PANEL PROBIT WITH FLEXIBLE CORRELATED EFFECTS: QUANTIFYING TECHNOLOGY SPILLOVERS IN THE PRESENCE OF LATENT HETEROGENEITY

MARTIN BURDA\textsuperscript{a,b} AND MATTHEW HARDING\textsuperscript{c,*}

\textsuperscript{a} Department of Economics, University of Toronto, Ontario, Canada
\textsuperscript{b} IES, Charles University, Prague, Czech Republic
\textsuperscript{c} Department of Economics, Stanford University, CA, USA

SUMMARY

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 (\textit{Journal of the American Statistical Association} 1993; 88: 669–679; in \textit{Bayesian Biostatistics}, Berry DA, Stangl DK (eds), Marcel Dekker: New York, 1996), and in Chib and Carlin (\textit{Statistics and Computing} 1999; 9: 17–26) 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. Copyright © 2012 John Wiley & Sons, Ltd.

Received 11 May 2010; Revised 5 March 2012

Supporting information can be found in the online version of this article.

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).

Fueled 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 recast 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 et al. (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, for example, 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 DP 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 et al. (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 USA.

We explore the possibility that R&D leads to two major externalities. On 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. A 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 DP 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 nonparametric estimation of an unknown function of the model covariates, while we model nonparametrically the distribution of the unobserved individual heterogeneity. Burda et al. (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 et al. (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 an 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 an 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 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 et al. (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, for example, 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 an application of the method to the estimation of the effect of technological spillovers and product market competition on innovation. Section 4 concludes. A series of Monte Carlo studies comparing the performance of our method with other existing approaches is presented in an online supplement to this article as supporting information.

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:

\[
\tilde{y}_{it} = x_{it}\beta + u_{it} \quad (1)
\]

\[
u_{it} = \tau_{i} + \lambda_{i} + \varepsilon_{it}
\]

\[
y_{it} = 1(\tilde{y}_{it} \geq 0) \quad (2)
\]

where \( x_{it} \) is a \( (1 \times K) \) vector of explanatory variables, \( \tau_{i} \) represents unobserved individual heterogeneity, \( \lambda_{i} \) captures latent time effects, and \( 1(C) \) denotes the indicator function, which takes the value one if the condition \( C \) is satisfied and zero otherwise. The term \( \tilde{y}_{it} \) 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 \( \tau_{i} \), a time-specific component \( \lambda_{i} \) and an idiosyncratic and transitory shock \( \varepsilon_{it} \). 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 \( \tau_{i}, \lambda_{i} \) or \( \varepsilon_{it} \). The model is specified in terms of the latent variable \( \tilde{y}_{it} \), not observed by the econometrician, who only observes the binary outcome variable \( y_{it} \).

Let \( \tilde{y} = (\tilde{y}_{11}, \ldots, \tilde{y}_{NT})' \), \( \tilde{y} = (\tilde{y}_{11}, \ldots, \tilde{y}_{N})' \), \( X_{i} = (x'_{11}, \ldots, x'_{NT})' \), and \( X = (X'_{1}, \ldots, X'_{N})' \), \( \lambda = (\lambda_{1}, \ldots, \lambda_{T})' \), \( \tau = (\tau_{11}, \ldots, \tau_{NT})' \), \( \varepsilon = (\varepsilon_{11}, \ldots, \varepsilon_{NT})' \) and let \( t \) denote a \( (T \times 1) \) vector of ones. Model (1) can thus be rewritten more compactly as \( \tilde{y}_{i} = X_{i}\beta + \tau_{i} + \lambda + \varepsilon_{i} \) for \( i = 1, \ldots, N \) or simply \( \tilde{y} = X\beta + \tau + \lambda + \varepsilon \). Following the notation in Geweke (2005), let the set-valued function

Copyright © 2012 John Wiley & Sons, Ltd.

J. Appl. Econ. (2012)
DOI: 10.1002/jae
\( C_{it} = c_i(y_{it}) \) with \( C_i = (-\infty, 0] \) if \( y_{it} = 0 \) and \( C_i = (0, \infty) \) if \( y_{it} = 1 \). Denote the collection \( C = \{ C_i : i = 1, \ldots, N; t = 1, \ldots, T \} \).

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 \( \tau_i, \lambda_i, \) and \( \varepsilon_{it} \). Its properties are given by the following Assumption.

