low-dimensional conditional density kernels, drawing one dimension at a time until until a single convergence state of the resulting Markov chain is attained. In many cases, this implies that Bayesian parameter estimation is substantially faster than SML. For example, in an empirical comparison study for a parametric multinomial probit model Bolduc, Fortin, and Gordon (1997) found the Bayesian approach about twice as fast and much simpler to implement, both conceptually and computationally, than the GHK method.

Moreover, the Bayesian Markov chain of parameter draws can be directly used for inference in analogy to a bootstrap sample. In contrast, frequentist SML procedures including GHK require additional estimation of the shape of the simulated likelihood around the argmax parameter value; this process is fraught with peril as integral likelihoods often suffer from multiple local modes or saddles (Knittel and Metaxoglou, 2008). Dealing with such features is avoided using the Bayesian approach. In a comparison study between a Bayesian approach and the frequentist SML approach for a class of parametric mixed logit models, Train (2001) finds the Bayesian approach to possess theoretical advantages from both a classical and Bayesian perspective. Additional benefits of Bayesian inference in latent variable models are discussed e.g. in Paap (2002).

The advantages of Bayesian methods become even more pronounced with increased dimensionality of the underlying problem. A nonparametric model for the distribution of unobserved heterogeneity, as considered in this paper, if estimated using the GHK approach, would necessitate maximization of a flexible functional form such as a series or kernel estimator involving a large number of parameter iterations. The high-dimensional likelihood integral would need to be numerically approximated to a sufficient degree of precision at each of these iterations, which may become computationally prohibitive for larger sample sizes. In contrast, the Bayesian conditional Gibbs sampling can be performed very accurately along each latent dimension whereby higher dimensionality of the problem does not diminish the precision of inference.

The remainder of the paper is organized as follows. Section 2 introduces our model and discusses the assumptions and sampling procedures. Section 3 presents a series of Monte-Carlo studies comparing the performance of our method with other existing approaches. Section 4 presents an application of the method to the estimation of the effect of technological spillovers and product market competition on innovation. Section 5 concludes.

2. Model

Consider a sample of binary responses y_{it}, for N individuals indexed by i, and T time periods indexed by t. We assume that the data are drawn from the following error-components model:

$$\widetilde{y}_{it} = \mathbf{x}_{it}\beta + u_{it} \tag{2.1}$$

$$u_{it} = \tau_i + \lambda_t + \epsilon_{it}$$

$$y_{it} = \mathbf{1}\left(\widetilde{y}_{it} \geq 0\right) \tag{2.2}$$

where $\mathbf{x}_{it}$ is a $(1 \times K)$ vector of explanatory variables, τ_i represents unobserved individual heterogeneity, λ_t captures latent time effects, and $\mathbf{1}(\mathcal{C})$ denotes the indicator function which takes the value one if the condition $\mathcal{C}$ is satisfied and zero otherwise. The term $\widetilde{y}_{it}$ can be thought of as a latent utility of individual

i at time t. In this error-components model the unobserved error u_{it} is decomposed into three parts: an individual specific error τ_i, a time specific component λ_t and an idiosyncratic and transitory shock ϵ_{it}. This structure of u_{it} allows for both the presence of individual heterogeneity and serial correlation in the residual while these components can still be separately identified. In this model we observe the covariates x_{it}, but not τ_i, λ_t or ϵ_{it}. The model is specified in terms of the latent variable $\widetilde{y}_{it}$, not observed by the econometrician, who only observes the binary outcome variable y_{it}.

Let $\widetilde{\mathbf{y}}_i = (\widetilde{y}_{i1}, \ldots, \widetilde{y}_{iT})'$, $\widetilde{\mathbf{y}} = (\widetilde{\mathbf{y}}_1', \ldots, \widetilde{\mathbf{y}}_N')'$, $\mathbf{X}_i = (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}')'$, and $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_N')'$, $\lambda = (\lambda_1, \ldots, \lambda_T)'$, $\tau = (\tau_1 \iota', \ldots, \tau_N \iota')'$, $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$, $\epsilon = (\epsilon_1', \ldots, \epsilon_T')'$ and let ι denote a $(T \times 1)$ vector of ones. Model (2.1) can thus be re-written more compactly as $\widetilde{\mathbf{y}}_i = \mathbf{X}_i \beta + \tau_i \iota + \lambda + \epsilon_i$ for $i = 1, \ldots, N$ or simply $\widetilde{\mathbf{y}} = \mathbf{X}\beta + \tau + \lambda + \epsilon$. Following the notation in Geweke (2005), let the set-valued function $C_{it} = c_{it}(\widetilde{y}_{it})$ with $C_{it} = (-\infty, 0]$ if $y_{it} = 0$ and $C_{it} = (0, \infty)$ if $y_{it} = 1$. Denote the collection $\mathbf{C} = \{C_{it} : i = 1, \ldots, N; t = 1, \ldots, T\}$.

The hierarchical structure of our model allows us to distinguish four different layers of parameters. The first layer corresponds to the structure of the error components τ_i, $\lambda_t$and ϵ_{it}. Its properties are given by the following Assumption:

ASSUMPTION 1. *The error components τ_i, λ_t, and ϵ_{ti}, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, are mutually independent conditionally on the $\boldsymbol{X}$.*

The second parameter layer characterizes the distributional properties of the first-layer parameters in Assumptions 2-4.

ASSUMPTION 2. *The variables $\tau_1, \ldots, \tau_N$ are independent, identically distributed*

$$\tau_i \sim F_\tau^0$$

where F_τ^0 is a continuous unknown distribution, conditionally on X_i and the other model parameters of primary economic interest.

Instead of imposing a parametric family model, F_τ^0 will be estimated as an infinite mixture of distributions using a Bayesian Dirichlet Process Mixture (DPM) model which we shall introduce below. Moreover, our sampling mechanism allows for joint posterior correlation of τ_i with other regressors. We do not impose any prior assumptions on this feature explicitly due to the absence of any initial information on this property.

Since τ_i is sampled conditional on X_i, such potential relationship is entirely data-driven. The next assumption specifies the prior distribution for the latent time effects:

ASSUMPTION 3. λ_t *is assumed to follow a stationary Gaussian autoregressive process*

$$\lambda_t = \rho_1 \lambda_{t-1} + \ldots + \rho_s \lambda_{t-s} + \eta_t,$$

with $\eta_t \sim N(0, \sigma_\eta^2)$. *Furthermore,* η_t *is independent of* ϵ_{ti} *for each* $t = 1, \ldots, T$.

Failure to account for serial correlation of the error term has potential negative consequences. In the Bayesian framework, the posterior distribution is a weighted average of the prior distribution and the parameter update learned from the data via the likelihood function. The latter is implied by the probit model specification (2.1). In our sampling scheme detailed below, the prior has weight $1/(T+1)$ while the likelihood information weights $T/(T+1)$. In samples with very small T this autoregressive specification for the prior impacts inference but the prior influence declines linearly with T. The autoregressive prior specification also facilitates learning about the posterior distribution of the latent time process hyperparameters ρ and σ_η that provide information about the nature of the persistence and volatility in the latent time error component.

The parametric nature of Assumption 3 renders the dynamic model specification potentially quite restrictive, especially in cases where the data-generating process follows some other form of dynamics. Nonetheless, panels of data in micro-econometric applications are typically characterized by large N and small T dimensions and hence a parametric model appears as a suitable way to capture the relatively limited amount of information conveyed by the time dimension. Conversely, the relatively rich informational content of the cross-sectional dimension lends itself to non-parametric modeling which we undertake in this paper.

Assumption 3 is stated conditional on a given lag order s. Model selection criteria can be further employed to determine the optimal lag order for a given dataset. A method of lag selection for the autoregressive model is discussed in Troughton and Godsill (1997).

The following assumption defines the probit structure of the model:

ASSUMPTION 4. $\epsilon_{it} \sim N(0, 1)$ *is a stochastic error component uncorrelated with any other regressor.*

Our proposed model builds on the traditional error-components framework due to its popularity in applied work. The random error to an observation, $u_{it} = \tau_i + \lambda_t + \epsilon_{it}$ is given by the sum of an individual effect τ_i, a time effect λ_t and an idiosyncratic shock ϵ_{it}. Variations on this framework can be readily incorporated into our model.

The third parameter layer in our model is formed by parameters of primary economic interest captured in the vector $\theta = (\beta', \sigma_\eta, \rho')'$. The assumptions on the prior distributions for this layer are specified as follows:

ASSUMPTION 5.

$$\beta \sim N(\underline{\beta}, \underline{\Sigma}_\beta) \tag{2.3}$$

$$\sigma_\eta^2 \sim IG(v_0, s_0) \tag{2.4}$$

$$\rho \sim Uniform(\Omega) \tag{2.5}$$

where $\Omega \subseteq \mathbb{R}^s$ is the stationarity region of the autoregressive process.

The fourth parameter layer is comprised of the remaining hyperparameters introduced in Assumptions 2-5. In order to fully characterize this layer, we will elaborate on the model specified for the distribution of the unobserved heterogeneity component. Assumption 2 implies the following model based on Neal (2000):

$$\tau_i|\psi_i \sim F_\tau(\psi_i) \tag{2.6}$$

$$\psi_i|G \sim G \tag{2.7}$$

$$G \sim DP(\alpha, G_0) \tag{2.8}$$

Thus, F_τ is specified as an infinite mixture of distributions $F_\tau(\psi)$ with the mixing distribution over ψ being G. Here, ψ_i are hyperparameters of the distribution $F_\tau(\psi_i)$ of τ_i drawn from a random probability measure G which itself is distributed according to a Dirichlet Process (DP) prior. The DP prior for G is indexed by two hyperparameters: a distribution G_0 that defines the "location" of the DP prior, and a positive scalar precision parameter α. The distribution G_0 may be viewed as a baseline prior that would be used in a typical parametric analysis. The flexibility of the DP prior model environment stems from allowing G to stochastically deviate from G_0. The precision parameter α determines the concentration of the prior for G around the DP prior location G_0 and thus measures the strength of belief in G_0. For large values of α, a sampled G is very likely to be close to G_0, and vice versa. Early important applications of the DP prior to economics were made in Chib and Hamilton (2002) and Hirano (2002).

By Assumption 2 the distribution F_τ is sampled conditional on the primary parameters of economic interest θ and on the regressors $\mathbf{X}$. This sampling framework gives us the flexibility to treat τ_i as nuisance parameters while at the same time allowing for the possibility of the individual effects being correlated with other right hand side variables. Following Arellano and Bonhomme (2009) we implicitly assume that the support of F_τ contains an open neighborhood of the true parameters θ.

The fourth parameter layer is thus formed by the hyperparameters $\{\psi_i\}_{i=1}^N$, G, α, G_0, $\underline{\beta}$, $\underline{\Sigma}_\beta$, v_0, and s_0. In our implementation, G_0, $\underline{\beta}$, $\underline{\Sigma}_\beta$, v_0, and s_0 are fixed, $\{\psi_i\}_{i=1}^N$, and α are sampled, while bypassing explicit sampling of G.

Let $\tau = \{\tau_i\}_{i=1}^N$, $\mu_{it} = \mathbf{x}_{it}\beta + \tau_i + \lambda_t$, and denote by $\Phi(\mu_{it})$ and $\phi(\mu_{it})$ the cdf and pdf of the Normal random variable with unity variance, respectively. Denote generically by $p(\cdot)$ a probability density or mass function and by $k(\cdot)$ a prior density function. The posterior of our model can then be expressed as

$$
\begin{aligned}
p(\widetilde{\mathbf{y}}, \tau, \beta, \lambda, \sigma_\eta^2, \rho|\mathbf{y}) \quad &\propto \quad p(\mathbf{y}|\widetilde{\mathbf{y}}, \tau, \beta, \lambda, \sigma_\eta^2, \rho, \psi, \alpha)p(\widetilde{\mathbf{y}}|\tau, \beta, \lambda, \sigma_\eta^2, \rho, \psi, \alpha) \\
&\quad \times k(\psi|\alpha)k(\alpha)k(\beta)k(\lambda)k(\sigma_\eta^2)k(\rho)
\end{aligned}
\tag{2.9}
$$

with $k(\rho)$, $k(\beta)$, and $k(\sigma_\eta^2)$ given in Assumption 5, $k(\lambda)$ in Section 5.5 in the Appendix, $k(\alpha)$ specified as in Escobar and West (1995), and $k(\psi|\alpha)$ given by (2.7-2.8). The remainder of the model is formulated similarly to Albert and Chib (1993) with the single index given by μ_{it}. Specifically, $p(\widetilde{\mathbf{y}}|\tau, \beta, \lambda, \sigma_\eta^2, \rho, \psi, \alpha) = \prod_i \prod_t \phi(\mu_{it})$ and $p(\mathbf{y}|\widetilde{\mathbf{y}}, \tau, \beta, \lambda, \sigma_\eta^2, \rho, \psi, \alpha)$ assigns probability mass one to $y_{it} = 1$ if $\widetilde{y}_{it} > 0$ and to $y_{it} = 0$ if $\widetilde{y}_{it} \leq 0$. Thus, y_{it} are independent Bernoulli random variables with $p_{it} = \Phi(\mu_{it})$.

