Space and Time Analysis of the Knowledge Production Function

Olivier Parent*
Department of Economics
University of Cincinnati
Cincinnati, OH 45221
olivier.parent@uc.edu

February 8, 2008

Abstract

Research and development-based growth models aim to explain the role of technological progress in the growth process. The role of knowledge production and intertemporal spillover effects are investigated using a panel data set covering 49 US states over the period 1994-2004. The aim is to estimate knowledge flows in the context of a space-time dynamic suggested by the knowledge production function. A space-time specification is set forth that can be applied to panel data models with random effects. We compare models that have been proposed recently in the panel data literature to provide a better understanding of how new ideas diffuse across space and time. The results indicate that strong intertemporal knowledge spillovers are present. These results are interpreted in light of the existing theoretical and empirical literature on endogenous growth.

KEYWORDS: Spatial correlation, dynamic panels, MCMC estimations, knowledge production function.

JEL: C11,C21,C23,O31,R11

*I would like thank James P. LeSage, Claude Lopez and Jeffrey Mills for their valuable comments.

1 Introduction

Research and development (R&D)-based models have focused on the functional form of the knowledge production function when modeling the rate of new knowledge generation. The spread of newly generated knowledge may have crucial implication for modeling technological change and economic growth. According to the knowledge production function, generation of new knowledge depends on the fraction of labor engaged in R&D and the existing stock of knowledge available to potential inventors. A sharp debate framed by the work of Romer (1990) and Jones (1995) has focused on how strongly the flow of new ideas is sensitive to the stock of ideas discovered in the past. But these theoretical models treat knowledge as completely diffuse within an economy.

There is an existing empirical literature on technology and knowledge diffusion mostly based on international diffusion flows (Coe and Helpman, 1995). Jaffe and Trajtenberg (2002) show that knowledge follows a complex diffusion process through geographic, institutional and technological spaces. Thus, the likelihood of researchers benefitting from previous inventions increases with proximity and may also vary with the passage of time.

Based on a theoretical model, this paper contributes to the empirical understanding of economic growth by estimating a knowledge production function that quantifies the strength of both intertemporal and inter-regional knowledge spillovers. We examine spatial and time-series patterns of inter-regional patenting to evaluate determinants of the flow of new knowledge. By evaluating patenting patterns using a panel dataset for 49 US states over the period 1994-2004, we contribute to an emerging literature on inter-regional knowledge spillovers. A dynamic model of production of new knowledge with spatial dependence is motivated and developed. Using a Bayesian approach, we provide a better understanding of how localization interacts with time.

Spatial panel data models deal with correlation across locations and usually suppose each period to be independent across the panel. Recent papers (Elhorst, 2004, Baltagi et al., 2007, Yu al., 2006 and Su and Yang, 2007) add the time dimension to the correlation. Focusing on the serial correlation, Elhorst (2004) and Baltagi et al. (2007) apply an extension of the Prais-Winsten transformation by taking into account a stationary process for the initial observations. Yu et al. (2007) do not model the observations of the first period and assume the process is conditional on this initial cross-section. As in Su and Yang (2007) who consider a dynamic model separating the spatial correlation in the error terms, we discuss the impact of the initial observations. Either these first observations are endogenous and the initial observations are approximated using the approach proposed by Bhargava and Sargan (1983) or these first observations are strictly exogenous and the model is set to be conditional on them. When the cross-section of the first period is not treated correctly, the estimator can be biased (Su and Yang, 2007). We focus only on panel data models with random effects. Fixed effects have been analysed when T is large in Yu et al. (2007) and for small T in Su and Yang (2007).

The remainder of the paper proceeds as follows. In Section 2, we review the econometric literature focusing on space-time models in panel data. We develop in Section 3 a parameterization and functional form of the knowledge production function. We extend the underlying model to incorporate production of new ideas in an inter-regional context. In Section 4, we analyse the dynamic model applied the the SAR panel data models and discuss the impact of the initial observations. In Sections 5, we turn to the development of an empirical model based on the use of patenting data. The principal empirical results are reviewed before summarizing the conclusions in Section 6.

2 The econometrics context

Spatial panel data models deal with correlation across locations but typically assume each period is independent across the panel. One exception can be found in work by Baltagi et al., 2007 and Elhorst (2004), where the disturbance structure of the panel data model is extended to allow for dependence across both time and space. Specifically, for a panel where the time index ranges over $t = 1, \ldots, T$ and the index i for ranges over the n regions,

$i = 1, \ldots, n$, we set $y_t = (y_{1t}, \ldots, y_{nt})'$, $x_t = (x_{1t}, \ldots, x_{nt})'$ and we can write the model as:

$$\begin{aligned} y_t &= x_t\beta + u_t \\ u_t &= \mu + \epsilon_t \\ \epsilon_t &= \rho W \epsilon_t + \nu_t \\ \nu_t &= \phi \nu_{t-1} + e_t \end{aligned} \tag{1}$$

where μ reflect an $n-$vector of random region-specific effects that are independent and identically distributed $N(0, \sigma_\mu^2)$, and assumed independent of ϵ. The matrix W is a known n by n spatial weight matrix whose diagonal elements are zero and ρ is a scalar coefficient $|\rho| < 1$, and W is also assumed to satisfy the condition that $I_n - \rho W$ is nonsingular. The scalar parameter $|\phi| < 1$, and represents time-wise serial correlation whereas ρ reflects spatial correlation. The disturbance $e_{it} \sim N(0, \sigma_e^2)$, and stationarity is assumed for the first period $\nu_{i,0} \sim N(0, \sigma_\nu^2/(1 - \phi^2))$.

This type of panel data model allows for serial and spatial dependence just for the error term ϵ. Only variation in the dependent variable y_t that is unexplained by the information set consisting of the explanatory variables and effects parameters is subjected to the serial and spatial dependence model. An additional point to note about the model structure is that space and time are separated.

Relaxing assumption of separation and including a new component that mixes the time lag and the spacial effects, a second set of dependence models have emerged that take the form (Yu et al., 2007):

$$y_t = \rho W y_t + \phi y_{t-1} + \theta W Y_{t-1} + x_t\beta + c + e_t \tag{2}$$

where y_t is an n by 1 cross-section of observations on the n regions at time $t = 1, \ldots, T$, and y_{t-1} is the vector from the previous time period. The $n-$vector e_t contains elements e_{it} that are assumed *i.i.d.* across both i and t with constant scalar variance σ_e^2. The $n-$vector c represents fixed individual effects parameter.

This model filters spatial dependence in the cross-sectional observations at time t using the n by n spatial weight matrix W and associated parameter ρ, as well as allowing for spatial diffusion effects with a one-period time lag reflected in the term: $\theta W y_{t-1}$, and autoregressive order one time dependence captured by: ϕy_{t-1}. An important difference between the model in (2) and that of Baltagi et al. (2007) and Elhorst (2004) in (1) is that the explanatory variables x_t must compete with the space-time filter (consisting of a spatial lag, time lag and space-time lag of the dependent variable) to explain variation in the dependent variable.

The model in (1) is useful when interest centers on controlling for heterogeneity as well as spatial and serial correlation in the disturbance process. In situations where interest focuses on spatial spillovers and diffusion over space through time, the model in (2) seems more appropriate. To see this distinction, note that, spatial spillovers and diffusion cannot take place in the model of (1). The impact of changes in the regressors for this model take the form: $\partial y_{ti}/\partial x_{ti} = \beta'$, $i = 1, \ldots, n$; $t = 1, \ldots, T$. So, a change in the regressors of the model in any region i at any time period will have an impact only on region i, not regions $j \neq i$. Further, the change will be equal for all regions and time periods.

