Employment concentration across U.S. counties

Klaus Desmet [a,b,*], Marcel Fafchamps [c]

[a] *Department of Economics, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain*
[b] *CEPR, United Kingdom*
[c] *CSAE, Department of Economics, University of Oxford, Manor Road, Oxford OX1 3UQ, United Kingdom*

Abstract

This paper examines the spatial distribution of jobs across U.S. counties between 1970 and 2000, and investigates whether sectoral employment is becoming more or less concentrated. The existing literature has found deconcentration (convergence) of employment across urban areas. Cities only cover a small part of the U.S. though. Using county data, our results indicate that deconcentration is limited to the upper tail of the distribution. The overall picture is one of increasing concentration (divergence). While this seemingly contradicts the well documented deconcentration in manufacturing, we show that these aggregate employment dynamics are driven by services. Non service sectors such as manufacturing and farming are indeed becoming more equally spread across space, but services are becoming increasingly concentrated.

JEL classification: R11; R12
Keywords: Spatial distribution of employment; Ergodic distributions; U.S. counties; Economic geography

1. Introduction

Economic activity is unevenly distributed across space. The interaction of positive and negative externalities creates intricate geographical patterns of city clusters and rural hinterland (Henderson, 1988; Fujita et al., 1999). Over time, these patterns evolve because of changes in preferences, production technologies and transport costs. As a result, the spatial distribution of employment adjusts as jobs are created in certain locations, and destroyed elsewhere. Understanding how economic activity is likely to be distributed through space in the future is important for policy makers at the national and local level.

* Corresponding author. Department of Economics, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain.
E-mail addresses: desmet@eco.uc3m.es (K. Desmet), marcel.fafchamps@economics.oxford.ac.uk (M. Fafchamps).

This paper describes the geographical evolution of jobs in the U.S. between 1970 and 2000, with the goal of understanding what the future spatial distribution of employment would look like if current tendencies were to continue. We use county-level employment in 13 different sectors ranging from farming to manufacturing and services and focus on the ergodic distribution of jobs.

Our work differs from the existing literature in a number of respects. First, rather than looking at income per capita or population, we are interested in employment. Many authors have studied whether standards of living in the U.S. are becoming more similar over time. For instance, Higgins et al. (2003) find a strong evidence of income convergence across counties. This is not entirely surprising, given the high degree of labor mobility in the U.S. (Blanchard and Katz, 1991). However, income convergence does not tell us anything about *where* economic activity is locating. Is the U.S. moving towards a situation with more or with less large- and medium-sized metropolitan counties? Are rural counties losing or gaining jobs? These are the kinds of questions we address in our paper.[1] This is similar to studying whether population is becoming more or less concentrated in space. In this respect, Beeson and DeJong (2002) are of particular interest. They find population divergence across counties, especially in the post-WWII period. Our work is complementary to theirs. By looking at employment, rather than population, we get additional insights from sectoral disaggregation.

Second, we examine the country as a whole, not just metropolitan areas. Most of the literature on the spatial organization of economic activity in the U.S. has focused on cities. One central finding of that line of research is that city growth is independent of city size, a phenomenon known as Gibrat's Law (Sutton, 1997). However, as pointed out by Beeson et al. (2001), limiting the analysis to urban areas introduces a selection bias, since cities are those areas which experienced high growth in the past. A recent paper by Eeckhout (2004) addresses this issue by revisiting Gibrat's Law using Census 'places'. In contrast to metropolitan areas, these data cover the entire size distribution, including small towns and villages. He confirms that growth is independent of size. However, 'places' still do not cover the entire U.S. In the 2000 Census they accounted for 74% of the population.

Our third point of departure with the existing literature is our methodology. Instead of relying on a single method whether β-convergence, σ-convergence, or ergodic distributions we develop a methodology that encompasses them all. Much of the existing work comparing geographical units is couched in terms of Barro's β-convergence: the underlying model is deterministic in nature (Barro, 1991; Mankiw et al., 1992). As first emphasized by Quah, evidence of β-convergence can yield a misleading picture, because it can arise even when countries or regions are getting further apart, and vice versa (Quah, 1993, 1996a; Durlauf and Quah, 1999). As a solution, Sala-i-Martin (1996) suggests studying distributions by looking at the evolution of the variance over time, a concept known as σ-convergence. Quah (1996b, 1997) goes one step further by focusing on the ergodic distribution. This refers to the long-term spatial distribution of economic activity that would arise if current transition probabilities would remain constant. The ergodic distribution is the distributional equivalent of the β coefficient in a standard Barro model: it predicts in which direction the process goes, should current structural factors remain unchanged. Of course, structural parameters may

[1] If labor and capital are not quite mobile, the distribution of GDP per capita can be regarded as capturing the distribution of economic activity across space. However, in a country like the U.S., where capital and workers are highly mobile, the dispersion of GDP per worker across geographical units is more a measure of dispersion in productivity than in economic activity per se.

change, in which case the direction of the process would change as well. The ergodic distribution is thus but a way of describing the current trend of the distribution.

In this paper we start by computing parametric and non-parametric versions of the unconditional and conditional β-convergence tests and explain how they can be understood as describing the expectation of the transition probability. We then compute two versions of the ergodic distribution. The first version is the stochastic equivalent of unconditional β-convergence: it assumes that all counties are inherently equivalent and could switch places with each other over time. The second version is the stochastic equivalent of conditional β-convergence: it conditions on county characteristics that are constant over time. It is our best estimate of how economic activity would be distributed across U.S. counties should current tendencies remain unchanged.

In addition, we also introduce a number of practical innovations when deriving the ergodic distribution. In particular, by computing the transition matrix from the smoothed conditional distribution rather than directly from the data, we get a better approximation of the ergodic distribution. This makes the results both more detailed and more accurate.[2] The methodology is easy to implement, and can be applied to any empirical study involving distribution dynamics.

We now turn to describing our main findings. Whereas recent work on metropolitan areas shows a tendency towards deconcentration, with total employment becoming more equally spread across *cities* (Chatterjee and Carlino, 2001; Carlino and Chatterjee, 2002), standard β-convergence tests using U.S. *county* data suggest the contrary, with jobs becoming more concentrated over recent decades (Desmet and Fafchamps, 2005). The analysis presented here resolves this apparent puzzle. Results show that, compared to the current distribution of total employment across counties, the ergodic distribution is a lot flatter, with the middle group thinning out. The overall picture that emerges is thus one of concentration (divergence), with lots of small- and medium-sized counties losing jobs to the more urban ones. At the upper tail of the distribution, however, the opposite is true, with large metro counties losing jobs in favor of intermediate-sized urban counties. In other words, there is deconcentration (convergence) in the upper part of the distribution, and concentration (divergence) in the distribution at large. This explains the opposing results of Chatterjee and Carlino (2001) and Desmet and Fafchamps (2005). Our findings confirm the results of Beeson and DeJong (2002) for population growth: for the post-WWII period they report divergence across most of the distribution, but convergence in the upper decile.

The increased concentration evident in total employment stands in contrast to what happened within the manufacturing sector. It is by now a stylized fact that since World War II manufacturing employment has become less concentrated, albeit at a slow pace (Dumais et al., 2002; Kim, 1995). Our data confirm this empirical regularity. Although manufacturing cannot account for the spatial dynamics of aggregate employment, services can. The main service sectors ‘retail’, ‘finance, insurance and real estate’ and ‘other services’ exhibit concentration (divergence) in the middle part of the distribution and deconcentration (convergence) in the upper tail. This is most patent in the case of ‘other services’, where we get ‘twin peaks’ a bimodal ergodic distribution.

That overall trends in the economy are driven by services should not come as a surprise, given their weight. However, the fact that services behave differently from the rest of the economy is interesting, because empirical work in economic geography has mainly focused on manufacturing. Our findings confirm that the much heralded demise of cities, epitomized by

[2] Moreover, contrary to Quah, who uses a highly complex programming language in Unix to obtain non-parametric kernel estimates of the transition matrix, we rely on simple Stata commands. The ado files are available from the authors upon request.

manufacturing jobs moving to less dense areas, is not occurring. The reason is the rise of the service industry (Kolko, 1999).

Though not the subject of this paper, our results have implications for the spatial dynamics of productivity and wages. Following Ciccone and Hall (1996), our findings suggest that sectors that have been deconcentrating, such as manufacturing, may have experienced spatial convergence in wages and productivity. In contrast, we would expect aggregate employment and services, which have been concentrating, to have seen increasing spatial divergence in wages and productivity.

2. Methodology

In this section we present a general framework to discuss the evolution of the distribution of an arbitrary geographical variable over time. Let our variable of interest be denoted Y_t^i where i stands for location and t for time. Variable Y_t^i could denote GDP per head, employment, or income, but for now this is of no importance. At each point in time we observe realizations of Y_t^i for each i. We want to know whether, over time, realizations of Y_t^i are becoming more 'alike' across all i's. This we call convergence. If realizations are becoming less alike, we call it divergence. We first discuss unconditional convergence; we then look at conditional convergence.

2.1. Unconditional convergence in a deterministic model

The growth convergence literature reverts around a β-convergence test meant to ascertain whether GDP per head across countries is converging towards a common value Y^*. This test is implemented via a regression of the form (Mankiw et al., 1992; Quah, 1993):

$$\log Y_{t+1}^i \quad \log Y_t^i = \alpha \quad \beta \log Y_t^i \tag{1}$$

to which an error term is added for estimation purposes. Eq. (1) can be rewritten as:

$$Y_{t+1}^i = e^{\alpha} (Y_t^i)^{1 \quad \beta} \tag{2}$$

Eq. (2) is a deterministic difference equation with two steady states:

$$Y^* = e^{\frac{\alpha}{\beta}}$$

$$Y_0 = 0$$

which, in the unconditional case, are the same for all i's. The stability of this deterministic system around the Y^* steady state depends on the sign of β: if $\beta<0$, the Y^* steady state is not stable and Y_t diverges from it. Standard convergence tests estimate Eq. (1) and examine whether β is positive or not.

One critique of this model is that it imposes too much structure on the law of motion of Y_{t+1}^i. In particular, it is unable to test for the presence of multiple steady states. In addition, it assumes that convergence is exponential. However, in general the linear approximation underlying Eq. (1) is only valid locally. It makes little sense to apply this approximation to observations which, according to the model, are very far from Y^*. A more satisfactory model is one that allows for nonlinearity:

$$\log Y_{t+1}^{i} = \phi(\log Y_{t}^{i}) \tag{3}$$

where $\phi(.)$ is an arbitrary smooth function. Eq. (3) can be estimated by standard non-parametric techniques. If function $\phi(.)$ cuts the 45° line more than once, the process driving Y_t^i has multiple equilibria. Each point at which $\phi(.)$ cuts the 45° line from above is stable; each point where it cuts from below is unstable. As it turns out, it is easier to graph the equivalent alternative model:

$$\begin{aligned}\log Y_{t+1}^{i} \;\; \log Y_{t}^{i} &= \phi(\log Y_{t}^{i}) \;\; \log Y_{t}^{i} \\ &= f(\log Y_{t}^{i})\end{aligned} \tag{4}$$

Estimates of Eq. (4) are presented in the empirical section. Evidence of multiple deterministic steady states is found for several sectors.

