

Co, Catherine Y.; Landon-Lane, John S.; Yun, Myeong-Su

Working Paper

Inter-state dynamics of invention activities, 1930 - 2000

Working Paper, No. 2005-11

Provided in Cooperation with:

Department of Economics, Rutgers University

Suggested Citation: Co, Catherine Y.; Landon-Lane, John S.; Yun, Myeong-Su (2005) : Inter-state dynamics of invention activities, 1930 - 2000, Working Paper, No. 2005-11, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:

<https://hdl.handle.net/10419/31281>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Inter-State Dynamics of Invention Activities, 1930–2000*

Catherine Y. Co

Economics, University of Nebraska at Omaha

cco@mail.unomaha.edu

John S. Landon-Lane

Economics, Rutgers University

lane@economics.rutgers.edu

Myeong-Su Yun

Economics, Tulane University

msyun@tulane.edu

April 3, 2005

*Corresponding author: John Landon-Lane, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA. Email: lane@econ.rutgers.edu Telephone: (732) 932-7691 Fax: (732) 932-7416. We thank an anonymous referee whose comments and suggestions have helped to improve the paper.

Inter-State Dynamics of Invention Activities, 1930–2000

Abstract

We study the dynamics of the cross-section distribution of patents per capita for the 48 continental U.S. states from 1930 to 2000 using a discrete-state Markov chain. We test for and find evidence in favor of the (knowledge) convergence hypothesis as we find that the distribution of patents is converging to a limiting distribution that is significantly more concentrated than its initial distribution. States in the extreme are more mobile and are more likely to move to the middle than states in the middle of the cross-sectional distribution and the rate of convergence to the limiting distribution is “slow.”

Keywords: Patent, U.S. States, Convergence

JEL Classification: O31; O18; C32

1 Introduction and Background

Knowledge is assumed to flow instantly between economies in neoclassical growth theory. If technology gaps exist initially, they will close very quickly. That is, if convergence tests on knowledge rates (e.g., patenting rates) are conducted, unequivocal evidence in favor of knowledge convergence would be found (See Caniëls (2000) for a review). Several papers have studied the temporal and spatial evolution of knowledge rates using U.S. patent data and have found evidence consistent with knowledge convergence.¹ Those that formally test whether patent activities have converged use a cross-section approach (β -convergence), a time series approach (stochastic convergence), or a combination of both.²

Although previous studies have contributed to our understanding of the convergence of patent activities, their econometric methodologies have some shortcomings. For example, β -convergence tests use cross-section data and Quah (1993b, p.429) points out that a negative coefficient on initial income levels when cross-section data is used is “perfectly consistent with the absence of convergence....” Hence, cross-section tests are “ill-designed to analyze data where some [regions] are converging and others are not.” (Bernard and Durlauf (1996, p.167)) Time series tests (i.e., unit root tests), on the other hand, have low power so that they tend to lead to under rejection of the unit root null hypothesis. That is, they are biased against convergence. One major limitation common to both cross-section and time-series approaches is that they do not account for cross-economy relationships. As Quah (1996, p. 147) points out, “estimating individual time

¹See, e.g., Schmookler (1966), Sokoloff (1988), Varga (1999), Suarez-Villa (2000), Co (2002), Johnson and Brown (2002), and Co and Wohar (2004).

²For example, Co (2002) studies the evolution of invention activities in the 48 continental U.S. states between 1963 and 1997 using “industry” level data, and finds that within any given industry, states with low initial patent per capita values grow faster than those with high initial per capita values (β -convergence), implying catch-up by initially lagging states. However, according to Carlino and Mills (1993), convergence requires both β -convergence and stochastic convergence. Stochastic convergence suggests that shocks to relative patent per capita are temporary. Using Carlino and Mills (1993)’s definition of convergence, Co and Wohar (2004) find that part of the U.S. is converging (in the β - and stochastic sense) using unit root tests.

series models one for each state, say, to permit differences across Maine and Oklahoma (e.g., Carlino and Mills (1993)) leaves undetected the co-movements across states.”

In order to overcome the above mentioned shortcomings, Quah (1993a, 1993b) emphasizes the importance of studying the dynamics of the cross-section distribution of a relevant economic variable such as income per capita. He argues that the behavior of the cross-section distribution of a relevant variable over time is more informative and convergence requires a narrowing cross-sectional distribution. Following Quah’s suggestions, we study the dynamics of the cross-section distribution of patents per capita for the 48 continental U.S. states from 1930 to 2000 applying Bayesian techniques to a discrete state Markov chain model.³

Figure 1: Cross-Sectional Distribution of Patents per Capita: Deviations from the Mean

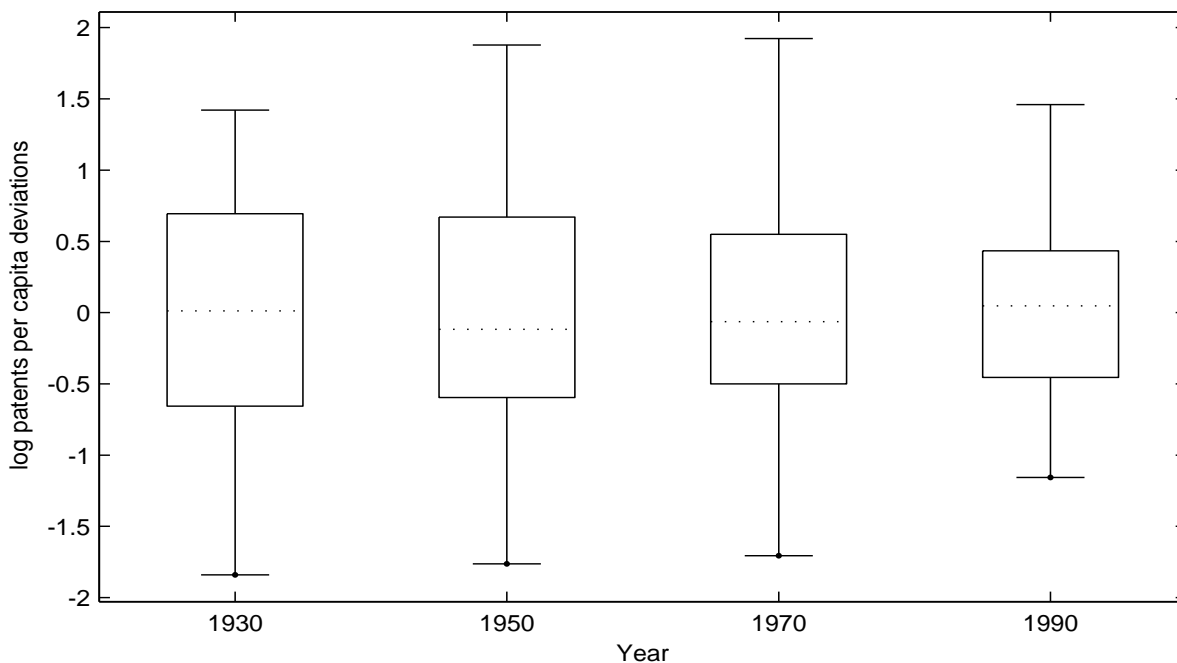


Figure 1 provides a quick overview of the cross-section distribution of log patent

³This paper also contributes to the debate on per capita income convergence since convergence (non-convergence) of inventive activities among U.S. states will promote the convergence (non-convergence) of per capita income among these areas.

per capita deviations for four years (1930, 1950, 1970 and 1990).⁴ For each year, the “box” represents the interquartile range while the “dotted” line represents the median of the data across the 48 states in our sample for that year. The “whiskers” of the boxplot depict the minimum and maximum observation for that year. It is clear from this figure that the distribution of log patents per capita, around its mean, is becoming more concentrated as time progresses. Evidence of this can be seen in the fact that the interquartile range (the length of the solid “box”) is getting smaller over time.

While this is suggestive of convergence, we employ formal testing strategies to examine whether invention activities across U.S. states are converging. In this paper we discuss two forms of convergence. First, we test for σ -convergence.⁵ That is, we test to see whether the cross-sectional distribution of log patents is becoming more concentrated over time. The statistic we use to measure concentration is the standard deviation of the cross-sectional distribution. Second, we characterize the convergence of the cross-sectional distribution of log patents to its limiting cross-sectional distribution. This approach is similar similar to that implemented by Quah (1993a, 1993b) and Johnson (2000).⁶ The intuition behind our tests is as follows: Invention activities across U.S. states are converging, in the σ -convergence sense, if the dispersion of the cross-section distribution of patents per capita diminishes over time. Thus, evidence of σ -convergence would indicate that the disparity in innovation activity would be expected to decrease over time.

Testing for both forms of convergence involves a comparison of the initial and

⁴A few remarks on the data we use in the analysis. First, we use patents per capita to refer to patents per 100,000 inhabitants. Second, we take a log transformation of patents per capita since the data show that the cross-section distribution is fat-tailed. Third, we subtract the average from the logged values in order to abstract from any trend or cyclical components. We call this transformed variable log patent per capita deviations hereafter.

⁵See Durlauf, Johnson and Temple (forthcoming) for an excellent discussion of the statistical and economical meanings of convergence.

⁶Quah (1993a, 1993b) uses Markov chains to study the evolution of the log relative income per capita distribution using data for 118 countries between 1962 and 1985. He finds no evidence of convergence. On the other hand, Johnson (2000) finds evidence in favor of convergence when he studies the evolution of the distribution of the log relative income per capita of the 48 continental U.S. states in 1948, 1963, 1978 and 1993.

limiting cross-sectional distribution of log patent per capita. The limiting distribution is obtained using a discrete state Markov chain model. We employ Bayesian methods to estimate a Markov chain model as these methods allow for exact inference on all functions of the underlying parameters of the model, including the limiting distribution implied by the Markov chain. We find that the limiting distribution of log patent per capita deviations has a substantially narrower dispersion than the initial distribution. This implies that, although knowledge gaps across U.S. states exist, these knowledge gaps are getting smaller. That is, there is evidence of convergence in patent activities across U.S. states. However, the rate of convergence to the limiting distribution is “slow.” These results are quite robust to the length of the transition period and the number of discrete classifications the data is partitioned into.