2.1. **Average Partial Effects**

In nonlinear models the estimated coefficients are only of limited interest by themselves. Instead, the average partial effects (APEs) are particularly useful for computing economic counterfactuals and widely used in applied work. In this section we describe how they are computed within the setup of our model. We utilize the classical concept of the APEs augmented with the latent variables. Let

$$
\begin{aligned}
m_{itk} &= \frac{\partial E[y_{it}|\mathbf{x}_{it}\beta, \tau_i, \lambda_t]}{\partial \mathbf{x}_{itk}} \\
&= \phi\left(\mathbf{x}_{it}\beta + \tau_i + \lambda_t\right)\beta_k
\end{aligned}
$$

denote the marginal effect of a change in $\mathbf{x}_k$, where $\phi(\cdot)$ denotes the standard normal density function. Define

$$
\widetilde{\gamma} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi\left(\mathbf{x}_{it}\beta + \tau_i + \lambda_t\right) \tag{2.10}
$$

The APE of $\mathbf{x}_k$ on $\mathbf{y}$ is then given by

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} m_{itk} = \widetilde{\gamma}\beta_k \tag{2.11}$$

We sample explicitly τ_i and λ_t throughout the MC iterations and hence can compute the APEs directly from the definition of $\widetilde{\gamma}$, as

$$\gamma = \frac{1}{NTS}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{S}\phi\left(\mathbf{x}_{it}\beta_s + \tau_{is} + \lambda_{ts}\right) \tag{2.12}$$

where s is the index over MC steps.

In both the application and the simulation study, we report the mean bias and means squared error of the estimated "APE scale" coefficient γ defined in (2.12). To obtain the APEs, γ is simply multiplied by each respective β_k.

3. Monte-Carlo Simulation

In order to highlight the potential advantages of our approach, we test the performance of our approach on a series of Monte Carlo studies based on simulated datasets. Our goal is to evaluate the robustness of our approach to relaxing the three main assumptions which are the main focus of our study. As our benchmark comparison models we chose the closest parametric alternatives, namely the fixed effects panel probit model with a full set of time dummies, and the random effects panel probit model augmented with the Chamberlain (1982) device incorporating the explanatory variables in all time periods, also with with a full set of time dummies. Both of these model are estimated in Stata.

All simulation results are reported in Tables 8-11. Keeping the same simulation design we vary the amount of data available at each run by varying N and T. We consider simulations where $N \in \{100, 300, 700, 1000\}$ and $T \in \{10, 20, 50\}$, each in one hundred replications. We then report the mean bias and root mean square error for each coefficient β_k in our simulation design.[4] Additionally, for our model we also report these statistics for the latent time process parameters ρ and σ_η.

We have found our approach to be computationally very efficient. All chains mix very well and appear to have converged within the burn-in section. The autocorrelation of β_k become statistically insignificant after

[4]In an earlier version of the paper we also ran simulations for a base-case pooled probit and a random coefficient panel probit augmented with the Mundlak version of the Chamberlain device. The results were substantially biased: the mean absolute deviation on the APE scale parameter was 0.295 in the former case and 0.209 in the latter case for the largest sample size.

10-15 lags while the ones of σ_η and ρ after 1-2 lags. In Tables 1-2 we present the mean absolute deviation and mean square errors for a simple classical data generating process where τ_i is uncorrelated with any of the included right hand side variables $\mathbf{x_k}$. We generate the data as follows. All $\mathbf{x_k}$ observations are drawn as $N(0, 1/9)$. The true parameter values are $\beta_0 = \{0, 1, 1, -1\}$, $\sigma_{\eta 0} = 0.5$, $\rho_0 = 0.5$. We draw the unobserved heterogeneity as an equal mixture between two Normal components, with means -2 and 2 and variance 1/5. The length of the MCMC chain is 5000 and we discard the first 4000 steps as the burn-in period. We initialize the algorithm by letting all observations be part of the same latent class, and allow the sampler to do further splits as required.

The *fixed effects probit* model with time dummies performs badly in terms of mean bias independently of the sample size. Typically we find mean bias in excess of 50%. The bias in the APE scale is also very large. Recall that the APE scale is bounded between 0 and approximately 0.4, thus a mean bias of 0.04 constitutes exceeds 10% deviation from the true value. The *random effects probit* with time effects improves somewhat in its performance regarding the β coefficients, albeit marginally, but performs worse in estimating the APEs, averaging some 40% bias. In contrast our nonparametric approach to modeling the distribution of unobserved heterogeneity adapts to any such distribution present in the data. As a result, the proposed *FLEP* model performs very well. Even in samples as small as $N = 100$ and $T = 10$ the bias in the coefficients is less than 5%. The bias in the APE scale is very small throughout.

In Tables 3-4 we report results for the simulation design where the individual effects are correlated with the right hand side variables. In this case we augment the random effects probit with the Chamberlain (1982) device which alleviates potential biases associated with such correlation. We employ the following simple classical simulation design. The variables $\mathbf{x_2}$ and $\mathbf{x_3}$ are drawn uniformly on $[-1, 1]$. We generate $\mathbf{x_1}$ to be correlated with τ_i using the following stylized approach to induce correlations. For a draw $u_i \sim U[0, 1]$, we generate τ_i and $\mathbf{x_1}$ as:

- 1st quarter of individuals: $\tau_i = 3 + u_i$ and $\mathbf{x_{1i}} = -4 - 0.2u_i$
- 2nd quarter of individuals: $\tau_i = -3 - u_i$ and $\mathbf{x_{1i}} = -1 - 0.2u_i$
- 3rd quarter of individuals: $\tau_i = 3 + u_i$ and $\mathbf{x_{1i}} = 1 - 0.2u_i$
- 4th quarter of individuals: $\tau_i = -3 - u_i$ and $\mathbf{x_{1i}} = 4 - 0.2u_i$

This yields a correlation coefficient between τ_i and $\mathbf{x_{1i}}$ close to -0.5 in each case. The true parameter values are $\beta_0 = \{0, 0.5, 0.5, 0.5\}$, $\sigma_{\eta 0} = 0.75$, $\rho_0 = 0.5$. The length of the MCMC chain is 5000 and we discard the first 4000 steps as the burn-in period. We initialize the algorithm by letting all observations be part of the same latent class, and allow the sampler to do further splits as required.

The coefficients estimates derived from running the *fixed effects probit* model with time dummies are severely biased. The bias appears to be very similar at different sample sizes. In fact the magnitudes of parameter biases with a sample of $N = 1000,\ T = 50$ are comparable to those estimated with $N = 100,\ T = 10$ observations. The estimate of the APE scale is also biased, in excess of 10%. The parameter estimates of the *random effects probit* with time dummies and the Chamberlain (1982) device are biased commensurately and some coefficients are estimated with a large mean square error. The the APE scale bias does not appear to diminish with increasing N. For sample sizes with $T = 50$ Stata on our 3 GHz PC could not deliver the necessary MC replication estimates in real time (days) and hence these results are not reported.

By contrast, our *FLEP* model proposed in this paper performs very well both in terms of bias and mean square error. The bias in the APE scale is also very small, indicating that the marginal effects will be nearly unbiased in almost all samples considered in this exercise. As anticipated, the nonparametric model for the unobserved heterogeneity component adapts to the multiple modes of its distribution dispersed in all quadrants of $\mathbf{x_{1i}}$. The standard errors in all cases are of comparable magnitude.

We further consider the case where the unobserved heterogeneity component is exactly Normally distributed with $\tau_i \sim N(0, \sigma_\tau^2)$, and hence the *random effects probit* should be asymptotically efficient. All other model attributes are kept as in the uncorrelated case. The comparison of the RE and FLEP output is reported in Table 10. The APE scale parameter and the intercept β_0 have smaller bias and RMSE under the FLEP. The partial effects β_1 to β_3 feature mixed results, with bias and RMSE smaller under the RE for some data sizes and under the FLEP for others. The standard errors, however, are smaller under the RE, except for β_0.

Our last simulation design investigates the effects of misspecification of the parametric models assumed for the time effects and the idiosyncratic component on the FLEP estimator. Two cases are considered in turn: (a) highly autocorrelated case with λ_t drawn from an $AR(2)$ model with $\rho_1 = 0.5$ and $\rho_2 = 0.4$, and (b) one-lag autocorrelation case with λ_t drawn from an $MA(1)$ model with $\rho = 0.5$. The output of these cases (Table 11) is of comparable magnitude relative to the well specified model.

To summarize our simulation results, we have found that estimating a large number of incidental parameters or restricting the model distribution of unobserved heterogeneity leads to extremely biased results in our parametric benchmark models. Moreover, the estimation of a large number of incidental parameters poses computational challenges, and can indeed lead to inconsistent estimates. The flexible latent effects probit model introduced in this paper performs very well and leads to nearly unbiased results for both the model parameters and the estimated average partial effects. These results hold well under specific types of misspecification.

4. Innovation and R&D Spillovers

4.1. The Role of Latent Effects in R&D Analysis

An ongoing puzzle in the economic literature on R&D concerns the relationship between innovation as measured by the patenting activity of firms and the spillover effects resulting from the strategic interactions between firms. Firms often interact in geographically delimited markets which leads to a localization of the spillover effects in terms of geographic distance (Griffith and Van Reenen, 2011). At the same time firms interact in more abstract spaces such as the technology space defined by the extent to which two firms are close to each other in terms of the underlying technology and the product market space defined by the extent to which two firms compete for the same product market (Bloom, Schankerman, and Van Reenen, 2010). Such spillover effects often have contradictory impacts on firm performance. While technological spillover effects may benefit a firm by enhancing the overall stock of knowledge to the firm, product market spillovers can lead to business stealing due to overlapping product offerings to consumers. These issues have been explored in the theoretical literature but have been very difficult to estimate empirically due to the presence of confounding latent effects which are particularly problematic in this setting.

On the one hand, innovation and the patenting activity of firms is likely to be influenced by unobserved firm level heterogeneity. Firm-specific differences in corporate culture, investment strategies, know-how, or brand name will arguably shape the different degrees of intensity of the innovation activity in firms. Griffith, Lee, and Van Reenen (2011) show that ignoring unobserved heterogeneity can have a large quantitative impact on our understanding of innovative activity. On the other hand, the econometric analysis of spillover effects is confounded by the presence of systematic common time effects reflecting the macroeconomic environment, technological trends, or global political events that also affect the resource allocation in firms R&D funding. When such common time effects dominate it is easy to falsely attribute the observed correlation in innovative activity to spillover effects.

The econometric analysis of innovation thus presents the econometrician with an important challenge in terms of consistently estimating the effect of spillover effects between firms while accounting for the presence of unobserved individual and time effects. The Bayesian model introduced in this paper presents a useful approach to the consistent estimation of spillover effects while accounting for the presence of latent individual and time effects. As we will show, our proposed model is not only computationally feasible to implement on large datasets but it is also superior to more traditional frequentist approaches in terms of its ability to correctly predict the incidence of innovative activity in firms.

In particular, we apply the method developed in this paper to the estimation of a patent equation on firm level data and test theoretical predictions on R&D spillovers. Since no economic theory is available that would recommend a particular distributional form for the unobserved heterogeneity we can take advantage of an important feature of our model, namely the ability to specify the distribution of the unobserved individual heterogeneity component non-parametrically. As we will show this expands the applicability of the analysis substantially since it allows us to investigate the presence of clustering in the unobserved effects and derive new economic insights into the innovative activities of firms in different industries. At the same time our model is flexible enough to account for the presence of potentially confounding time factors which if not properly accounted will induce spurious correlations in innovative activity not attributable to the spillover effects under consideration.

4.2. **Data**

We employ data from a recent study of firm level R&D by Bloom, Schankerman, and Van Reenen (2010) (denoted by BSV for the rest of this section). BSV collected firm level accounting data, such as sales, from the US Compustat database. This data were then matched to the NBER US Patent and Trademark Office data containing detailed information on granted US patents, yielding an unbalanced panel of 729 firms with observations recorded between 1980 and 2001.

BSV investigate two major spillover effects of R&D, technological and product market spillovers. One the one hand, R&D may increase the productivity of firms using similar technology. A firm can benefit from the R&D conducted by another firm in the same technology area. On the other hand, it can have a product market rivalry effect, which is detrimental to social welfare. Using the firm level information available, BSV attempt to map the location of each firm in both the technology and product space, by comparing information on patents and information on sales across firms.

Following BSV we measure the technological closeness between firms using information on patents for each firm. All available patents are allocated into $\kappa = 1, \ldots, 425$ different technological classes. If we then let $T_i = (T_{i\kappa})$ denote a vector where each element represents the average share of patents of firm i in technological class κ over the period 1980 to 2001, we can define technological closeness ($Tech$) between two firms i and j by the uncentered correlation between the allocations for the two firms:

$$Tech_{i,j} = \frac{T_i T_j'}{(T_i T_i'')^{1/2} (T_j T_j')^{1/2}}. \tag{4.1}$$

The degree of technology spillover $SpillTech$ is then measured as the technology distance weighted average of the R&D stock of all other firms at each point in time:

$$SpillTech_{i,t} = \sum_{j,i \neq j} Tech_{i,j} R_{j,t}, \tag{4.2}$$

where $R_{j,t}$ is the stock of R&D of firm j at time t computed from the expenditure on R&D data available in the accounting statements recorded by U.S. Compustat.