2.2. Unconditional convergence in a stochastic model

As Durlauf and Quah (1999) have emphasized, the approach to convergence based on a deterministic model is hardly appropriate because it fails to recognize that in practice Y_i^t is stochastic. To illustrate this point, let us return for a moment to the linear model (1), to which we add a stochastic term u_t^i:

$$\log Y_{t+1}^{i} \;\; \log Y_{t}^{i} = \alpha \;\; \beta \log Y_{t}^{i} + u_{t}^{i} \tag{5}$$

Defining $y_t^i \equiv \log Y_t^i$, Eq. (5) can be rewritten more simply as:

$$y_{t+1}^{i} \;\; y_{t}^{i} = \alpha \;\; \beta y_{t}^{i} + u_{t}^{i}$$

In this case, y_t^i never actually settles anywhere permanently so there is no steady state in the deterministic sense and thus no β-convergence. As Quah (1993) has shown, β in this context measures the speed at which y_t^i reverts to the mean.

When y_i^t is stochastic, a more adequate representation of its evolution over time is:

$$f_{t+1}(y_{t+1}) = \int_{\infty}^{\infty} g(y_{t+1}|y_t) f_t(y_t)\,\mathrm{d}y_t \tag{6}$$

where $f_t(y_t)$ denotes the (unconditional) distribution of y_t at time t across all i's and $g(y_{t+1}|y_t)$ denotes its transition probability. Here and in the remainder of this section, we assume the transition probability to be constant over time. This is of course an oversimplification. We revisit this issue in the empirical section. Eq. (6) is itself a deterministic law of motion. Provided that certain conditions are satisfied (Stokey et al., 1989; Luenberger, 1979), this system has a steady state or time-invariant distribution $f(y_{t+1})$ to which it converges.[3] This time-invariant distribution is called the ergodic distribution. It is the distribution $f(y_{t+1})$ that satisfies:[4]

$$f(y_{t+1}) = \int_{\infty}^{\infty} g(y_{t+1}|y_t) f(y_t)\,\mathrm{d}y_t \tag{7}$$

[3] In empirical applications, the most important issue that arises with respect to the existence of a non-degenerate ergodic distribution is that of detrending. This is discussed in detail below.

[4] There might be multiple solutions to Eq. (7) and thus multiple ergodic distributions. Multiple solutions do not arise in our empirical analysis and are ignored here.

β-convergence corresponds to the case when the ergodic distribution is degenerate with a mass point at y^*. In general, $f(y_{t+1})$ is not degenerate. If the ergodic distribution $f(y_{t+1})$ is more concentrated has lower variance than the current distribution $f_t(y_t)$, we conclude that there is convergence, and vice versa. This is but a straightforward extension of the concept of σ-convergence introduced by Sala-i-Martin (1996). The advantage of dealing with the ergodic distribution is that we can extrapolate current transition probabilities to the indefinite future, hence obtaining a clearer picture of what these probabilities imply for the long-term.

2.3. Empirical implementation

In the empirical implementation, we begin by estimating linear and nonlinear β-convergence models and look for possible evidence of multiple deterministic steady states. We then turn to the stochastic approach and derive ergodic distributions. Computing the ergodic distribution involves three steps: (i) calculating $f_t(y_t)$ and $f_{t+1,t}(y_{t+1}, y_t)$; (ii) deriving $g(y_{t+1}|y_t)$ from the fact that the conditional distribution is the joint distribution divided by the marginal distribution:

$$g(y_{t+1}|y_t) = \frac{f_{t+1,t}(y_{t+1}, y_t)}{f_t(y_t)}$$

and (iii) obtaining the ergodic distribution by solving Eq. (7) for a constant $f(.)$. As illustrated by Quah (1996b), steps (i) and (ii) are easily handled by non-parametric techniques: $f_t(y_t)$ and $f_{t+1,t}(y_{t+1}, y_t)$ are estimated by fitting a kernel density to the data, and $g(y_{t+1}|y_t)$ is obtained by dividing one by the other. For the third step, it is difficult to work with Eq. (7) directly. The standard approach in practice is to discretize the space of possible values y into N discrete cells Γ_k, with $k=\{1,\ldots, N\}$. Formally, the probability of being in cell Γ_k at time t is:

$$p_{kt} \equiv \Pr(y_t \in \Gamma_k)$$

The transition probability of moving from cell Γ_k to cell Γ_m over one time period is denoted:

$$a_{km} \equiv \Pr(y_{t+1} \in \Gamma_k | y_t \in \Gamma_m)$$

The ergodic distribution is then a set of interval probabilities p_k that solves:[5]

$$p_k = \sum_m a_{km} p_m$$

---

[5] In matrix form we have:

$$p \quad A_p$$

$$(I-A)p \quad 0$$

It looks like the above system only has a solution of the form $p \quad 0$, but this is an illusion. Matrix A does not have full rank since, by definition of a probability, each column sums to 1. To find p, one needs to drop one row of A and to add the requirement that:

$$\sum_i p_i \quad 1$$

We obtain a system of the form:

$$\begin{bmatrix} 1-a_{11} & \ldots & -a_{1N} \\ \ldots & 1-a_{ii} & \ldots \\ -a_{N\ 1,1} & \ldots & -a_{N\ 1,N} \\ 1 & \ldots & 1 \end{bmatrix} [p] \quad \begin{bmatrix} 0 \\ \ldots \\ 0 \\ 1 \end{bmatrix}$$

$$Bp \quad b$$

The modified system can be solved by inverting:

$$p \quad B^{\ 1} b$$

In a linear system such as this, the ergodic distribution is in general unique (Luenberger, 1979).

Because step (iii) involves discretization, it is common to discretize steps (i) and (ii) as well and to compute the transition matrix $[a_{km}]$ directly from the data. This approach, however, fails to take advantage of the smoothing properties of kernel densities. For this reason, transition matrices used in practice are usually very coarse e.g., with four or five intervals only. The resulting ergodic distribution is too rough to draw precise conclusions about convergence. In this paper, we obtain a better approximation of the ergodic distribution by postponing discretization until step (iii) and computing the transition matrix from the smoothed conditional distribution $g(y_{t+1}|y_t)$ rather than directly from the data.[6]

An important practical detail is to make the data mean stationary. Our objective in computing the ergodic distribution is to get an idea of the long-term distribution of y_t^i around its mean, should the transition matrix remain unchanged. If $E[y_t^i]$ changes over time, failing to subtract the mean will lead to biased results. To see why, suppose that we have two time periods, t and $t+s$, with $E[y_{t+s}^i]>E[y_t^i]$. A transition matrix computed on the raw data will tend to produce a degenerate ergodic distribution with a mass point at the highest value of y.[7] To avoid degenerate ergodic distributions, it is essential that the mean of the detrended variable does not change over time. Simply subtracting the mean $E[y_t^i]$ from y^t in each time period would do the job. In practice, however, it will be convenient to use the following equivalent form instead:[8]

$$\begin{aligned} w_t^i &= y_t^i \\ w_{t+s}^i &= y_{t+s}^i \quad E[y_{t+s}^i] + E[y_t^i] \end{aligned} \tag{8}$$

This transformation offers the advantage of presenting all distributions in terms of y_t^i around its mean at time t, thereby facilitating visual interpretation. As is clear from Eq. (8), this formulation amounts to removing any stochastic linear trend in the data.[9] Applying steps (i) to (iii) to the detrended variable yields the transition matrix and ergodic distribution of w_t^i. The resulting ergodic distribution represents what would be the long-term distribution of y_t^i around its mean, should the current transition matrix remain unchanged.

2.4. Conditional convergence

Until now we have implicitly assumed that the distribution of y_t^i is the same for each location i, irrespective of their inherent, time-invariant characteristics X^i. In many situations, this is an unrealistic assumption: part of the variation in y_t^i across i's is due to differences in their X^i. This variation does not disappear over time. Consequently, if we fail to control for X^i we may falsely conclude that y_t^i is not converging when in fact it is. Correct inference about convergence therefore requires that we condition on X^i. We call this approach conditional convergence.

We decompose the variation in detrended y_t^i into two parts: that due to $X^i\theta$ and that due to a first order stochastic process z_t^i. The data generation process takes the form:

$$w_t^i = X^i\theta + z_t^i$$

[6] Quah (1996a,b, 1997) computes the transition matrix (steps i and ii) using non-parametric techniques. But as far as we can tell, when computing the ergodic distribution he uses a crudely discretized transition matrix with a small number of cells only.

[7] By the same token, if $E[y_{t+s}^i]<E[y_t^i]$, the ergodic distribution will tend to a mass point at the lowest value y.

[8] This is equivalent to adding a constant $E[y_t^i]$ to $y_t^i-E[y_t^i]$, $y_{t+s}^i-E[y_{t+s}^i]$, etc.

[9] An alternative would be to assume a stochastic exponential trend. In that case, the appropriate transformation would be $w_{t+s}^i \quad y_{t+s}^i E[y_t^i]/E[y_{t+s}^i]$. Using this alternative does not change our results qualitatively.

with $E[Xz]=0$ and:

$$f_{t+1}(z_{t+1}) = \int_{\infty}^{\infty} g(z_{t+1}|z_t)f_t(z_t)\mathrm{d}z_t \tag{9}$$

with corresponding ergodic equation

$$f(z_{t+1}) = \int_{\infty}^{\infty} g(z_{t+1}|z_t)f(z_t)\mathrm{d}z_t \tag{10}$$

Although we do not observe z_t^i directly, we can obtain a consistent estimate by first estimating θ from a pooled regression of the form

$$\begin{aligned} w_t^i &= X^i\theta + z_t^i \\ &\dots \\ w_{t+s}^i &= X^i\theta + z_{t+s}^i \end{aligned} \tag{11}$$

and then using $\hat{\theta}$ to compute

$$\begin{aligned} \hat{z}_t^i &= w_t^i \;\; X^i\hat{\theta} \\ &\dots \\ \hat{z}_{t+s}^i &= w_{t+s}^i \;\; X^i\hat{\theta} \end{aligned} \tag{12}$$

The ergodic distribution of $\hat{z}_t^i$ can then be estimated through steps (i) (iii) as detailed before.