We also find that there is some evidence that the cross-sectional distribution of log patents per capita contain multiple modes. This suggests that there may be “innovation” clubs. However, consistent with our finding of σ -convergence, we find that the modes are moving together over time.

Convergence, in the σ -convergence sense, also requires that there be significant cross-section mobility over short- and long-horizons (See, e.g., Quah (1992)). When we study the mobility of states in the patents per capita distribution, we find that states initially at the high (low) end of the patents per capita distribution have a greater probability of moving down (up) of the cross-section distribution. Thus, there is considerable evidence of “movement to the middle” in the cross-sectional distribution of patents per capita.

The rest of the paper proceeds as follows. The next section provides a detailed discussion of the empirical method used. Data and specifications for Markov chain model are discussed in Section 3. The results are presented and analyzed in Section 4. Finally, Section 5 provides some policy implications and concluding comments.

2 Empirical Methodology

Since the early work of Champernowne (1953) and Prais (1955), discrete state Markov chain models have been used to model the dynamics of cross-sectional income distribution and social standing over time. In testing for convergence of the cross-sectional patents per capita distribution, we are testing whether 1) the cross-sectional distribution of log patents per capita has attained its stationary state and 2) if not, we then test whether there is evidence that the cross-sectional distribution is becoming more concentrated around its mean over time. Closely related to the idea of convergence is mobility. If we find that the cross-sectional distribution is becoming more concentrated over time we would like to characterize how states are moving in the cross-sectional distribution. Is it just the states in the extreme of the distribution that are moving or is the pattern of mobility common across all initial levels of patents per capita?

The discrete-state Markov chain model described in Champernowne (1953) is a natural way to answer the questions stated above. In the discrete-state Markov chain model, the data is divided up into a finite number of mutually exclusive classifications. According to Champernowne (1953), these classifications should be equal in length except for, possibly, the first and last classification which can be open sets that cover the tails of the distribution. The choice of the total number of classifications is somewhat dependent on the number of cross-section observations in the data. The size of the classifications should be small enough to allow the cross-sectional distribution to be studied in some detail but not so small so as to make the number of observations in each classification too small.⁷

Let y_{it} be the natural log of patents per capita (in deviation from the mean) for

⁷Note that we choose to break data into a finite number of discrete absolute classes rather than using quantiles. In order to test for convergence we compare the limiting cross-sectional distribution with the initial cross-sectional distribution. If we use quantiles then the cross-sectional distribution, by definition, is in its stationary state so that the limiting cross-sectional distribution would be identical to the initial cross-sectional distribution. Thus breaking the data into quantiles is inappropriate for what we are trying to test for.

U.S. state i in period t . Let $\mathcal{C} = \{c_1, \dots, c_M\}$ be the set of classifications that cover the interval $(-\infty, \infty)$; the range of possible values that y_{it} can take. For $j = 2, \dots, M - 1$ the classification c_j is defined as

$$c_j = (c_{lj}, c_{uj}], \quad (1)$$

where $c_{lj} = c_{uj-1}$. For $j = 1$ and $j = M$ the classification definitions are

$$\begin{aligned} c_1 &= (-\infty, c_{1u}] \\ c_M &= (c_{Ml}, \infty) \end{aligned} \quad (2)$$

respectively. The upper and lower limits for each classification definition are chosen so that the interval length, $\ell_j = c_{uj} - c_{lj}$, is constant for $j = 2, \dots, M - 1$.

Let $s_{it} \in \{1, \dots, M\}$ represent the classification that state i is a member of in period t . That is, $s_{it} = k$ if and only if $y_{it} \in c_k$. Finally, let π_{jt} be the unconditional probability that $s_{it} = j$, that is

$$\pi_{jt} = Pr [s_{it} = j] \equiv Pr [y_{it} \in c_j]. \quad (3)$$

Therefore, the distribution of log patent per capita deviations across states in period t can be summarized by

$$\pi_t = (\pi_{1t}, \dots, \pi_{Mt})', \quad (4)$$

which is the unconditional probability that a state is a member of one of the M classifications.

In order to test for convergence we need to describe the dynamics of π_t over time. We do this by assuming that π_t follows a Markov process. That is, we assume that the probability that a state is a member of classification j in period t , for example, is a function of what classification it was a member of in previous periods. In particular, we

assume that the probability distribution, π_t , follows a first-order Markov chain.⁸

The first-order Markov chain assumption implies that

$$Pr [s_{it}|s_{it-1}, s_{it-2}, \dots, s_{it-k}] = Pr [s_{it}|s_{it-1}] \quad k = 2, 3, \dots \quad (5)$$

Therefore, the probability that a state is in classification j in period t is only a function of what classification it was in the previous period. Let

$$p_{jk} = Pr [s_{it} = k | s_{it-1} = j] \quad (6)$$

be the probability that a state moves from classification j in period $t - 1$ to classification k in period t . Then the first-order Markov assumption implies that, for any state i in our sample,

$$\pi_{kt} = Pr [s_{it} = k] = \sum_{j=1}^M Pr [s_{it} = k | s_{it-1} = j] Pr [s_{it-1} = j] = \sum_{j=1}^M p_{jk} \pi_{jt-1}. \quad (7)$$

Therefore, the cross-section probability distribution, π_t , evolves over time according to

$$\pi'_t = \pi'_{t-1} P, \quad (8)$$

where $P = [p_{jk}]$ is the probability transition matrix associated with the Markov chain.

The transition probability matrix, P , has a number of interesting properties. First, P is row stochastic. That is, each row of P sums to one. This implies that all eigenvalues of P have modulus less than or equal to 1. Second, at least one eigenvalue is exactly equal to 1, and the left eigenvector associated with the eigenvalue of 1 is the limiting distribution of the Markov chain given by (8). Furthermore, if there is only one eigenvalue

⁸The first-order assumption follows from our definition of the transition period to be 10 years in length. In this case, we are assuming that the level of patents per capita in a state today is unaffected by the level of patents per capita in the same state 20 or more years ago.

of P equal to 1, the limiting distribution is unique (Geweke, Marshall and Zarkin 1986).

Consider the Markov chain given in (8) and let π_0 be the initial probability distribution of log patent per capita deviations. Then, applying (8) t times, we see that the probability distribution in period t is

$$\pi'_t = \pi'_0 P^t. \quad (9)$$

Then the limiting distribution, π^* is

$$\pi^{*'} = \lim_{t \rightarrow \infty} \pi'_t = \lim_{t \rightarrow \infty} \pi'_0 P^t. \quad (10)$$

The limiting distribution is also referred to as the ergodic distribution or the invariant distribution and satisfies the following relationship:

$$\pi^{*'} = \pi^{*'} P. \quad (11)$$

2.1 Exact Inference in the First Order Markov Model

We are interested in estimating the unknown parameters of the first order discrete state Markov chain model described in (8). In particular, we are interested in estimating π_0 , the initial probability distribution of log patents per capita across the states, and P , the probability transition matrix. All other functions of interest, such as the limiting distribution and measures of convergence and mobility are just functions of these two parameter vectors with some of these functions being highly non-linear functions of the primal parameter vectors.

In order to perform statistical tests of our various hypotheses (defined later) we need to be able to characterize the distribution of these non-linear functions of interest. There are a number of ways of estimating (8). While maximum likelihood estimates are easy to calculate, obtaining the variance of these estimates and the variance of non-linear

functions of these estimates is difficult (Geweke et al. 1986). In a recent study Buchinsky and Hunt (1999) used the bootstrap to estimate the variances. A simpler approach is to use Bayesian methods described in Geweke et al. (1986). Using Bayesian methods we are able to characterize the full distribution of all functions of interest, including highly non-linear functions of the primal parameter vector of our model. We are also able to include any prior information we have available to us.

Let $Y_{NT} = \{\{y_{it}\}_{i=1}^N\}_{t=0}^T$ represent the data that is observed for N states across T time periods. Let $S_{NT} = \{\{s_{it}\}_{i=1}^N\}_{t=0}^T$ represent the classifications that is observed for the N states across T periods. The information contained in S_{NT} is all that is needed from Y_{NT} to estimate (8). For each state i , we observe the initial classification that it is a member of and all subsequent classifications it is a member of for $t = 1, \dots, T$. The probability of observing this is

$$Pr[\{s_{it}\}_{t=0}^T | \pi_0, P] = \pi_{s_{i0}} \prod_{t=1}^T p_{s_{it-1}s_{it}}, \quad (12)$$

so that the probability of observing the classification history of all N states is

$$Pr[S_{NT} | \pi_0, P] = \prod_{i=1}^N \pi_{s_{i0}} \prod_{t=1}^T p_{s_{it-1}s_{it}}. \quad (13)$$

Let n_{j0}^0 represent the number of states in classification j in the initial period and let n_{jk} represent the total number of times a state is observed transiting from classification j to classification k . Then $n^0 = (n_{10}^0, \dots, n_{M0}^0)'$ and $\mathcal{N} = [n_{jk}]$ are sufficient statistics so that the likelihood function (data density) is

$$p(S_{NT} | \pi_0, P) \propto \prod_{j=1}^M \pi_{0j}^{n_{j0}^0} \prod_{k=1}^M p_{jk}^{n_{jk}}. \quad (14)$$

Given the form of this kernel, a natural conjugate prior for the parameters $\theta = (\pi_0, P)'$

is

$$\begin{aligned} (\pi_{01}, \dots, \pi_{0M}) &\sim Di_M(a_0) \\ (p_{j1}, \dots, p_{jM}) &\sim Di_M(A(j)) \quad j = 1, \dots, M, \end{aligned} \tag{15}$$

where $Di_M(a)$ refers to a multivariate-Beta (Dirichlet) of order $M - 1$ indexed by the parameter vector $a = (a_1, \dots, a_M)$. The multivariate-Beta distribution of order $M - 1$ has density

$$p(x|a) = \frac{\Gamma\left(\sum_{j=1}^M a_j\right)}{\prod_{j=1}^M \Gamma(a_j)} \prod_{j=1}^M x_j^{(a_j-1)} \tag{16}$$

where $a_j > 0$ for all $j = 1, \dots, M$ and $x \in \{x : x_j > 0 \ (j = 1, \dots, M), \sum_{j=1}^M x_j = 1\}$.