In contrast, the model in (2) does allow for spatial spillovers that take a fairly complicated form. Here we have for the rth regressor: $\partial y_t/\partial x_t^{(r)} = (I_n - \rho W)\beta^{(r)}$, where the superscript (r) denotes the rth column of the matrix x_t and the rth element from the vector β (see Pace and LeSage (2007) for a detailed exposition of this idea). The diagonal of this n by n matrix reflects own-partial derivatives and the off-diagonal represents cross-partial derivatives. Both the own- and cross-partial derivatives allow for spatial spillovers. A change in the regressor for region i will potentially impact all other regions $j \neq i$, where this spillover impact is captured by the off-diagonal elements of the n by n matrix of partial derivative impacts. The main diagonal reflects own-partials, that is $\partial y_{ti}/\partial x_{ti}$, which also has a feedback loop. To see this, note that we can express $(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \ldots$, where higher powers of the matrix W denote higher-order neighbors. For example, W^2 will identify neighbors to W, the first-order neighbors to region i. W_3 will represent neighbors to these neighbors of the neighbors, and so on. Although the matrix W has zeroes on the diagonal, the matrices W^2, W^3, etc. do not, reflecting the fact that region i will be a second-order neighbor to itself. This leads to a situation where a feedback loop exists that is captured by the own-partials or main diagonal elements.

The model also allows for diffusion over space with time and time dependence, which can be seen by considering $\partial y_t/\partial y_{t-1} = (I_n - \rho W)^{-1}(I_n\phi + \theta W)$. Again, changes in past period values in region i will have a potential impact on all regions $j = 1, \ldots, n$, with the main diagonal elements of the n by n matrix reflecting own-partials and off-diagonal elements cross-partials.

Since our interest centers on the space-time dynamics of knowledge spillovers, we will focus on special cases of the model in (2). Based on a theoretical model developed in the next section, we consider two models, one that we label $Model$ 1 that allows for space and time diffusion only after a lag of one period:

$$\begin{aligned} y_t &= \phi y_{t-1} + \theta W y_{t-1} + \iota_n \alpha + x_t \beta + \eta_t \\ \eta_t &= \mu + \varepsilon_t \end{aligned} \tag{3}$$

where α is the intercept, ι_n is an $n \times 1$ column vector of 1 and μ is $n \times 1$ column vector of random effects with $\mu_i \sim N(0, \sigma_\mu^2)$. The disturbance vector $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$ is *i.i.d.* across i and t with zero mean and variance σ_ε^2. In this model the impact of changing the regressors takes the form: $\partial y_t/\partial x_t = \beta$ as in the conventional independence panel data model of (1). The time dynamic takes the form: $\partial y_t/\partial y_{t-1} = I_n\phi + \rho W$, so that changes in region i in previous periods can exert an impact on other regions $j \neq i$ in the next time period.

Yu et al. (2007) develop a dynamic SAR model that leads to a simultaneous dependence structure, where changes in the regressors exert an immediate simultaneous impact on all other regions in the model. $Model$ 2 is a special case where spatial dependence occurs only at a time lag of one period:

$$\begin{aligned} y_t &= \phi y_{t-1} + \rho W y_t + \iota_n \alpha + x_t \beta + \eta_t \\ \eta_t &= \mu + \varepsilon_t. \end{aligned} \tag{4}$$

This specification assumes the spatial effects are instantaneous, consistent with most of the models used in the spatial econometric literature. Partial derivative impacts from changes in the rth explanatory variable in this model take the form: $\partial y_t / \partial x_t^{(r)} = (I_n - \rho W)\beta^{(r)}$. The time dynamics of this model are reflected in $\partial y_t / \partial y_{t-1} = (I_n - \rho W)^{-1}\phi$.

A key point here is that small changes in the estimates for the scalar dependence parameters as well as the coefficients on the explanatory variables can lead to large changes in the space-time impacts of these models due to the non-linear nature of the own- and cross-partial derivatives. We explore this issue using a reciprocal misspecification experiment that focuses on how changes in treatment of the initial period observations impact the estimates of the model.

3 Inter-regional Knowledge Production Function

As in Jones (2002) we consider a world consisting of N separate economies or regions. They differ because of different endowments and allocations, but they have the same production possibilities. Within an economy or region, all agents are identical. The economies evolve independently in all respects except one: they share ideas.

In each economy or region, individuals can produce a consumption-capital good that we will call output. Total output $Y_i(t)$ produced at time t for the economy i is given by

$$Y_i(t) = K_i(t)^\alpha (A_i(t) L_i(t))^{1-\alpha}, \tag{5}$$

where $K_i(t)$ is physical capital, $L_i(t)$ is the total quantity of human capital employed to produce output, and $A_i(t)$ is the total stock of ideas available to this economy. We assume $0 < \alpha < 1$. Notice that there are constant returns to scale and that A is measured in units of Harrod-neutral productivity. Human capital factors are among the primary determinants of the production and diffusion process of innovation.

In the model proposed by Jones (2002), ideas represent the only link between economies; there is no trade in goods, and capital and labor are not mobile. Ideas created anywhere in the world are available to be used in any economy or region. New ideas are produced by researchers, using a knowledge production function (Jones, 1995).

However, as recently proposed by Ertur and Koch (2007), the technological interdependence can be through the level of technological progress. The level of technology in a country depends on the level of technology in other countries. Since each country has different access to this international technology they use an $N \times N$ connectivity matrix W where each element takes values $0 \leq w_{ij} \leq 1$, and the main diagonal elements $w_{ii} = 0$. Here we propose a dynamic specification where the stock of knowledge in the surrounding regions has spillover effects on the growth rate of ideas in region i:

$$\frac{\dot{A}_i(t)}{A_i(t)} = \delta L_i(t)^{\lambda} A_i(t)^{\gamma-1} \prod_{j \neq i} A_j(t)^{\psi w_{ij}}, \tag{6}$$

According to equation (6), the number of new ideas produced at any point in time is driven by the number of researchers and the existing stock of ideas in region i as well as in surrounding regions. We allow $\lambda < 1$ to capture the possibility of duplication in research. For now, we assume $\gamma < 1$ and $\psi < 1$, but stability conditions are discussed in more detail later.

Under the assumption that $\gamma < 1$, we can define a situation in which all variables grow at constant rates (possibly zero), so the dynamics of this economy lead to a stable balanced growth path. Letting g_X denote the growth rate of some variable X along the balanced growth path, we have:

$$\begin{aligned}
& \dot{g}_{A_i(t)} = 0 \\
\iff 0 \quad = \quad & \delta \lambda \dot{L}_i(t) L_i(t)^{\lambda-1} A_i(t)^{\gamma-1} \prod_{j \neq i} A_j(t)^{\psi w_{ij}} + \\
& \delta(\gamma - 1) L_i(t)^{\lambda} \lambda \dot{L}_i(t) \dot{A}_i(t) A_i(t)^{\gamma-2} \prod_{j \neq i} A_j(t)^{\psi w_{ij}} + \\
& \delta L_i(t)^{\lambda} \lambda \dot{L}_i(t) \dot{A}_i(t) A_i(t)^{\gamma-2} \prod_{j \neq i} \psi w_{ij} \dot{A}_j(t) A_j(t)^{(\psi w_{ij}-1)} \\
\iff 0 \quad = \quad & \lambda \frac{\dot{L}_i(t)}{L_i(t)} + (\gamma - 1) \frac{\dot{A}_i(t)}{A_i(t)} + \sum_{j \neq i} \psi w_{ij} \frac{\dot{A}_j(t)}{A_j(t)} \\
\iff g_A \quad = \quad & \frac{\lambda n}{(1-\gamma)} + \frac{\psi}{(1-\gamma)} W g_A \\
\iff g_A \quad = \quad & \left[I_n - \frac{\psi}{(1-\gamma)} W \right]^{-1} \frac{\lambda n}{(1-\gamma)},
\end{aligned}$$

where g_A is a vector of steady state growth rates, and n is the vector of steady state growth rates of researchers. The long-run growth rate of the economy exists only if the matrix $B = \left[I_n - \frac{\psi}{(1-\gamma)}W\right]$ is invertible. Since the matrix W is row-normalized, its maximum eigenvalue is equal to unity and B is invertible if and only if $\psi_{min}^{-1} < \frac{\psi}{(1-\gamma)} < 1$ where ψ_{min} is the minimum eigenvalue of W. If $\psi = 0$, there is no knowledge spillover between regions and the equilibrium growth rate of the stock of ideas corresponds to the one introduced by Jones (1995).