**Assumption 1.** The error components \( \tau_i, \lambda_i, \) and \( \varepsilon_{it} \), for \( i = 1, \ldots, N \) and \( t = 1, \ldots, T \), are mutually independent conditionally on the \( X \).

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:

\[
\tau_i \sim F^0 \]

where \( F^0 \) 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, \( F^0 \) 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 \( \tau_i \) 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 \( \tau_i \) 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.** \( \lambda_i \) is assumed to follow a stationary Gaussian autoregressive process:

\[
\lambda_i = \rho_1 \lambda_{i-1} + \ldots + \rho_s \lambda_{i-s} + \eta_i
\]

with \( \eta_i \sim N\left(0, \sigma^2_\eta\right) \). Furthermore, \( \eta_i \) is independent of \( \varepsilon_{it} \) 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 (1). In our sampling scheme detailed below, the prior has weight \( 1/(T+1) \) while the likelihood information has weight \( T/(T+1) \). 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 \( \rho \) and \( \sigma_\eta \) 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 $\tau_i$, a time effect $\lambda_t$ and an idiosyncratic shock $\epsilon_{it}$. 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 $\theta = (\beta', \sigma^2, \rho')'$. The assumptions on the prior distributions for this layer are specified as follows.

**Assumption 5.**

\[
\beta \sim N(\beta, \sum \beta) \quad (3)
\]

\[
\sigma^2 \sim IG(v_0, s_0) \quad (4)
\]

\[
\rho \sim \text{Uniform}(\Omega) \quad (5)
\]

where $\Omega \subseteq R^p$ 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. It implies the following model based on Neal (2000):

\[
\tau_i | \psi_i \sim F_i(\psi_i) \quad (6)
\]

\[
\psi_i | G \sim G \quad (7)
\]

\[
G \sim DP(\alpha, G_0) \quad (8)
\]

Thus $F_i$ is specified as an infinite mixture of distributions $F_i(\psi_i)$, with the mixing distribution over $\psi$ being $G$. Here, $\psi_i$ are hyperparameters of the distribution $F_i(\psi_i)$ of $\tau_i$ drawn from a random probability measure $G$, which itself is distributed according to a 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 $\alpha$. 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 $\alpha$ 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 $\alpha$, 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 $F_i$ is sampled conditional on the primary parameters of economic interest $\theta$ and on the regressors $X$. This sampling framework gives us the flexibility to treat $\tau_i$ 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 $F_i$ contains an open neighborhood of the true parameters $\theta$. 

Copyright © 2012 John Wiley & Sons, Ltd.

*M. Burda and M. Harding*

*J. Appl. Econ.* (2012)

DOI: 10.1002/jae
The fourth parameter layer is thus formed by the hyperparameters \( \{ \psi_i \}_{i=1}^N, G, \alpha, G_0, \beta, \sum \beta, v_0, \) and \( s_0. \) In our implementation, \( G_0, \beta, \sum \beta, v_0, \) and \( s_0 \) are fixed, \( \{ \psi_i \}_{i=1}^N, \) and \( \alpha \) are sampled, while bypassing explicit sampling of \( G. \)

Let \( \tau = \{ \tau_i \}_{i=1}^N, \mu_{it} = x_{it} \beta + \tau_i + \lambda_i, \) 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

\[
p(y_i \mid \tau, \beta, \lambda, \sigma^2_\eta, \rho, \psi, \alpha) = p(y_i \mid \tau, \beta, \lambda, \sigma^2_\eta, \rho, \psi, \alpha) k(\psi) k(\lambda) k(\sigma^2_\eta) k(\rho)
\]

with \( k(\rho) \), \( k(\beta) \), and \( k(\sigma^2_\eta) \) given in Assumption 5, \( k(\lambda) \) in Section 5.5 in the Appendix, \( k(\alpha) \) specified as in Escobar and West (1995), and \( k(\psi | \tau) \) given by (7–8). The remainder of the model is formulated similarly to Albert and Chib (1993) with the single index given by \( \mu_{it}. \) Specifically, \( p(y_i \mid \tau, \beta, \lambda, \sigma^2_\eta, \rho, \psi, \alpha) = \prod \phi(\mu_{it}) \) and \( p(y_i \mid \tau, \beta, \lambda, \sigma^2_\eta, \rho, \psi, \alpha) \) assigns probability mass one to \( y_{it} = 1 \) if \( \tilde{y}_{it} > 0 \) and to \( y_{it} = 0 \) if \( \tilde{y}_{it} \leq 0. \) 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 are 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

\[
m_{itk} = \partial E[y_{it} \mid x_{it} \beta, \tau_i, \lambda_i] \bigg/ \partial x_{itk} = \phi(x_{it} \beta + \tau_i + \lambda_i) \beta_k
\]