The prior for the initial cross-sectional distribution, π_0 , and for each row of P , $P(j)$, is multivariate-Beta of order $M - 1$ indexed by the parameter vectors a_0 and $A(j) = (A_{j1}, \dots, A_{jM})$ respectively. Thus the prior is a product of $(M + 1)$ independent multivariate-Beta distributions. The posterior density can easily be shown to be the product of $(M + 1)$ multivariate-Beta densities with

$$\begin{aligned} \pi_0|S_{NT} &\sim Di_M(a_0 + n^0) \\ (p_{j1}, \dots, p_{jM})|S_{NT} &\sim Di_M(A_j + \mathcal{N}_j) \text{ for } j = 1, \dots, M. \end{aligned} \tag{17}$$

The parameters indexing the posterior density are $a_0 + n^0$ for the component of the posterior coming from the initial distribution and $A + \mathcal{N}$ for the component coming from the probability transition matrix, P .

Using the method described in Devroye (1986, pp 593-596) it is easy to make independent and identically distributed (i.i.d.) draws from the posterior distribution $p((\pi_0, P)|S_{NT})$. Given that we can obtain i.i.d. draws from the posterior, $p((\pi_0, P)|S_{NT})$, it is then simple to exactly characterize the posterior distribution of any (well-defined) function of interest.

Let $\theta = (\pi'_0, \text{vec}(P)')'$ and let $\{\theta^k\}_{k=1}^K$ be i.i.d. draws from $p(\theta|S_{NT})$. Then for

any function, $g(\theta)$ such that

$$E(g(\theta)|S_{NT}) = \bar{g} < \infty, \text{ and } E((g(\theta) - \bar{g})^2|S_{NT}) = \sigma_g^2 < \infty, \quad (18)$$

it follows from the strong law of large numbers that

$$\begin{aligned} K^{-1} \sum_{k=1}^K g(\theta^k) &= \bar{g}_K \xrightarrow{a.s.} \bar{g} \text{ and} \\ K^{-1} \sum_{k=1}^K (g(\theta) - \bar{g}_K)^2 &\xrightarrow{a.s.} \sigma_g^2. \end{aligned} \quad (19)$$

Thus we are able to fully characterize the distribution of any (well-behaved) function of the underlying parameters that we are interested in. Details pertaining to a number of these functions appear in the next section.

2.2 Functions of Interest

Apart from characterizing the posterior distribution of π_0 and P , we are mainly interested in the following functions. The first function of interest is the limiting distribution implied by P . The limiting distribution of the first-order discrete state Markov chain satisfies (11) and is calculated as the left normalized eigenvector associated with the eigenvalue equal to 1.

Let s_0 be the random variable that represents the classification a state is a member of in the initial period and let s_∞ be the random variable that represents the classification that a state is a member of in the limiting distribution. Then the measure of concentration that we use is the standard deviation, $\sigma(s)$, of the cross sectional distribution of s . For a given discrete probability distribution, π , the standard deviation is

$$\sigma(s) = \sqrt{Var(s)} = \sqrt{\sum_{k=1}^M (k - E(s))^2 \pi_k}, \quad (20)$$

where

$$E(s) = \sum_{k=1}^M k\pi_k. \quad (21)$$

Using (20) we find evidence of convergence (whether it be absolute or conditional) if

$$\sigma(s_\infty) < \sigma(s_0). \quad (22)$$

Thus, a function of interest that is useful for testing for convergence is

$$g(\pi_0, P) = \sigma(s_\infty) - \sigma(s_0). \quad (23)$$

Apart from testing for convergence in the cross-sectional distribution, π , we are also interested in measures of mobility since these measures may shed light on how convergence is achieved. These mobility measures are functions of the transition probability matrix, P . Shorrocks (1978) and Geweke et al. (1986) describe in detail a number of important mobility indices for Markov chain models. We report two sets of mobility indices in this paper.

The first set of mobility indices are aggregate measures of mobility. Shorrocks (1978) defines

$$\mathcal{M}_S = \frac{M - \text{tr}(P)}{M - 1}, \quad (24)$$

which measures overall mobility. Aggregate measures of upward and downward mobility are

$$\mathcal{M}_U = (M - 1)^{-1} \sum_{k=1}^{M-1} \mathcal{M}_{pU}(k), \quad (25)$$

and

$$\mathcal{M}_D = (M - 1)^{-1} \sum_{k=2}^M \mathcal{M}_{pD}(k). \quad (26)$$

Gang, Landon-Lane and Yun (2004) show that Shorrocks' measure can be decomposed into its upward and downward components. That is, $\mathcal{M}_S = \mathcal{M}_U + \mathcal{M}_D$.

The second set of indices report the probability of moving up or down the distribution conditional on the current class. These indices are:

$$\mathcal{M}_U(j) = \sum_{k=j+1}^M p_{jk} \quad (27)$$

and

$$\mathcal{M}_D(j) = \sum_{k=1}^{j-1} p_{jk}. \quad (28)$$

These two indices describe the probability of moving to a higher (lower) classification in the next period given the state is in classification j this period.

Finally, we are also interested in the speed of convergence to the limiting distribution. Geweke et al. (1986) shows that the modulus of the second largest eigenvalue of P is an upper bound on the rate of convergence to the limiting distribution of the Markov chain. Let $\{\lambda_1, \dots, \lambda_M\}$ be the set of eigenvalues of P such that $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_M|$. Then, $|\lambda_2|$ is an upper bound to the rate of convergence of π_t to π^* and

$$h = \frac{-\log 2}{\log |\lambda_2|} \quad (29)$$

is the half-life of the Markov chain.

3 Data and Specifications for Modeling

3.1 Data

We use utility patent count data from 1930 to 2000.⁹ According to the U.S. Patent and Trademark Office (USPTO), utility patents are patents “issued for the invention of a new and useful process, machine, manufacture, or composition of matter, or a new and useful improvement thereof...” Patents are granted to inventors (who can then assign or license the patent) and the data are tabulated using the reported state residence of the inventors. In cases where there are multiple inventors, the residence of the first-named inventor is used by the USPTO to allocate each patent’s geographic origin.

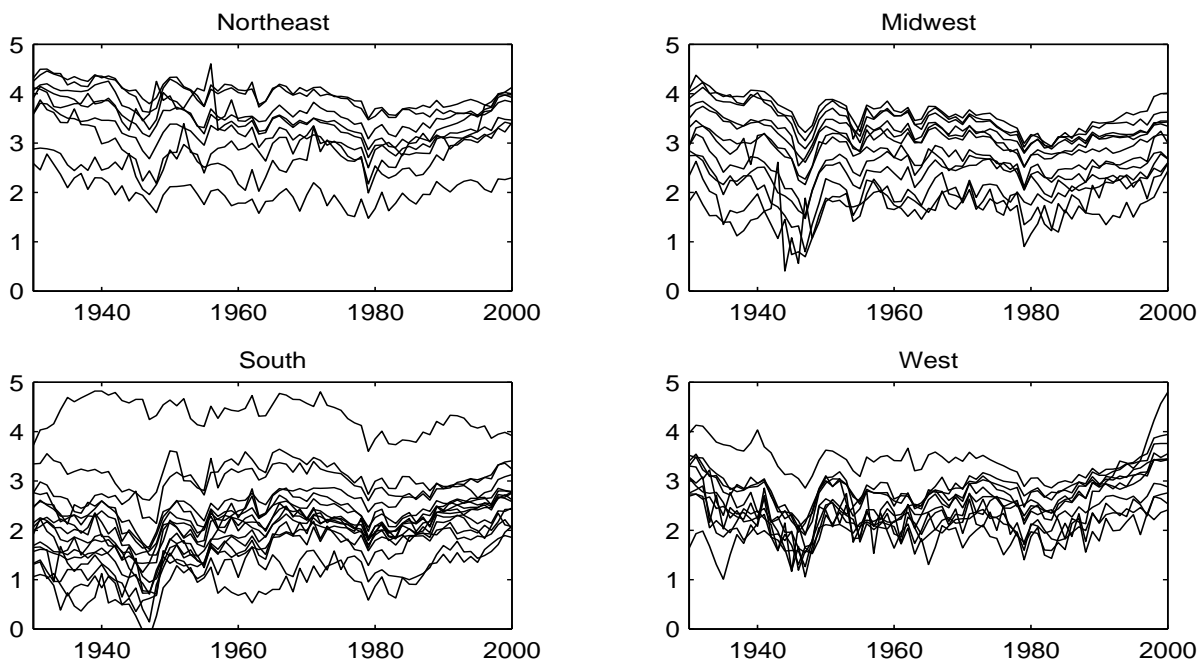
The patent count data are adjusted using state population. Ideally, total patents need to be adjusted by the number of R&D scientists and engineers in the state; however, the geographical distribution of the number of doctoral scientists and engineers by state is not available annually and, as far as we know, available data series (at the state level) started only in the late 1970s. The use of patent per 100,000 inhabitants is appropriate: Co (2002) reports that the correlations between state patents per 1,000 R&D scientist and engineer and state patents per 100,000 inhabitants for periods when data are available are around 0.70.