We log-linearize equation (6) around g_A where A_{it} and $H_{A_{it}}$ are growing at constant rates. In this case, the Taylor expansion corresponds to:

$$\begin{aligned}
\frac{\dot{A}_i(t)}{A_i(t)} &= \left.\frac{\dot{A}_i(t)}{A_i(t)}\right|_{g_A} + g_A \left.\left(\frac{\dot{A}_i(t)}{A_i(t)}\right)'\right|_{g_A} [\log(g_A) - \log(g_A)] \\
&= g_A(1 - \log(g_A/\delta)) + \\
&\quad g_A \left[\lambda \log(L_i(t)) - (1-\gamma)\log(A_i(t)) + \sum_{(j \neq i)} \psi w_{ij} \log(A_j(t))\right]
\end{aligned} \tag{7}$$

Focusing on the specific case where $\psi = 0$, Jones (2002) underlines several sources of misspecification in estimating the model in (7). Reverse causality may occur and a more complex autoregressive structure for the distributed lag of research could be elaborated. Jones (2002) makes the important point that $\log(L_i(t))$ and $\log(A_i(t))$ are cointegrated.

We can rewrite (7) as

$$\begin{aligned}
\log(A_i(t+1)) &= g_A(1 - \log(g_A/\delta)) + \\
&\quad g_A \left[\lambda \log(L_i(t)) - (1-\gamma-1/g_A)\log(A_i(t)) + \sum_{(j \neq i)} \psi w_{ij} \log(A_j(t))\right] \\
\log(A_i(t+1)) &= \phi \log(A_i(t)) + \sum_{(j \neq i)} \theta w_{ij} \log(A_j(t)) + \alpha + \beta \log(L_i(t)),
\end{aligned} \tag{8}$$

where $\phi = -g_A(1-\gamma) + 1$, $\theta = g_A\psi$, $\alpha = g_A(1 - \log(g_A/\delta))$ and $\beta = g_A\lambda$.

This diffusion process is similar to an autoregressive model where spatial interaction occurs with a lag of one period. In the next section we will focus on the two different specifications presented in the previous section.

4 Spatial dynamic panel data models

From the theoretical model developed in (8), we first consider the $Model$ 1 as shown in (3). Using matrix notation, let $Y = (y_1', \ldots, y_T')'$, $Y_{-1} = (y_0', \ldots, y_{T-1}')'$ and $X = (x_1', \ldots, x_T')'$. We can write $Model$ 1 as

$$\begin{aligned} Y &= \phi Y_{-1} + \theta(I_t \otimes W)Y_{-1} + \iota_{nT}\alpha + X\beta + \eta \\ \eta &= U\mu + \varepsilon \end{aligned} \tag{9}$$

where $U = \iota_t \otimes I_n$.

The variance-covariance matrix of η has the form $E(\eta\eta') = \Omega$ with

$$\Omega = \sigma_\mu^2(J_T \otimes I_n) + \sigma_\varepsilon^2(I_T \otimes I_n), \tag{10}$$

where $J_T = \iota_T\iota_T'$.

If we assume that y_1 is taken as exogenous, the likelihood function is conditional on y_1. Setting $u = Y - (\phi I_n + \theta W)Y_{-1} - X\beta - \iota_{nT}\alpha$, the log-likelihood function of the complete sample size of T is:

$$\log L(\xi) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\log|\Omega| - \frac{1}{2}u'\Omega^{-1}u \tag{11}$$

where $\xi = (\beta', \alpha, \sigma_\varepsilon^2, \sigma_\mu^2, \phi, \theta)'$.

Baltagi et al. (2007) focusing on a serial and spatial correlation in the error term introduce a Prais-Winsten transformation and therefore assume that $u_0 \sim N(0, \sigma_u^2/(1-\phi^2)I_n)$, where $|\phi| < 1$. Since we propose a dynamic model, we have to assume the explanatory variables to be strictly exogenous and generated by a stationary time process. We follow the estimation procedure used by Bhargava and Sargan (1983).

By substitution of the spatial dynamic panel data model described by (10), the observation y_t is equal to:

$$
\begin{aligned}
y_t &= \sum_{s=0}^{+\infty} A^s y_{t-s} + \sum_{s=0}^{+\infty} A^s x_{t-s} + \\
&\quad (I_n - A)^{-1}(\iota_N \alpha + \mu) + \sum_{s=0}^{+\infty} A^s \varepsilon_{t-s} \qquad (12)
\end{aligned}
$$

where $A = \phi I_n + \theta W$

The stationary assumption $|A| < 1$ implies the first right-hand side variable approaches zero as s approaches infinity. Since $\sum_{s=0}^{+\infty} x_{-s}$ is not observed, $Var(y_0)$ is undetermined. Bhargava and Sargan (1983) suggest predicting y_0 by replacing all the exogenous regressors by $\bar{x} = (T+1)\sum_{t=0}^{T} x_t$. The optimal predictor $\tilde{x}$ under the stationary assumption is $\pi_0 \iota_n + \bar{x}\pi_1 + \xi$, where $\xi \sim N(0, \sigma_{\xi_0}^2 I_n)$. For calculation simplicity, we will assume $\sigma_\xi^2 = \sigma_{\xi_0}^2 \sigma_\varepsilon^2$, were $\sigma_{\xi_0}^2$ is a variance parameter to be estimated. Then the initial observation y_0 can be approximated by $y_0 = \tilde{y}_0 + \xi_0$ where $\tilde{y}_0 = \pi_0 \iota_n + \bar{x}\pi_1$ and

$$
\xi_0 = \xi + \sum_{s=0}^{+\infty} A^s \varepsilon_{-s}. \qquad (13)
$$

Thus,

$$
\begin{aligned}
w_{11} = E(\xi_0 \xi_0') &= \left[\sigma_\xi^2 I_n + (I_n - AA')^{-1}\right] \sigma_\varepsilon^2 + [(I_n - A)'(I_n - A)]^{-1} \sigma_\mu^2 \qquad (14) \\
w_{21} = E(\xi_0 \eta') &= \sigma_\mu^2 (I_n - A)^{-1} \iota_T' \otimes I_n \qquad (15)
\end{aligned}
$$

The joint distribution of $y_T, \ldots, y_1, y_0$ is derived from (3), (11), and (13). Denoting by $\Omega^\star$ the $n(T+1) \times n(T+1)$ variance matrix of $u = (\xi_0, \eta')'$, we see that $\Omega^\star$ has the following form:

$$
\Omega^\star = \begin{pmatrix} w_{11} & w_{21}' \\ w_{21} & \Omega \end{pmatrix},
$$

where Ω, w_{11} and w_{21} are defined in (10), (14) and (15), respectively.