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

\[
\tilde{\gamma} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \phi(x_{it} \beta + \tau_i + \lambda_i)
\]

The APE of \( x_{itk} \) on \( y \) is then given by

\[
\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T m_{itk} = \tilde{\gamma} \beta_k
\]

We sample explicitly \( \tau_i \) and \( \lambda_i \) throughout the MC iterations and hence can compute the APEs directly from the definition of \( \tilde{\gamma}, \) as

---

*FLEXIBLE PANEL PROBIT

\[
\gamma = \frac{1}{NTS} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{S} \phi(x_{it} \beta_s + \tau_{is} + \lambda_{ts})
\] (12)

where \(s\) is the index over MC steps.

In both the application and the simulation study (included in supporting information), we report the mean bias and means squared error of the estimated ‘APE scale’ coefficient \(\gamma\) defined in (12). To obtain the APEs, \(\gamma\) is simply multiplied by each respective \(\beta_s\).

### 3. INNOVATION AND R&D SPILLOVERS

#### 3.1. 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 et al., 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 et al., 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 are 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 et al. (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 nonparametrically. 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.

3.2. Data

We employ data from a recent study of firm-level R&D by Bloom et al. (2010) (denoted by BSV for the rest of this section). BSV collected firm-level accounting data, such as sales, from the US Compustat database. These 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 to \( k = 1, \ldots, 425 \) different technological classes. If we then let \( T_i^k \) denote a vector where each element represents the average share of patents of firm \( i \) in technological class \( k \) over the period 1980–2001, we can define technological closeness \((Tech)\) between two firms \( i \) and \( j \) by the uncentered correlation between the allocations for the two firms:

\[
Tech_{i,j} = \frac{T_i T_j'}{(T_i T_i')^{1/2} (T_j T_j')^{1/2}}
\]

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:

\[
SpillTech_{i,t} = \sum_{j \neq i} Tech_{i,j} R_{j,t}
\]

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 US 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.

### 3.3. 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 (1) and Assumptions 1-5, the joint posterior density can be decomposed into the following Gibbs blocks:

1. $\beta | \tau, \lambda, \psi, \theta, \tilde{y}, y, X.$

2. $\tilde{y} | \tau, \lambda, \psi, \theta, y, X.$

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

4. $\tau_i | \psi, \theta, \lambda, \tilde{y}, y, X$ for each $i$.

5. $\lambda | \tau, \psi, \theta, \tilde{y}, y, X.$

6. $\sigma^2 | \tau, \lambda, \psi, \theta, \tilde{y}, y, X.$

7. $\rho | \tau, \lambda, \psi, \theta, \tilde{y}, y, X.$

Table I. Summary statistics. All variables are in logarithms and lagged by one period

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpillTech (Jaffe)</td>
<td>9.554</td>
<td>1.142</td>
<td>4.838</td>
<td>11.707</td>
</tr>
<tr>
<td>SpillSIC (Jaffe)</td>
<td>7.272</td>
<td>2.323</td>
<td>-6.024</td>
<td>11.154</td>
</tr>
<tr>
<td>SpillTech (Mah.)</td>
<td>11.314</td>
<td>0.847</td>
<td>8.235</td>
<td>13.156</td>
</tr>
<tr>
<td>SpillSIC (Mah.)</td>
<td>8.513</td>
<td>1.660</td>
<td>0.356</td>
<td>11.559</td>
</tr>
<tr>
<td>R&amp;D Stock</td>
<td>3.030</td>
<td>3.026</td>
<td>2.513</td>
<td>10.765</td>
</tr>
<tr>
<td>Sales</td>
<td>6.230</td>
<td>1.962</td>
<td>0</td>
<td>12.103</td>
</tr>
<tr>
<td>Patenting</td>
<td>0.540</td>
<td>0.498</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Firms</td>
<td>729</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total obs.</td>
<td>12,928</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 nonparametric version with the restricted parametric version of the same model we can gain additional insights into the importance and advantages of using a flexible nonparametric specification over more traditional parametric approaches. Posterior means are reported for these two techniques, obtained from chains of total length of 10,000 MC steps with a 5000 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.