Similarly, the distance between firms in the product market can be computed by decomposing each firm's sales by the respective four digit industry code. Most firms are multi-product firms with reported sales in an average of 5.2 different industry codes. The sample of firms spans a total of 762 different industries. The distance between firms in the product market is then measured as the uncentered correlation between the allocation of sales activity of firms into industries. The degree of product market spillovers ($SpillSIC$) is computed as the product market distance weighted average of the R&D stock of all other firms.

The above definition of technology and product market spillovers are based on the Jaffe (1986) distance measure. BSV note as its drawback the implicit assumption of technology spillovers only occurring within the same product technology class. Patent class categorizations however are extremely narrow. As BSV illustrate the Patent Office distinguishes between "arithmetic processing and calculating" and "processing architectures and instruction processing" when they both may refer to very similar computer technology. Moreover, categorization into patent classes may well be subject to measurement error. As such it is worthwhile to investigate additional distance measures which take into account the fact that technological spillovers may occur across patent classes. One such option is the Mahalanobis distance which allows spillovers to occur across multiple patent classes but weighs their importance by the extent to which a firm is active across different patent classes. A Mahalanobis measure of the product market spillovers is constructed similarly. We enrich our analysis by adding a Mahalanobis version of the $SpillTech$ and $SpillTech$ variables, which will serve as a subsequent robustness check on our baseline specifications and guard against measurement errors resulting from the more narrow Jaffe variable construction.

The dataset contains two additional variables of interest. The first is the R&D stock which has already been mentioned above. The second is a firm specific measure of industry sales (*Sales*). This variable uses the same *SIC* weighting technique as *SpillSIC* but applied to rival firm sales.

The dependent variable of interest *Patenting* is a binary variable denoting whether or not firm i filed at least one patent in year t. The data summary statistics are given in Table 1. All independent variables are expressed in logarithms and have been lagged by one period to remove simultaneity concerns.

Variable	Mean	S.D.	Min	Max
SpillTech (Jaffe)	9.554	1.142	4.838	11.707
SpillSIC (Jaffe)	7.272	2.323	-4.602	11.154
SpillTech (Mah.)	11.314	0.847	8.235	13.156
SpillSIC (Mah.)	8.513	1.660	-0.356	11.559
R&D Stock	3.030	3.026	-2.513	10.765
Sales	6.230	1.962	0	12.103
Patenting	0.540	0.498	0	1
Firms	729			
Total Obs	12928			

TABLE 1. Summary statistics. All variables are in logarithms and lagged by one period.

4.3. **Econometric Implementation**

We implement the model developed in Section 2 to the estimation of R&D spillover effects on patenting activity using the data described above. We use a Bayesian Gibbs sampling scheme (the precise implementation details of drawing from individual Gibbs blocks are given in the Appendix). Under the Model (2.1) and Assumptions 1-5, the joint posterior density can be decomposed into the following Gibbs blocks:

(1) $\beta|\tau, \lambda, \psi, \theta_{/\beta}, \widetilde{\mathbf{y}}, \mathbf{y}, \mathbf{X}$

(2) $\widetilde{\mathbf{y}}|\tau, \lambda, \psi, \theta, \mathbf{y}, \mathbf{X}$

(3) Update the assignments of τ_i to latent classes by alternating between the SAMS (Dahl, 2005) and Algorithm 7 (Neal, 2000), which includes sampling $\{\psi_i\}_{i=1}^N$

(4) $\tau_i|\psi, \theta, \lambda, \widetilde{\mathbf{y}}, \mathbf{y}, \mathbf{X}$ for each i

(5) $\lambda|\tau, \psi, \theta, \widetilde{\mathbf{y}}, \mathbf{y}, \mathbf{X}$

(6) $\sigma^2_\eta|\tau, \lambda, \psi, \theta_{/\sigma^2_\eta}, \widetilde{\mathbf{y}}, \mathbf{y}, \mathbf{X}$

(7) $\rho|\tau, \lambda, \psi, \theta_{/\rho}, \widetilde{\mathbf{y}}, \mathbf{y}, \mathbf{X}$

The Bayesian model described above contains a non-parametric specification of the individual effects and we will denote it as FLEP (flexible latent effects probit). It is possible to estimate a restricted parametric version of the same model by imposing the condition that in the unobserved individual effects are Normally distributed. We shall label this version of the model as PLEP (parametric latent effects probit). By comparing the results of different specifications of the unrestricted non-parametric version with the restricted parametric version of the same model we can gain additional insights into the importance and advantages of using a flexible non-parametric specification over more traditional parametric approaches. Posterior mens are reported for these two techniques, obtained from chains of total length of 10,000 MC steps with a 5,000 burn-in section.

Additionally we implement two frequentist approaches to the estimation of spillover effects. First, we implement the fixed effects probit model with time dummies (denoted by FE). As we shall see this approach suffers from serious computational limitations in large data. Second, we implement the random effects probit model with time dummies and the Chamberlain (1982) device (which we denote by RE).

Each estimation technique was applied to the two different specifications of the econometric model: one using the Jaffe distance measure and one using the Mahalanobis distance measure. Below we shall discuss the empirical results in detail and perform additional econometric robustness checks .

4.4. **Empirical Results**

In a simple model of R&D BSV show that it is possible to derive a number of theoretical implications of these two spillover effects. If we assume that the production of knowledge is exogenous then we would not expect to find an effect of market rivalry on patent counts. Empirically, this means that the coefficient on $SpillSIC$ should be close to zero. The presence of positive market spillover effects may however indicate endogenous patenting activity. Thus, we can investigate the extent to which strategic patenting activity is consistent with the evidence in the data.

At the same time we expect the marginal effect of technology spillovers on patent counts to be positive. The production of knowledge benefits from the innovation activity in a firm conditional on its own R&D stock. Empirically this implies that we should expect the coefficient on $SpillTech$ to positive and significant.

We will test these predictions using our model and several alternative benchmark models that are commonly applied in the empirical literature. Recall that we define the dependent variable to be one if the given firm registered a patent during the particular year or not. We can think of this case as an indicator of innovation for a given firm-year dyad. We then regress this indicator on the measures of technological and product market spillovers discussed above, $SpillTech$ and $SpillSIC$. In order to control control for observed firm level heterogeneity, we include two additional variables. One corresponds to firm sales $Ln(Sales)$, while the other corresponds to the pre-existing stock of R&D available within the firm $Ln(R\&D\ stock)$. Furthermore, we lag all right hand side variables by one period so as to remove the possibility of contemporaneous effects.

4.4.1. ***Partial Effects***

Estimation results on the partial effects and the latent common time component are reported in Table 2.

	Jaffe Distance				Mahalanobis Distance			
	FE probit	RE probit	PLEP	FLEP	FE probit	RE probit	PLEP	FLEP
Ln(SpillTech)	-0.247	0.251*	0.389*	0.364*	-0.125	0.660*	0.548*	0.641*
	(0.233)	(0.072)	(0.047)	(0.046)	(0.332)	(0.104)	(0.029)	(0.061)
Ln(SpillSIC)	0.073	0.011	0.021	0.037	-0.179	-0.200	-0.035*	-0.024
	(0.074)	(0.059)	(0.021)	(0.019)	(0.134)	(0.102)	(0.013)	(0.031)
Ln(R&D Stock)	0.091*	0.144*	0.304*	0.313*	0.095*	0.143*	0.267*	0.304*
	(0.044)	(0.037)	(0.028)	(0.027)	(0.044)	(0.037)	(0.016)	(0.031)
Ln(Sales)	0.377*	0.288*	0.204*	0.204*	0.383*	0.292*	0.100*	0.177*
	(0.050)	(0.044)	(0.027)	(0.022)	(0.050)	(0.043)	(0.013)	(0.023)
APE scale	0.203	0.187	0.169	0.170	0.203	0.184	0.220	0.169
ρ			0.654*	0.645*			0.670*	0.552*
			(0.142)	(0.145)			(0.146)	(0.164)

TABLE 2. Estimation of Patent Equation

Notes: Standard errors are reported in brackets. Coefficients significant at 5% confidence level are marked with an asterisk. All independent variables are lagged by one period. All regressions include a constant and a dummy for observations where lagged R&D stock is zero.

In the absence of endogenous patenting activity we should see the marginal effect of technology spillovers $SpillTech$ on patenting activity to be positive while the marginal effect of product market spillovers $SpillSIC$ to be zero.

Across the various model specifications we find the effect of market rivalry to be small and statistically insignificant, with the exception of PLEP for the Mahalanobis distance. Moreover, the effect changes sign depending on which distance measure is used. The evidence presented therefore does not reject the hypothesis of exogenous knowledge production. Moreover, across all specifications we observe positive and significant effects of the lagged R&D stock and the lagged sales, which is consistent with basic economic intuition.

If we use the FE probit model the estimated coefficient on *SpillTech* is not statistically significant. Moreover, the estimate appears to indicate a negative effect of technology spillovers, which contradicts economic theory. RE, PLEP and FLEP predict a positive and statistically significant effect of technology spillovers. The estimated magnitude differs however for each method. Note that while the results are qualitatively very similar for both the Jaffe distance measure and the Mahalanobis distance, they are quantitatively different.

It is noteworthy that both RE and FLEP produce results consistent with economic theory. Thus, it is worth investigating further which model performs better on other fronts. The quantitative difference in the estimated coefficients indicates that these models may in fact produce very different predictions. From a policy perspective we would like to know which model to use for more accurate predictions. To verify this point we contrasted the outcomes predicted by each method in both models with the actual outcomes observed in the data (Table 3).[5] In the Jaffe distance model, the FLEP predicted correctly 86% of outcomes and incorrectly 14% of outcomes, while the RE predicted correctly 79% of outcomes and incorrectly 21% of outcomes. In the Mahalanobis distance model, the FLEP predicted correctly 86% of outcomes and incorrectly 14% of outcomes, while the RE predicted correctly 81% of outcomes and incorrectly 19% of outcomes. On average, the RE has thus 48% higher prediction error rate than the FLEP.

		RE		FLEP	
	Predicted	0	1	0	1
	Actual				
Jaffe Distance	0	4473	1470	5026	917
	1	1190	5795	911	6074
Mahalanobis Distance	0	4591	1352	5020	923
	1	1148	5837	911	6074

Table 3. Actual vs predicted outcomes for RE and FLEP in the Jaffe distance model and Mahalanobis distance model.

[5]This tabulation is often termed a 'misclassification' (or 'confusion') matrix. The diagonal elements contain correctly predicted outcomes while the off-diagonal ones contain incorrectly predicted (confused) outcomes (Kohavi and Provost, 1998).

The latent effects models also indicate the existence of a time factor, measured as having moderate persistence over time with an autocorrelation coefficient of approximately 0.5. Recall that equation (2.10) implies that the APE scale depends not only on the estimated β coefficients but also on the estimates of the latent individual and time variables. A more precise estimate of the distribution of unobserved heterogeneity should improve the estimates of the APEs. The results in Table 2 show that the flexible latent effects probit model estimates a marginal effect for technology spillovers that is substantially larger than the marginal effect estimated by the random effects probit model. Such differences in the estimated quantities of interest may lead to very different policy implications.

4.4.2. ***Unobserved Firm Heterogeneity***

We have noted above that both FLEP and PLEP produce quantitatively similar results for both distance measures.[6] A key advantage of FLEP is that does not impose the normality constraint on the unobserved heterogeneity. Furthermore, FLEP is the only model that allows us to uncover a non-parametric estimate of the distribution of firm heterogeneity. There is no sound economic reason to assume that this distribution is normal and in fact we would expect that different types of production processes have very different forms of unobserved heterogeneity which impacts patenting activity. In our application, the distribution of heterogeneity is shown to have a multimodal clustering structure as plotted in Figure 1. These clusters may reflect the presence of missing variables important for characterizing innovation, such as firm culture or investment strategy. In Figure 1 we can easily discern several major clustering structures in each distance model, labeled by numbers in square boxes, corresponding to the major modes of the distribution. In the Appendix we show that this clustering is robust to the choice of the DPM prior hyperparameter.

In the FLEP output in Figure 1, each clustering structure is composed of draws of the firm-specific unobserved heterogeneity component τ_i which we can use to further analyze the composition of each cluster. Table 4 lists the SIC code names for 20 firms whose τ_i was most frequently drawn within each given clustering structure. Thus, for the Jaffe distance model, the lowest unobserved heterogeneity component group (Clustering 1) is composed e.g. of 'meat packing plants', 'blowers and fans', or 'department stores'; the medium unobserved heterogeneity component group (Clustering 2) includes 'food and kindred products', 'footwear', and 'electrical industrial apparatus'; while the high unobserved heterogeneity component group (Clustering 3) features 'semiconductors and related devices', 'electronic components', or 'commercial physical research'. The cluster

[6]Nonetheless, PLEP estimated statistically significant product market spillovers, which was not confirmed by any other model specification.