Figure 2 depicts the patents per capita data for each state. Each line represents the natural log of patents per capita for a state. Though the data do not have apparent linear time trend, there is a distinctive non-linear pattern. For example, during the 1940’s and again during the late 1970’s there appears to be a significant drop in patents per

⁹Patent counts prior to 1963 include utility, design and plant patents. Data from 1930 to 1962 are from the U.S. Patent and Trademark Office, *Technology Assessment and Forecast, Seventh Report*. U.S. Department of Commerce, Arlington, VA: U.S. Dept. of Commerce, Patent and Trademark Office, Office of Technology Assessment and Forecast, March 1977. Data from 1963 to 1998 are from the U.S. Patent and Trademark Office, Information Product Division, *Patenting Trends in the U.S., 1998, State/Country Report- All Years, 1963-1998*. CD-Rom issued July 2000. Data for 1999 and 2000 are from the U.S. Patent and Trademark Office, Patent Counts by Country/State and Year, Utility Patents, January 1, 1963 - December 31, 2001.

capita across all states in the sample.¹⁰

Figure 2: Log Patents per Capita: By Region



The national sample mean of patents per capita for each period is depicted in Figure 3. It is clear from this figure that there is a cycle component in the patent data. To focus on the cross-sectional distribution we abstract from trends and/or cyclical components in the data by subtracting from each period the national sample mean patents per capita. The resulting patents per capita in deviations from mean form is depicted in Figure 4. These are the data we use in the subsequent empirical analysis.

¹⁰This analysis does not attempt to explain what causes this aggregate behavior but rather we attempt to characterize the dynamics of the patent activities around the aggregate behavior. See, e.g., Kortum and Lerner (1998) for a recent study of the aggregate behavior of patenting.

Figure 3: Mean Log Patents per Capita

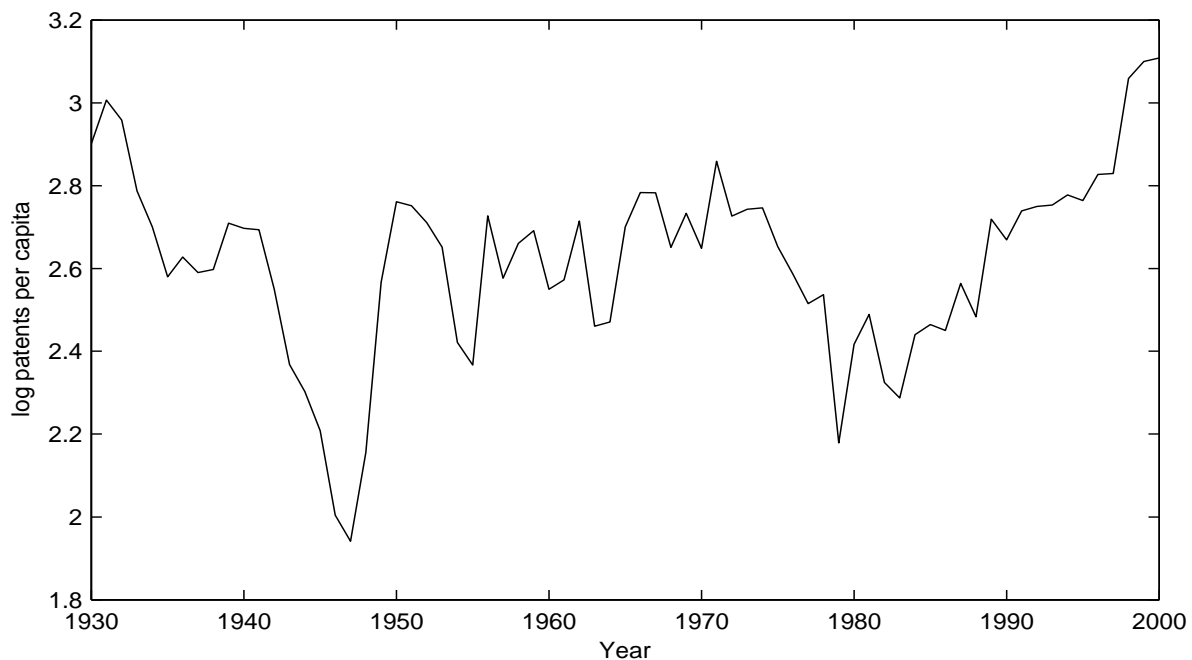
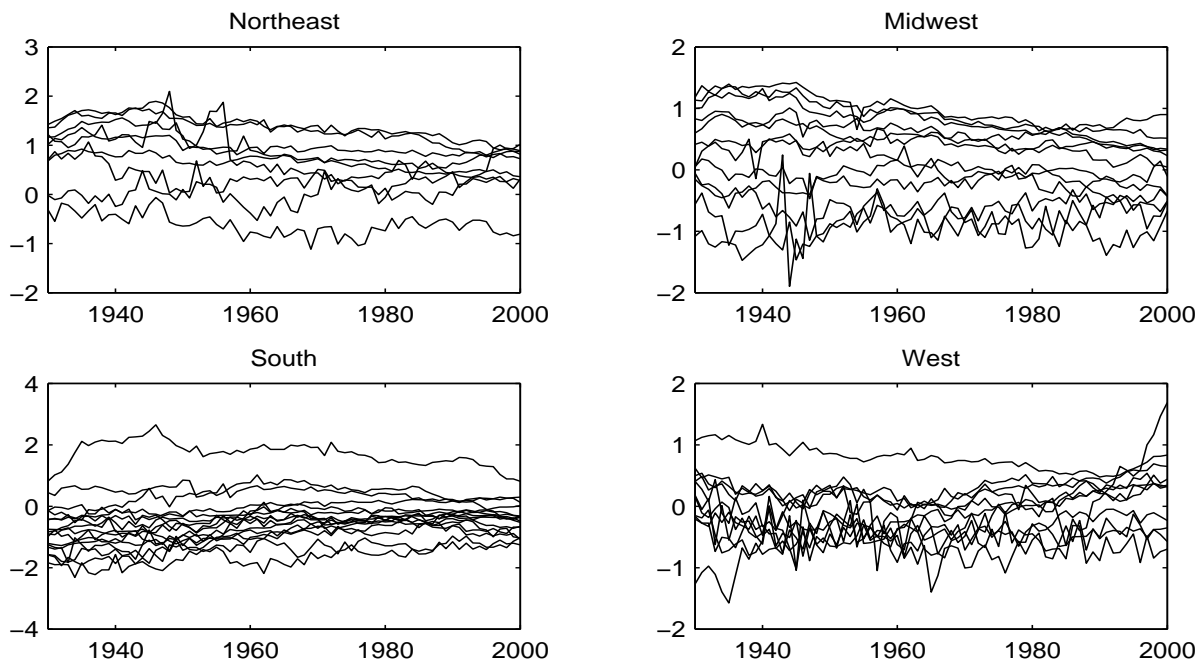


Figure 4: Log Patents per Capita Deviations: By Region



3.2 Definition of Patent Classifications and Transition Period

We use log patent per capita deviations, depicted in Figure 4. By definition, the mean of the data is 0 for each period. For this paper we divide the data in 9 mutually exclusive classifications. The classification definitions are described in Table 1. The middle 7 classifications have a common length with the middle classification, classification number 5, containing the value 0. Therefore, for example, $\pi_{2,1930}$ is the unconditional probability that the value of the log patent per capita deviations is between -1.3 and -0.93 from the mean of the distribution in 1930. The classifications used divide the interval $[-1.3, 1.3]$ into seven equally spaced intervals. The first and last interval cover all observations lower than -1.3 and higher than 1.3 respectively. This interval is chosen because approximately 90% of all observations (across states and across time) fall between -1.3 and 1.3.

Table 1: Classification Definitions

Classification Number	Lower Bound	Upper Bound
1	-5	-1.3
2	-1.3	-0.92857
3	-0.92857	-0.55714
4	-0.55714	-0.18571
5	-0.18571	0.18572
6	0.18572	0.55715
7	0.55715	0.92857
8	0.92857	1.3
9	1.3	5

The transition period is also an important component of the model definition. The shorter the transition period the less time a state has to change its patent classification. For short transition periods we would not expect to observe too many transitions of more than one classification. In this case we would expect to see a lot of 0's off the main and first diagonals in the data transition matrix, \mathcal{N} . A longer transition period would allow states to adjust their policies and incentives and thus we would expect to see more

transitions. However, the longer the transition period the less the number of transition periods we would observe.

The choice of the transition length needs to take into account several lags: first, it takes time for policy makers to change and adopt new R&D policies (policy lag); second, inventors may respond to policies with some lag (response lag); third, there may be delays between the conduct of R&D and the filing of patents (application lag); fourth, review of the patent application by the USPTO suggests a lag in the granting of patents (grant lag).

Policy lag (at the state level) does not play a crucial role in our choice of the transition length. This is because for the majority of the time period covered, U.S. science and technology (S&T) policy is mostly conducted at the federal level. For example, the first initiative to form S&T advisory bodies for state governors only started in the 1960s (See, e.g., Carnegie Commission on Science, Technology and Government (1992) for details.) It is reasonable to suppose that the policy lag (at the federal level) is related to the political cycle.¹¹

If there is no response lag from inventors, and since there appears to be no significant lag between R&D and the filing of patents (See, e.g. Hall, Griliches and Hausman (1986)), the transition length should at the minimum allow for a two-year grant lag. Popp, Juhl and Johnson (2003) find a mean (median) lag of 28 months (23 months) from application to grant using U.S. patents granted between 1976 and 1996.

Taking all the above into account, we believe that a 10-year transition period is appropriate. This give us observations on 7 ten-year transitions. As a robustness test we

¹¹We should note that the time it takes for legislations to pass Congress and be signed by the President varies substantially; it partly depends on a legislation's comprehensiveness and controversy surrounding it (from the authors' conversation with a legal expert). For example, days after his re-election in November 1944 President Franklin Roosevelt asked his *de facto* science advisor, Vannevar Bush, to explore how lessons learned from the organization of science and engineering during the Second World War can be used in peacetime. Bush issued his report in July 1945; by the end of July, Senator Magnuson introduced legislation to implement the recommendations contained in the Bush report. Debates to reconcile this legislation with an earlier legislation introduced in February 1944 by Senator Kilgore ensued between 1945 and 1950. The National Science Foundation was finally created in May 1950. See National Science Board (2000) for details.

also conduct similar analysis for 1, 5, 20 and 35-year transition periods. For the most part, the qualitative nature of the results are similar to those we report below.