3.4. 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 \( \text{SpillSIC} \) 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 \( \text{SpillTech} \) to be 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: \( \text{SpillTech} \) and \( \text{SpillSIC} \). In order to control for observed firm-level heterogeneity, we include two additional variables. One corresponds to firm sales \( \text{Ln}(\text{Sales}) \), while the other corresponds to the pre-existing stock of R&D available within the firm \( \text{Ln}(\text{R&D stock}) \). Furthermore, we lag all right-hand-side variables by one period so as to remove the possibility of contemporaneous effects.

3.4.1. Partial Effects

Estimation results on the partial effects and the latent common time component are reported in Table 2. In the absence of endogenous patenting activity we should see the marginal effect of technology spillovers \( \text{SpillTech} \) on patenting activity to be positive and the marginal effect of product market spillovers \( \text{SpillSIC} \) 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).1 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.

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 (10) implies that the APE scale depends not only on the estimated $\beta$ 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.

---

1 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).
3.4.2. Unobserved Firm Heterogeneity

We have noted above that both FLEP and PLEP produce quantitatively similar results for both distance measures. A key advantage of FLEP is that it does not impose the normality constraint on the unobserved heterogeneity. Furthermore, FLEP is the only model that allows us to uncover a nonparametric 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 impact 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 $t_i$, which we can use to further analyze the composition of each cluster. Table 4 lists the SIC code names for 20 firms whose $t_i$ 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, for example, 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 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 further investigate the extent to which our theoretical predictions are satisfied for each cluster.

Using the FLEP output we can also explore the evolution of the draws of the unobserved heterogeneity parameters $t_i$ for specific individual companies in order to investigate whether 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 $t_i$. If a draw of $t_i$ 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 $t_i$ values observed in the estimated distribution of unobserved heterogeneity may contain relevant information since it establishes a fairly tight

---

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

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

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

<table>
<thead>
<tr>
<th>Clustering 1</th>
<th>Clustering 2</th>
<th>Clustering 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>4923  Gas Transmission and Distribution</td>
<td>3841  Surgical and Medical Instruments</td>
<td>2835  Diagnostic Substances</td>
</tr>
<tr>
<td>2851  Paints and Allied Products</td>
<td>3661  Telephone and Telegraph Apparatus</td>
<td>2840  Cosmetics</td>
</tr>
<tr>
<td>3060  Fabricated Rubber Products</td>
<td>3825  Instruments To Measure Electricity</td>
<td>3714  Motor Vehicle Parts and Accessories</td>
</tr>
<tr>
<td>3440  Fabricated Structural Metal</td>
<td>3577  Computer Peripheral Equipment</td>
<td>2761  Manifold Business Forms</td>
</tr>
<tr>
<td>2011  Meat Packing Plants</td>
<td>4011  Railroads, Line-haul Operating</td>
<td>2390  Misc. Fabricated Textile</td>
</tr>
<tr>
<td>2731  Book Publishing</td>
<td>3590  Misc. Industrial machinery</td>
<td>3523  Process Control Instruments</td>
</tr>
<tr>
<td>3561  Pumps and Pumping Equipment</td>
<td>3537  Industrial Trucks and Tractors</td>
<td>3572  Computer Storage Devices</td>
</tr>
<tr>
<td>3640  Electric Lighting and Wiring Equipment</td>
<td>6324  Hospital and Medical Service Plans</td>
<td>3674  Semiconductors and Related Devices</td>
</tr>
<tr>
<td>2030  Canned, Frozen, and Preserved Fruit</td>
<td>2000  Food and Kindred Products</td>
<td>3679  Electronic Components</td>
</tr>
<tr>
<td>3533  Oil and Gas Field Machinery</td>
<td>3669  Communications Equipment</td>
<td>3420  Handtools</td>
</tr>
<tr>
<td>3663  Radio and TV Communications Eqpt</td>
<td>3310  Steel Works, Blast Furnaces</td>
<td>2842  Sanitation Goods</td>
</tr>
<tr>
<td>3944  Games, Toys, and Children’s Vehicles</td>
<td>3140  Footwear, Except Rubber</td>
<td>3990  Misc. Manufacturing Industries</td>
</tr>
<tr>
<td>3621  Motors and Generators</td>
<td>3620  Electrical Industrial Apparatus</td>
<td>8731  Commercial Physical Research</td>
</tr>
<tr>
<td>2253  Knit Outerwear Mills</td>
<td>2834  Pharmaceutical Preparations</td>
<td>3613  Switchgear and Switchboard Apparatus</td>
</tr>
<tr>
<td>3579  Office Machines</td>
<td>3711  Motor Vehicles and Car Bodies</td>
<td>3670  Electronic Components and Accessories</td>
</tr>
<tr>
<td>3743  Railroad Equipment</td>
<td>4931  Electric and Other Services Combined</td>
<td>3530  Material Handling Equipment</td>
</tr>
<tr>
<td>3490  Miscellaneous Fabricated Metal Products</td>
<td>3021  Rubber and Plastics Footwear</td>
<td>3829  Measuring and Controlling Devices</td>
</tr>
<tr>
<td>3564  Blowers and Fans</td>
<td>2911  Petroleum Refining</td>
<td>8731  Commercial Physical Research</td>
</tr>
<tr>
<td>2522  Office Furniture, Except Wood</td>
<td>3569  General Industrial Machinery</td>
<td>2821  Plastics Materials and Resins</td>
</tr>
<tr>
<td>5311  Department Stores</td>
<td>3690  Misc. Electrical Machinery</td>
<td>3861  Photographic Equipment and Supplies</td>
</tr>
</tbody>
</table>
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 $t_i$ 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 $t_i$ 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.