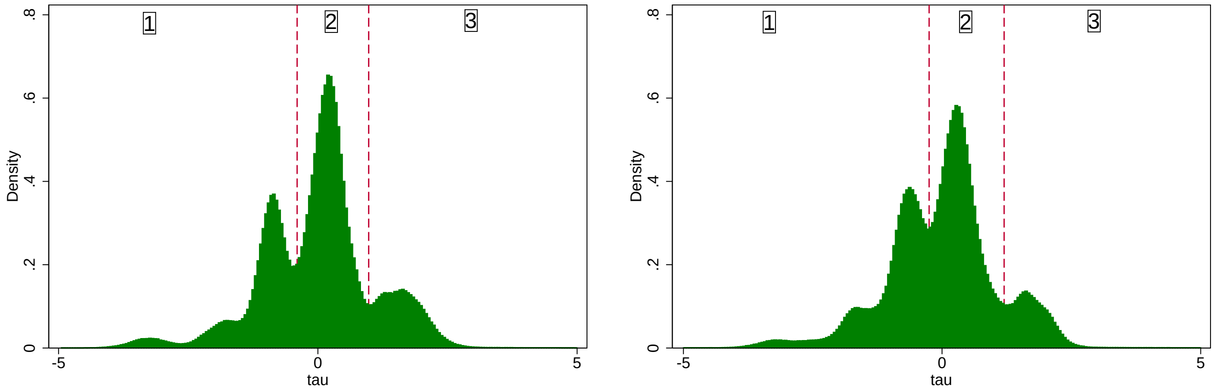


Figure 1. Flexible latent effects probit (FLEP) distribution of unobserved heterogeneity for the Jaffe distance model (left) and Mahalanobis distance model (right).

composition is very similar in the Mahalanobis distance model and hence not reported here. Uncovering such cluster structures of firms that behave similarly in terms of their unobserved characteristics can provide important insights for industry analysts and policy makers analyzing firms' R&D behavior. Below we shall further investigate the extent to which our theoretical predictions are satisfied for each cluster.

Clustering 1		Clustering 2		Clustering 3	
4923	Gas Transmission and Distribution	3841	Surgical and Medical Instruments	2835	Diagnostic Substances
2851	Paints and allied products	3661	Telephone and Telegraph Apparatus	2840	Cosmetics
3060	Fabricated rubber products	3825	Instruments To Measure Electricity	3714	Motor Vehicle Parts and Accessories
3440	Fabricated Structural Metal	3577	Computer Peripheral Equipment	2761	Manifold Business Forms
2011	Meat Packing Plants	4011	Railroads, Line-haul Operating	2390	Misc Fabricated Textile
2731	Book Publishing	3590	Misc industrial machinery	3823	Process Control Instruments
3561	Pumps and Pumping Equipment	3537	Industrial Trucks and Tractors	3572	Computer Storage Devices
3640	Electric lighting and wiring equipment	6324	Hospital and Medical Service Plans	3674	Semiconductors and Related Devices
2030	Canned, frozen, and preserved fruit	2000	Food and Kindered Products	3679	Electronic Components
3533	Oil and Gas Field Machinery	3669	Communications Equipment	3420	Handtools
3663	Radio and T.v. Communications Eqpt	3310	Steel Works, Blast Furnaces	2842	Sanitation Goods
3944	Games, Toys, and Children's Vehicles	3140	Footwear, Except Rubber	3990	Misc Manufacturing Industries
3621	Motors and Generators	3620	Electrical Industrial Apparatus	8731	Commercial Physical Research
2253	Knit Outerwear Mills	2834	Pharmaceutical Preparations	3613	Switchgear and Switchboard Apparatus
3579	Office Machines	3711	Motor Vehicles and Car Bodies	3670	Electronic Components and Accessories
3743	Railroad Equipment	4931	Electric and Other Services Combined	3530	Material Handling Equipment
3490	Miscellaneous fabricated metal products	3021	Rubber and Plastics Footwear	3829	Measuring and Controlling Devices
3564	Blowers and Fans	2911	Petroleum Refining	8731	Commercial Physical Research
2522	Office Furniture, Except Wood	3569	General Industrial Machinery	2821	Plastics Materials and Resins
5311	Department Stores	3690	Misc Electrical Machinery	3861	Photographic Equipment and Supplies

Table 4. Most frequent members of clustering structures for the Jaffe distance model.

Using the FLEP output we can also explore the evolution of the draws of the unobserved heterogeneity parameters τ_i for specific individual companies in order to investigate if the draws are concentrated or show large variances as the Markov Chain progresses. Overall, we have found the draws to be remarkably stable indicating a clear tendency of the model to associate each firm with a narrow range of draws of τ_i. If a draw of τ_i jumps to a different cluster, it does not stay there long and returns shortly back to its long term average. This is an important indicator that the clustering of τ_i values observed in the estimated distribution of unobserved heterogeneity may contain relevant information since it establishes a fairly tight link between firms and different modes of the distribution of heterogeneity. To exemplify we plot the draws of τ_i for the first five firms for the model of positive patents model in Figure 2.

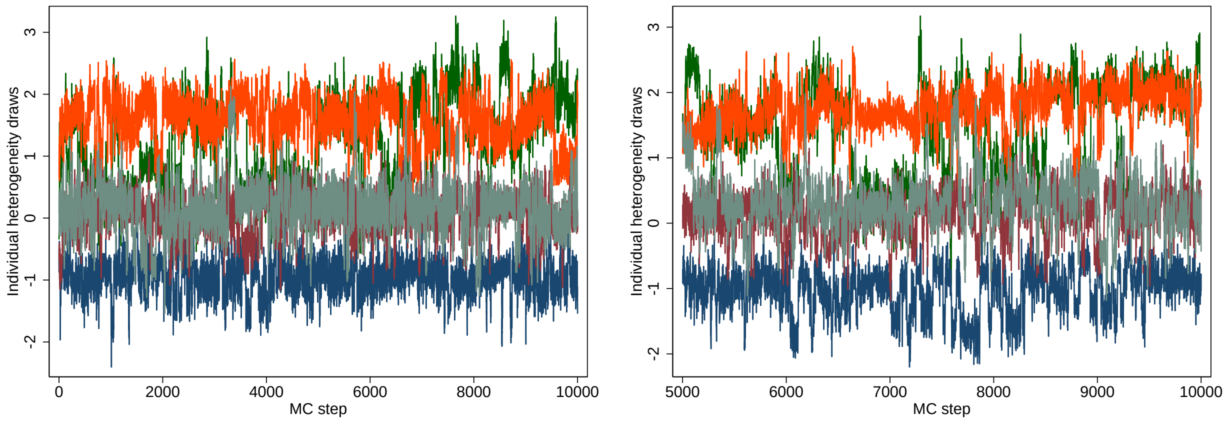


FIGURE 2. Draws of τ_i for the first five companies. Jaffe distance model (left) and Mahalanobis distance model (right).

The estimated distribution of the unobserved heterogeneity provides valuable economic information which can be used to analyze the data further and test additional economic hypotheses of interest. One such hypothesis is that there are unobserved industry level factors driving innovation. In order to investigate this hypothesis, we plot the average sampled value of the unobserved heterogeneity component τ_i by each SIC code for the two estimated models of innovation in Figure 3. The absence of any discernible pattern in the graphs suggests that the unobserved heterogeneity component is not driven by industry factors but is rather firm-specific at the individual level. Indeed, further examination of individual τ_i revealed large differences in company types even within SIC categories. It appears that associating the unobserved heterogeneity with industry categories and attempting to capture it for example by industry indicator variables may obscure important differences among firms regarding their innovation activity. When we re-estimated the models using industry dummies in addition to the variables introduced above, the resulting changes were negligible. The nonparametric density estimates of the unobserved heterogeneity were almost identical to the ones previously discussed. This further highlights the benefit of the FLEP model in tracking unobserved heterogeneity at the individual level.

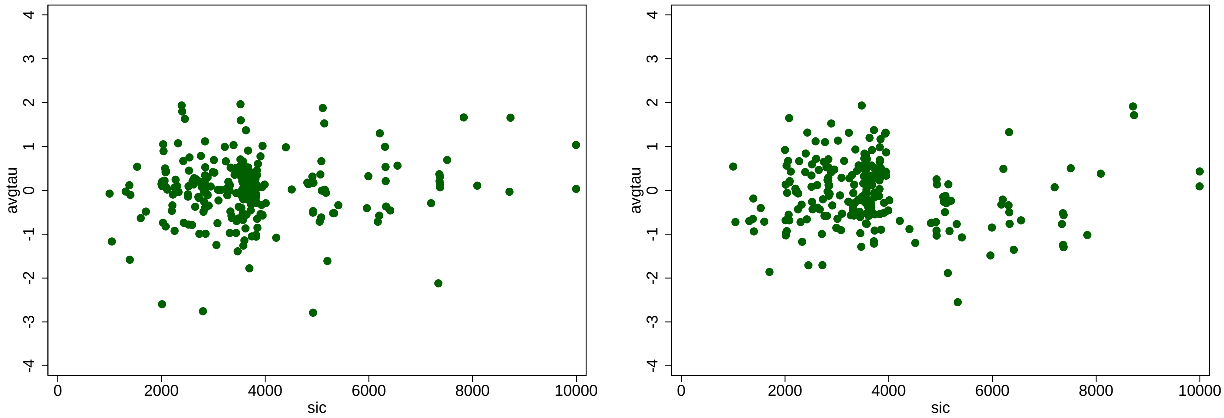


FIGURE 3. Average τ_i of companies for each SIC. Jaffe distance model (left) and Mahalanobis distance model (right).

4.4.3. ***Cluster-based Partial Effects***

Given that we have established the presence of three major clusters in the distribution of firm heterogeneity, we can revisit our original model and re-estimate it separately for each cluster. This allows us to investigate the extent to which the strength of the spillover effects varies across groups of firms. As BSV emphasize an important robustness check for the economic model is to verify the extent to which the results hold across groups of firms. If they do not, this may indicate that the estimated spillover effects are spuriously generated by pooling across different types of firms.

The summary statistics for the firms in each cluster are given in Table 5. It is interesting to notice that the extent of patenting activity varies substantially across clusters. The first cluster corresponding to negative values for the firm level heterogeneity has a low degree of patenting activity, while the third cluster corresponding to positive values of the firm level heterogeneity has a high degree of patenting activity. The summary statistics for the observable variables are however fairly similar across clusters which indicates that the unobserved heterogeneity has an important role to play.

The results of the patent equation estimation with FLEP are given in Table 6 for each cluster, respectively. The presence of technology spillover effects is confirmed for each cluster individually. The effect of product

	Variable	Jaffe Distance Mean	S.D.	Min	Max	Mahalanobis Distance Mean	S.D.	Min	Max
Cluster 1	*SpillTech*	9.516	1.029	4.838	11.707	11.252	0.801	8.519	13.156
	SpillSIC	7.157	2.420	-4.602	11.154	8.417	1.685	1.220	11.559
	R&D Stock	4.112	1.996	-0.888	10.231	2.378	2.909	-1.482	10.231
	Sales	6.143	1.922	1.098	11.760	6.144	1.946	1.098	11.760
	Patenting	0.262	0.440	0	1	0.291	0.454	0	1
	Firms	219				260			
	Total Obs	3916				4593			
Cluster 2	*SpillTech*	9.621	1.181	4.935	11.531	11.377	0.868	8.252	13.110
	SpillSIC	7.388	2.283	-4.321	11.053	8.585	1.656	-0.356	11.355
	R&D Stock	4.694	2.219	-2.513	10.765	3.511	3.082	-2.513	10.765
	Sales	6.286	2.023	0.693	12.103	6.320	2.000	0	12.103
	Patenting	0.622	0.484	0	1	0.649	0.477	0	1
	Firms	415				402			
	Total Obs	7361				7155			
Cluster 3	*SpillTech*	9.346	1.195	5.017	11.411	11.172	0.860	8.235	12.840
	SpillSIC	7.032	2.233	-2.700	10.981	8.4523	1.565	2.631	11.268
	R&D Stock	4.318	1.698	0.085	8.306	4.0012	1.530	0.085	8.161
	Sales	6.187	1.764	0	10.430	6.0130	1.752	1.386	10.430
	Patenting	0.832	0.373	0	1	0.8457	0.361	0	1
	Firms	95				67			
	Total Obs	1651				1180			

Table 5. Summary statistics for individual clusters. All variables are in logarithms and lagged by one period.

market rivalry continues to be statistically negligible in for each cluster. These results show that the economic model is thus robust to unobserved heterogeneity.

We can also perform one additional robustness check. If the FLEP model has correctly identified each cluster we should be able to estimate the model reasonably well using RE by sub-setting the data for each cluster. If heterogeneity is driving the results of the model once we condition on a cluster RE should perform similar to FLEP.[7] The robustness check results are reported in Table 7, using subsets of data corresponding to each cluster. RE and FLEP perform similarly in terms of predictions. It is important to remember however that this exercise can only be performed in post-estimation, conditional on the given cluster. A-priori we can

[7]Note that FLEP also controls for the presence of time factors which are ignored by the RE model. If however firm specific heterogeneity dominates in the data, then we would expect both models to have a similar performance.

never be sure about the structure of the distribution of the unobserved heterogeneity, which emphasizes the importance of using a flexible model when addressing unobserved heterogeneity.