3.3 Prior Distributions

We define natural conjugate priors for the parameters π_0 and P . In fact we define $M + 1$ independent prior distributions for π_0 and the M rows of P . Each prior has the same form. The general form of the prior for π_0 is

$$p(\pi_0) = Di_{M-1}(\pi_0|a_0), \quad (30)$$

and the general form of the prior for the j^{th} row of P , $P_{j\cdot}$, is

$$p(P_{j\cdot}) = Di_{M-1}(P_{j\cdot}|A_{j\cdot}), \quad (31)$$

where a random vector, $x = (x_1, \dots, x_M) \sim Di_{M-1}(x|\alpha)$ if

$$x \in \{x : 0 < x_i < 1 \text{ and } \sum_{i=1}^M x_i = 1\}$$

$$\alpha = (\alpha_1, \dots, \alpha_M), \quad \alpha_i > 0 \quad i = 1, 2, \dots, M \text{ and} \quad (32)$$

$$p(x) = \frac{\Gamma\left(\sum_{i=1}^M \alpha_i\right)}{\prod_{i=1}^M \Gamma(\alpha_i)} x_1^{\alpha_1-1} \dots x_M^{\alpha_M-1}.$$

Bernado and Smith (1994, pp 134-135) show that the mean and variance of each element of x is

$$E(x_i) = \frac{\alpha_i}{\sum_{k=1}^M \alpha_k} \quad (33)$$

$$Var(x_i) = \frac{E(x_i)(1 - E(x_i))}{1 + \sum_{k=1}^M \alpha_k}.$$

The prior in this context has two important uses. First, it allows us to explicitly state our prior beliefs about the parameters of the model. Second, it allows us to “fill in” zero elements of the data transition matrix, \mathcal{N} . In our application, with classification definitions given in Table 1, we do not observe any transition from classification 1 to classification 9 in any single period. Hence, the data transition matrix, \mathcal{N} , has a zero in the 9th column of the 1st row. In the definition of the Dirichlet distribution of order $M - 1$ given in (32) we see that the parameter vector that indexes the distribution cannot contain any zero elements. The posterior distribution, given in (17), is the product of $M + 1$ Dirichlet distribution where the posterior of the j^{th} row of P , $P(j)$, is indexed by $A(j) + \mathcal{N}(j)$, the sum of the j^{th} rows of A and \mathcal{N} . Thus, as long as the prior parameter, $A(j)$, does not contain any 0’s, the posterior parameter will not contain any zero elements.

The prior parameter vectors used in this analysis are chosen to produce the prior moments reported in Table 2. This prior is chosen to represent a large degree of prior uncertainty about the true values of π_0 and P . The prior for the initial distribution, π_0 is symmetric around the middle classification with the highest prior probability being placed on a state being in classification 5, the classification containing 0. The lowest prior probability is placed on the extreme classifications, 1 and 9, and the probability declines exponentially from the middle to the extreme classifications. However, the prior standard deviation is quite large for each parameter of π_0 reflecting our prior uncertainty towards the true values for π_0 . The prior for P is symmetric around the main diagonal. That is, our prior is that the most likely event is that a state will stay in its current classification. The prior means also reflect our prior belief that the probability of moving from the current classification to another classification is lower the larger the move. Again, the prior standard deviations are quite large, reflecting our prior uncertainty over the true values of P . This prior specification is designed to have minimal impact on the posterior distribution so that the results are determined by the data.

Table 2: Prior Moments

	Classification								
	1	2	3	4	5	6	7	8	9
π_0	0.018 (0.105)	0.039 (0.151)	0.083 (0.215)	0.175 (0.297)	0.370 (0.377)	0.175 (0.297)	0.083 (0.215)	0.039 (0.151)	0.018 (0.105)
	0.528 (0.415)	0.250 (0.360)	0.118 (0.268)	0.056 (0.191)	0.026 (0.133)	0.012 (0.092)	0.006 (0.063)	0.003 (0.044)	0.001 (0.030)
	0.200 (0.320)	0.423 (0.396)	0.200 (0.320)	0.094 (0.234)	0.045 (0.165)	0.021 (0.115)	0.010 (0.080)	0.005 (0.055)	0.002 (0.038)
	0.086 (0.222)	0.183 (0.305)	0.387 (0.384)	0.183 (0.305)	0.086 (0.222)	0.041 (0.156)	0.019 (0.108)	0.009 (0.075)	0.004 (0.052)
	0.039 (0.152)	0.083 (0.216)	0.177 (0.299)	0.374 (0.379)	0.177 (0.299)	0.083 (0.216)	0.039 (0.152)	0.019 (0.106)	0.009 (0.073)
P	0.018 (0.105)	0.039 (0.151)	0.083 (0.215)	0.175 (0.297)	0.370 (0.377)	0.175 (0.297)	0.083 (0.215)	0.039 (0.151)	0.018 (0.105)
	0.009 (0.073)	0.019 (0.106)	0.039 (0.152)	0.083 (0.216)	0.177 (0.299)	0.374 (0.379)	0.177 (0.299)	0.083 (0.216)	0.039 (0.152)
	0.004 (0.052)	0.009 (0.075)	0.019 (0.108)	0.041 (0.156)	0.086 (0.222)	0.183 (0.305)	0.387 (0.384)	0.183 (0.305)	0.086 (0.222)
	0.002 (0.038)	0.005 (0.055)	0.010 (0.080)	0.021 (0.115)	0.045 (0.165)	0.094 (0.234)	0.200 (0.320)	0.423 (0.396)	0.200 (0.320)
	0.001 (0.030)	0.003 (0.044)	0.006 (0.063)	0.012 (0.092)	0.026 (0.133)	0.056 (0.191)	0.118 (0.268)	0.250 (0.360)	0.528 (0.415)

Note: Numbers reported are prior means and, in parentheses, prior standard deviations.

4 Results

Using a transition period of ten years we divided the 48 states in our sample into the nine classifications given in Table 1. The length of all but the first and last classification is equal to 0.3714 in the log scale. These classifications are quite large and so there needs to be an appropriate amount of time to allow a state to transition from its current level of patents per capita. Thus we looked at patents per capita data for the states in our sample for the years 1930, 1940, 1950, 1960, 1970, 1980, 1990, and 2000. This gives us a total of seven transitions.

Before estimating the first-order Markov chain model in (8), we perform a number of diagnostic tests. We use Bayesian model comparison methods to perform exact model

comparison tests. These tests are performed using the posterior odds ratio. Consider two competing models, Ψ_1 and Ψ_2 . Then, conditional on the model and the observed sample, the posterior distribution can be written as

$$p(\theta|S_{NT}, \Psi_i) \propto p(\theta|\Psi_i)p(S_{NT}|\theta, \Psi_i), \quad (34)$$

for each model Ψ_i . The marginal likelihood is then defined as

$$p(\Psi_i|S_{NT}) = \int_{\Theta} p(\theta|S_{NT}, \Psi_i)d\theta, \quad (35)$$

and can be interpreted as the probability of the model conditional on the data. It can also be shown to represent the cumulative out-of-sample prediction performance of model Ψ_i on the observed data, S_{NT} .

Models are then compared via the Bayes factor, which is defined as

$$BF_{ij} = \frac{p(\Psi_i|S_{NT})}{p(\Psi_j|S_{NT})}. \quad (36)$$

A model, Ψ_i is “superior” to another model, Ψ_j , if the Bayes factor in favor of model Ψ_i over model Ψ_j , BF_{ij} is greater than 1. In this case, model Ψ_i is more likely than model Ψ_j conditional on the observed data. That is, $p(\Psi_i|S_{NT}) > p(\Psi_j|S_{NT})$. In what follows we actually report the log Bayes factor, $\log(BF_{ij})$, rather than the Bayes Factor. In this case, model i is superior to model j if $\log(BF_{ij}) > 0$.

4.1 Diagnostic Tests

We perform a number of diagnostic tests on our Markov chain model given in (8). Implicit in this formulation is that the transition probability matrix, P , is stable across the whole sample. An alternative model would be that there was a structural break and that the

model is

$$\pi'_t = \pi'_{t-1}P, \tag{37}$$

where

$$P = \begin{cases} P_1 & t \leq s \\ P_2 & t > s \end{cases}.$$

Another way to write this model is

$$\pi'_t = \begin{cases} \pi'_0 P_1 & t \leq s \\ \pi'_0 P_1^s P_2^{t-s} & t > s \end{cases}. \tag{38}$$

Thus, to test whether there is a structural break at period s we would estimate (38) under the restriction that $P_1 = P_2$ (the no break case) and without any restriction on P_1 and P_2 . The method of estimating (38) is a very simple extension to the method of estimating (8) described in Section 2.1.

Table 3 contains the log Bayes factors in favor of the structural break model over the no-structural break model for each period of the sample. As there is an exact analytical expression for the log marginal likelihood for the Markov chain model, the log Bayes factors are not reported with any standard errors. The log Bayes factor is strongly in favor of the no break model for all periods. Thus, there is no evidence of instability of P in the Markov chain over the sample period.