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

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Jaffe distance</th>
<th>Mahalanobis distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Cluster 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpillTech</td>
<td>9.516</td>
<td>1.029</td>
</tr>
<tr>
<td>SpillSIC</td>
<td>7.157</td>
<td>2.420</td>
</tr>
<tr>
<td>R&amp;D Stock</td>
<td>4.112</td>
<td>1.996</td>
</tr>
<tr>
<td>Sales</td>
<td>6.143</td>
<td>1.922</td>
</tr>
<tr>
<td>Patenting</td>
<td>0.262</td>
<td>0.440</td>
</tr>
<tr>
<td>Firms</td>
<td>219</td>
<td></td>
</tr>
<tr>
<td>Total obs.</td>
<td>3916</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpillTech</td>
<td>9.621</td>
<td>1.181</td>
</tr>
<tr>
<td>Sales</td>
<td>6.286</td>
<td>2.023</td>
</tr>
<tr>
<td>Patenting</td>
<td>0.622</td>
<td>0.484</td>
</tr>
<tr>
<td>Firms</td>
<td>415</td>
<td></td>
</tr>
<tr>
<td>Total obs.</td>
<td>7361</td>
<td></td>
</tr>
<tr>
<td><strong>Cluster 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpillTech</td>
<td>9.346</td>
<td>1.195</td>
</tr>
<tr>
<td>SpillSIC</td>
<td>7.032</td>
<td>2.233</td>
</tr>
<tr>
<td>R&amp;D Stock</td>
<td>4.318</td>
<td>1.698</td>
</tr>
<tr>
<td>Patenting</td>
<td>0.832</td>
<td>0.373</td>
</tr>
<tr>
<td>Firms</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Total obs.</td>
<td>1651</td>
<td></td>
</tr>
</tbody>
</table>

Table VI. Estimation of patent equation by cluster with FLEP

<table>
<thead>
<tr>
<th>Variable</th>
<th>Jaffe distance</th>
<th>Mahalanobis distance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>Ln(SpillTech)</td>
<td>0.681*</td>
<td>0.529*</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Ln(SpillSIC)</td>
<td>0.026</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Ln(R&amp;D Stock)</td>
<td>0.219*</td>
<td>0.345*</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Ln(Sales)</td>
<td>0.157*</td>
<td>0.182*</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>APE scale</td>
<td>0.192</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>(0.147)</td>
<td>(0.150)</td>
</tr>
</tbody>
</table>

Note: Standard errors are reported in parentheses. 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.
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. 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 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.