The unconditional log likelihood has the following form:

$$\log L(\xi) = -\frac{n(T+1)}{2}\log(2\pi) - \frac{1}{2}\log|\Omega^\star| - \frac{1}{2}u^{\star\prime}\Omega^\star u^\star \quad (16)$$

where $\xi = (\beta', \alpha, \sigma_\varepsilon^2, \sigma_\mu^2, \phi, \theta)'$ and $u^\star = (y_0' - \pi_0 \iota_n' - \pi_1' \bar{x}', u')'$ and $u = Y - (\phi I_n + \theta W)Y_{-1} - X\beta - \iota_{nT}\alpha$.

As explained in Section 2, $Model$ 2 represents a second specification that allows for a time lag and simultaneous spatial effect as expressed in (4). The idea is to compare these two models in order to give a better understand of how new ideas are diffused across space and time. The first model captures the space and time diffusions only trough the previous period whereas the second model separates the spatial effects from the time effects. Thus, the latter specification assumes the spatial effects are instantaneous, consistent with most of the models used in the spatial econometric literature.

The recent literature about space-time modeling in panel data focuses mostly on stationary processes, where the stationary constraint is imposed by assuming y_0 is endogenous. Elhorst (2007) introduced the idea of extending the Prais-Winsten transformation to the case of a space-time model. As in the case of the first model, we follow the estimation procedure proposed by Bhargava and Sargan (1983).

By substitution of the spatial dynamic panel data model described by (4), the observations y_t are equal to:

$$\begin{aligned} By_t &= (\phi B^{-1})^s A y_{t-(s+1)} + \sum_{s=0}^{+\infty}(\phi B^{-1})^s x_{t-s} + \\ &\quad (I_n - (\phi B^{-1}))^{-1}(\iota_N \alpha + \mu) + \sum_{s=0}^{+\infty}(\phi B^{-1})^s \varepsilon_{t-s}. \end{aligned} \quad (17)$$

where $B = I_N - \rho W$.

Assuming the process started a long time ago, the stationary assumption $|\phi B^{-1}| < 1$ implies the first right-hand side variable approaches zero as s approaches infinity. These stationary assumption are equivalent to $|\phi| < 1 - \rho$ if $\rho \geq 0$ and $|\phi| < 1 - \rho\psi_{min}$ if $\rho < 0$. This allows $B^{-1}(I_n - (\phi B^{-1}))^{-1} = (B - \phi I_n)^{-1}$ to be invertible and therefore the first period corresponds to:

$$y_0 \quad = \quad B^{-1}\sum_{s=0}^{+\infty}(\phi B^{-1})^s x_{-s} + (B-\phi I_n)^{-1}(\iota_N \alpha + \mu) + B^{-1}\sum_{s=0}^{+\infty}(\phi B^{-1})^s \varepsilon_{-s}. \qquad (18)$$

Since $\sum_{s=0}^{+\infty} x_{-s}$ is not observed, $Var(y_0)$ is undetermined. Bhargava and Sargan (1983) suggest predicting y_0 by replacing all the exogenous regressors by $\bar{x} = (T+1)\sum_{t=0}^{T} x_t$. And the optimal predictor $\tilde{x}$ under the stationary assumption is $\pi_0 \iota_n + \bar{x}\pi_1 + \xi$, where $\xi \sim N(0, \sigma_{\xi_0}^2 I_n)$. For calculation simplicity, we will assume $\sigma_\xi^2 = \sigma_{\xi_0}^2 \sigma_\varepsilon^2$, were $\sigma_{\xi_0}^2$ is a variance parameter to be estimated. Then the initial observation y_0 can be approximated by $y_0 = \tilde{y}_0 + \xi_0$ where $\tilde{y}_0 = \pi_0 \iota_n + \bar{x}\pi_1$ and

$$\xi_0 = \xi + (B-\phi I_n)^{-1}\mu + B^{-1}\sum_{s=0}^{+\infty}(\phi I_n + \rho W)^s \varepsilon_{-s}. \qquad (19)$$

Thus,

$$\begin{aligned} w_{11} = E(\xi_0 \xi_0') \quad &= \quad \left\{\sigma_{\xi_0}^2 I_n + \left[B'B - \phi^2 B'(B'B)^{-1}B)'\right]^{-1}\right\}\sigma_\varepsilon^2 + \\ &\quad\; \left[(B-\phi I_n)'(B-\phi I_n)\right]^{-1}\sigma_\mu^2 \qquad (20) \\ w_{21} = E(\xi_0 \eta') \quad &= \quad \sigma_\mu^2 (I_n - \phi B^{-1})^{-1}\iota_T' \otimes I_n \qquad (21) \end{aligned}$$

Denoting by $\Omega^\star$ the $n(T+1) \times n(T+1)$ variance matrix of $u = (\xi_0, \eta')'$, we see that $\Omega^\star$ has the following form:

$$\Omega^\star = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & \Omega \end{pmatrix},$$

where Ω, w_{11} and $w_{12} = w_{21}$, are defined in (10), (20) and (21), respectively.

The unconditional log likelihood has the following form:

$$\log L(\xi) \quad = \quad -\frac{n(T+1)}{2}\log(2\pi) - \frac{1}{2}\log|\Omega^\star| - \frac{1}{2}u^{\star\prime}\Omega^\star u^\star \qquad (22)$$

where $\xi = (\beta', \alpha, \sigma_\varepsilon^2, \sigma_\mu^2, \phi, \rho)'$ and $u^\star = (y_0' - \pi_0 \iota_n' - \pi_1' \bar{x}', u')'$ and $u = Y - (\phi I_n + \rho W)Y_{-1} - X\beta - \iota_{nT}\alpha$.

Su and Yang (2007) have shown that when the cross-section of the first period is exogenous but treated incorrectly as endogenous, the estimator can be biased. Estimator are also biased when y_0 is endogenous but treated incorrectly as exogenous. We will present for each model both treatments for the cross-section of the first period.

5 Estimation Method

Even if most of the studies used the Quasi-Maximum Likelihood (QML) estimation, the most difficult step is to maximize the concentrated log-likelihood function of θ and ϕ (see Su and Yang, 2007; Yu et al. 2007).

The MCMC method can handle the problem of local optima, that often arise in space-time modeling. In the presence of local optima, conventional likelihood methods may provide misleading inference whereas the ability of Bayesian MCMC methods to directly sample from the posterior can avoid some of these problems (Lesage and Pace, 2007).

Bayesian inference is based on the joint posterior distribution of the parameters given the data. Due to the hierarchical structure of (6), its unnormalized form is easily derived as:

$$\begin{aligned} p(\alpha, \beta, \sigma_\mu^2, \sigma_\varepsilon^2, \theta, \phi, \pi_0, \pi_1, \sigma_{\xi_0}^2 | y) \quad \propto \quad & p(y|\beta, \alpha, \mu, \sigma_\varepsilon^2, \phi, \theta, \pi_0, \pi_1, \sigma_{\xi_0}^2) p(\mu|\sigma_\mu^2) p(\sigma_\mu^2) p(\beta) \\ & p(\alpha)p(\theta)p(\sigma_\varepsilon^2)p(\phi)p(\pi_0)p(\pi_1)p(\sigma_{\xi_0}^2) \qquad (23) \end{aligned}$$

where the expression $p(y|.)$ is the likelihood function and $p(.)$ are the prior distributions.

Note that we propose a sampling scheme discarding the first period. Direct evaluation of the joint posterior distribution involves multidimensional numerical integration and is not computationally feasible. We use MCMC sampling methods which involve generating sequential samples from the complete set of conditional posterior distributions detailed in Appendix A.