Another important assumption of (8) is the assumption of the non-stationarity of the Markov chain. That is, (8) assumes that the limiting distribution, π^* is different to π_0 . This test of whether the data is in its limiting distribution is fundamental to testing for evidence of convergence. The stationary model is (8) with the restriction that $\pi_0 = \pi^*$. Table 3 contains the log Bayes factor in favor of the stationary finite state Markov chain model over the non-stationary (no-break) Markov chain model. The log Bayes factor in favor of the stationary model over the non-stationary model is -8.78 which

Table 3: Diagnostic Tests for Markov Chain Model

Test	log Bayes factor	Std. error
Break in 1940	-281.392	–
Break in 1950	-317.857	–
Break in 1960	-326.764	–
Break in 1970	-331.013	–
Break in 1980	-321.259	–
Break in 1990	-286.657	–
Stationary Model	-8.78	(0.13)

Note: In all cases the model in the denominator of the Bayes factor is the non-stationary first-order Markov chain model given in (8).

implies that the stationary Markov-chain model is inferior to the non-stationary Markov chain model. The standard error reported in Table 3 is the numerical standard error made in calculating the log marginal likelihood of the stationary model.¹² Therefore, there is evidence to suggest that the cross-sectional distribution of patents per capita across the states of the U.S. is not in its limiting or stationary state. The results in this section leads us to estimate (8) in order to fully characterize the convergence (divergence) properties of the patents per capita data.

4.2 Testing for σ -Convergence in State Patent Data

To test for σ -convergence we estimate (8) and use the estimates to characterize the posterior distribution of the initial cross-sectional distribution of patents, π_{1930} , and the limiting distribution π^* , implied by the estimated transition probability matrix, P , for our sample.

¹²Unlike the non-stationary first order Markov chain model given in (8), the posterior distribution of the stationary first-order Markov chain model is not a known distribution so we cannot make i.i.d. draws from this posterior. Draws from the posterior distribution of the stationary first-order Markov chain model are made using Markov Chain Monte Carlo methods. The marginal likelihood is then numerically approximated using the method of Gelfand and Dey (1994).

The observed cross-sectional distribution for 1930 and the observed data transition matrix for our sample can be found in Table 4. We see that most of the observed transitions are from one classification to an adjacent classification. There are very few movements of two or more classifications. This observed data is then combined with the priors defined in Table 2. The prior has the effect of filling in the 0's in the data transition matrix with positive, albeit small, values.

Table 4: Summary of Data: Patents (1930-2000)

	Classification								
	1	2	3	4	5	6	7	8	9
n_{1930}^0	4	5	4	4	11	5	6	7	2
	10	4	0	0	0	0	0	0	0
	0	15	9	2	0	0	0	0	0
	0	4	22	15	0	0	0	0	0
	0	2	9	35	14	2	0	0	0
\mathcal{N}	0	0	3	16	28	10	0	0	1
	0	0	0	2	9	30	6	1	0
	0	0	0	0	0	13	29	2	1
	0	0	0	0	0	0	9	10	4
	0	0	0	0	0	0	3	4	12

The data combined with the priors are used to estimate the model. Posterior moments of π_{1930} , the initial distribution, P , the probability transition matrix, and π^* , the invariant or limiting distribution implied by P are reported in Table 5.

It is clear that the limiting distribution is different from the initial distribution. There appears to have been a “shift left” of the distribution of patents. The probability of being a member of classifications 3 and 4 are significantly higher in the limiting distribution (at 0.144 and 0.258, respectively) than in the initial distribution (at 0.083 and 0.085, respectively). Also, the probability of being a member of one of the two extreme classifications is lower in the limiting distribution, π^* , than in the initial distribution, π_0 .

Table 5: Posterior Moments of Markov chain model of Patents

	Classification								
	1	2	3	4	5	6	7	8	9
π_{1930}	0.082 (0.039)	0.103 (0.043)	0.083 (0.040)	0.085 (0.039)	0.230 (0.060)	0.105 (0.044)	0.125 (0.047)	0.145 (0.050)	0.041 (0.028)
	0.706 (0.117)	0.287 (0.116)	0.004 (0.016)	0.002 (0.010)	0.001 (0.008)	0.000 (0.003)	0.000 (0.003)	0.000 (0.002)	0.000 (0.001)
	0.004 (0.013)	0.572 (0.094)	0.344 (0.090)	0.078 (0.051)	0.001 (0.005)	0.000 (0.004)	0.000 (0.003)	0.000 (0.002)	0.000 (0.001)
	0.001 (0.006)	0.099 (0.046)	0.535 (0.077)	0.362 (0.074)	0.001 (0.005)	0.001 (0.004)	0.000 (0.002)	0.000 (0.001)	0.000 (0.001)
	0.000 (0.002)	0.033 (0.023)	0.145 (0.044)	0.562 (0.062)	0.226 (0.053)	0.033 (0.023)	0.000 (0.003)	0.000 (0.002)	0.000 (0.001)
P	0.000 (0.002)	0.000 (0.003)	0.052 (0.029)	0.275 (0.058)	0.482 (0.065)	0.172 (0.049)	0.001 (0.004)	0.000 (0.002)	0.017 (0.017)
	0.000 (0.002)	0.000 (0.002)	0.001 (0.004)	0.043 (0.029)	0.187 (0.055)	0.622 (0.068)	0.125 (0.046)	0.021 (0.020)	0.001 (0.003)
	0.000 (0.001)	0.000 (0.002)	0.000 (0.002)	0.001 (0.003)	0.001 (0.005)	0.287 (0.066)	0.641 (0.071)	0.046 (0.031)	0.023 (0.022)
	0.000 (0.001)	0.000 (0.002)	0.000 (0.003)	0.001 (0.004)	0.001 (0.007)	0.002 (0.009)	0.387 (0.099)	0.435 (0.100)	0.174 (0.076)
	0.000 (0.001)	0.000 (0.001)	0.000 (0.002)	0.000 (0.005)	0.001 (0.005)	0.001 (0.008)	0.156 (0.080)	0.211 (0.090)	0.631 (0.108)
π^*	0.003 (0.007)	0.058 (0.031)	0.144 (0.047)	0.258 (0.056)	0.178 (0.039)	0.188 (0.053)	0.111 (0.049)	0.028 (0.018)	0.032 (0.023)

Note: Numbers reported are posterior mean and posterior standard deviation (in parentheses) using 10,000 draws from the posterior distribution.

This is most evident in classification 1 where the probability of membership drops from 0.082 to 0.003. There is also evidence of movement towards to center of the distribution at the upper end of the distribution. The probability of being a member of classification 8 and 9 falls from 0.186 in π_{1930} to 0.060 in π^* .

To formally test for σ -convergence we compare the measure of concentration for the limiting distribution to that of the initial distribution. The definition of our measure of concentration is given in (20) and is equal to the standard deviation of the distribution. The results of this analysis can be found in Table 6. In this table we report posterior

moments for the concentration measure of the initial distribution, $\sigma(s_{1930})$, the concentration measure of the limiting distribution, $\sigma(s_\infty)$, and the difference in concentration measures, $\sigma(s_\infty) - \sigma(s_{1930})$.

Table 6: Posterior Measures of Concentration

Statistic	Mean	Standard Error	95 % HPD
$\sigma(s_{1930})$	2.257	0.1630	[1.934, 2.576]
$\sigma(s_\infty)$	1.613	0.1512	[1.331, 1.923]
$\sigma(s_\infty) - \sigma(s_{1930})$	-0.644	0.2230	[-1.069, -0.192]

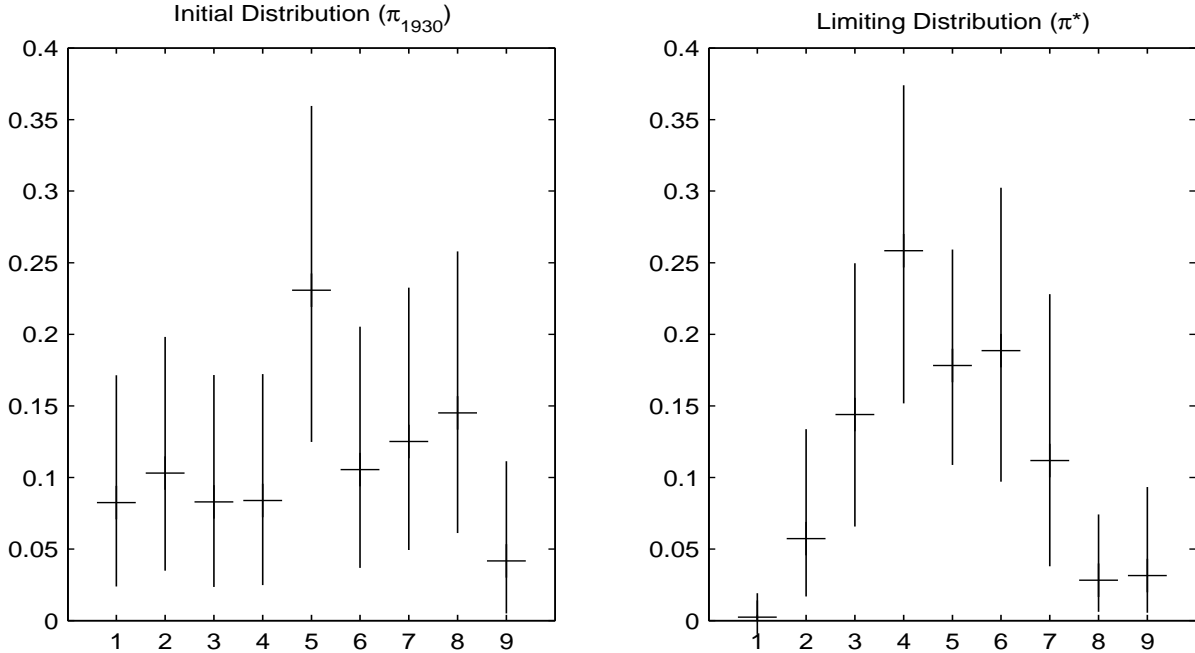
The 95% highest posterior density (HPD) region for the difference in concentration measures is the interval [-1.069, -0.192]. This does not contain 0 so that there is considerable evidence that the concentration measure is lower for the limiting distribution than for the initial distribution. The posterior probability that the standard deviation of the limiting cross-sectional distribution is lower than the standard deviation of the initial cross-sectional distribution of log patents per capita is 0.998.¹³ We take these results as strong evidence that the cross-sectional distribution is becoming more concentrated over time. Thus, we conclude that there is significant evidence of convergence in patents per capita across the states in our sample.