The long-term distribution of w_t^i is obtained by combining the variation due to $X^i\beta$ with the ergodic distribution of z_t^i. Let this distribution be written $f_Y(y)$. We have:

$$\begin{aligned} w^i &= X^i\hat{\theta} + z^i \\ &= h^i + z^i \end{aligned}$$

This is a standard problem in statistics. The general formula in the discrete case is (Mood et al., 1974, p. 186):

$$\begin{aligned} f_W(w) &= \sum_{w} f_Z(w \;\; h)f_H(h) \\ &= \sum_{X\hat{\theta}} f_Z(w \;\; X\hat{\theta})f_H(X\hat{\theta}) \end{aligned} \tag{13}$$

where $f_Z(.)$ is the ergodic distribution function of z^i. Applying this formula to the data yields the conditional ergodic distribution of w_t^i.[10]

[10] To compute the long-term probability of a particular value of $w \;\; W$, we proceed as follows. Say we have 3000 values of $X^i\hat{\theta}$, each with frequency 1.

1. Outer loop: let $w \;\; W$.
   (a) Inner loop: Take a specific value of $X^i\hat{\theta}$. We have $f_W(X^i\hat{\theta}) \quad \frac{1}{3000}$.
   (b) Compute $\hat{z}^i \;\; W - X^i\hat{\theta}$.
   (c) Obtain $f_Z(W - X^i\hat{\theta})$ using the ergodic distribution of $\hat{z}$. This is just the frequency of the discretized $\hat{z}^i$ interval in which $W - X^i\hat{\theta}$ happens to fall.
   (d) Repeat for all values of $X^i\hat{\theta}$ and take the sum of $f_Z(W - X^i\hat{\theta})$ divided by 3000. This yields the probability that $w \;\; W$, which we have written $f_W(W)$. End of inner loop.
2. Repeat for all values W to obtain all values of $f_W(w)$.

Given that the algorithm is based on a discretization, we renormalize probabilities $f_w(w)$ so that they exactly sum to 1.

2.5. Relation to β convergence

It is useful to illustrate how our approach to conditional convergence relates to the standard β-convergence literature and how conditional convergence can be implemented in the nonlinear β-convergence model. Since Eq. (9) represents a first-order stochastic process, there exists an equivalent representation of Eq. (9) of the form:

$$z^i_{t+1} = \phi(z^i_t) + e^i_{t+1} \tag{14}$$

where $\phi(.)$ is an arbitrary smooth function.[11] If $\phi(z^i_t) = \rho z^i_t$, we can write:

$$w^i_{t+1} = X^i\theta + z^i_{t+1}$$

$$= X^i\theta + \rho z^i_t + e^i_{t+1}$$

$$= X^i\theta + \rho(w^i_t - X^i\theta) + e^i_{t+1}$$

which can be rewritten:

$$w^i_{t+1} - w^i_i = (1 - \rho)X^i\theta + (\rho - 1)w^i_t + e^i_{t+1} \tag{15}$$

If w^i_t stands for log GDP per head, then the deterministic version of Eq. (15) is the standard conditional convergence model (Barro, 1991; Mankiw et al., 1992). What we estimate in this paper is a generalized version of model (15) where we replace the fixed parameter ρ with a smooth function $\phi(.)$ to yield:

$$w^i_{t+1} = X^i\theta + \phi(w^i_t - X^i\theta) + e^i_{t+1} \tag{16}$$

In the deterministic version of Eq. (16) the shape of function $\phi(.)$ captures the way in which w^i_{t+1} converges to its steady state $X^i\theta$. Eq. (16) can thus be seen as a generalization of the standard MRW model in which we do not impose linearity around the steady state and let the data tell us how rapidly the process converges depending on how far it is from its steady state. It can also identify the presence of multiple (deterministic) steady states and determine which ones are stable. As we will see, however, this approach to convergence is insufficiently informative when the true data generation process is stochastic because the shape of $\phi(.)$ by itself tells us little about σ-convergence. We therefore also compute conditional ergodic distributions.

Function $\phi(.)$ can be estimated by replacing, in Eq. (16), $w^i_t - X^i\theta$ with z^i_t (or a consistent estimate of it). After replacement, this boils down to applying a standard kernel regression to:

$$\hat{z}^i_{t+1} = \phi(\hat{z}^i_t) + e^i_{t+1} \tag{17}$$

In the unconditional case, we simply replace $\hat{z}^i_t$ and $\hat{z}^i_{t+1}$ with w^i_t and w^i_{t+1}.

3. The data

We now turn to the empirical implementation. As discussed in the Introduction, our goal is to predict what the future distribution of economic activity over space would look like, should

[11] For our illustration, it is enough to assume that the errors e^i_{t+1} are not autocorrelated, but they need not be homoskedastic.

current tendencies persist. We use job figures as a proxy for economic activity. County-level sectoral employment data come from the Regional Economic Information System (REIS) compiled by the U.S. Bureau of Economic Analysis (BEA). We use employment data for 1970 to 2000 in thirteen sectors, covering the entire economy: farming; agricultural services; mining; construction; manufacturing; transportation and utilities; wholesale; retail; FIRE (finance, insurance and real estate); other services; federal government; military; and state and local government. We focus on the contiguous U.S. because we believe that, over the period under investigation, labor and capital mobility towards Alaska and Hawaii were lower than now. Pooling them with the contiguous U.S. may therefore not be appropriate for our purpose. We are left with 3071 counties. Sectoral employment data are missing for some counties, either because they are unavailable or because they are not disclosed.[12]

Because the distribution of employment levels is approximately log-normal, we focus our analysis on the log of employment. By dramatically reducing heteroskedasticity, this limits the role of outliers and increases the robustness of our results. If we were to perform the analysis in employment levels, a handful of urban counties with a lot of employment would dominate the analysis. Our focus is on all counties.

To control for county-specific time-invariant characteristics, we use data on county area, latitude, and longitude from the U.S. Geological Survey. Counties are assumed to be centered at their county seat. The average county size is 2491 km^2, corresponding to an average diameter of approximately 50 km (30 mi).[13] Counties vary considerably in size, however: the coefficient of variation of county area is 1.36. Western counties in particular tend to be larger than their eastern counterparts. Dummies are also created to control for whether a county is on a large body of water, such as a lake or ocean, or for whether it is on the border with Canada or Mexico. In particular, we include dummies for: the Atlantic ocean; the Pacific ocean; the Great Lakes; the gulf of Mexico; the Mexican border; and the Canadian border. Information of proximity to borders and water was compiled from detailed maps provided by the American Automobile Association (AAA). Latitude and longitude are also included as regressors. Finally, given that economic activity in the U.S. is concentrated on the Atlantic and the Pacific seaboards, we add dummies for counties located in states on the East coast or the West coast.

4. Results

4.1. σ convergence and β convergence

To get a feel for whether jobs have become more or less concentrated across space, Table 1 reports the standard deviation of detrended log employment at the county level in 1970, 1980, 1990 and 2000. A decreasing standard deviation reflects log employment becoming more equally spread across counties, a phenomenon known as σ-convergence. An increasing standard deviation points to employment becoming more concentrated in space, with some counties having lots of jobs and some having very few. As can be seen, for total employment the tendency has been towards more concentration (divergence). This increasing concentration of aggregate employment can be seen even more clearly in Fig. 1, which plots the same standard deviation at an annual frequency. At the

[12] For some counties sectoral employment is not revealed in order not to violate employer confidentiality. For other counties sectoral employment is simply reported as ‘less than 10’; in those cases we set employment equal to 5.

[13] This approximation obviously underestimates the actual diameter, since counties are not perfect circles. It is nevertheless useful as a ballpark figure.

Table 1
Standard deviations of sectoral employment in 1972 and 1992 in logs

Standard deviation of log employment				
Sector	1970	1980	1990	2000
Total	1.33	1.37	1.43	1.46
Farming	.93	.84	.84	.86
Agricultural services	1.35	1.41	1.32	1.26
Mining	1.72	1.94	1.83	1.93
Construction	1.54	1.49	1.61	1.49
Manufacturing	1.99	1.88	1.80	1.71
Transportation/utilities	1.59	1.55	1.56	1.55
Wholesale	1.78	1.65	1.70	1.64
Retail	1.42	1.49	1.59	1.60
FIRE	1.55	1.60	1.64	1.60
Other services	1.51	1.56	1.62	1.64
Federal civilian	1.57	1.58	1.61	1.58
Military	1.55	1.56	1.54	1.55
State/local	1.29	1.35	1.36	1.37

Source: REIS, Bureau of Economic Analysis.

sectoral level, there is a clear difference between service and non-service sectors. Most services ('retail', 'FIRE', and 'other services') have become more concentrated; most other sectors, such as 'manufacturing' and 'farming', have exhibited deconcentration (convergence).

Fig. 2 shows a scatter plot of the log difference in total employment between 1970 and 2000 on the log of employment in 1970. At first sight it is difficult to see any pattern in this cloud of points. To get a clearer picture, Table 2 reports the results of a standard linear regression of annual employment growth between 1970 and 2000 on initial log employment in 1970 the standard test of unconditional β-convergence. A positive coefficient on initial employment points to concentration (divergence), whereas 14 a negative coefficient indicates deconcentration (convergence). Our findings from Table 1 are confirmed. There is concentration of employment at the aggregate level and in most service sectors ('retail' and 'other services'), and deconcentration in the other sectors. This suggests that services are driving aggregate employment dynamics. This

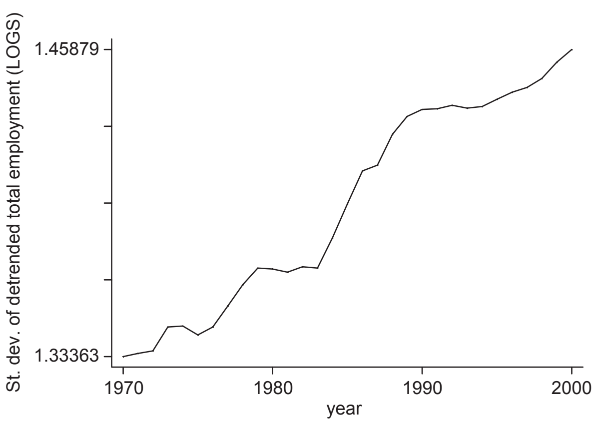

Fig. 1. Standard deviation of detrended total employment 1970 2000 (logs).

Fig. 2. Growth in total employment on initial log employment (1970 2000).

should not come as a surprise, given the weight of services in the economy: in 1970 'retail' and 'other services' already made up 34% of total employment; by 2000 this share had grown to 48%.