6 Empirical Results

To estimate $Model$ 1 and $Model$ 2, we must measure the stock of ideas. Observable measures of new ideas at a regional or international level are never perfect. We organize the analysis by focusing on observed number of U.S. domestic patents, a useful indicator of the state level of realized innovation for a given period. We estimate the knowledge production function using a dataset of patenting activity and its determinants from 1994 to 2004 across 49 states.[1] The data include the granted patents per capita for each state in each year and measures of the factor inputs into ideas production. We review in Appendix B the definition and summary statistics associated with each of the measures. Figure 1 represents the average growth rates of granted patent per 100,000 inhabitants across the US over the period 1994-2004. Highest growth rates are located in the West regions, west of the Midwest and in New-England with the largest value for the state of Idaho. Regions with the lowest growth rates are located in East North Central and East South Central and the state of West Virginia has the lowest value. Figure 2 shows the evolution of the granted patent per capita over time. Interestingly we note that over the last decade, regions with the highest level of granted patent per capita are also those who have the highest growth rates (Idaho, Vermont, California, Oregon,etc.). And regions with the lowest level of innovation are those with the lowest growth rates (Mississippi, Arkansas, West Virginia, Louisiana, Alabama, etc). These observations may suggest that convergence in the United States proceeds among geographically neighboring states rather than among distant states.

Skilled labor $L_i(t)$ for each state i at the period t is measured with two different explanatory variables, $LPost_i(t)$, the number of Postdoctorates in Science and Engineering, and $ExpRD_i(t)$, the total Research and development expenditures as a percentage of the gross state product. Total R&D expenditures are calculated by adding all the sources of funds: industry, public and private non-profit institutes and universities. For $Model$ 1, the stock

[1] The District of Columbia is treated as a state and the states of Alaska and Hawaii are omitted.

of granted patents per worker for state i at the period t results in the following regression:

$$\begin{aligned}
\log(A_{it}) &= \alpha + \phi \log(A_{i,t-1}) + \sum_{(j \neq i)} \theta w_{ij} \log(A_{j,t-1}) + \\
&\quad \beta_1 \log(LPost_{i,t-1}) + \beta_2 \log(ExpRD_{i,t-1}) + \eta_t \\
\eta_t &= \mu + \varepsilon_t. \qquad (24)
\end{aligned}$$

For $Model$ 2 we allow for a time lag and simultaneous spatial effect:

$$\begin{aligned}
\log(A_{it}) &= \alpha + \phi \log(A_{i,t-1}) + \sum_{(j \neq i)} \rho w_{ij} \log(A_{j,t}) + \\
&\quad \beta_1 \log(LPost_{i,t}) + \beta_2 \log(ExpRD_{i,t}) + \eta_t \\
\eta_t &= \mu + \varepsilon_t. \qquad (25)
\end{aligned}$$

Note that from the theoretical model we have derived in the first section, there is a lag of one period for the explanatory variables. As previously explained, explanatory variables and the dependent variable have to be cointegrated in this case (Jones, 2002). Note that this lag is omitted in the second specification.

These two models allow us to perform empirical analyses dissecting the drivers of domestic innovative capacity and evaluate the impact of the diffusion process of neighboring activities on regional innovative performance. Table 1 and Table 2 report the regression results for both specifications.[2] Consistent with the ideas-based growth literature, the results suggests that the level of innovation is influenced powerfully its level of effort devoted to the ideas sector.

Expenditures in R&D have a more permanent impact on the growth process if a highly skilled workforce eases the adoption of new technologies. Advanced regions in technology indeed often have strong links with education, especially at the post doctoral level. Thus, high education should lead to a faster rate of technological progress via improvements in the quality of the workforce. However for both model, the effects of the variable $LPost$ is not significant. Focusing on the expenditure in R&D, results reveal a strong positive influence

[2] Estimation results are based on a simulated chain where the first 5,000 samples are discarded as a 'burn-in' period, followed by 15,000 iterations.

Table 1: Estimation Results for $Model$ 1 - 20,000 iterations

parameter	Post. mean	s. d.	Low. 0.05	Up. 0.95	Post. mean	s. d.	Low. 0.05	Up. 0.95
$Const$	-0.054	0.157	-0.364	0.136	0.382	0.060	0.289	0.485
$ExpRD$	0.096	0.035	0.020	0.172	0.070	0.027	0.025	0.112
$LPost$	0.021	0.022	-0.014	0.061	-0.013	0.014	-0.037	0.010
σ_ε^2	0.069	0.039	0.043	0.118	0.015	0.001	0.013	0.017
σ_μ^2	0.050	0.020	0.022	0.087	0.054	0.013	0.037	0.08
θ	0.358	0.025	0.321	0.396	0.265	0.016	0.245	0.293
ϕ	0.530	0.086	0.409	0.643	0.564	0.016	0.536	0.589

Note: The last four column correspond to estimates treating y_1 exogenously, whereas the other column corresponds to estimates treating y_1 endogenously.

Table 2: Estimation Results for $Model$ 2 - 20,000 iterations

parameter	Post. mean	s. d.	Low. 0.05	Up. 0.95	Post. mean	s. d.	Low. 0.05	Up. 0.95
$Const$	0.036	0.061	-0.043	0.150	0.347	0.108	0.167	0.505
$ExpRD$	0.061	0.023	0.024	0.102	0.037	0.019	0.005	0.066
$LPost$	0.010	0.015	-0.013	0.036	-0.005	0.012	-0.026	0.015
σ_ε^2	0.045	0.011	0.028	0.066	0.013	0.000	0.011	0.014
σ_μ^2	0.006	0.006	0.001	0.019	0.058	0.014	0.036	0.086
ρ	0.323	0.014	0.302	0.353	0.302	0.012	0.285	0.324
ϕ	0.580	0.031	0.537	0.626	0.571	0.032	0.521	0.611

Note: The last four column correspond to estimates treating y_1 exogenously, whereas the other column corresponds to estimates treating y_1 endogenously.

on the innovative activities.

Focusing on the random effects, estimation results depicted in Figure 3 show strong positive effect in the Western and Northeaster regions. Idaho, which has the largest positive value, is a relatively small state with a growing science and technology sector which amounts for over 25% of the State's total revenue. The combination of a very high-valued output and low level of skilled labor and postdoctoral students results in the large values for the random coefficient. Similar observation can be relevant for the State of Vermont with leading innovative firms like IBM. On the other side states like West Virginia or Louisiana have the lowest growth rate of innovative activities despite the large number of postdoctoral students. This could explain the negative values for the random effects. It is difficult for technological laggards to progress as the technological leading states. For most regions, a substantial amount of technical effort is devoted towards imitative or absorptive activity rather than the production of innovative activities. Skills availability in the western parts of the U.S. meets the needs of the labour market. In these regions, innovative economic activities are helped by the presence of high number of research or scientist per capita.

Local spillover effects are related to the presence of the knowledge stock of neighboring regions. It requires that regions possess the ability to absorb and to adopt new technologies (see Parent and LeSage, 2008). R&D activities can increase the incidence of technology diffusion by enhancing the regional absorptive capacity. Our result confirm that R&D expenditure can directly increases the level of regional innovative activities. Therefore increasing R&D absorptive capacity can facilitate technology spillovers from other regions.

We propose two different specification to measure the impact of neighboring regions. In $Model$ 1, we assume that new ideas may not be created immediately in response to a change in the level of neighboring innovative activities. Spatial dependence is these effects may not occur immediately in response to a change in the number of scientists of researchers. The relationship between the stock of ideas $\log(A_{it})$ and $\log(A_{i,t-1})$, represents the relationship between ideas discovered today and the number of ideas previously revealed. Note that the sum of the parameters $\phi + \theta$ is smaller than unity, revealing that the process is stationary with the time effect having a greater weight than the spatial dependence. This evidence reveals the possibility of knowledge diffusion across space and time. The second model which is usually used in spatial econometrics assume simultaneous spatial dependence. The strength of the spatial dependence ρ in $Model$ 2 is positive indicating a strong effect among geographically adjacent area on innovative activities. This result confirms that technological innovation at a regional level cannot be assumed statistically independent because of the presence of similar observations among neighboring regions (Anselin et al. 1997). The spatial effects are modeled using a spatial weight matrix W based on contiguity between regions. Interactions between regions are spatially limited and reveal that localized spillovers effects lead to regional clusters with persistently different levels of innovative activities. These results have been confirmed by Keller (2002) who has underlined the importance of geographic distance for technology diffusion.