	Jaffe Distance			Mahalanobis Distance		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
Ln(SpillTech)	0.681*	0.529*	0.720*	0.852*	0.610*	0.897*
	(0.077)	(0.043)	(0.097)	(0.069)	(0.040)	(0.124)
Ln(SpillSIC)	0.026	0.035	0.016	-0.031	-0.008	-0.066
	(0.026)	(0.019)	(0.054)	(0.027)	(0.017)	(0.069)
Ln(R&D Stock)	0.219*	0.345*	0.290*	0.198*	0.332*	0.385*
	(0.035)	(0.022)	(0.088)	(0.033)	(0.024)	(0.117)
Ln(Sales)	0.157*	0.182*	0.347*	0.169*	0.177*	0.205*
	(0.028)	(0.019)	(0.068)	(0.024)	(0.017)	(0.086)
APE scale	0.192	0.170	0.067	0.191	0.167	0.054
ρ	0.671*	0.620*	0.274*	0.468*	0.689*	0.514*
	(0.147)	(0.150)	(0.214)	(0.197)	(0.130)	(0.196)

Table 6. Estimation of Patent Equation by Cluster with FLEP.

Notes: Standard errors are reported in brackets. Coefficients significant at 5% confidence level are marked with an asterisk. All independent variables are lagged by one period. All regressions include a constant and a dummy for observations where lagged R&D stock is zero.

			RE		FLEP	
		Predicted	0	1	0	1
		Actual				
Jaffe Distance	Cluster 1	0	2668	219	2674	213
		1	412	617	426	603
	Cluster 2	0	2212	567	2195	584
		1	480	4102	497	4085
	Cluster 3	0	213	64	214	63
		1	26	1348	26	1348
Mahalanobis Distance	Cluster 1	0	2987	269	2987	270
		1	481	856	471	866
	Cluster 2	0	1959	546	1922	583
		1	463	4187	443	4206
	Cluster 3	0	145	37	147	35
		1	20	978	19	979

Table 7. Actual vs predicted outcomes for RE and FLEP in the Jaffe distance model and Mahalanobis distance model for each cluster.

5. Conclusion

This paper introduced a new Bayesian semi-parametric approach to the estimation of the probit model in panel data with unobserved heterogeneity. The proposed model substantially improved on current benchmark methods by relaxing three assumptions that are often either ignored or treated in an ad-hoc fashion in empirical work. First, we modeled unobserved individual effects using a flexible nonparametric form with desirable local adaptability properties. Second, we allowed for the unobserved heterogeneity to be correlated with the observables. Finally, our model incorporated common latent time effects.

We employed a combination of recent powerful sampling algorithms in order to draw from a Dirichlet Process Mixture model specified for the unobserved heterogeneity component. We evaluated the proposed model in a number of Monte Carlo simulations along with existing fixed and random effects model alternatives. The underlying parameters are shown to be estimated with high precision in the proposed model, unlike for the benchmark cases. The simulations highlight the benefit of using the flexible proposed model when the underlying heterogeneity is not well approximated by a parametric distributional form.

We applied the proposed method to the estimation of a patent equation in the presence of both technological and product market spillover effects. We showed that technological innovation is subject to substantial firm-level heterogeneity which persists within individual industries. We have showed that innovation depends in an important way on technology spillovers but that there is little evidence in favor of product market spillover effects. On the basis of the estimated firm-level heterogeneity we also showed that unobserved heterogeneity is heavily clustered and that the clustering matters when making in-sample predictions of patenting activity.

References

Abrevaya, J. (1999): "Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable," *Journal of Econometrics*, 93(2), 203 – 228.

Albert, J., and S. Chib (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88(422), 669–679.

Albert, J., and S. Chib (1996): "Bayesian modeling of binary repeated measures data with application to crossover trials," in *Bayesian Biostatistics*, ed. by D. A. Berry, and D. K. Stangl. Marcel Dekker, New York.

Allenby, G., and P. Lenk (1994): "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, 89, 1218–1231.

Arellano, M., and S. Bonhomme (2009): "Robust Priors in Nonlinear Panel Data Models," *Econometrica*, 77(2), 489–536.

Arellano, M., and J. Hahn (2006): "A Likelihood-Based Approximate Solution To The Incidental Parameter Problem In Dynamic Nonlinear Models With Multiple Effects," Working paper, CEMFI.

Berry, S. T., and P. A. Haile (2009): "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Cowles Foundation Discussion Papers 1718, Cowles Foundation, Yale University.

Bloom, N., M. Schankerman, and J. Van Reenen (2010): "Identifying Technology Spillovers and Product Market Rivalry," NBER Working Paper 13060.

Bolduc, D., B. Fortin, and S. Gordon (1997): "Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques," *International Regional Science Review*, 20(1-2), 77101.

Burda, M., M. C. Harding, and J. A. Hausman (2008): "A Bayesian Mixed Logit-Probit Model for Multinomial Choice," *Journal of Econometrics*, 147(2), 232246.

Burda, M., R. Liesenfeld, and J.-F. Richard (2011): "Bayesian Analysis of a Probit Panel Data Model with Unobserved Individual Heterogeneity and Autocorrelated Errors," *International Journal of Statistics and Management Systems*, 6(1-2), 1–21.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46.

Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches, and M. Intriligator. Amsterdam: North-Holland.

Chib, S. (1993): "Bayes Estimation of Regressions with Autoregressive Errors: A Gibbs Sampling Approach," *Journal of Econometrics*, 58, 275–294.

Chib, S., and B. Carlin (1999): "On MCMC Sampling in Hierarchical Longitudinal Models," *Statistics and Computing*, 9, 17–26.

Chib, S., and B. Hamilton (2002): "Semiparametric bayes analysis of longitudinal data treatment models," *Journal of Econometrics*, 110, 67–89.

Chib, S., and I. Jeliazkov (2006): "Inference in Semiparametric Dynamic Models for Binary Longitudinal Data," *Journal of the American Statistical Association*, 101(474), 685–700.

Dahl, D. B. (2005): "Sequentially-Allocated Merge-Split Sampler for Conjugate and Nonconjugate Dirichlet Process Mixture Models," Technical report, Texas A&M University.

Escobar, M. D., and M. West (1995): "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Fernandez-Val, I. (2007): "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models," working paper, Boston University.

Geweke, J. (1991): "Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints,," in *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, ed. by E. M. Keramidas. Fairfax: Interface Foundation of North America, Inc.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. Wiley.

Griffith, R., L. S., and J. Van Reenen (2011): "Is distance dying at last? Falling home bias in fixed effects models of patent citations," *Quantitative Economics*, 2(2), 211–250.

Gu, Y., D. G. Fiebig, E. Cripps, and R. Kohn (2009): "Bayesian estimation of a random effects heteroscedastic probit model," *Econometrics Journal*, 12(2), 324–339.

Hajivassiliou, V. (1990): "Smooth Estimation Simulation of Panel Data LDV Models," Working paper, Yale University.

Hausman, J. A., and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49(6), 1377–98.

Hirano, K. (2002): "Semiparametric bayesian inference in autoregressive panel data models," *Econometrica*, 70, 781–799.

Jaffe, A. B. (1986): "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, 76(5), 984–1001.

Keane, M. (1990): "A Computationally Efficient Practical Simulation Estimator for Panel Data, with Applications to Estimating Temporal Dependence in Employment and Wages," working paper, University of Minnesota.

Knittel, C. R., and K. Metaxoglou (2008): "Estimation of random coefficient demand models: challenges, difficulties and warnings," working paper, NBER.

Kohavi, R., and F. Provost (1998): "Glossary of terms. Editorial for the special issue on application of machine learning and the knowledge of discovery process," *Machine Learning*, 30, 271274.

Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*. Blackwell, Malden, MA.

Li, T., and X. Zheng (2008): "Semiparametric Bayesian Inference for Dynamic Tobit Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, 23, 699–728.

Neal, R. (2000): "Markov Chain Sampling Methods for Dirichlet Process Mixture Models," *Journal of Computational and Graphical Statistics*, 9(2), 249–265.

Paap, R. (2002): "What are the Advantages of MCMC based Inference in Latent Variable Models?," *Statistica Neerlandica*, 56(1), 2–22.

Rossi, P. (2010): *bayesm: Bayesian Inference for Marketing/Micro-econometrics,* R package version 2.2-3.

Train, K. (2001): "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit," working paper, Department of Economics, University of California, Berkeley.

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

Troughton, P., and S. Godsill (1997): "Bayesian model selection for time series using Markov chain Monte Carlo," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 5, pp. 3733–3736.

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

6. Appendix

6.1. Sampling β

In this block we apply the method of Albert and Chib (1993) to the recentered latent variable $\widetilde{y}_{it}^{*} = \widetilde{y}_{it} - \tau_i - \lambda_t$. The joint conditional density of $(\beta, \widetilde{\mathbf{y}}^*)$ is given by

$$p(\beta, \widetilde{\mathbf{y}}^* | \tau, \lambda, \psi, \theta_{/\beta}, \widetilde{\mathbf{y}}, \mathbf{y}, \mathbf{X}) \propto \exp\left[-\frac{1}{2}(\beta - \underline{\beta}')\underline{\Sigma}_{\beta}^{-1}(\beta - \underline{\beta})\right] \exp\left[-\frac{1}{2}(\widetilde{\mathbf{y}}^* - \mathbf{X}\beta)'(\widetilde{\mathbf{y}}^* - \mathbf{X}\beta)\right]$$

yielding a closed form of the conditional posterior for β which facilitates direct sampling from $\beta|\cdot \sim N\left(\overline{\beta}, \overline{\Sigma}\right)$ where

$$\begin{aligned} \overline{\beta} &= \overline{\Sigma}\left(\underline{\Sigma}^{-1}\underline{\beta} + \mathbf{X}'\widetilde{\mathbf{y}}^*\right) \\ \overline{\Sigma} &= \left(\underline{\Sigma}^{-1} + (\mathbf{X}'\mathbf{X})\right)^{-1} \end{aligned}$$

In the application, we specify the hyperparameter values $\underline{\beta} = 0$ and $\underline{\Sigma}_{\beta} = 10I$ where I is the identity matrix. This specification is aimed at rendering the prior for β sufficiently diffuse.

6.2. Sampling $\widetilde{y}_{it}$

Here we benefit from the second step of the Albert and Chib (1993) procedure, augmented by τ_i and λ_t. Thus, sample directly

$$\begin{aligned} \widetilde{y}_{it}|\cdot &\sim N(v_{it}, 1) \\ v_{it} &= \mathbf{x}_{it}\beta + \tau_i + \lambda_t \end{aligned}$$

truncated by 0 from the left if $y_{it} = 1$ and from the right if $y_{it} = 0$.

6.3. Updating Latent Class Assignments

For this block we utilize a hybrid sampler that alternates between the non-conjugate version of the Sequentially Allocated Split-Merge (SAMS) sampler of Dahl (2005), and Algorithm 7 of Neal (2000). This approach is suggested by Dahl (2005) as optimally combining the virtues of each method: the ability to move large blocks of elements among latent classes in one step for the former, and one-at-a-time allocations of individual elements among latent classes for the latter.

The SAMS sampler is based on an alternative expression of the model (2.6)-(2.8) in terms of a set partition $\pi = \{S_1, \ldots, S_q\}$ for $S_0 = \{1, \ldots, n\}$ in addition to the latent class parameters $\phi = \{\phi_{S1}, \ldots, \phi_{Sq}\}$ where

ϕ_S is associated with component S. The set partition π for S_0 is a set of subsets $S_1, \ldots, S_q$ such that (1) $\cup_{S\in\pi} S = S_0$, (2) $S^i \cap S^j = \emptyset$ for all $S^i \neq S^j$, and (3) $S \neq \emptyset$ for all $S \in \pi$. Using this notation, the model (2.6)-(2.8) can be recast as (Dahl, 2005):

$$\tau_i|\pi, \phi \sim F_\tau(\phi_S^i) \tag{6.1}$$

$$\psi|\pi \sim \prod_{S\in\pi} G_0(\phi_S) \tag{6.2}$$

$$\pi \sim b \prod_{S\in\pi} \eta_0 \Gamma(|S|) \tag{6.3}$$

where $|S|$ is the number of elements of the component S. The sampling scheme works as follows: In each MC iteration, uniformly select a pair of distinct indices i and j. If i and j belong to the same component in π, say S, propose π^* by splitting S. Otherwise, i and j belong to different components in π, say S^i and S^j. Propose π^* by merging S^i and S^j. In each case, compute the Metropolis-Hastings (MH) ratio $a(\pi^*, \phi^*|\pi, \phi)$ and accept the new latent class configuration π^* with probability given by this ratio. We derive the MH ratio for our model in the following Section.

Algorithm 7 of Neal (2000), which we utilize in every alternate MC step, is based on limiting probabilities of a latent class finite mixture model with the number of classes tending to infinity. The sampling procedure itself is built around drawing with a stochastic number of mixture components or classes whose number and size varies at each MC iteration. Denote by c a label of a generic latent class with membership count N_c. Given the current state of the system, τ_i are first re-assigned into latent classes with labels c_i whereby new classes can be created and old ones may vanish. The probabilities of class assignment for the τ_i are proportional to the likelihood of τ_i conditional on the current draw of the class parameters ψ_c. Second, the class parameters ψ_c are updated in a standard way for each class separately. If we specify F_τ as an infinite mixture of Normals, then $\psi_c = (\mu_{\tau c}, \sigma_{\tau c}^2)$ are the moments of the Normal density.