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

<table>
<thead>
<tr>
<th>Predicted</th>
<th>RE</th>
<th>FLEP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2668</td>
<td>219</td>
</tr>
<tr>
<td>1</td>
<td>412</td>
<td>617</td>
</tr>
<tr>
<td>Cluster 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2212</td>
<td>567</td>
</tr>
<tr>
<td>1</td>
<td>480</td>
<td>4102</td>
</tr>
<tr>
<td>Cluster 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>213</td>
<td>64</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>1348</td>
</tr>
<tr>
<td>Cluster 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2987</td>
<td>269</td>
</tr>
<tr>
<td>1</td>
<td>481</td>
<td>856</td>
</tr>
<tr>
<td>Cluster 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1959</td>
<td>546</td>
</tr>
<tr>
<td>1</td>
<td>463</td>
<td>4187</td>
</tr>
<tr>
<td>Cluster 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>145</td>
<td>37</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>978</td>
</tr>
</tbody>
</table>

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. 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 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.

4. 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 DP 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 presented in the online supporting information 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

---

3 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 similar performance.
to substantial firm-level heterogeneity which persists within individual industries. We have shown 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.

ACKNOWLEDGEMENTS

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 Econometrics 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).

REFERENCES


Arellano M, Hahn J. 2006. A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. Working paper. CEMFI.


APPENDIX A

Sampling $\beta$

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

$$p(\beta, \tilde{y}^* | \tau, \lambda, \psi, \theta, \tilde{y}, \mathbf{X}) \propto \exp\left[-\frac{1}{2}(\beta - \mu)^\prime \Sigma_\beta^{-1} (\beta - \mu) - \frac{1}{2}(\tilde{y}^* - \mathbf{X}\beta) (\tilde{y}^* - \mathbf{X}\beta) \right]$$

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

$$\bar{\Sigma} = \left( \Sigma^{-1} + (\mathbf{X}^\prime \mathbf{X}) \right)^{-1}$$

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

Sampling $\tilde{y}_i$

Here we benefit from the second step of the Albert and Chib (1993) procedure, augmented by $\tau_i$ and $\lambda_i$. Thus we sample directly

$$\tilde{y}_i | \cdot \sim N(v_i, 1)$$

$$v_i = \mathbf{x}_i \beta + \tau_i + \lambda_i$$

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

Updating Latent Class Assignments

For this block we utilize a hybrid sampler that alternates between the non-conjugate version of the 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 (6)–(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 $\phi = \{\phi_{S_1}, \ldots, \phi_{S_q}\}$, where $\phi_S$ is associated with component $S$. The set partition $\pi$ for $S_0$ is a set of subsets $S_1, \ldots, S_q$ such that (1) $\cup_{S \in \pi} S = S_0$, (2) $S' \cap S'' = \emptyset$ for all $S' \neq S''$, and (3) $S \neq \emptyset$ for all $S \in \pi$. Using this notation, the model (6)–(8) can be recast as (Dahl, 2005)
Given the current state of the system, to the likelihood of a vector of ones as explanatory variables in place of a hypothetical parameters \( \phi \). For an \( N \) of Normals, then sampling. In this case, for every \( v \) of the DP prior, \( \pi \). 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 \( \pi \) of a latent class by the SAMS sampler before the Neal procedure continues to alleviate the in configuration. This is then rapidly split into 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.

Sampling \( \tau_i \)

Let \( \tilde{y}_i^{**} = \tilde{y}_i - X_i \beta = \lambda \). Then

\[
\tilde{y}_i^{**} = \tau_i I + \epsilon_i
\]

Consider for the moment the case \( \tau_i \sim N(\tau, \sigma^2) \); 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 \( \tau_i \) and a \( (T \times 1) \) vector of ones as explanatory variables in place of a hypothetical \( X_i \). Using standard latent regression results (see, for example, Lancaster, 2004)
\[ p(\tau_i | \cdot) = \phi(\bar{\tau}_i, \sigma^2_{\tau_i}) \]  
\[ \bar{\tau}_i = \sigma^2_{\tau_i} \left( \frac{\sigma^{-2}_{\tau} T + \sum_{t=1}^{T} \tilde{y}_i^*}{\tau} \right) \]  
\[ \sigma^2_{\tau_i} = \left( \frac{\sigma^{-2}_{\tau} + T}{\tau} \right)^{-1} \]

Since \( \tau_i \sim N(\mu_{tc}, \sigma^2_{tc}) \) given a previous assignment to the latent class \( c \), let \( \tau = \mu_{tc}, \sigma^2_{\tau} = \sigma^2_{tc} \) and sample \( \tau_i \) directly from (18).