These specifications suggest that key elements that are amenable to policy change in regional economics are significant in explaining innovative productivity across regions over the past decade. This empirical analysis has addressed the question to what extent the geographical proximity affects the diffusion process of innovative activities over time. In particular, we show that regions located in the West and Northeast have a higher stock of technological knowledge at their disposal than more centrally located regions.

7 Conclusion

This paper reveals some of the key findings of recent research on regional innovative capacity. Our findings suggest (a) that patenting activities are well-characterized by different economic factors which may be affected by public policy and (b) that the United States has experienced substantial space and time dependence on the regional innovative capacity over the last decade.

We introduce a new endogenous growth model framework containing technological change to incorporate spatially structured technological diffusion over time. This can be interpreted as reflecting spillovers that arise in the dynamic process that governs technological growth over time. A Bayesian approach is then proposed in order to estimates the spatial diffusion of innovative activities in a dynamic framework. An empirical study based on US domestic granted patents across 49 states over the period 1994-2004 reveals the importance of proximity for technology diffusion and discuss how innovative activities are limited by distance. In future work, we hope to extend this framework, to provide further evidence about the relationship between the regional innovation infrastructure and R&D productivity in individual industrial clusters using a Vector Autoregressive specification with spatial dependence.

References

[1] Anselin, L., A. Varga and Z. Acs. 1997. "Local Geographic Spillovers Between University Research and High Technology Innovations," *Journal of Urban Economics*, **42**

422-448.

[2] Baltagi B.H., S.H.Song, B.C. Jung and W. Koh (2007). “Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data”, *Journal of Econometrics*, **140**, 5-51.

[3] Bhargava A., and J.D. Sargan (1983). “Estimating dynamic random effects models from panel data covering short time periods”, *Econometrica* 51, 1635-1659.

[4] Elhorst P. (2004). “Serial and Spatial Error Dependence in Space-Time Models”. In: A. Getis, J. Mur and H.G. Zoller (eds.). *Spatial Econometrics and Spatial Statistics*, pp. 176-193. Palgrave-MacMillan

[5] Ertur C. and W. Koch (2007). “The Role of Human Capital and Technological Interdependence in Growth and Convergence Processes: International Evidence”, *Journal of Applied Econometrics*, 22:6, 1033-1062.

[6] Hsiao, C. (2003). *Analysis of Panel Data*. 2nd edition. Cambridge University Press.

[7] Jaffe A. B. and Trajtenberg, M. (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*, Cambridge, Mass.: MIT Press.

[8] Jones C.I. (1995). “R&D-Based Models of Economics Growth”, *Journal of Political Economy*, **103**, 759-582.

[9] Jones C.I. (2002). “Sources of U.S. Economic Growth in a World of Ideas ”, *American Economic Review*, **92**, 220-239.

[10] Keller W. (2002). “Geographic Localization of International Technology Diffusio”, *American Economic Review*, **92** 120-142.

[11] LeSage, J. and R.K. Pace (2007). “A Matrix Exponential Spatial Specification”, *Journal of Econometrics*, **140**, 190-214 .

[12] Pace R.K. and LeSage J.P. (2007). “Interpreting spatial econometric models”, working paper.

[13] Parent O. and J. LeSage (2008). "Using Constraints on the Variance Structure in the Conditional Autoregressive Specification to Model Knowledge Spillovers", *Journal of Applied Econometrics*, forthcoming.

[14] Romer P.M. (1990). "Endogenous Technological Change", *Journal of Political Economy*, **98**, 71102.

[15] Su L. and Z. Yang (2007). "QML Estimation of Dynamic Panel Data Models with Spatial Errors", Singapore Management University manuscript.

[16] Yu J., R. de Jong and L.F. Lee (2006). "Quasi-Maximum Likelihood Estimators For Spatial Dynamic Panel Data With Fixed Eects When Both n and T Are Large", working paper.

Appendix A

For both models, we suppose that the prior distribution of $p(\alpha, \beta, \sigma_\mu^2, \sigma_\varepsilon^2, \psi, \phi, \pi_0, \pi_1, \sigma_{\xi_0}^2)$ are all a priori independent, where $\psi = \theta$ for $Model$ 1 and $\psi = \rho$ for $Model$ 2. We estimate separately the intercept term α and the parameters β assuming a non-hierarchical prior of the independent Normal-Gamma variety. Thus,

$$\begin{aligned} \alpha &\sim n(\alpha_0, M_\alpha^{-1}) \\ \beta &\sim n(\beta_0, M_\beta^{-1}) \\ \sigma_\varepsilon^{-2} &\sim G(v_0/2, S_0/2) \end{aligned} \tag{26}$$

Focusing first on $Model$ 1, since the first period does not depend on β, combining the normal prior of β in (26) with the likelihood of β defined by (16), we obtain using standard results that

$$\begin{aligned} p(\beta|y, \sigma_\varepsilon^2, \mu, \phi, \theta, \alpha) &\propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}(\tilde{V} - X\beta)'(\tilde{V} - X\beta)\right\} \exp\left\{-\frac{1}{2}(\beta - \beta_0)' M_\beta (\beta - \beta_0)\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\beta - \beta_1)'(\sigma_\varepsilon^{-2} X'X + M_\beta)(\beta - \beta_1)\right\} \end{aligned}$$

where $\tilde{V} = Y - (\phi I_n + \theta W)Y_{-1} - \iota_{nT}\alpha - U\mu$, $\beta_1 = [\sigma_\varepsilon^{-2}X'X + M_\beta]^{-1}[\sigma_\varepsilon^{-2}X'X\widehat{\beta} + M_\beta\beta_0]$ and $\widehat{\beta} = (X'X)^{-1}X'\tilde{V}$. That is $p(\beta|y,\sigma_\varepsilon^2,\mu,\phi,\theta,\alpha) \sim N(\beta_1, [\sigma_\varepsilon^{-2}X'X + M_\beta]^{-1})$.

Simulating the constant term α is also straightforward. We adopt the same procedure to draw from the posterior distribution of the intercept term. Therefore $p(\alpha|y,\beta,\sigma_\varepsilon^2,\mu,\phi,\theta) \sim N(\alpha_1, [\sigma_\varepsilon^{-2}nT+M_\alpha]^{-1})$, where $\alpha_1 = [\sigma_\varepsilon^{-2}nT+M_\alpha]^{-1}[\sigma_\varepsilon^{-2}nT\widehat{\alpha}+M_\alpha\alpha_0]$ and $\widehat{\alpha} = (nT)^{-1}\iota'_{nT}(\tilde{Y}-\mu - X\beta)$, with $\tilde{Y} = Y - AY_{-1}$ and $A = \phi I_n + \theta W$.