For updating ψ in the Algorithm 7 scan, we specify F_τ as a mixture of Normals with $\psi = (\mu_\tau, \sigma_\tau^2)$. Since for all τ_i that fall into one latent class it holds that $\tau_i \sim N(\mu_{\tau c}, \sigma_{\tau c}^2)$ we can apply result B (p. 300) of Train (2003) to each latent class separately: for an $IG(s_0, v_0)$ prior, the posterior of $\sigma_{\tau c}^2$ is given by $IG(s_1, v_1)$ with $v_1 = v_0 + N_c$ and $s_1 = (v_0 s_0 + N_c \overline{s}_{c_i})/(v_0 + N_c)$ where $\overline{s}_c = N_c^{-1} \sum_{i=1}^{N_c} \tau_i^2$. We utilize a diffuse IG prior. Analogously, to sample $\mu_{\tau c}$ we use result A of Train (2003) applied to each latent class. The hyperparameter of the DP prior α is sampled according to the scheme of Escobar and West (1995).

The iteration between the samplers of Dahl (2005) and Neal (2000) alleviates the influence of particular starting values. The SAMS sampler is capable of re-allocating large blocks of data to one of the latent classes while the Neal algorithm addresses the individual by individual allocation to latent classes. This allows us to initialize the procedure with a unique parametric component. This is then rapidly split into classes by the SAMS sampler before the Neal procedure continues to fine-tune the posterior draws.

6.4. Sampling τ_i

Let $\widetilde{\mathbf{y}}_i^{**} = \widetilde{\mathbf{y}}_i - \mathbf{X}_i\beta - \lambda$. Then

$$\widetilde{\mathbf{y}}_i^{**} = \tau_i \iota + \epsilon_i$$

Consider for the moment the case $\tau_i \sim N(\underline{\tau}, \sigma_\tau^2)$; it will be used as a building block in the DP prior sampling. In this case, for every i we have one latent regression with one parameter τ_i and a $(T \times 1)$ vector of ones as explanatory variables in place of a hypothetical $\mathbf{X}_i$. Using standard latent regression results (see e.g. Lancaster, 2004),

$$\begin{aligned} p(\tau_i|\cdot) &= \phi\left(\overline{\tau}_i, \overline{\sigma}_{\tau i}^2\right) \\ \overline{\tau}_i &= \overline{\sigma}_{\tau i}^2 \left(\sigma_\tau^{-2}\underline{\tau} + \sum_{i=1}^{T} \widetilde{\mathbf{y}}_i^{**}\right) \\ \overline{\sigma}_{\tau i}^2 &= \left(\sigma_\tau^{-2} + T\right)^{-1} \end{aligned} \tag{6.4}$$

Since $\tau_i \sim N(\mu_{\tau c}, \sigma_{\tau c}^2)$ given a previous assignment to the latent class c, let $\underline{\tau} = \mu_{\tau c}$, $\sigma_\tau^2 = \sigma_{\tau c}^2$ and sample τ_i directly from (6.4).

6.5. Sampling λ

Let

$$\widetilde{\mathbf{y}}_{i\lambda} = \widetilde{\mathbf{y}}_i - \mathbf{X}_i\beta - \tau_i\iota$$

Then the joint density implied for $\widetilde{\mathbf{y}}_\lambda = (\widetilde{\mathbf{y}}_{1\lambda}, \ldots, \widetilde{\mathbf{y}}_{N\lambda})$ by the recentered probit model conditional on λ is

$$f(\widetilde{y}_\lambda|\lambda, \cdot) = (2\pi)^{-NT/2} \det(I_T)^{-N/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\left[\widetilde{\mathbf{y}}_{i\lambda}{}' I_T^{-1} \widetilde{\mathbf{y}}_{i\lambda} - 2\lambda' I_T^{-1} \widetilde{\mathbf{y}}_{\lambda i} + \lambda' I_T^{-1}\lambda\right]\right\}$$

while the prior density specified by Assumption 3 takes the form

$$f(\lambda) = (2\pi)^{-T/2} \det(I_T)^{-1/2} \exp\left\{-\frac{1}{2}\lambda'\Omega_\lambda^{-1}\lambda - 2\lambda'\Omega_\lambda^{-1}\Lambda\rho + \rho'\Lambda'\Omega_\lambda^{-1}\Lambda\rho\right\}$$

where $\Lambda_t = (\lambda_{t-1}, \ldots, \lambda_{t-s})$, $\Lambda = (\Lambda_1', \ldots, \Lambda_T')'$, $\rho = (\ \rho_1, \ldots, \rho_s)$ and Ω_λ is the covariance matrix associated with the autoregressive process. Hence we can sample λ directly from $N(\overline{\lambda}, \overline{\Sigma}_\lambda)$ where

$$\begin{aligned}
\overline{\lambda} &= \overline{\Sigma}_\lambda \left(\Omega_\lambda^{-1} \underline{\lambda} + \sum_{i=1}^{N} \widetilde{\mathbf{y}}_{\lambda i} \right) \\
\overline{\Sigma}_\lambda &= \left(\Omega_\lambda^{-1} + N \times I_T \right)^{-1}
\end{aligned}$$

For ease of implementation we restrict ourselves to the $AR(1)$ specification with a single autoregressive parameter ρ in the application. In this case,

$$\begin{aligned}
\Omega_\lambda &= \gamma_0 \begin{bmatrix} \rho^0 & \rho^1 & \rho^2 & \cdots & \rho^{T-1} \\ \rho^1 & \rho^0 & \rho^1 & \cdots & \rho^{T-2} \\ \rho^2 & \rho^1 & \rho^0 & & \rho^{T-3} \\ \vdots & \vdots & & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & \rho^0 \end{bmatrix} \\
\gamma_0 &= \frac{\sigma_\eta^2}{1-\rho^2}
\end{aligned}$$

6.6. **Sampling** ρ

Note that for the $AR(1)$ process,

$$p(\lambda_t | \lambda_{t-1}, \cdot) \propto \begin{cases} \exp\left(-\frac{(1-\rho^2)}{2\sigma_\eta^2} \lambda_1^2\right), & t = 1 \\ \exp\left(-\frac{1}{2\sigma_\eta^2} (\lambda_t - \rho\lambda_{t-1})^2\right), & t = 2, ..., T \end{cases}$$

and hence

$$\begin{aligned}
p(\rho|\lambda) &= \exp\left(-\frac{1}{2\sigma_\eta^2}\left[(1-\rho^2)\lambda_1^2 + \sum_{t=2}^{T}(\lambda_t - \rho\lambda_{t-1})^2\right]\right) \\
&= \exp\left(\frac{1}{2\sigma_\eta^2}\left[\rho^2\left(\sum_{t=2}^{T}\lambda_{t-1}^2 - \lambda_1^2\right) - 2\rho\sum_{t=2}^{T}\lambda_t\lambda_{t-1} + \sum_{t=1}^{T}\lambda_t^2)\right]\right)
\end{aligned}$$

Matching this expression with a Gaussian kernel $\exp\left(-\frac{1}{2\sigma^2}\left[\rho^2 - 2\rho\mu + \mu^2\right]\right)$ yields

$$\begin{aligned}
\overline{\sigma}_\rho^2 &= \sigma_\eta^2 \left(\sum_{t=2}^{T-1} \lambda_t^2\right)^{-1} \\
\overline{\mu}_\rho &= \frac{\overline{\sigma}_\rho^2}{\sigma_\eta^2} \sum_{t=2}^{T} \lambda_t \lambda_{t-1} \\
&= \left(\sum_{t=2}^{T-1} \lambda_t^2\right)^{-1} \sum_{t=2}^{T} \lambda_t \lambda_{t-1}
\end{aligned}$$

We can therefore sample ρ directly from $N(\overline{\mu}_\rho, \overline{\sigma}_\rho^2)$ truncated at -1 and 1 to preserve stationarity. Extension to $AR(p)$ will amend the likelihood function $p(\lambda_t|\lambda_{t-1}, \cdot)$ but the derivation would be similar. The approach for sampling the $AR(p)$ parameters conditional on the initial observations is presented in Chib (1993).

6.7. Sampling σ_η^2

For this block we use the result derived in Burda, Liesenfeld, and Richard (2011) which adapts the standard result on sampling univariate variances (given e.g. by result B, p. 300, of Train, 2003) to the likelihood of the variance of the AR process. Conditional on λ and ρ, the likelihood function of σ_η^2 takes the form

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

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. We can then sample directly from the posterior

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

where

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

In the application, the prior for σ_η^2 will be specified as diffuse with $s_0 \to 0$ and $v_0 = 0$.

6.8. The SAMS sampler

In this Section, we explicitly derive the form of the MH ratio for our case. For a general description, see Dahl (2005). Let k be the successive values in random permutations of the indices in S. In our model, the MH ratio is given by

$$a(\pi^*, \phi^*|\pi, \phi) = \min\left[1, \frac{p(\pi^*, \phi^*|y)}{p(\pi, \phi|y)} \frac{q(\pi, \phi|\pi^*, \phi^*)}{q(\pi^*, \phi^*|\pi, \phi)}\right]$$

If the proposal involves a split, $q(\pi^*, \phi^*|\pi, \phi)$ is the split probability and $q(\pi, \phi|\pi^*, \phi^*) = 1$ is the merge probability. If the proposal involves a merge, the roles of $q(\pi^*, \phi^*|\pi, \phi)$ and $q(\pi, \phi|\pi^*, \phi^*)$ are reversed.

Consider e.g. proposal for a split:

$$q(\pi^*, \phi^*|\pi, \phi) = \prod_{k=1}^{N} P\left(k \in S^i | S^i, S^j, \phi, y\right) P(\phi_{S^i})$$

The first term is given in equation (13) in Dahl (2005). The second term $P(\phi_{S^i})$ is the proposal density of the new ϕ_{S^i}. The merge probability is

$$q(\pi, \phi|\pi^*, \phi^*) = 1$$

By Bayes theorem,

$$p(\pi, \phi|y) \propto p(y|\pi, \phi)p(\pi, \phi) \tag{6.5}$$

where $p(y|\pi, \phi)$ is the likelihood

$$p(y|\pi, \phi) = \prod_{i=1}^{n} p(y_i|\phi_{S^i})$$

and $p(\pi, \phi)$ is the prior

$$p(\pi, \phi) = p(\phi|\pi)p(\pi) \tag{6.6}$$

where

$$\begin{aligned} p(\phi|\pi) &= {}_{S\in\pi}F_0(\phi_S) \\ p(\pi) &= b_{S\in\pi}\eta_0\Gamma(|S|) \\ b^{-1} &= \prod_{i=1}^{n}\Gamma\left(\eta_0 + i - 1\right) \end{aligned}$$

Note that for a split of a class S^s into S^i and S^j,

$$\frac{p(y|\pi^*, \phi^*)}{p(y|\pi, \phi)} = \frac{\prod_{t=1}^{|S^i|} p(y_t|\phi_{S^i}) \prod_{t=1}^{|S^j|} p(y_t|\phi_{S^j})}{\prod_{t=1}^{|S^s|} p(y_t|\phi_{S^s})} \tag{6.7}$$

where the index t in $p(y_t|\phi_{S^i})$ refers to elements of the class S^i. Similarly, for a merge of classes S^i and S^j into S^s,

$$\frac{p(y|\pi^*, \phi^*)}{p(y|\pi, \phi)} = \frac{\prod_{t=1}^{|S^s|} p(y_t|\phi_{S^s})}{\prod_{t=1}^{|S^i|} p(y_t|\phi_{S^i}) \prod_{t=1}^{|S^j|} p(y_t|\phi_{S^j})}$$

i.e. the inverse of the ratio of split probabilities. Note that for a split we can use the stored values of the likelihood evaluations from the allocation of k into S^i and S^j. Hence only two additional likelihood evaluations $p(y_i|\phi_{S^i})$ and $p(y_j|\phi_{S^j})$ that initiated the split need to be performed for obtaining the ratio $\frac{p(y|\pi,\phi)}{p(y|\pi^*,\phi^*)}$. For a merge, only $2|S^i| + |S^j|$ likelihood evaluations need to be performed, which for small classes can be substantially less than the sample size n. In the same spirit, for computing prior components for a split

$$\frac{p(\phi^*|\pi^*)}{p(\phi|\pi)} = \frac{F_0(\phi_{S^i})F_0(\phi_{S^j})}{F_0(\phi_{S^s})}$$

and

$$\frac{p(\pi^*)}{p(\pi)} = \frac{\Gamma(|S^i|)\Gamma(|S^j|)}{\Gamma(|S^s|)}$$

while for a merge

$$\frac{p(\phi^*|\pi^*)}{p(\phi|\pi)} = \frac{F_0(\phi_{S^s})}{F_0(\phi_{S^i})F_0(\phi_{S^j})}$$

and

$$\frac{p(\pi^*)}{p(\pi)} = \frac{\Gamma(|S^s|)}{\Gamma(|S^i|)\Gamma(|S^j|)}$$

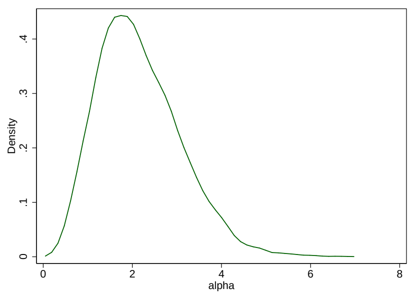
Thus, using (6.5), the ratio of the p-terms for a split becomes

$$\frac{p(\pi^*, \phi^*|y)}{p(\pi, \phi|y)} = \frac{\prod_{t=1}^{|S^i|} p(y_t|\phi_{S^i}) \prod_{t=1}^{|S^j|} p(y_t|\phi_{S^j})}{\prod_{t=1}^{|S^s|} p(y_t|\phi_{S^s})} \frac{F_0(\phi_{S^i})F_0(\phi_{S^j})}{F_0(\phi_{S^s})} \frac{\Gamma(|S^i|)\Gamma(|S^j|)}{\Gamma(|S^s|)} \tag{6.8}$$

while for a merge this ratio is given by the inverse of the expression in (6.8).