**Sampling \( \lambda \)**

Let

\[ \tilde{y}_l = \tilde{y}_i - X_i \beta - \tau_i t \]

Then the joint density implied for \( \tilde{y}_l = (\tilde{y}_1, \ldots, \tilde{y}_N) \) by the recentered probit model conditional on \( \tilde{\lambda} \) is

\[ f(\tilde{y}_l | \tilde{\lambda}, \cdot) = (2\pi)^{-NT/2} \det(I_T)^{-N/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{N} \left[ \tilde{y}_i^T I_T^{-1} \tilde{y}_i - 2\tilde{\lambda}^T I_T^{-1} \tilde{y}_i + \tilde{\lambda}^T I_T^{-1} \tilde{\lambda} \right] \right\} \]

while the prior density specified by Assumption 3 takes the form

\[ f(\tilde{\lambda}) = (2\pi)^{-T/2} \det(I_T)^{-1/2} \exp \left\{ -\frac{1}{2} \tilde{\lambda}^T \Omega^{-1} \tilde{\lambda} - 2\tilde{\lambda}^T \Omega^{-1} \Lambda \rho + \rho^T \Lambda \Omega^{-1} \rho \right\} \]

where \( \Lambda_t = (\lambda_{t-1}, \ldots, \lambda_{t-s}), \Lambda = (\Lambda_1, \ldots, \Lambda_T), \rho = (\rho_1, \ldots, \rho_s) \) and \( \Omega_\lambda \) is the covariance matrix associated with the autoregressive process. Hence we can sample \( \tilde{\lambda} \) directly from \( N(\tilde{\lambda}, \tilde{\Sigma}_\lambda) \) where

\[ \tilde{\lambda} = \tilde{\Sigma}_\lambda \left( \Omega^{-1} \lambda + \sum_{i=1}^{N} \tilde{y}_{i,i} \right) \]

\[ \tilde{\Sigma}_\lambda = (\Omega^{-1}_\lambda + N \times I_T)^{-1} \]

For ease of implementation we restrict ourselves to the AR(1) specification with a single autoregressive parameter \( \rho \) in the application. In this case
Note that for the AR(1) process

\[
p(\lambda_t | \lambda_{t-1}, \rho) \propto \begin{cases} 
\exp \left( -\frac{1}{2\sigma^2} (\lambda_t - \rho \lambda_{t-1})^2 \right), & t = 1 \\
\exp \left( -\frac{1}{2\sigma^2} (\lambda_t - \rho \lambda_{t-1})^2 \right), & t = 2, \ldots, T
\end{cases}
\]

and hence

\[
p(\rho | \lambda) = \exp \left( -\frac{1}{2\sigma^2} \left[ (1 - \rho^2) \lambda_1^2 + \sum_{i=2}^{T} (\lambda_i - \rho \lambda_{i-1})^2 \right] \right)
= \exp \left( \frac{1}{2\sigma^2} \left[ \rho^2 \left( \sum_{i=2}^{T} \hat{\lambda}_i^2 - \lambda_1^2 \right) - 2\rho \sum_{i=2}^{T} \hat{\lambda}_i \hat{\lambda}_{i-1} + \sum_{i=1}^{T} \hat{\lambda}_i^2 \right] \right)
\]

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

\[
\hat{\sigma}_\rho^2 = \sigma_\eta^2 \left( \sum_{i=2}^{T} \hat{\lambda}_i^2 \right)^{-1}
\]
\[
\hat{\mu}_\rho = \frac{\hat{\sigma}^2_\rho}{\sigma^2_\eta} \sum_{i=2}^{T} \hat{\lambda}_i \hat{\lambda}_{i-1}
= \left( \sum_{i=2}^{T} \hat{\lambda}_i^2 \right)^{-1} \sum_{i=2}^{T} \hat{\lambda}_i \hat{\lambda}_{i-1}
\]

We can therefore sample \( \rho \) directly from \( N\left( \hat{\mu}_\rho, \hat{\sigma}_\rho^2 \right) \) 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).
Sampling $\sigma^2_\eta$

For this block we use the result derived in Burda et al. (2011), which adapts the standard result on sampling univariate variances (given, for example, by result B, p. 300 of Train, 2003) to the likelihood of the variance of the AR process. Conditional on $\lambda$ and $\rho$, the likelihood function of $\sigma^2_\eta$ takes the form