Figure 5 depicts the 95% highest posterior density regions for the initial and limiting cross-sectional distribution of log patent per capita deviations. The vertical lines for each classification represents the length of the HPD while the horizontal line represents the posterior mean. While there is a fair amount of uncertainty in the posterior estimates there does appear to be multiple modes in both the initial and limiting cross-sectional distribution. In 1930 there clearly appears to be a mode situated in the fifth patent classification. There also appears, albeit with somewhat less certainty, modes situated at the second and eighth patent class as well. Thus there is some evidence that

¹³This posterior probability measure is calculated by counting the number of times $\sigma(s_\infty) < \sigma(s_{1930})$ for each of the i.i.d draws from the posterior distribution. A total of 10,000 i.i.d draws is made in total.

there are “innovation” clubs present in 1930.

Figure 5: Initial and Limiting Distribution of log Patents



When we look at the limiting distribution we also see some evidence of multiple modes. Although the limiting distribution has become more concentrated over time, it appears to have two modes in classes 4 and 6. This bimodality is suggestive of some form of convergence clubs perhaps associated with the presence of inputs to innovation in the states. For example, a smaller proportion of states belonging in class 4 (1 out of 12 states) and a larger proportion of states belonging in class 6 (4 out of 12 states) have higher R&D intensities than the U.S. rate of 2.45% in 2000. In addition, a smaller proportion of states belonging in class 4 (2 out of 12 states) and a larger proportion of states belonging in class 6 (6 out of 12 states) have more doctoral scientists per 100,000 inhabitants compared with the U.S. rate of 186 in 2001. Similar observations can be made using other input indicators such as the number of doctoral engineers, the number of science and engineering (S&E) doctorates awarded, and the number of S&E postdoctorates and

S&E graduate students in doctorate-granting institutions.¹⁴ We should note, however, that these modes are much closer together compared to modes observed in the initial period. Thus, this reinforces our previous result that the cross-sectional distribution has become more concentrated.

Posterior moments of the measures of mobility can be found in Tables 7 and 8. These report overall measures of mobility and measures of mobility for each individual classification. Table 7 reports overall measures of mobility and measures of speed of convergence to the limiting distribution. The measure proposed by Shorrocks (1978), \mathcal{M}_S , is the inverse of the harmonic mean of the expected length of stay in each classification. The estimated value of this measure is 0.477 which equates to an average length of stay, over all classifications, of 2.09 ten-year periods. This overall measure of mobility is also decomposed into its upward and downward components, \mathcal{M}_U and \mathcal{M}_D . We see that there is no significant difference between the upward and downward components. This should not be surprising as we are using data that is in deviations from the mean. We would not expect there to be any shift in the location of the distribution and so would not expect to see a significant difference in the upward and downward mobility measures overall.

Table 7: Measures of Overall Mobility for the Markov Chain

\mathcal{M}_S	\mathcal{M}_U	\mathcal{M}_D	$ \lambda_2 $	half-life (h)
0.477	0.240	0.236	0.894	6.525
(0.032)	(0.026)	(0.024)	(0.023)	(1.714)

Note: The numbers in parentheses are standard errors.

While we do not expect to see a difference in the overall upward and downward mobility measures we would expect differences in upward and downward mobility for the individual measures. Table 8 contains these individual mobility measures with $\mathcal{M}_U(j)$

¹⁴Use footnote 9 for the source of these observations.

($\mathcal{M}_D(j)$) representing the probability of moving up (down) the cross-sectional distribution conditional on being in classification j . We find that for classifications 5 through 8, conditional on moving, there is a significantly greater probability that a state will move to a lower classification than to a higher classification and that the probability of moving down is higher the further away a state is from the middle of the distribution.¹⁵ For classifications 2 through 4, however, the opposite is true. There is a greater probability to move to a higher classification than to move to a lower classification with the probability of moving to a higher classification being greater the further a state is away from the middle of the distribution. Therefore, we are seeing movements to the middle of the distribution rather than movements to the extremes of the distribution. Furthermore, this movement is from all parts of the cross-sectional distribution and not just from the extremes of the distribution.

Table 8: Measures of Mobility for the Markov Chain

	Classification								
	1	2	3	4	5	6	7	8	9
$\mathcal{M}_U(j)$	0.294 (0.117)	0.423 (0.094)	0.364 (0.074)	0.259 (0.056)	0.190 (0.051)	0.147 (0.050)	0.070 (0.038)	0.174 (0.076)	—
$\mathcal{M}_D(j)$	—	0.004 (0.013)	0.101 (0.047)	0.178 (0.048)	0.328 (0.061)	0.231 (0.060)	0.289 (0.066)	0.391 (0.099)	0.369 (0.108)

Note: Numbers in parentheses are posterior standard deviations obtained from using 10,000 draws from the posterior distribution.

Finally, the modulus of the second largest eigenvalue is reported in Table 7. The posterior mean of $|\lambda_2|$ is 0.894. As this is significantly different from 1, we conclude that the limiting distribution we characterized above is unique. The posterior mean of the half-life, h , is estimated to be 6.525. This equates to a half-life of about 65 years so

¹⁵For example, if a state is initially in the 8th patent classification, the probability of moving to a lower class is 0.391 with a standard error of 0.099 while the probability of moving to a higher class is only 0.174 with a standard error of 0.076.

that we conclude that the speed of convergence of the cross-sectional distribution to the unique limiting distribution is “slow.”

Therefore, we find that the cross-sectional distribution of log patents per capita deviations is converging slowing to a unique limiting distribution that is significantly more concentrated than the initial cross-sectional distribution in 1930. The mobility present in the data reflects a movement to the middle of the cross-sectional distribution with the greatest mobility at the extremes.

4.3 Robustness of Our Results

The results reported above were for a Markov chain with nine states with a transition period of ten years. Nine classifications was chosen so that the class lengths were short enough to allow for us to observe states making transitions between patent classifications. The time period was chosen to be ten years so that we could allow enough time for any change in state policy to affect the innovation activity. While we believe that this parameterization of the model is an appropriate one, Table 9 reports the 95% HPD for $\sigma(s_\infty) - \sigma(s_{1930})$ for other combinations of classification numbers and transition periods. In this experiment, the interval $[-1.3, 1.3]$ was broken up into 3, 5, 7, 9, and 11 equal length intervals leading to 5, 7, 9, 11, and 13 patent classifications respectively. A discrete state Markov chain with a transition period of 5, 10 and 20 years respectively were estimated using analogous priors to the one reported in Section 3.3 above.

It is clear from these results that when the time period is long enough and the patent classification is fine enough there is strong evidence of σ -convergence in the data. In all but two cases, five year transition with five and seven classifications respectively, the 90% or higher HPD for $\sigma(s_\infty) - \sigma(s_{1930})$ does not include 0. Clearly when the length of the classification is large, five years is not long enough to observe many transitions. The transition probability matrix, P , is then essentially diagonal. We believe that this is not a failing of the model but rather a failure to define the appropriate transition

Table 9: $\sigma(s_\infty) - \sigma(s_{1930})$: 95% Highest Posterior Density Regions

Number of bins	Time Period		
	5 years	10 years	20 years
5	[-0.295, 0.108]	[-0.405, 0.031] ^a	[-0.475, 0.023] ^b
7	[-0.523, 0.107]	[-0.692, -0.021]	[-0.956, -0.258]
9	[-0.887, -0.023]	[-1.069, -0.192]	[-1.293, -0.276]
11	[-1.198, -0.090]	[-1.395, -0.268]	[-1.709, -0.488]
13	[-1.559, -0.220]	[-1.870, -0.464]	[-2.193, -0.616]
Decile	[-1.066, -0.069]	[-1.265, -0.167]	[-1.444, -0.368]

^a90 % HPD does not include 0

^b92 % HPD does include 0

period for the patent classifications that are used. We conclude that the results reported in Section 4 are therefore robust to the number of patent classifications and transition periods as long as the transition period is long enough to allow states to change their patent classification.

The approach we have followed in defining the patent classifications was to define classifications of equal length. The interval the the middle classifications were define over, $[-1.3, 1.3]$ was chosen that that 90% of the observations in 1930 fell in the middle seven classifications in the case where we defined nine patent classifications. The consequence of this approach is the have an initial distribution that has “thin” tails in the sense that the most observations fall in the middle classifications. An alternative approach would have been to define classifications so that the probability that a randomly chosen state in 1930 is in any one classification is equal across classifications. In Table 9 the row labeled decile contains the 95% HPD for $\sigma(s_\infty) - \sigma(s_{1930})$ where the patent classifications are defined based on the population deciles in 1930. The classification definitions then remain unchanged throughout the other time periods. The results show that the 95% HPD for $\sigma(s_\infty) - \sigma(s_{1930})$ does not include 0 for a 5, 10 or 20 year transition. Therefore, our result of σ -convergence in patents across U.S. states is robust to defining the classifications to be equal in length or equal in probability in the initial period.

Finally, in this paper we study the log patent distribution. We subtract the national mean of log patents for each year and then define the patent classifications over the log patents deviation from mean distribution. We do this to abstract away from the cyclical nature of the patent series.