The increasing concentration of aggregate county employment stands in contrast with the observed employment deconcentration across metropolitan areas (Chatterjee and Carlino, 2001; Carlino and Chatterjee, 2002). This suggests deconcentration across large counties, and concentration across smaller sized counties. A quick and easy way of checking this is to split up our sample into two groups: the 200 counties with more than 82,215 workers in 1970, and the remaining 2871 counties. Although the overlap is not perfect, almost all of the 200 largest counties are classified as 'urban' by the Office of Management and Budget. As expected, for the group of large metro counties we find deconcentration across the board. Table 3 shows a negative coefficient on initial log employment for all sectors. In contrast, for the group of smaller counties Table 4 shows total employment becoming more concentrated. Again, there is a dichotomy across sectors: concentration in services, and deconcentration in the rest of the economy. Summing up, standard convergence analysis indicates that the tendency towards employment deconcentration

Table 2
Sectoral employment growth on initial sectoral employment (all 3074 counties)

Dependent variable: annual growth rate in sectoral employment 1970 2000

	Total	Farming	Ag serv	Mining	Constr	Manuf	Trans/util
Emp1970	0.0014	−0.0044	−0.0057	−0.0047	−0.0036	−0.0059	−0.0037
	(6.74)	(21.59)	(12.42)	(10.56)	(12.49)	(26.11)	(14.06)
Constant	0.0047	0.0220	0.0733	0.0227	0.0506	0.0540	0.0420
	(2.56)	(15.72)	(34.25)	(11.27)	(28.60)	(32.38)	(26.07)
Adjusted R^2	0.0143	0.0246	0.092	0.0649	0.0533	0.1968	0.0675

	Wholesale	Retail	FIRE	Other serv	Fed civ	Milit	State/loc
Emp197	−0.0048	0.0022	−0.0007	0.0010	−0.0019	−0.0021	0.0004
	(17.35)	(9.16)	(2.69)	(4.20)	(9.16)	(10.73)	(2.08)
Constant	0.0556	0.0039	0.0290	0.0260	0.0143	0.0012	0.0159
	(35.65)	(2.22)	(18.48)	(15.14)	(13.80)	(1.14)	(11.89)
Adjusted R^2	0.1019	0.0265	0.0022	0.0057	0.0263	0.0358	0.0011

Absolute values of *t*-statistics in brackets.

Table 3
Sectoral employment growth on initial sectoral employment 1970 2000 (200 largest counties)

Dependent variable: annual growth rate in sectoral employment 1970 2000

	Total	Farming	Ag serv	Mining	Constr	Manuf	Trans/util
Emp1970	−0.0037	−0.0007	−0.0058	−0.0083	−0.0058	−0.0114	−0.0074
	(3.00)	(1.45)	(4.52)	(6.05)	(3.79)	(8.85)	(5.69)
Constant	0.0649	−0.0049	0.0893	0.0571	0.0746	0.1170	0.0859
	(4.26)	(1.48)	(10.41)	(6.70)	(5.27)	(8.55)	(7.17)
Adjusted R^2	0.0387	0.0056	0.1120	0.1868	0.0642	0.2807	0.1362
	Wholesale	**Retail**	**FIRE**	**Other serv**	**Fed civ**	**Milit**	**State/Loc**
Emp1970	−0.0100	−0.0066	−0.0037	−0.0031	−0.0058	−0.0058	−0.0028
	(7.87)	(5.19)	(3.09)	(2.95)	(7.22)	(5.91)	(2.80)
Constant	0.1122	0.0889	0.0604	0.0716	0.0525	0.0308	0.0456
	(9.64)	(6.77)	(5.27)	(6.38)	(7.67)	(3.73)	(4.57)
Adjusted R^2	0.2344	0.1153	0.0412	0.0372	0.2046	0.1456	0.0333

Absolute values of *t*-statistics in brackets.

only holds for a limited group of high employment metropolitan counties. For the rest of the distribution, concentration seems to be the norm. This finding confirms the results of Beeson and DeJong (2002) for population.

4.2. Unconditional kernel regressions

Splitting up the sample into two parts, and running linear regressions on each part, is a rather rudimentary way of dealing with nonlinearities. A more appropriate way of capturing the richness of the dynamics is to run nonlinear kernel regressions on the entire sample. Because we are interested in long-run trends, not in trade cycles, we focus on 10-year intervals, i.e., we only use data from 1970, 1980, 1990 and 2000.

Table 4
Sectoral employment growth on initial sectoral employment (2874 smallest counties)

Dependent variable: annual growth rate in sectoral employment 1970 2000

	Total	Farming	Ag serv	Mining	Constr	Manuf	Trans/util
Emp1970	0.0020	−0.0054	−0.0092	−0.0053	−0.0041	−0.0063	−0.0055
	(7.46)	(23.78)	(16.23)	(10.61)	(11.16)	(23.44)	(16.08)
Constant	−0.0005	0.0285	0.0867	0.0232	0.0533	0.0560	0.0510
	(0.22)	(18.51)	(34.66)	(10.99)	(24.70)	(29.74)	(26.04)
Adjusted R^2	0.0187	0.1645	0.1619	0.0727	0.0461	0.1753	0.0929
	Wholesale	**Retail**	**FIRE**	**Other serv**	**Fed civ**	**Milit**	**State/loc**
Emp1970	−0.0061	0.0036	−0.0011	0.0007	−0.0025	−0.0018	0.0009
	(17.00)	(11.49)	(3.21)	(2.26)	(9.41)	(7.30)	(3.76)
Constant	0.0616	−0.0052	0.0311	0.0278	0.0170	−0.0002	0.0125
	(32.50)	(2.36)	(15.62)	(12.89)	(13.31)	(0.17)	(7.45)
Adjusted R^2	0.1054	0.044	0.0035	0.0015	0.0296	0.0179	0.0046

Absolute values of *t*-statistics in brackets.

The unconditional estimating equation is of the form:

$$w^i_{t+10} = \phi(w^i_t) + e^i_{t+1} \tag{18}$$

where w^i_t is (detrended) log employment in year t in county i. The estimation uses an Epanechnikov kernel with optimal bandwidth.[14] To facilitate interpretation, we have plotted the annual employment *growth* as a function of initial log employment. In this case, a negative slope indicates deconcentration (convergence) and a positive slope indicates concentration (divergence). When the curve cuts the horizontal axis from above, we have a stable equilibrium; when it cuts it from below, we have an unstable equilibrium.

Fig. 3 plots the results for total employment and by sector. The graphs also report a robust 95% confidence interval around the kernel regression.[15] We start by looking at the picture for total employment (‘Total’). The curve is upward sloping across much of the distribution, and cuts the horizontal axis somewhere in the middle. This suggests that at intermediate values of initial employment, forces exist that push total employment away towards the extremes. In other words, the middle part of the distribution exhibits divergence in the deterministic sense: if employment starts off below the middle equilibrium, the county is on average predicted to lose jobs, and is expected to converge towards the low steady state. In contrast, if employment starts off above the middle equilibrium, the county is expected to gain jobs and converge towards the high steady state. Note that the slope of the estimated curve turns negative towards the upper end of the distribution. This suggests convergence amongst metropolitan counties, a result in line with Chatterjee and Carlino (2001). These findings can be quantified. An estimated 52% of the counties, i.e., all counties with less than 7720 jobs in 1970, is predicted to slowly empty out; the remaining half is predicted to gain jobs and end up in the high steady state. Regarding the upper tail, the 8% largest counties corresponding to those with more than 60,818 jobs in 1970 exhibit convergence.

Turning to individual sectors, we see that for most of the non-service sectors such as ‘manufacturing’ and ‘construction’ the slope tends to be negative. This suggests deconcentration (convergence). The deterministic steady state is where the curve cuts the horizontal axis. In contrast, for the service sectors ‘retail’, ‘FIRE’, and ‘other services’ the picture resembles that of aggregate employment. In ‘other services’, for instance, there is a steady state with low retail employment and a steady state with high retail employment, with the middle group disappearing. More specifically, the model predicts that 62% of the counties will end up in the low steady state, and the remaining 38% will end up in the high steady state. Note that government employment at the state and local level looks much like the other service sectors. Finally, ‘farming’ exhibits convergence across much of the distribution, but divergence in the upper tail. This tells us that some of the larger farming counties are becoming increasingly specialized.

[14] The size of the optimal bandwidth is obtained by cross-validation as follows. Pick a bandwidth κ. For each observation w^i, estimate a κ bandwidth kernel regression that omits observation w^i. Let $\hat{w}^{i\kappa}$ denote the fitted value from this local kernel regression. Obtain the residual $\hat{u}^{i\kappa}$ $w^i - \hat{w}^{i\kappa}$. Repeat this procedure for all observations and compute the sum of squared residuals S^κ. The optimal bandwidth is the value of κ that minimizes S^κ (Silverman, 1986).

[15] The 95% confidence interval is equal to the kernel estimate plus and minus 1.96 times the robust standard error of the intercept in each kernel regression (Silverman, 1986). Since each standard error is robustly estimated and the standard error of the intercept varies across the sample, this approach is robust to the presence of heteroskedasticity in the data.

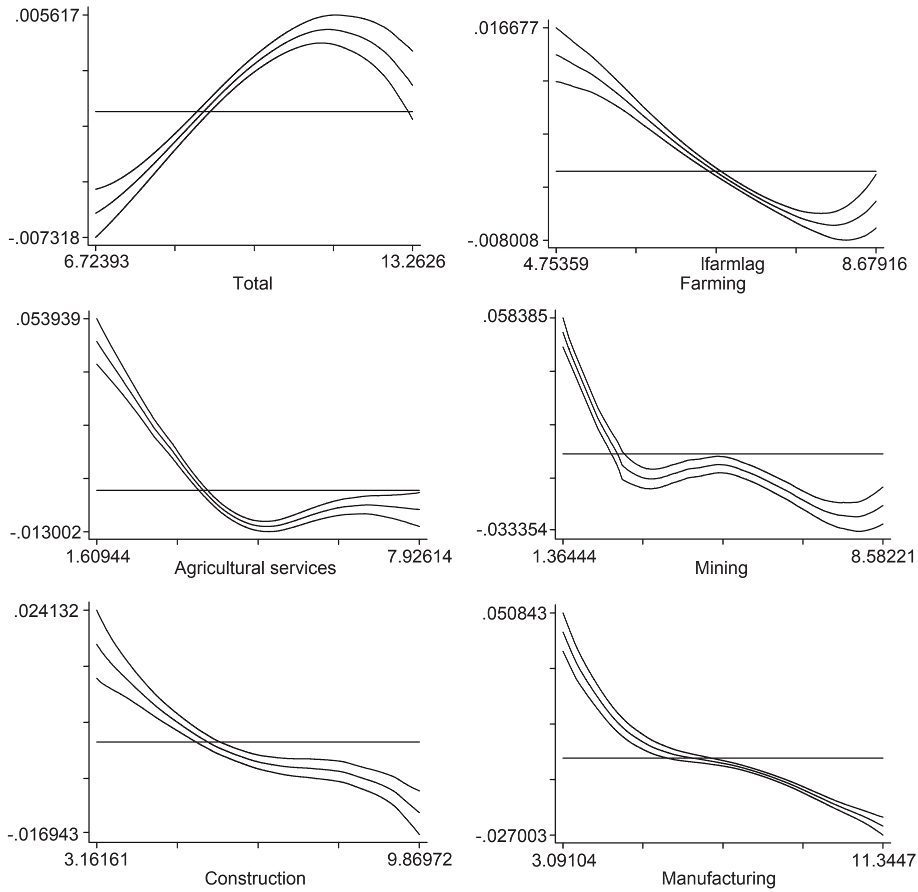

Fig. 3. Unconditional kernel regressions. Employment growth on initial log employment (1970 2000). 95% confidence intervals based on robust standard errors.