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

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

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

where $m_0$ is a normalizing constant. We can then sample directly from the posterior

$$L\left(\sigma^2_\eta | \lambda, \theta / \sigma^2_\eta\right) \propto L\left(\sigma^2_\eta | \lambda, \theta / \sigma^2_\eta\right) k\left(\sigma^2_\eta\right)$$

$$\propto \frac{1}{\sigma^{(T+w+1)/2}} \exp\left[-\frac{1}{2\sigma^2_\eta} \left(\lambda_1^2 + \sum_{t=2}^{T} (\lambda_t - \rho \lambda_{t-1})^2 + v_0 s_0\right)\right]$$

$$= IG(v_1, s_1)$$

where

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

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

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

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

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, for example, proposal for a split:
The first term is given in equation (13) in Dahl (2005). The second term \( P(\phi_{y}) \) is the proposal density of the new \( \phi_{y} \). The merge probability is

\[
q(\pi, \phi|\pi^{*}, \phi^{*}) = 1
\]

By Bayes’ theorem

\[
p(\pi, \phi|y) \propto p(y|\pi, \phi)p(\pi, \phi)
\]

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

\[
p(y|\pi, \phi) = \prod_{i=1}^{n} p(y_{i}|\phi_{y})
\]

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

\[
p(\pi, \phi) = p(\phi|\pi)p(\pi)
\]

where

\[
p(\phi|\pi) = \prod_{S \in \pi} F_{0}(\phi_{S})
p(\pi) = b \prod_{S \in \pi} \eta_{0}\Gamma(|S|)
b^{-1} = \prod_{i=1}^{n} \Gamma(\eta_{0} + i - 1)
\]

Note that for a split of a class \( S^{x} \) into \( S^{i} \) and \( S^{j} \)

\[
\frac{p(y|\pi^{*}, \phi^{*})}{p(y|\pi, \phi)} = \frac{\prod_{i=1}^{S^{i}} p(y_{i}|\phi_{S^{i}}) \prod_{j=1}^{S^{j}} p(y_{j}|\phi_{S^{j}})}{\prod_{i=1}^{S^{x}} p(y_{i}|\phi_{S^{x}})}
\]

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

\[
\frac{p(y|\pi^{*}, \phi^{*})}{p(y|\pi, \phi)} = \frac{\prod_{i=1}^{S^{i}} p(y_{i}|\phi_{S^{i}}) \prod_{j=1}^{S^{j}} p(y_{j}|\phi_{S^{j}})}{\prod_{i=1}^{S^{x}} p(y_{i}|\phi_{S^{x}}) \prod_{i=1}^{S^{x}} p(y_{i}|\phi_{S^{x}})}
\]

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.
Thus, using (19), the ratio of the \( p \)-terms for a split becomes

\[
\frac{pt(\phi^+ | \pi^+)}{pt(\phi | \pi)} = \frac{F_0(\phi_S) F_0(\phi_S)}{F_0(\phi_S)}
\]

and

\[
\frac{p(\pi^+)}{p(\pi)} = \frac{\Gamma(|S^i|) \Gamma(|S^i|)}{\Gamma(|S^i|)}
\]

while for a merge

\[
\frac{pt(\phi^+ | \pi^+)}{pt(\phi | \pi)} = \frac{F_0(\phi_S)}{F_0(\phi_S) F_0(\phi_S)}
\]

and

\[
\frac{p(\pi^+)}{p(\pi)} = \frac{\Gamma(|S^i|)}{\Gamma(|S^i|) \Gamma(|S^i|)}
\]

Thus, using (19), the ratio of the \( p \)-terms for a split becomes

\[
\frac{pt(\pi^*, \phi^+ | y)}{pt(\pi, \phi | y)} = \frac{\prod_{i=1}^{[S]} pt(y_i | \phi_S) \prod_{i=1}^{[S]} pt(y_i | \phi_S) F_0(\phi_S) F_0(\phi_S) \Gamma(|S^i|) \Gamma(|S^i|)}{\prod_{i=1}^{[S]} pt(y_i | \phi_S) F_0(\phi_S)} \Gamma(|S^i|)
\]

(22)

while for a merge this ratio is given by the inverse of the expression in (22).

**DPM Prior Hyperparameter**

In order to explore the distribution of unobserved heterogeneity, \( \tau_n \), 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 \( \alpha \) that controls the extent to which the DP 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.