Our test of σ -convergence involves testing whether the standard deviation of the limiting cross-sectional distribution is significantly lower than the standard deviation of the initial cross-sectional distribution. For a continuous random variable, it does not matter whether you calculate the standard deviation of the data in log levels or in deviations from log mean. However we use in this paper the standard deviation of a discrete random variable to test for σ -convergence. In defining patent classifications over the cross-sectional distributions of log patent deviations we are effectively normalizing the data each year relative to the national mean. Therefore the classification definitions, in relation to the log levels, are different than they would be if we defined equal length classifications over the log level patent data. In order to check whether our results are affected by subtracting the national mean of log patent series we constructed a discrete state Markov chain with nine classifications whose length were equal in terms of the log level patent distribution. The 95% HPD for $\sigma(s_\infty) - \sigma(s_{1930})$ using this classification was $[-0.800, -0.010]$. Thus, our result is not affected by subtracting the mean from the log patent distribution. We also tested whether taking logs had an effect on the result. The 95% HPD for $\sigma(s_\infty) - \sigma(s_{1930})$ is $[-1.199, -0.106]$ for a nine classification discrete state Markov chain defined over the patent distribution in levels. Therefore taking logs does not appear to have an effect on the basic result that the cross-sectional distribution of patents in becoming more concentrated over time.

5 Implications and Concluding Comments

We have studied the dynamics of the cross-section distribution of patents per capita for the 48 continental U.S. states from 1930 to 2000 applying Bayesian techniques to a discrete Markov chain model. Empirical findings show that invention activities across states are converging, in a σ -convergence sense, albeit at a slow rate. A necessary but not sufficient condition for σ -convergence is mobility in states' standing in the cross-section patents per capita distribution. Indeed, we find that some initially invention-lagging areas catch up and some invention-leading areas lose their initial leads. For example, in 1930 North Carolina and Arizona are invention-lagging states. North Carolina belonged in class 2 while Arizona belonged in class 4. By 2000, North Carolina has moved up to class 5 while Arizona has moved up to class 6. Invention-leading states like Illinois and Michigan moved down from class 8 in 1930 to class 6 in 2000. A sufficient condition for convergence is for the limiting distribution of log patent per capita deviations to have a substantially narrower dispersion than the distribution in 1930. We find that there is evidence that the cross-sectional distribution has indeed become more concentrated over time.

The finding of slow convergence is consistent with the empirical observation that knowledge spillover is spatially-mediated.¹⁶ That is, although knowledge diffuses across space, the effectiveness of knowledge externalities tend to dissipate with distance; hence, inventions would tend to cluster spatially. This clustering tendency contributes to the slow rate of convergence. The slow rate of convergence is also partly because technology initiatives at the state level is a fairly recent phenomenon. For example, S&T advisory bodies for state governors only started in the 1960s; state cooperative technology programs (such as technology development and technology financing) that complement federal, private and academic R&D investments only started in the late 1970s (See, Coburn

¹⁶See, e.g., Jaffe, Trajtenberg and Henderson (1993), Feldman (1994), and Audretsch and Feldman (1996).

and Berglund (1995) and Carnegie Commission on Science, Technology and Government (1992)).

The observed slow rate of invention convergence is a challenge to state policy makers to pursue more aggressive policies related to knowledge creation. This is because large differences still characterize states' knowledge creating capabilities. For example, in 2000, total R&D as a share of gross state product is between 0.32% (Wyoming) and 5.81% (Michigan), with a mean of 2.07% and a standard deviation of 1.46%. The number of doctoral scientists and engineers per 100,000 inhabitants in 1999 is between 111.30 (Florida) and 566.46 (Delaware), with a mean of 224.31 and a standard deviation of 102.25.¹⁷

One way to close these gaps is for state governments to increase their direct contribution to university R&D. According to the most recent estimate, in 1965, U.S. state governments funded only 0.90% of total U.S. R&D, mostly going to universities; this increased to only 1.18% in 1995 (See, National Science Foundation (1999)). Interestingly, in 1953 (earliest available data), state and local governments funded 14.7% of academic R&D; this has dropped to 7.3% in 2000 (See, National Science Board (2002)). Clearly, states (some more than others) need to reconsider their role as a source of funding for university R&D especially in light of strong evidence that industry also benefit from university R&D (See, e.g., Jaffe (1989) and Jaffe and Trajtenberg (2002)).

A states' role in technology development is not limited to direct monetary support to R&D. A number of studies have shown that a skilled workforce and a dependable infrastructure are attributes firms value when making R&D (and production) location decisions. Increasing support to education and infrastructure development would also contribute to the narrowing of knowledge gaps.

¹⁷These are tabulated from the National Science Foundation's *Science and Engineering State Profiles: 2000-2001, 1999-2000* available at <http://www.nsf.gov/sbe/srs/sepro/start.htm>.

References

- Audretsch, D., and M. Feldman (1996) 'R&D Spillovers and the Geography of Innovation and Production.' *American Economic Review* 86(3), 630–640
- Bernado, J.M., and A.F.M. Smith (1994) *Bayesian Theory* (Chichester, UK: John Wiley and Sons)
- Bernard, A. B., and S. N. Durlauf (1996) 'Interpreting Tests of the Convergence Hypothesis.' *Journal of Econometrics* 71(1/2), 161–173
- Buchinsky, M., and J. Hunt (1999) 'Wage Mobility in the United States.' *Review of Economics and Statistics* 81(3), 351–368
- Caniëls, M. C. J. (2000) *Knowledge Spillovers and Economic Growth* (Cheltenham: Edward Elgar)
- Carlino, G., and L. Mills (1993) 'Are U.S. Regional Incomes Converging? A Time Series Analysis.' *Journal of Monetary Economics* 32(2), 335–346
- Carnegie Commission on Science, Technology and Government (1992) 'Science, technology, and the states in america's third century.' New York: Carnegie Commission
- Champernowne, D. G. (1953) 'A Model of Income Distribution.' *Economic Journal* 63(250), 318–351
- Co, C. Y. (2002) 'Evolution of the Geography of Innovation: Evidence from Patent Data.' *Growth and Change* 33(4), 393–423
- Co, C. Y., and M. Wohar (2004) 'Technological Convergence Across U.S. Regions and States.' *Economics of Innovation and New Technology* 13(2), 101–126
- Coburn, C., and D. Berglund (1995) *Partnerships: A Compendium of State and Federal Cooperative Technology Programs* (Columbus: Battelle)

- Devroye, L. (1986) *Non-uniform Random Variate Generation* (New York:Springer-Verlag)
- Durlauf, S., P. Johnson, and J. Temple (forthcoming) ‘Growth econometrics.’ In *Handbook of Economic Growth*, ed. P. Aghion and S. Durlauf (Amsterdam: North Holland)
- Feldman, M. (1994) *The Geography of Innovation* (Dordrecht: Kluwer)
- Gang, I., J.S. Landon-Lane, and M-S. Yun (2004) ‘Good and Bad Mobility: A Note on Directional Mobility Measures.’ Mimeo, Department of Economics, Rutgers University.
- Gelfand, A.E., and D.K. Dey (1994) ‘Bayesian model choice: Asymptotics and exact calculations.’ *Journal of the Royal Statistical Society Series B*
- Geweke, J. F., C. Marshall, and G. A. Zarkin (1986) ‘Exact Inference for Continuous Time Markov Chain Models.’ *Review of Economic Studies* 53(4), 653–69
- Hall, B. H., Z. Griliches, and J. A. Hausman (1986) ‘Patent and R and D: Is There a Lag.’ *International Economic Review* 27(2), 265–283
- Jaffe, A. B. (1989) ‘Real Effects of Academic Research.’ *American Economic Review* 79(5), 957–970
- Jaffe, A. B., and M. Trajtenberg (2002) *Patents, Citations, & Innovations: A Window on the Knowledge Economy* (Cambridge: The MIT Press)
- Jaffe, A. B., M. Trajtenberg, and R. Henderson (1993) ‘Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.’ *Quarterly Journal of Economics* 108(3), 577–598
- Johnson, D., and A. Brown (2002) ‘How the West was Won: Regional and Industrial Inversion in U.S. Patent Activity.’ Mimeo, Wellesley College

- Johnson, P. A. (2000) ‘A Nonparametric Analysis of Income Convergence Across the US States.’ *Economics Letters* 69(2), 219–223
- Kortum, S., and J. Lerner (1998) ‘Stronger Protection or Technological Revolution: What is Behind the Recent Surge in Patenting?’ *Carnegie-Rochester Conference Series on Public Policy* 48, 247–304
- National Science Board (2000) *Science and Engineering Indicators – 2000* (Arlington: NSF)
- (2002) *Science and Engineering Indicators – 2002* (Arlington: NSF)
- National Science Foundation (1999) *What is the State Government’s Role in the R&D Enterprise?* (Arlington: NSF)
- Popp, D., T. Juhl, and D. K.N. Johnson (2003) ‘Time in Purgatory: Determinants of the Grant Lag for U.S. Patent Application.’ NBER Working Paper 9518
- Prais, S.J. (1955) ‘Measuring Social Mobility.’ *Journal of the Royal Statistical Society, Series A* Part I(118), 56–66
- Quah, D. T. (1992) ‘International Patterns of Growth: II. Persistence, Path Dependence, and Sustained Take-Off in Growth Transition.’ Mimeo, London School of Economics.
- (1993a) ‘Empirical Cross-Section Dynamics in Economic Growth.’ *European Economic Review* 37(2/3), 426–434
- (1993b) ‘Galton’s Fallacy and Tests of the Convergence Hypothesis.’ *Scandinavian Journal of Economics* 95(4), 427–443
- (1996) ‘Aggregate and Regional Disaggregate Fluctuations.’ *Empirical Economics* 21(1), 131–159

- Schmookler, J. (1966) *Invention and Economic Growth* (Cambridge: Harvard)
- Shorrocks, A. F. (1978) 'The Measurement of Mobility.' *Econometrica* 46, 1013–1024
- Sokoloff, K. (1988) 'Inventive Activity in Early Industrial America: Evidence from Patent Records, 1790-1864.' *Journal of Economic History* 48(4), 813–850
- Suarez-Villa, L. (2000) *Invention and the Rise of Technocapitalism* (Lanham, MD: Rowman & Littlefield)
- Varga, A. (1999) 'Time-Space Patterns of U.S. Innovation: Stability or Change?' In *Innovation, Networks and Localities*, ed. M.M. Fischer and L. Suarez-Villa (Berlin: Springer) pp. 215–234