6.9. DPM Prior Hyperparameter

In order to explore the distribution of unobserved heterogeneity, τ_i, in our patent models we need to make sure that its behavior is not implicitly restricted by the estimation procedure or some other deep model parameters. One parameter that is of concern to us is the smoothing parameter α that controls the extent to which the Dirichlet Process draws mixture distributions that are more or less "similar" to the Normal baseline parametric distribution G_0. In the limiting case of $\alpha \to \infty$ the mixture distribution becomes equivalent to G_0, while in the other extreme $\alpha \to 0$ the mixture distribution limits to a convolution of density kernels centered at each data point without any influence of the DP prior. The posterior distribution estimate for both models is plotted in Figure 4. The distributions are concentrated around a mode of 2 indicating a strong influence of data relative to the baseline prior distribution.



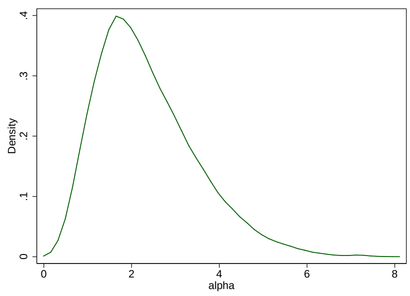


FIGURE 4. Density of draws of α, Jaffe distance model (left) and Mahalanobis distance model (right).

	N	T	APE	β_0	β_1	β_2	β_3	σ_η	ρ	APE	β_0	β_1	β_2	β_3	σ_η	ρ	β_0	β_1	β_2	β_3	σ_η	ρ
			Mean Absolute Deviation							RMSE							S.E.					
Fixed Effects Probit	100	10	0.047	-1.109	0.408	0.337	-0.427	-	-	0.052	2.205	0.649	0.603	0.559	-	-	0.726	0.333	0.334	0.333	-	-
	300	10	0.052	-0.717	0.352	0.307	-0.301	-	-	0.055	2.219	0.391	0.374	0.359	-	-	0.645	0.181	0.180	0.181	-	-
	700	10	0.052	-0.602	0.316	0.309	-0.299	-	-	0.055	2.224	0.344	0.336	0.325	-	-	0.632	0.117	0.117	0.116	-	-
	1000	10	0.054	-0.604	0.322	0.311	-0.345	-	-	0.056	2.001	0.335	0.326	0.357	-	-	0.625	0.098	0.098	0.098	-	-
	100	20	0.019	-1.450	0.248	0.189	-0.228	-	-	0.021	2.256	0.320	0.263	0.331	-	-	0.585	0.198	0.197	0.197	-	-
	300	20	0.023	-1.241	0.170	0.154	-0.145	-	-	0.025	2.162	0.205	0.194	0.191	-	-	0.508	0.109	0.109	0.109	-	-
	700	20	0.023	-1.208	0.150	0.164	-0.155	-	-	0.024	2.003	0.172	0.185	0.171	-	-	0.503	0.072	0.072	0.071	-	-
	1000	20	0.025	-1.117	0.154	0.157	-0.153	-	-	0.026	2.400	0.168	0.166	0.162	-	-	0.502	0.059	0.059	0.059	-	-
	100	50	0.001	-1.954	0.074	0.106	-0.116	-	-	0.004	2.384	0.138	0.150	0.156	-	-	0.450	0.113	0.113	0.113	-	-
	300	50	0.004	-1.429	0.076	0.065	-0.073	-	-	0.005	2.227	0.104	0.083	0.095	-	-	0.364	0.063	0.063	0.063	-	-
	700	50	0.006	-2.217	0.069	0.053	-0.065	-	-	0.006	2.571	0.082	0.067	0.076	-	-	0.372	0.041	0.041	0.041	-	-
	1000	50	0.005	-1.757	0.072	0.062	-0.061	-	-	0.005	2.471	0.083	0.073	0.068	-	-	0.357	0.034	0.034	0.034	-	-
Random Effects Probit	100	10	0.055	-0.234	0.110	0.049	-0.134	-	-	0.059	1.150	0.351	0.343	0.281	-	-	0.470	0.295	0.289	0.289	-	-
	300	10	0.110	0.043	0.088	0.054	-0.048	-	-	0.111	1.065	0.155	0.170	0.161	-	-	0.281	0.159	0.157	0.156	-	-
	700	10	0.145	-0.068	0.066	0.059	-0.052	-	-	0.146	1.107	0.122	0.117	0.109	-	-	0.203	0.101	0.102	0.101	-	-
	1000	10	0.157	-0.016	0.071	0.066	-0.088	-	-	0.158	0.908	0.099	0.101	0.114	-	-	0.159	0.084	0.084	0.083	-	-
	100	20	0.044	0.139	0.104	0.051	-0.087	-	-	0.046	1.253	0.207	0.162	0.219	-	-	0.382	0.180	0.180	0.181	-	-
	300	20	0.211	0.191	0.022	0.039	0.002	-	-	0.211	0.768	0.082	0.097	0.102	-	-	0.214	0.100	0.100	0.100	-	-
	700	20	0.212	0.200	0.041	0.039	-0.003	-	-	0.213	0.560	0.091	0.100	0.080	-	-	0.147	0.067	0.066	0.066	-	-
	1000	20	0.218	-0.059	0.012	0.014	-0.029	-	-	0.218	0.842	0.075	0.047	0.044	-	-	0.124	0.056	0.056	0.056	-	-
FLEP	100	10	0.001	0.004	0.025	-0.046	-0.060	0.114	-0.309	0.013	0.163	0.370	0.386	0.334	-0.010	0.581	0.163	0.369	0.384	0.328	0.228	0.492
	300	10	0.001	0.014	0.024	-0.015	0.031	0.067	-0.231	0.008	0.079	0.191	0.193	0.196	-0.006	0.540	0.078	0.189	0.193	0.194	0.202	0.489
	700	10	0.001	0.010	-0.001	-0.016	0.023	0.074	-0.223	0.005	0.055	0.121	0.127	0.126	-0.005	0.526	0.054	0.121	0.126	0.123	0.200	0.477
	1000	10	0.000	-0.007	0.007	0.013	-0.017	0.102	-0.263	0.004	0.048	0.101	0.106	0.104	-0.005	0.554	0.047	0.101	0.105	0.103	0.186	0.488
	100	20	-0.003	0.016	0.084	0.025	-0.070	0.162	-0.185	0.010	0.127	0.257	0.229	0.271	0.046	0.382	0.125	0.243	0.228	0.262	0.161	0.335
	300	20	-0.000	-0.003	0.021	0.002	0.003	0.104	-0.089	0.005	0.059	0.137	0.143	0.139	-0.003	0.328	0.059	0.135	0.143	0.139	0.131	0.316
	700	20	-0.001	0.002	-0.007	0.006	0.001	0.065	-0.096	0.004	0.041	0.092	0.097	0.085	-0.002	0.343	0.041	0.092	0.097	0.085	0.123	0.329
	1000	20	0.001	-0.001	-0.010	-0.003	0.007	0.034	-0.062	0.006	0.031	0.090	0.077	0.083	-0.002	0.332	0.031	0.090	0.077	0.082	0.119	0.326
	100	50	0.001	0.012	-0.054	0.108	-0.150	0.110	-0.018	0.006	0.044	0.123	0.159	0.212	0.004	0.185	0.042	0.110	0.117	0.150	0.077	0.185
	300	50	-0.001	0.001	0.022	0.010	-0.016	0.062	-0.075	0.003	0.034	0.090	0.078	0.083	-0.001	0.214	0.034	0.087	0.077	0.082	0.087	0.200
	700	50	-0.000	0.000	0.011	-0.004	-0.008	0.030	-0.049	0.002	0.024	0.058	0.056	0.055	-0.001	0.198	0.024	0.057	0.056	0.054	0.079	0.192
	1000	50	-0.000	0.000	0.011	0.001	0.001	0.029	-0.030	0.003	0.020	0.054	0.053	0.049	-0.000	0.203	0.020	0.053	0.053	0.049	0.076	0.201

TABLE 8. Monte Carlo output, X and τ_i uncorrelated

	N	T	APE	β_0	β_1	β_2	β_3	σ_η	ρ	APE	β_0	β_1	β_2	β_3	σ_η	ρ	β_0	β_1	β_2	β_3	σ_η	ρ
			Mean Absolute Deviation							RMSE							S.E.					
Fixed Effects Probit	100	10	0.059	4.049	-0.278	-0.243	1.707	-	-	0.064	5.521	0.365	0.349	1.722	-	-	1.212	0.208	0.209	0.207	-	-
	300	10	0.076	3.618	-0.362	-0.375	1.651	-	-	0.079	4.750	0.381	0.393	1.655	-	-	0.789	0.108	0.109	0.109	-	-
	700	10	0.076	2.953	-0.363	-0.359	1.616	-	-	0.080	4.450	0.373	0.368	1.618	-	-	0.670	0.071	0.071	0.070	-	-
	1000	10	0.082	3.271	-0.379	-0.402	1.634	-	-	0.085	4.315	0.383	0.406	1.635	-	-	0.702	0.059	0.059	0.059	-	-
	100	20	0.058	4.142	-0.395	-0.391	1.612	-	-	0.060	5.165	0.408	0.414	1.617	-	-	1.355	0.122	0.122	0.122	-	-
	300	20	0.061	3.397	-0.430	-0.424	1.567	-	-	0.062	4.623	0.437	0.429	1.568	-	-	0.581	0.068	0.068	0.068	-	-
	700	20	0.067	3.842	-0.432	-0.440	1.562	-	-	0.067	4.191	0.434	0.443	1.563	-	-	0.491	0.044	0.044	0.044	-	-
	1000	20	0.066	3.300	-0.446	-0.443	1.559	-	-	0.067	3.577	0.448	0.445	1.560	-	-	0.474	0.037	0.037	0.037	-	-
	100	50	0.052	4.055	-0.434	-0.450	1.549	-	-	0.053	4.784	0.441	0.455	1.551	-	-	0.497	0.071	0.071	0.071	-	-
	300	50	0.056	3.682	-0.463	-0.473	1.529	-	-	0.057	4.024	0.464	0.475	1.530	-	-	0.353	0.040	0.040	0.040	-	-
	700	50	0.057	3.535	-0.474	-0.475	1.535	-	-	0.057	3.953	0.475	0.476	1.535	-	-	0.318	0.026	0.026	0.026	-	-
	1000	50	0.059	3.469	-0.475	-0.474	1.525	-	-	0.059	3.742	0.475	0.474	1.525	-	-	0.285	0.022	0.022	0.022	-	-
Random Effects Probit	100	10	0.039	-0.044	-0.439	-0.417	1.533	-	-	0.042	1.934	0.469	0.454	1.543	-	-	0.540	0.176	0.174	0.176	-	-
	300	10	0.095	-0.015	-0.492	-0.501	1.522	-	-	0.096	1.497	0.500	0.509	1.524	-	-	0.308	0.093	0.094	0.093	-	-
	700	10	0.118	-0.172	-0.492	-0.488	1.495	-	-	0.119	1.426	0.495	0.493	1.496	-	-	0.222	0.061	0.062	0.061	-	-
	1000	10	0.136	-0.259	-0.500	-0.517	1.510	-	-	0.136	1.457	0.501	0.519	1.512	-	-	0.194	0.051	0.052	0.052	-	-
	100	20	0.024	-0.123	-0.463	-0.459	1.544	-	-	0.025	1.723	0.472	0.474	1.548	-	-	0.511	0.114	0.112	0.113	-	-
	300	20	0.075	0.087	-0.461	-0.488	1.502	-	-	0.076	2.057	0.465	0.491	1.503	-	-	0.282	0.064	0.064	0.064	-	-
	700	20	0.122	-0.918	-0.512	-0.510	1.479	-	-	0.123	1.304	0.513	0.513	1.479	-	-	0.176	0.041	0.041	0.041	-	-
	1000	20	0.141	-0.063	-0.530	-0.500	1.489	-	-	0.141	1.001	0.530	0.501	1.489	-	-	0.173	0.034	0.034	0.034	-	-
FLEP	100	10	-0.001	-0.005	-0.034	0.049	0.013	0.063	-0.246	0.012	0.147	0.125	0.226	0.215	-0.011	0.543	0.147	0.121	0.220	0.215	0.324	0.484
	300	10	0.000	-0.001	-0.030	-0.011	0.018	-0.038	-0.230	0.008	0.084	0.123	0.124	0.126	-0.005	0.514	0.084	0.120	0.123	0.125	0.294	0.459
	700	10	-0.001	0.004	-0.003	0.015	-0.003	-0.008	-0.224	0.006	0.056	0.061	0.085	0.075	-0.005	0.528	0.056	0.061	0.083	0.075	0.276	0.478
	1000	10	-0.001	0.005	-0.004	-0.016	0.012	-0.049	-0.270	0.006	0.045	0.059	0.067	0.080	-0.004	0.544	0.044	0.059	0.065	0.079	0.278	0.473
	100	20	-0.001	0.004	-0.015	0.037	0.040	0.118	-0.166	0.009	0.106	0.119	0.161	0.162	-0.005	0.386	0.105	0.118	0.157	0.157	0.210	0.348
	300	20	-0.001	0.002	-0.021	0.016	0.009	0.025	-0.063	0.006	0.059	0.129	0.088	0.090	-0.003	0.328	0.059	0.128	0.087	0.090	0.207	0.322
	700	20	0.000	0.006	-0.016	-0.000	0.001	0.019	-0.064	0.005	0.036	0.125	0.062	0.058	-0.002	0.321	0.035	0.125	0.062	0.058	0.169	0.315
	1000	20	0.001	0.006	-0.011	-0.001	0.001	-0.020	-0.091	0.006	0.031	0.059	0.054	0.056	-0.002	0.315	0.030	0.058	0.054	0.056	0.181	0.301
	100	50	-0.003	0.003	0.021	0.028	0.028	0.148	-0.113	0.006	0.073	0.071	0.098	0.096	0.017	0.227	0.073	0.068	0.094	0.092	0.146	0.197
	300	50	-0.001	0.005	0.006	0.005	0.007	0.033	-0.069	0.004	0.036	0.040	0.054	0.055	-0.001	0.213	0.036	0.039	0.054	0.054	0.112	0.202
	700	50	-0.001	0.004	0.003	0.002	0.010	-0.009	-0.011	0.002	0.025	0.027	0.036	0.036	-0.001	0.191	0.025	0.027	0.036	0.035	0.123	0.191
	1000	50	-0.000	0.002	0.002	0.004	0.002	-0.006	-0.030	0.002	0.022	0.022	0.031	0.029	-0.001	0.185	0.022	0.022	0.031	0.029	0.108	0.182