4.3. Unconditional ergodic distributions

Interpreting kernel regressions results as indications of convergence relies on the implicit assumption that y_t^i can reasonably be approximated by a deterministic model, with all counties ending up near one (or several) steady states. Put differently, such an interpretation implicitly requires that the distribution over time converges to mass points on the steady states. In the case of a single steady state, this is equivalent to σ-convergence.[16] To investigate convergence in the distribution itself, we need to examine the ergodic distribution. If this distribution converges to one mass point, we have σ-convergence. A β-convergence analysis will then reveal where this mass point is.

[16] The kernel regression basically describes the conditional mean of the transition matrix, averaging across rows. σ-convergence depends not only on the conditional mean but also on dispersion around this mean. It is easy to construct transition matrices that display β-convergence but no σ-convergence.

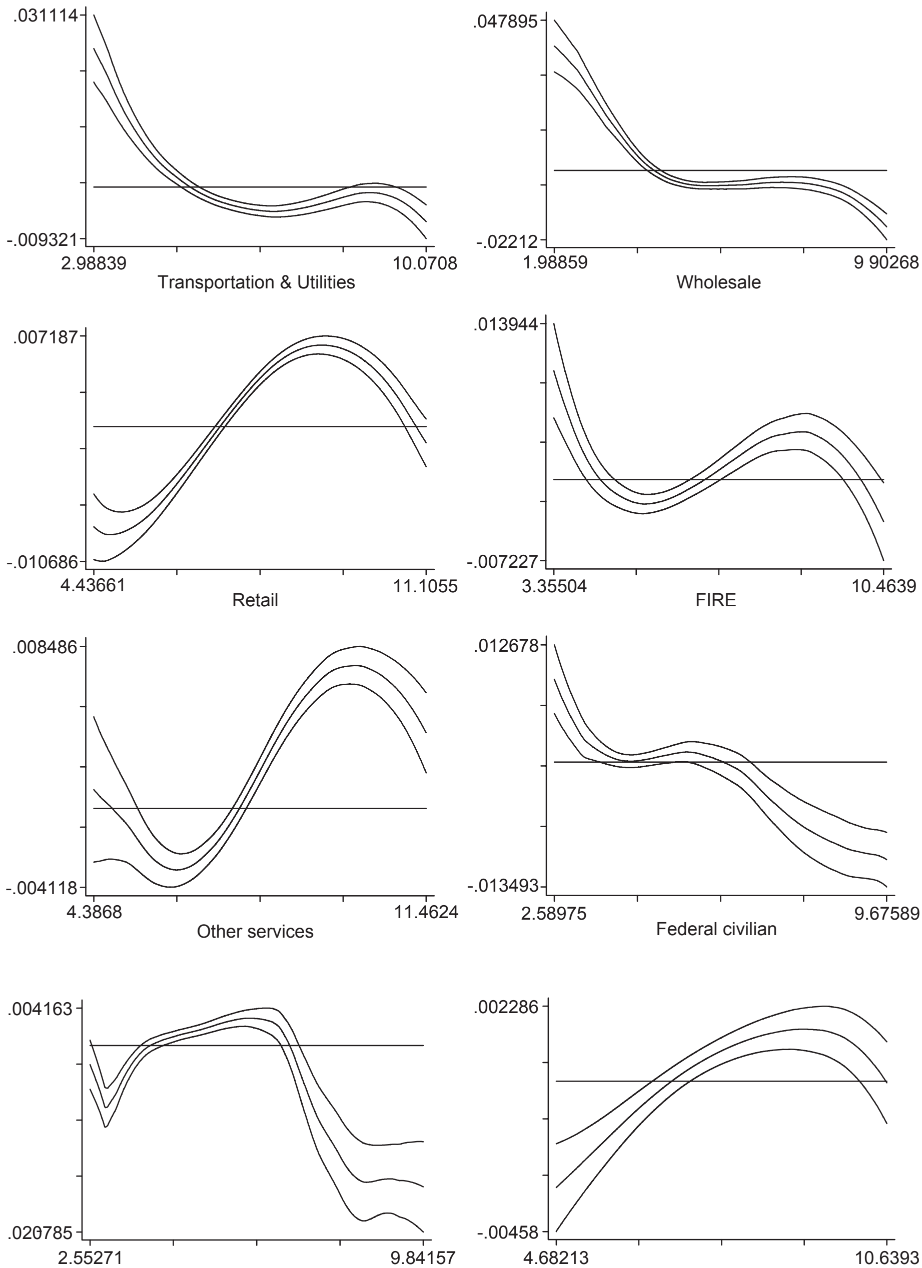

Fig. 3 (*continued*).

It is important to understand the objective and limitations of such an exercise. First, the ergodic distribution is nothing but a convenient way of depicting the trend in the shape of a distribution. For the actual distribution to converge to the ergodic distribution, the transition matrix would have to remain unchanged. We are not saying this will be the case.[17] In fact we expect transition matrices to vary over time with changes in technology (e.g., air conditioning and move to the sunbelt) and in preferences (e.g., concerns about pollution and move of industry out of urban centers). The ergodic distribution depicts what the distribution would look like in the long run if current conditions remain the same. Second, the ergodic distribution represents how y_t^i evolves around its mean. It does not say anything about the mean itself. With population growth and immigration, we expect U.S. total employment to continue to grow but our focus is on the geographical distribution of employment, not on the trend. Third, in order to obtain a reasonable approximation of the transition matrix, it is important that the data be distributed evenly. Kernel smoothing does a poor job of approximating highly skewed distributions. As mentioned before, this is the reason why we focus on the log of employment.

Fig. 4 plots the ergodic distributions for all sectors, and compares them to the distributions of employment in 1970 and 2000. The method used to derive the ergodic distribution is described in the methodology section. Log employment data is detrended using Eq. (8), that is, we estimate the ergodic distribution of:

$$w_{t+s}^i = y_{t+s}^i \quad E[y_{t+s}^i] + E[y_0^i]$$

where, as before, y is log employment and y_0^i is employment in the first year for which we have the data, that is, 1970. The ergodic distribution is obtained by inverting ten-year ahead transition matrices computed using data from 1970, 1980, 1990 and 2000. An appropriate kernel bandwidth is selected as follows. We begin by calculating the kernel bandwidth that minimizes the mean integrated squared error for univariate log employment densities.[18] In all cases, the optimal bandwidth is between 0.08 and 0.11 in relative terms suggesting that a bandwidth of 0.1 is a good choice.[19] This is the bandwidth we use to calculate the bivariate kernel from which the transition matrix is extracted.

If the ergodic distribution is tighter than the distributions in 1970 and 2000, this suggests deconcentration in log employment: based on current trends, counties are predicted to look more alike in the future than in the past. If the ergodic distribution has a thinner upper tail than the actual employment distributions for 1970 and 2000, this means fewer high employment counties in the future. In contrast, if the ergodic distribution has a ‘hat shape’ instead of a ‘bell shape’, with more mass on high employment values, there will be more concentration in the future if current trends continue: the number of high employment counties is predicted to increase, while the number of intermediate employment counties is expected to fall.

Results shown in Fig. 4 fall basically into three categories. First, some sectors exhibit hat shapes or even twin peaks: ‘total’, ‘retail’, ‘FIRE’ and ‘other services’. This suggests increasing concentration. Second, some sectors, such as ‘farming’, ‘manufacturing’ and ‘construction’, exhibit (slightly skewed) bell shapes. In the case of ‘farming’ and ‘manufacturing’ the ergodic distribution is tighter, indicating a tendency towards further convergence. In contrast, in the case of ‘construction’ the ergodic distribution is not tighter, suggesting there will not be any further convergence in the future. Third, the government sectors also give bell shapes, but with a

[17] Computing an ergodic distribution is like computing the trend in (the mean of) a variable: computing the trend does not imply that the researcher believes the trend will remain the same forever. It is just a way of representing a tendency in the data at a moment in time.

[18] This procedure yields an optimal bandwidth if the data are Gaussian, which is approximately the case here. The Stata 9 kdensity command is used to calculate the optimized bandwidth.

[19] A relative bandwidth of 0.1 means that observations covering 10% of the range of the data are used in each kernel regression.

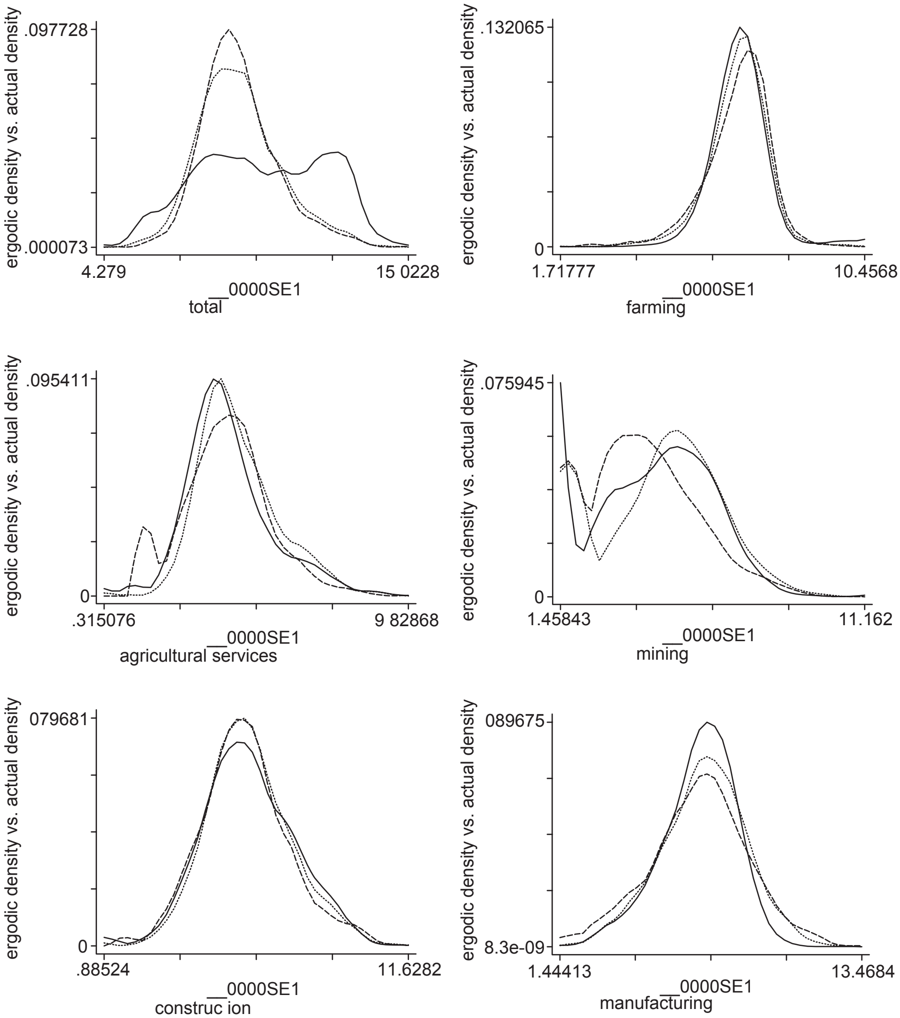

Fig. 4. Unconditional ergodic distributions. Ergodic distributions are given by the full-curve ( ). For comparison purposes, the distributions in 1970 are represented by the dashed curve (---), whereas the distributions in 2000 are given by the dotted curves (…).

distinctly fatter upper tail. This suggests an increasing presence of a small number of counties with a high level of public employment.