To estimate the random effects, we use a proper prior and we assume that, for $i = 1,\ldots,n$, $\mu_i \sim N(0,\sigma_\mu^2)$ with μ_i and μ_j being independent of one another for $i \neq j$. A hierarchical structure of the prior arises because we treat σ_μ^2 as unknown parameters which require its own prior. We define the error terms as the $nT \times 1$ vector $u^\star = (y_0' - \pi_0\iota_n' - \pi_1'\bar{x}', u')'$ where $u = Y - (\phi I_n + \theta W)Y_{-1} - X\beta - \iota_{nT}\alpha$. We obtain the following posterior distribution $p(\mu|y,\beta,\alpha,\sigma_\varepsilon^2,\theta,\phi) \sim N(m_1, [\sigma_\varepsilon^{-2}U'\Omega_0^{-1}U + \sigma_\mu^{-2}I_n]^{-1})$, where $m_1 = [\sigma_\varepsilon^{-2}U'\Omega_0^{-1}U + \sigma_\mu^{-2}I_n]^{-1}[\sigma_\varepsilon^{-2}U'\Omega_0^{-1}u^\star]$, where

$$\Omega_0 = \begin{pmatrix} \sigma_\xi^2 I_n + (I_n - AA')^{-1} & 0 \\ 0 & I_{n(T-1)} \end{pmatrix}. \tag{27}$$

For the hierarchical structure of the random term, we specify a Gamma distribution for the precision parameter, $\sigma_\mu^{-2} \sim G(v_1/2, S_1/2)$. Thus, the posterior distribution for σ_μ^{-2} corresponds to

$$\begin{aligned} p(\sigma_\mu^{-2}|y,\sigma_\varepsilon^2,\mu,\phi,\theta) &\propto (\sigma_\mu^{-2})^{n/2}\exp\left\{-\frac{\sigma_\mu^{-2}}{2}\mu'\mu\right\}(\sigma_\mu^{-2})^{v_1/2-1}\exp\left\{-\frac{\sigma_\mu^{-2}}{2}S_1\right\} \\ &\propto (\sigma_\mu^{-2})^{(v_1+n)/2-1}\exp\left\{-\frac{\sigma_\mu^{-2}}{2}(\mu'\mu + S_1)\right\} \end{aligned}$$

that is $p(\sigma_\mu^{-2}|y,\gamma,\mu,\theta) \sim G([v_1+n]/2, [\mu'\mu + S_1]/2)$.

Using also a Gamma distribution for the precision parameter, $\sigma_\varepsilon^2 \sim G(v_0/2, S_0/2)$, the conditional posterior is given by:

$$
\begin{aligned}
\sigma_\varepsilon^2|y,\gamma,\mu,\theta &\sim G(\overline{\tau},\overline{H}) \qquad (28)\\
\overline{\tau} &= (T+v_0)/2\\
\overline{H} &= (u^{\star\prime}\Omega_0^{-1}u^\star + S_0)/2,
\end{aligned}
$$

we adopt improper prior for ϕ and θ which are uniformly distributed over the real line.

$$
\theta|\mu,\sigma_\varepsilon^2,\phi,y \propto |\sigma_\varepsilon|^{-n(T+1)}\left|\sigma_{\xi_0}^2 I_n + \left[I_n - (\phi I_n+\theta W)(\phi I_n + \theta W)'\right]^{-1}\right|^{-1}
$$
$$
\exp\left(-\frac{1}{2\sigma_\varepsilon^2}e'\Omega_0^{-1}e\right). \qquad (29)
$$

This is not reducible to a standard distribution, so we adopt a Metropolis-Hastings step during the MCMC sampling procedures that relies on a random walk proposal with normally distributed increments, $\theta_{new} = \theta_{old} + \gamma N(0,1)$. The acceptance probability is calculated as the ratio of (29) evaluated at the old and new candidate draws. The proposal turning parameter was systematically incremented or decremented when the acceptance rate moved below 0.40 or above 0.60, which lead to an acceptance rate close to 0.50 after a burn-in period. We compute the log-determinant term using a direct sparse matrix LU decomposition approach described in Pace and Barry (1997) that produces vectorized grids of values for this over the domain of support for θ and ϕ. We draw ϕ using the same sampling procedure.

Relating to the first period, parameters π_0 and π_1 are generated the same way we draw α and β. Assuming a Normal prior for $\pi_1 \sim N(b_1, T_1^{-1})$, the posterior distribution corresponds to $p(\pi_1|y_0,\sigma_\varepsilon^2,\phi,\theta,\pi_0,\sigma_{\xi_0}^{-2}) \sim N(\tilde{b}_1,\tilde{T}_1)$, where $\tilde{T}_1 = (\bar{x}'w_{11}^{-1}\bar{x} + T_1)$ with w_{11} defined in (14) and $\tilde{b}_1 = \tilde{T}_1^{-1}\left[\bar{x}'w_{11}(y_0-\pi_0) + T_1 b_1\right]$.

We use a Normal prior for $\pi_0 \sim N(b_0, T_0^{-1})$. Therefore the posterior distribution is equivalent to $p(\pi_0|y_0,\sigma_\varepsilon^2,\phi,\theta,\pi_1,\sigma_{\xi_0}^{-2}) \sim N(\tilde{b}_0,\tilde{T}_0)$ where $\tilde{T}_0 = (\iota_n' w_{11}^{-1}\iota_n + T_0)$ and $\tilde{b}_0 = \tilde{T}_0^{-1}\left[\iota_n' w_{11}(y_0 - \bar{x}\pi_1) + T_0 b_0\right]$.

We use an additional Metropolis Hastings step in order to generate $\sigma_{\xi_0}^2$. A Gamma distribution is introduced as a prior for $\sigma_{\xi_0}^{-2} \sim G(v_2/2, S_2/2)$. Thus the posterior is proportional to

$$\begin{aligned} \sigma_{\xi_0}^{-2}|y_0, \sigma_\varepsilon^2, \phi, \theta, \pi_1 \quad &\propto \quad (\sigma_{\xi_0}^{-2})^{v_2/2}(\sigma_\varepsilon^{-2})^{n/2} \left|\sigma_{\xi_0}^2 I_n + \left[I_n - (\phi I_n + \theta W)(\phi I_n + \theta W)'\right]^{-1}\right|^{-1} \\ &\exp\left\{-\frac{e'\Omega_0^{-1}e}{2\sigma_\varepsilon^2} - \frac{S_2}{2\sigma_{\xi_0}^2}\right\} \end{aligned} \tag{30}$$

where Ω_0^{-1} is defined in (27).

The MCMC sampling procedures relies again on a random walk proposal with normally distributed increments, $\sigma_{\xi_0 new}^2 = \sigma_{\xi_0 old}^2 + \gamma N(0,1)$. The acceptance probability is calculated as the ratio of (30) evaluated at the old and new candidate draws.

For each of these Metropolis-Hastings algorithm, we need to satisfy the stationarity assumption $\left|\sigma_{\xi_0}^2 I_n + [I_n - (\phi I_n + \theta W)(\phi I_n + \theta W)']^{-1}\right| < 1$.

The estimation method for $Model$ 2 is similar if we replace θ by ρ and $\tilde{V} = Y - \phi Y_{-1} - \rho WY - \iota_{nT}\alpha - U\mu$. We obtain the same posterior distribution for the parameters β, α and the parameters for the first cross section π_0, π_1 and $\sigma_{\xi_0}^2$.

For the posterior distribution of the random coefficient μ, for and for the parameters σ_{mu}^2 and σ_ε^2, we need to redefine the variance matrix Ω_0 using equations 14 and 15.

We adopt again an improper prior for ρ and its posterior distribution is given by

$$\begin{aligned} \rho|\phi, \mu, \sigma_\varepsilon^2, \phi, y \quad &\propto \quad |\sigma_\varepsilon|^{-n(T+1)}|B|^{-T} \left|\sigma_{\xi_0}^2 I_n + \left[B'B - \phi^2 B'(B'B)^{-1}B)'\right]^{-1}\right|^{-1} \\ &\exp\left(-\frac{1}{2\sigma_\varepsilon^2} e'\Omega_0^{-1}e\right). \end{aligned} \tag{31}$$

The same MCMC sampling procedures that relies on a random walk proposal with normally distributed increments is implemented to generate draws from the conditional distribution of ρ and ϕ.