TABLE 9. Monte Carlo output, X and τ_i correlated

	N	T	APE	β_0	β_1	β_2	β_3	σ_η	ρ	APE	β_0	β_1	β_2	β_3	σ_η	ρ	β_0	β_1	β_2	β_3	σ_η	ρ
			Mean Absolute Deviation							RMSE							S.E.					
Random Effects Probit	100	10	0.003	-0.027	-0.007	0.037	-0.029	-	-	0.010	0.858	0.132	0.155	0.166	-	-	0.162	0.150	0.152	0.152	-	-
	300	10	0.012	-0.071	0.006	0.012	-0.011	-	-	0.012	0.806	0.082	0.081	0.074	-	-	0.091	0.086	0.086	0.086	-	-
	700	10	0.015	-0.016	-0.007	0.007	-0.001	-	-	0.016	0.820	0.060	0.054	0.061	-	-	0.060	0.056	0.056	0.056	-	-
	1000	10	0.016	0.091	0.009	0.005	-0.002	-	-	0.016	0.752	0.047	0.046	0.053	-	-	0.050	0.047	0.047	0.047	-	-
	100	20	-0.008	-0.501	0.086	-0.017	-0.033	-	-	0.009	0.957	0.116	0.078	0.091	-	-	0.164	0.104	0.102	0.103	-	-
	300	20	0.004	-0.182	0.008	-0.006	-0.003	-	-	0.006	1.001	0.069	0.054	0.062	-	-	0.095	0.059	0.059	0.059	-	-
	700	20	0.015	-0.160	-0.006	-0.014	-0.006	-	-	0.015	0.891	0.036	0.042	0.033	-	-	0.061	0.039	0.039	0.039	-	-
	1000	20	0.015	0.054	0.002	-0.010	-0.006	-	-	0.016	0.785	0.033	0.033	0.033	-	-	0.050	0.032	0.032	0.032	-	-
	100	50	-0.009	-0.337	-0.004	0.077	-0.028	-	-	0.009	0.829	0.094	0.097	0.061	-	-	0.169	0.065	0.065	0.065	-	-
	300	50	0.004	0.236	0.027	0.013	-0.007	-	-	0.005	0.590	0.035	0.048	0.032	-	-	0.087	0.037	0.037	0.037	-	-
	700	50	0.012	-0.055	0.005	-0.007	0.006	-	-	0.012	0.604	0.023	0.018	0.023	-	-	0.057	0.024	0.024	0.024	-	-
	1000	50	0.014	0.268	0.001	0.004	0.002	-	-	0.014	1.177	0.020	0.033	0.026	-	-	0.056	0.020	0.020	0.020	-	-
FLEP	100	10	-0.008	0.007	0.001	0.049	-0.038	0.146	-0.241	0.016	0.060	0.204	0.223	0.223	-0.007	0.526	0.060	0.204	0.217	0.219	0.206	0.468
	300	10	-0.003	0.001	-0.000	0.012	-0.017	0.104	-0.196	0.008	0.041	0.119	0.112	0.113	-0.005	0.504	0.041	0.119	0.112	0.112	0.179	0.464
	700	10	-0.001	0.004	-0.002	0.004	-0.002	0.104	-0.260	0.005	0.028	0.078	0.076	0.081	-0.005	0.543	0.028	0.078	0.076	0.081	0.180	0.477
	1000	10	-0.001	-0.001	0.010	0.006	-0.004	0.072	-0.250	0.005	0.019	0.065	0.064	0.069	-0.005	0.536	0.019	0.064	0.064	0.069	0.176	0.474
	100	20	-0.006	0.009	0.029	0.022	-0.030	0.050	-0.156	0.010	0.116	0.151	0.142	0.154	-0.004	0.384	0.115	0.148	0.140	0.151	0.475	0.351
	300	20	-0.001	-0.006	0.012	-0.004	-0.008	0.088	-0.104	0.005	0.025	0.088	0.077	0.086	-0.003	0.369	0.024	0.087	0.077	0.086	0.131	0.354
	700	20	-0.000	-0.000	-0.004	-0.013	-0.005	0.028	-0.044	0.003	0.017	0.052	0.056	0.052	-0.002	0.328	0.017	0.052	0.055	0.051	0.116	0.325
	1000	20	-0.000	0.000	0.003	-0.009	-0.005	0.042	-0.139	0.003	0.015	0.046	0.046	0.046	-0.002	0.378	0.015	0.046	0.045	0.045	0.120	0.351
	100	50	-0.006	-0.004	0.013	0.019	-0.020	0.047	-0.061	0.008	0.036	0.095	0.098	0.101	-0.001	0.212	0.036	0.094	0.096	0.099	0.150	0.203
	300	50	-0.002	-0.002	0.017	0.007	-0.007	0.029	-0.037	0.004	0.017	0.054	0.052	0.054	-0.001	0.186	0.017	0.051	0.052	0.054	0.084	0.182
	700	50	-0.001	0.001	0.004	0.007	-0.005	0.027	-0.026	0.002	0.011	0.032	0.033	0.034	-0.001	0.200	0.011	0.032	0.032	0.033	0.074	0.198
	1000	50	-0.001	0.002	0.004	0.006	-0.002	0.033	-0.032	0.002	0.008	0.030	0.031	0.028	-0.000	0.196	0.008	0.030	0.030	0.028	0.069	0.193

TABLE 10. Comparison of RE and FLEP for $\tau_i \sim N(0, \sigma_\tau^2)$

	N	T	APE	β_0	β_1	β_2	β_3	σ_η	ρ	APE	β_0	β_1	β_2	β_3	σ_η	ρ	β_0	β_1	β_2	β_3	σ_η	ρ
			Mean Absolute Deviation							RMSE							S.E.					
$\lambda_t \sim AR(2)$	100	10	-0.001	0.002	0.032	-0.012	-0.003	0.238	-0.545	0.011	0.150	0.311	0.348	0.361	-0.019	0.764	0.150	0.309	0.347	0.361	0.334	0.534
	300	10	-0.001	0.001	-0.000	0.011	0.010	0.198	-0.588	0.008	0.079	0.180	0.181	0.207	-0.016	0.810	0.079	0.180	0.180	0.207	0.287	0.557
	700	10	0.001	0.001	-0.000	-0.006	0.016	0.126	-0.639	0.004	0.051	0.121	0.129	0.121	-0.013	0.860	0.051	0.121	0.128	0.120	0.294	0.575
	1000	10	0.002	0.003	0.012	-0.006	0.004	0.165	-0.581	0.005	0.045	0.104	0.102	0.102	-0.015	0.814	0.045	0.104	0.101	0.102	0.299	0.570
	100	20	-0.004	-0.008	0.020	0.062	-0.008	0.220	-0.319	0.009	0.098	0.230	0.239	0.236	0.078	0.541	0.098	0.230	0.231	0.236	0.234	0.438
	300	20	-0.001	-0.001	0.014	0.005	-0.038	0.169	-0.323	0.005	0.056	0.130	0.128	0.139	-0.010	0.573	0.056	0.130	0.128	0.134	0.205	0.474
	700	20	-0.000	-0.002	-0.000	-0.008	0.003	0.157	-0.180	0.003	0.039	0.082	0.087	0.085	-0.009	0.529	0.039	0.082	0.087	0.085	0.243	0.498
	1000	20	-0.000	-0.006	0.005	0.002	0.005	0.102	-0.277	0.003	0.035	0.075	0.071	0.075	-0.008	0.575	0.034	0.075	0.071	0.075	0.248	0.504
	100	50	-0.003	-0.003	0.017	0.035	-0.034	0.221	0.070	0.006	0.057	0.146	0.140	0.143	0.047	0.289	0.057	0.145	0.135	0.139	0.180	0.281
	300	50	-0.001	-0.000	0.014	0.015	-0.008	0.105	0.091	0.003	0.048	0.078	0.084	0.076	-0.004	0.283	0.048	0.077	0.083	0.075	0.241	0.268
	700	50	-0.000	0.003	0.008	0.005	-0.012	0.100	0.097	0.003	0.023	0.060	0.060	0.056	0.018	0.285	0.023	0.060	0.060	0.055	0.126	0.268
	1000	50	0.000	-0.002	-0.008	-0.000	0.005	0.074	0.077	0.003	0.019	0.048	0.047	0.049	-0.004	0.292	0.019	0.047	0.047	0.048	0.166	0.282
$\lambda_t \sim MA(1)$	100	10	-0.002	-0.049	0.026	0.037	-0.109	0.178	-0.652	0.012	0.201	0.365	0.366	0.371	-0.010	0.804	0.195	0.364	0.365	0.355	0.222	0.472
	300	10	0.002	-0.000	0.011	0.004	0.023	0.088	-0.598	0.008	0.075	0.182	0.203	0.207	-0.006	0.772	0.075	0.182	0.203	0.205	0.188	0.488
	700	10	0.001	-0.011	-0.031	0.012	0.006	0.102	-0.550	0.005	0.058	0.119	0.136	0.122	-0.005	0.721	0.057	0.115	0.135	0.122	0.176	0.467
	1000	10	0.001	-0.001	0.006	-0.022	-0.000	0.151	-0.563	0.004	0.040	0.103	0.105	0.101	-0.007	0.730	0.040	0.103	0.103	0.101	0.195	0.466
	100	20	-0.005	-0.000	0.092	0.077	-0.066	0.162	-0.526	0.010	0.120	0.244	0.271	0.275	0.020	0.620	0.120	0.226	0.260	0.267	0.151	0.329
	300	20	-0.000	-0.009	0.015	0.030	-0.007	0.102	-0.498	0.005	0.060	0.141	0.147	0.148	-0.003	0.607	0.060	0.140	0.144	0.148	0.127	0.347
	700	20	-0.000	-0.003	0.002	0.013	0.002	0.088	-0.605	0.003	0.040	0.093	0.091	0.086	-0.003	0.695	0.040	0.093	0.090	0.086	0.120	0.343
	1000	20	-0.000	0.004	0.012	0.000	0.016	0.070	-0.527	0.003	0.032	0.077	0.075	0.078	-0.003	0.635	0.032	0.076	0.075	0.077	0.126	0.355
	100	50	-0.003	0.002	0.061	0.046	-0.053	0.162	-0.533	0.006	0.066	0.153	0.167	0.181	0.009	0.573	0.066	0.140	0.160	0.174	0.098	0.210
	300	50	-0.002	-0.005	0.025	0.010	-0.031	0.097	-0.500	0.004	0.035	0.094	0.080	0.088	0.005	0.541	0.034	0.091	0.079	0.082	0.086	0.207
	700	50	-0.000	-0.001	-0.000	0.005	-0.006	0.076	-0.506	0.003	0.022	0.058	0.058	0.057	0.006	0.549	0.022	0.058	0.058	0.057	0.082	0.211
	1000	50	0.000	0.003	0.001	0.010	0.002	0.056	-0.511	0.002	0.023	0.046	0.047	0.044	-0.001	0.550	0.022	0.046	0.046	0.044	0.081	0.204

TABLE 11. Monte Carlo output, FLEP under misspecification