These results by and large confirm our findings in the kernel regressions. Total employment is becoming concentrated over time. This means that U.S. counties are becoming more differentiated in terms of employment, with more counties with little if any employment, more counties with high employment, and fewer with intermediate employment. This phenomenon at the aggregate level is a reflection of increased concentration in ‘retail’, ‘FIRE’ and ‘other services’, and to a smaller extent in public employment. All other sectors are predicted to either remain at their current level of concentration or to become less concentrated.

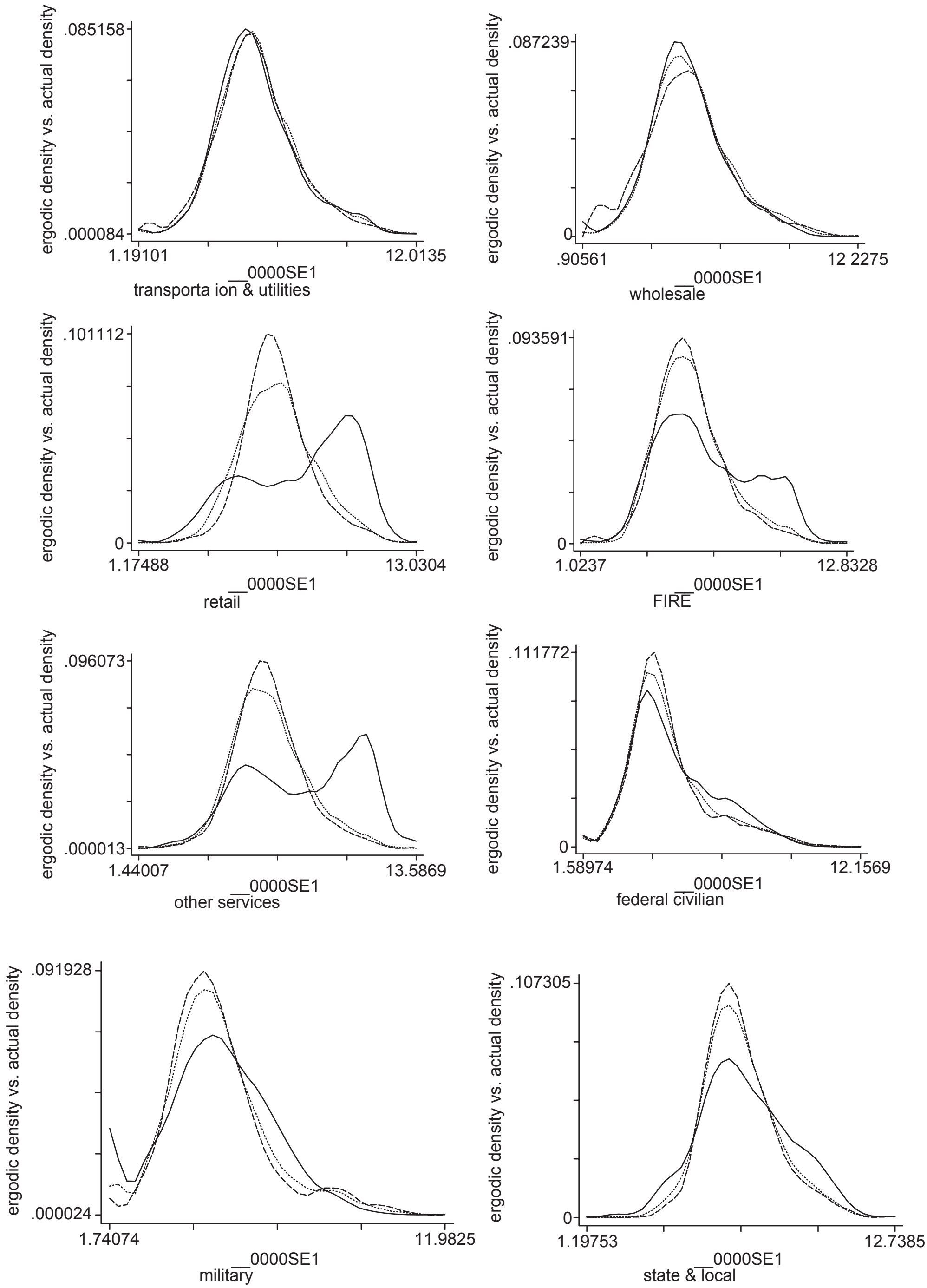

Fig. 4 (*continued*).

If we compare Fig. 4 with Fig. 3, we note that hat-shaped ergodic distributions arise whenever the kernel regression has two clearly identified stable equilibria. This is not surprising: the two deterministic steady states can be thought of as attracting points to which employment realizations tend. Stochastic shocks, however, ensure that employment does not settle in either steady state. Intuitively, the hat shape of the ergodic distribution results from the ‘mixing’ of the distributions around each of the two deterministic steady states. The kernel regression, however, fails to predict the extent of mixing and is therefore less informative. In contrast, whenever the kernel regressions suggest convergence to a single steady state, the ergodic distribution exhibits a bell shape. The kernel regression does not, however, indicate whether future concentration will differ from current concentration. This information is only obtained by calculating the ergodic distribution. We therefore see that kernel regressions which are themselves a generalization of standard β-convergence tests are less informative than the ergodic distribution in identifying the direction of change.

As mentioned before, in computing the ergodic distributions, we have used a 10-year ahead transition matrix. One question is whether our results are robust to that choice. To address this issue, Fig. 5 shows the ergodic distribution of total employment using a 5-year ahead transition matrix and a 15-year ahead transition matrix. As can be seen, qualitatively our results go through: total employment is becoming increasingly concentrated. The ergodic distribution changes its shape, however. In particular, using a 5 year interval to estimate the transition matrix seems to lead to less divergence. This makes sense: in as far as short term shocks are mean reverting (e.g., trade cycles), one would expect shorter time lags to lead to more convergence. This is the Galton fallacy argument as revisited by Quah (1993). In contrast, those mean reverting short term shocks are largely absent once we move to 10-or 15-year lags, so that we get more evidence of divergence.

As already mentioned, we do not really view the ergodic distribution as a reliable picture of the way the world will look like in the future; instead, we believe ergodic distributions are useful as a way of visualizing current trends. Be that as it may, it is still interesting to analyze how long it takes to get ‘close’ to the ergodic distribution. Focusing on total employment, Fig. 6 shows what the transition matrices imply about where the distribution would be in 2050, 2100 and 2200, if

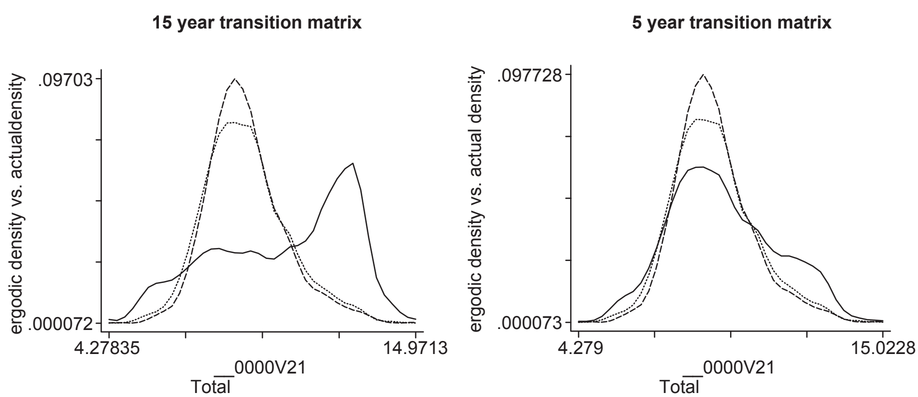

Fig. 5. Unconditional ergodic distributions of total employment: effect of time intervals. In Fig. 4 the ergodic distributions are computed using a 10-year ahead transition matrix. In this figure we show the ergodic distributions of total employment using a 15-year and a 5-year ahead transitions matrix.

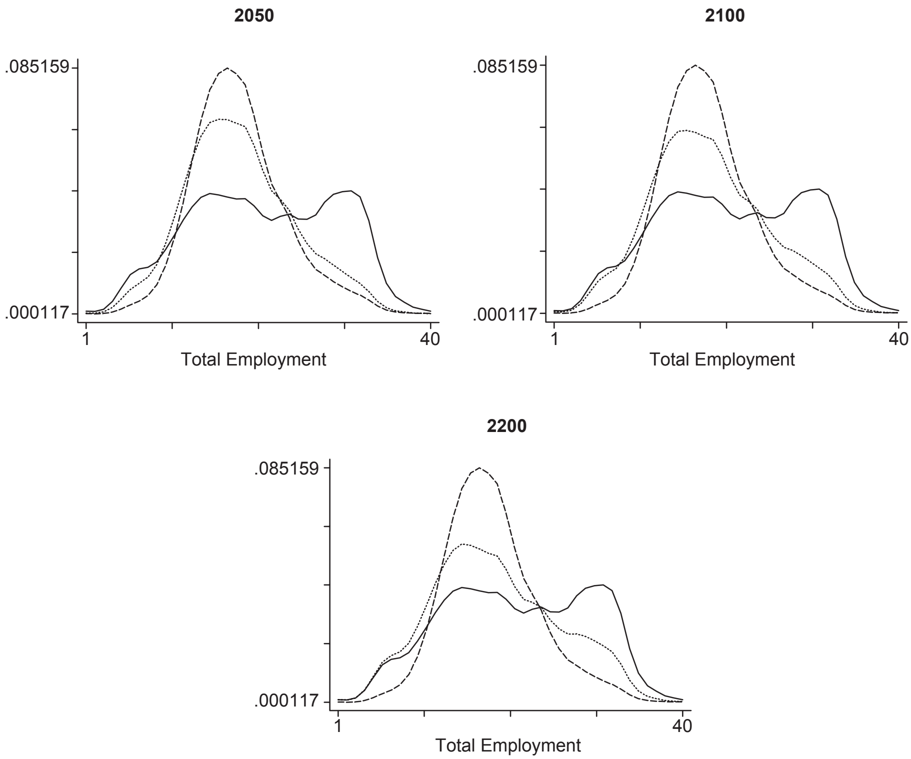

Fig. 6. Estimated distributions of total employment in 2050, 2100, 2200. Ergodic distributions are given by the full-curve ( ). The distributions in 1970 is represented by the dashed curved (---), whereas the distributions in 2050, 2100, 2200 are given by the dotted curves (...).

current trends were to continue.[20] As can be seen, change is slow. This is consistent with the fact that the transition matrix is tightly distributed around its diagonal.