Appendix B

The dependent variable corresponds to the number of U.S. patents granted by the U.S. Patent and Trademark Office during the period 1994-2004.[3] We assume that ideas production in a given year is reflected in research activities undertaken previously. National Patterns of R&D Resources and can be downloaded from the National Science Foundation Website. They describe and analyze patterns of research and development (R&D) in the United States every year since 1994. Data related to Postdoctoral studies are derived from the fall 2005 National Science Foundation-National Institutes of Health Survey of Graduate Students and Postdoctorates in Science and Engineering. The measure of output is Gross State Product (GSP) in manufacturing from the BEA Website.

Table 3: Descriptive statistics

Variable	mean	std	min	max
A	24.37	4.60	3.85	135.98
$ExpRD$	23.24	0.59	2.86	94.78
$Lpost$	586.3	129.6	17	5008

Figures

[3] Data available at the U.S. Department of Commerce, U.S. Patent and Trademark Office. *Patent Counts By Country/State And Year: Utility Patents: January 1, 1963 - December 31, 2006*

Figure 1: Average growth rates of granted patent per 100,000 inhabitants across the US (1994-2004)

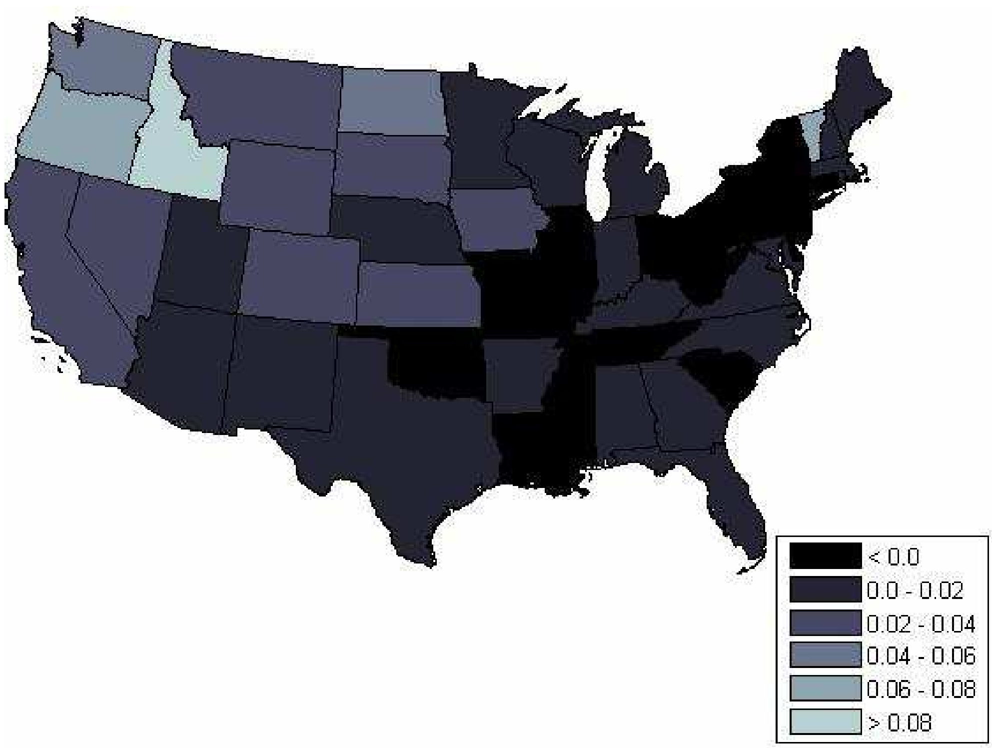

Figure 2: Granted patent per 100,000 inhabitants across the US (1994-2004)

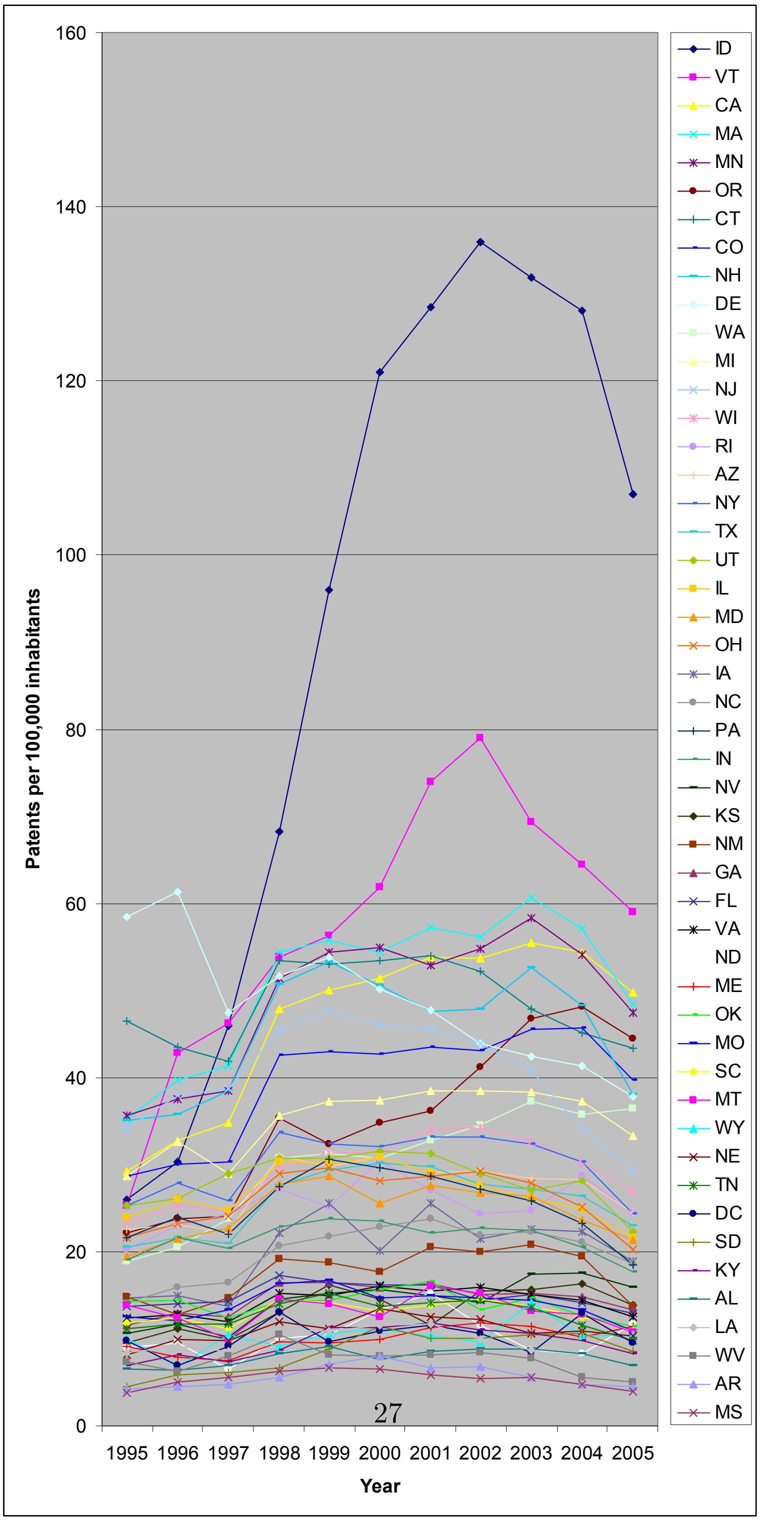

Figure 3: Estimation results for the random effects (missing values are not significant at the 95% HPDI)