4.4. Conditional kernel regressions and ergodic distributions

The analysis presented so far may be misleading if the distribution of employment across counties partly reflects time-invariant differences. Ergodic distributions computed without conditioning on these differences may underestimate the magnitude of stochastic shocks and thus misrepresent the long-term distribution of employment across counties.

To deal with this problem, we turn to the conditional model discussed in Section 2. As explained there, we first run a pooled regression of county employment on county characteristics. This regression has the form (11). The X^i characteristics include a variety of geographical features for which we have data.[21] We then obtain the $\hat{z}_t^i$ using (12). These $\hat{z}_t^i$ are then used to calculate a new set of kernel regressions and a new set of ergodic distributions using formula (13). The

[20] Fig. 6 is obtained by iterating on the transition matrix using: $p_{t+s}\ \ A^s p_t$, where p_t is a vector representing the frequency distribution of log employment and A is the transition matrix.

[21] The complete list was given in the data section (Section 3).

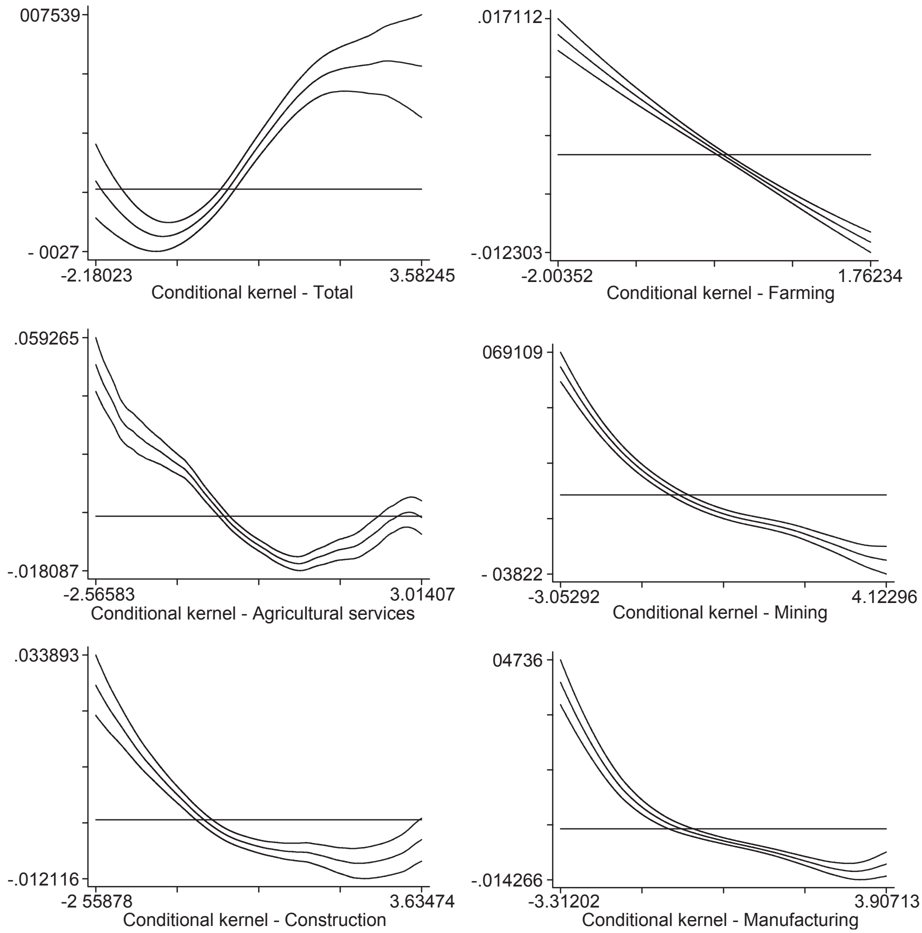

Fig. 7. Conditional kernel regressions. Employment growth on initial log employment (1970 2000). 95% confidence intervals based on robust standard errors.

methodology followed is the same as for the unconditional case, except that it is applied to $\hat{z}_t^i$ instead of to detrended log-employment w_t^i.[22] We use data of 1970, 1980, 1990 and 2000.

Fig. 7 plots the outcome of the kernel regressions of the $\hat{z}_t^i$ obtained using Eq. (17). As before, kernel bandwidth is optimized using cross-validation. As explained in Section 2, this approach is equivalent to the standard conditional β-convergence approach, except that it allows for nonlinearities. Using this kernel regression to draw inference about convergence in employment implicitly assumes that the time-varying component of w_t^i can be approximated by a deterministic process.

Comparing Fig. 7 with Fig. 3, we again get divergence in the middle part of the distribution of 'total' employment. There is some evidence of convergence in the lower tail, whereas the upper tail is now flat. Turning to individual sectors, our previous results are confirmed. The non-service

[22] Since $\hat{z}_t^i$ is estimated using detrended log employment w_t^i it is itself detrended.

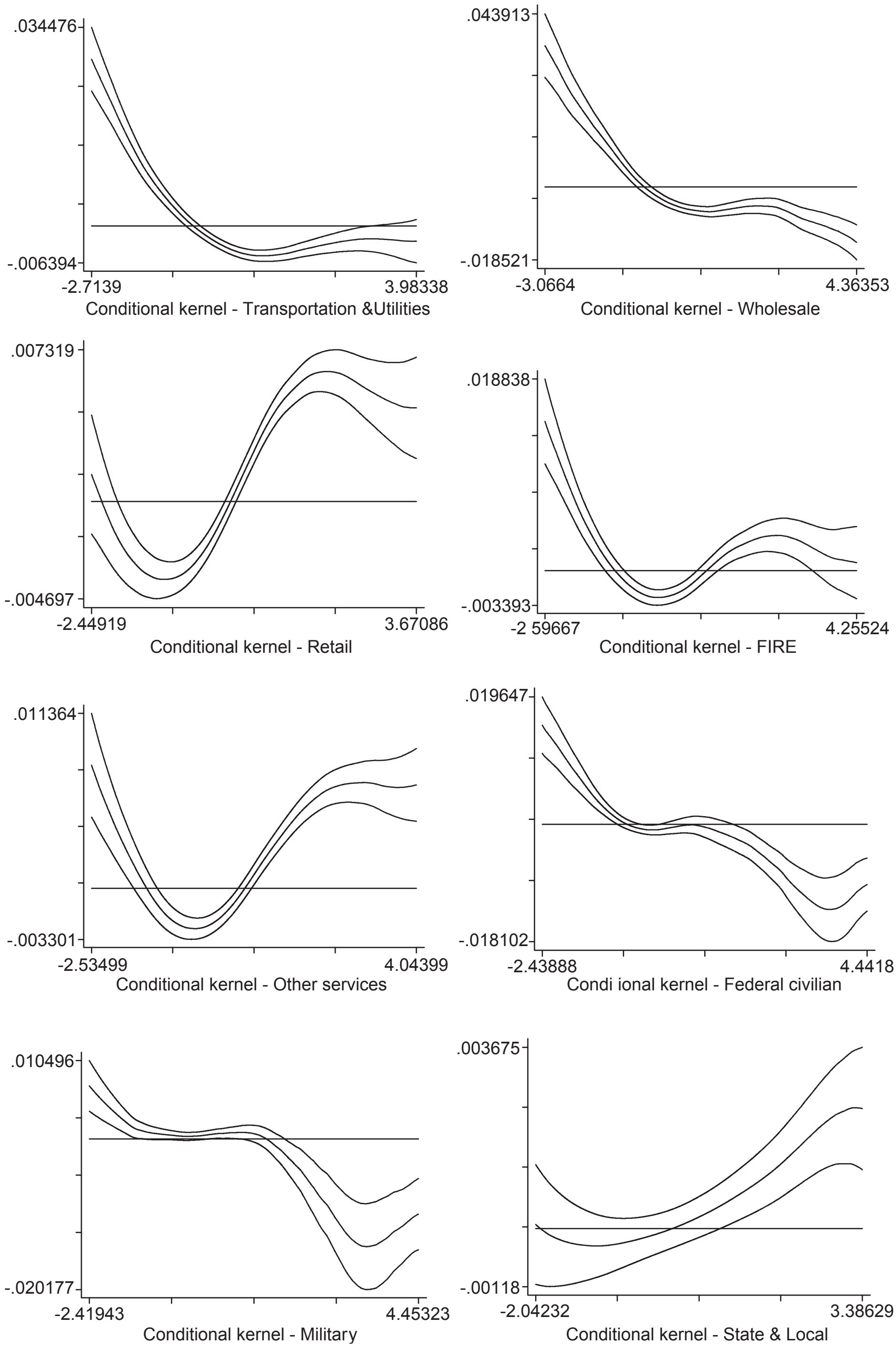

Fig. 7 (*continued*).

sectors exhibit mostly deconcentration (convergence), whereas most of the service sectors 'retail', 'FIRE', 'other services' and 'local and state employment' are becoming more concentrated over time. In other words, conditioning on time-invariant county characteristics does not change the basic story: there is concentration at the aggregate level, and this concentration is driven by the service sectors. This implies that observed concentration is not due to the geographical differences between counties on which we conditioned.

Conditional ergodic distributions are presented in Fig. 8. As explained in the methodology section, these ergodic distributions are constructed by regressing w_t^i on X^i to remove the time-invariant part $X^i\theta$, calculating the transition matrix of $\hat{z}_t^i$, obtaining the ergodic distribution of $\hat{z}_t^i$,

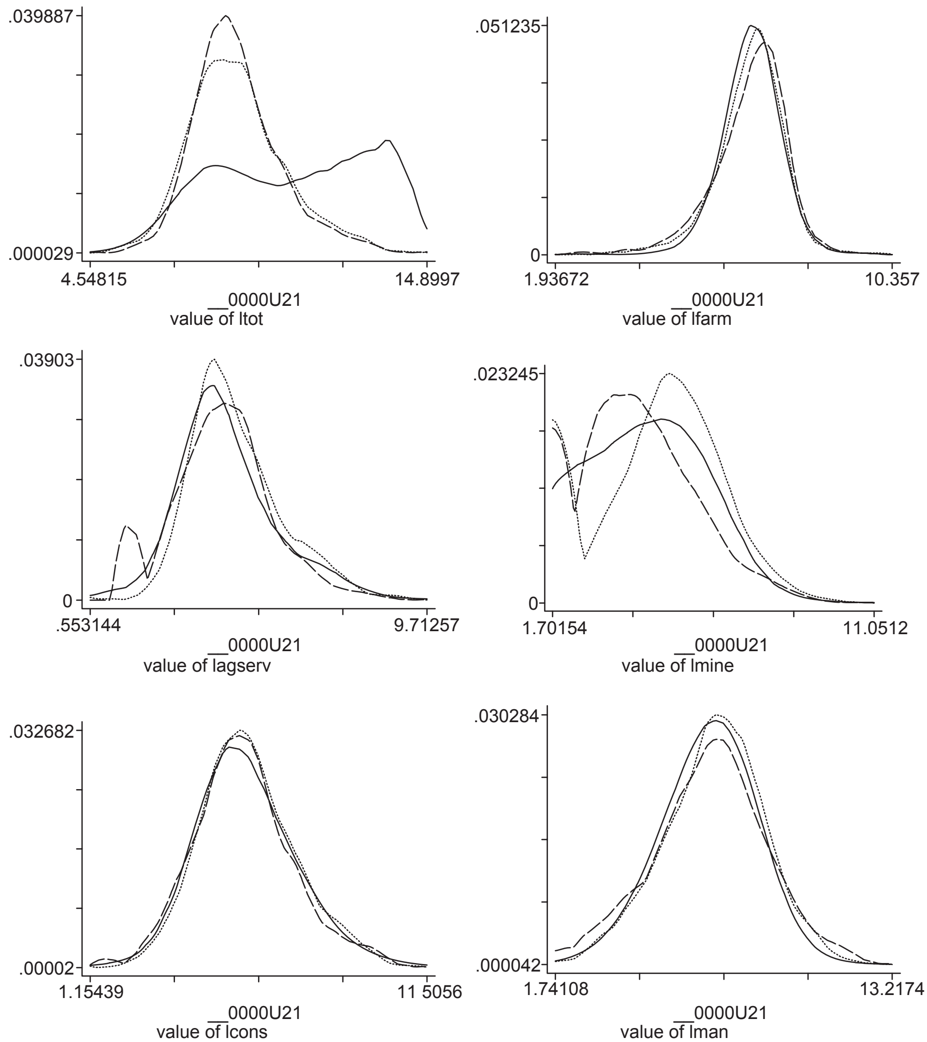

Fig. 8. Conditional ergodic distributions. Ergodic distributions are given by the full-curve ( ). For comparison purposes, the distributions in 1970 are represented by the dashed curve (---), whereas the distributions in 2000 are given by the dotted dots (...).

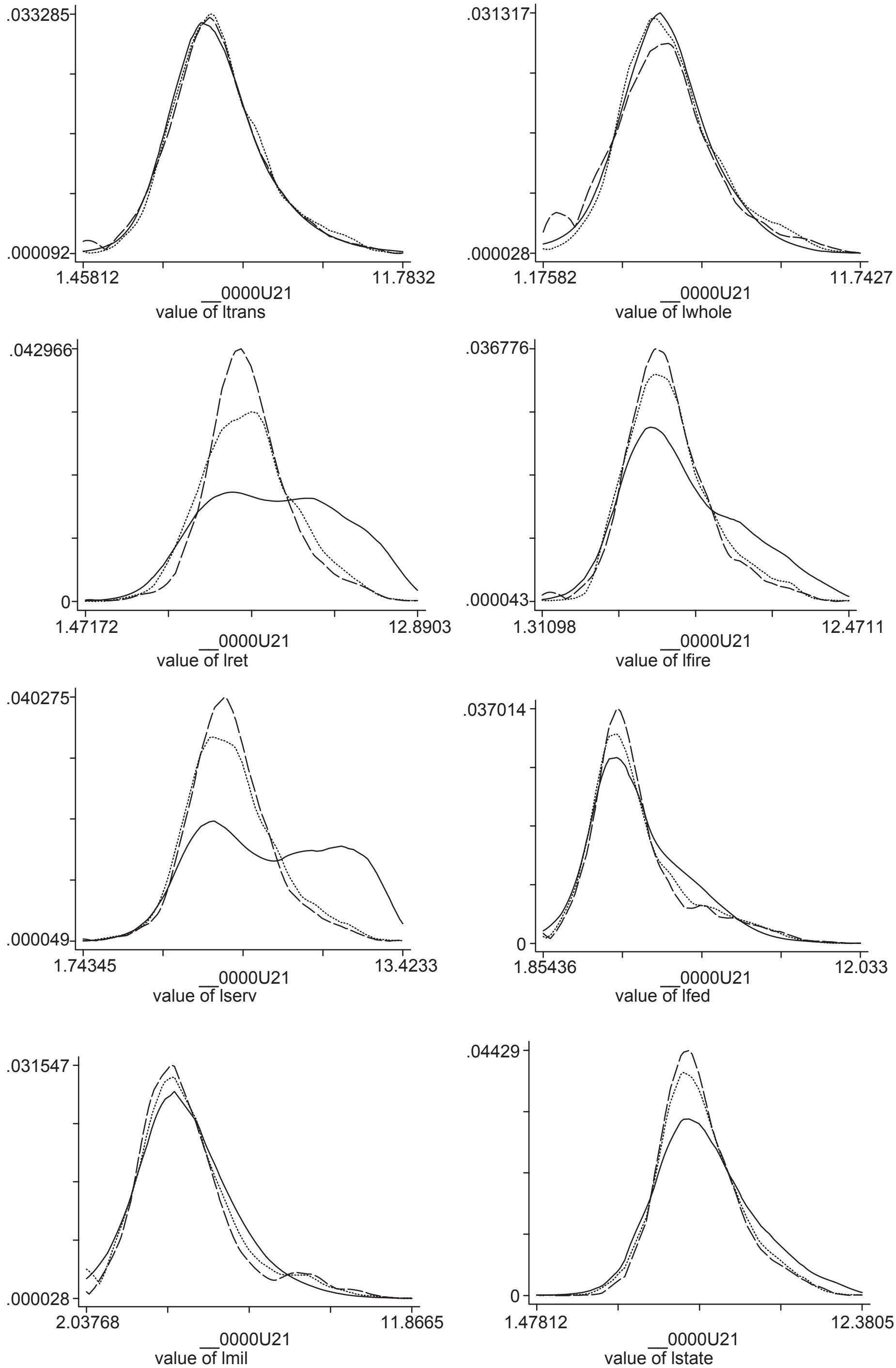

Fig. 8 (*continued*).

and finally adding $X^i\hat{\theta}$ back in using the switch in variable procedure detailed in footnote 10. The optimal bandwidth is chosen following the same methodology as in the unconditional case.

The basic pattern is similar to that depicted in Fig. 4: ‘total’ employment, ‘retail’, ‘FIRE, and ‘other services’ exhibit a hat-shaped ergodic distribution, suggesting increasing concentration. Taken together, these results suggest increasing concentration services. The results for non-service sectors are also broadly similar to Fig. 4, although in several instances the conditional ergodic distributions display less change than the unconditional ones. This suggests distributions are changing little over time, once geographical distributions are taken into account. This is especially clear in the manufacturing sector. Whereas the unconditional ergodic distribution has become tighter indicating further convergence this is no longer true for the conditional ergodic distribution.

5. Concluding remarks

In this paper we examined how the distribution of employment across U.S. counties is likely to evolve if current concentration and deconcentration forces remain unchanged. To do so, we developed a methodology borrowing from the work of Quah and building upon the literature on β- and σ-convergence. We computed non-parametric β-convergence regressions, conditional and unconditional. Using non-parametric methods, we also computed detailed ergodic distributions for total employment and sectoral employment across U.S. counties.

Our results suggest that employment is becoming increasingly concentrated across counties. Although very large metro counties may be losing jobs, the proportion of counties with modal employment is decreasing in favor of medium to high employment counties. More specifically, the 8% largest counties exhibit deconcentration; the remaining 82% exhibit concentration. This result is consistent with deconcentration across urban areas (Chatterjee and Carlino, 2001) and concentration across U.S. counties (Desmet and Fafchamps, 2005). It also confirms the results of Beeson and DeJong (2002) of population divergence across counties. Whether the overall picture is one of concentration or deconcentration is not entirely obvious. In terms of the number of counties, concentration holds the upper hand. However, in terms of the number of people, deconcentration dominates, since the 8% highest employment counties accounted for nearly two thirds of total employment in 1970.

There are important differences across sectors. As in the rest of the literature, we find deconcentration in manufacturing. Deconcentration is also the norm in other non-service sectors. However, service activities are becoming more concentrated, in particular ‘retail’ trade, ‘finance, insurance and real estate’, and ‘other services’. Given the importance of these sectors, they drive the evolution of the spatial distribution of total employment. Limiting the focus of analysis to manufacturing is misleading. The U.S. is a service economy, and services are behaving very differently from the other sectors.

Although we have limited our analysis to employment, our findings may shed light on the spatial dynamics of productivity and wages across the United States. Using county-level data, Ciccone and Hall (1996) conclude that doubling employment density leads to a 6% increase in productivity. Similar numbers have been found in subsequent studies by Harris and Ioannides (2000) for U.S. metropolitan areas and by Ciccone (2002) for European regions. Here we have looked at employment levels rather than at employment density. But our qualitative results remain basically unchanged if density is taken as the dependent variable. Following the insights of Ciccone and Hall (1996), we would expect 23 sectors that have been deconcentrating such as manufacturing to have experienced a fall in spatial productivity (and wage) differences. The opposite should have occurred for aggregate employment and services.

These predictions about productivity and wages are speculative, and warrant further investigation. They are based on a world in which employment dynamics are driven by changing agglomeration and congestion effects on the production side. An example of such approach can be found in Chatterjee and Carlino (2001) who argue that rising aggregate employment causes congestion costs to rise faster in more dense areas, leading to deconcentration of jobs. However, other forces such as congestion on the consumption side or a change in people's preferences may also be at work. In that case, the picture may be more complex. For instance, soaring house prices in urban areas could be consistent with densely populated areas losing employment but experiencing rising wages. Similarly, if people have an increasing preference to live in low density areas (Beale, 1977), this may lead to deconcentration of employment but increasing wage differentials.

On the methodological side, our research shows the importance of using non-parametric methods and of looking at the entire distribution, not just at cities. It also demonstrates that β-convergence tests, even when done non-parametrically, are not sufficiently informative. Computing the ergodic distribution associated with a given set of transition probabilities is more useful to understand spatial trends. Moreover, our approach is able to condition on time-invariant characteristics in a way that is fully consistent with standard analysis of conditional β-convergence. The methodology developed here can easily be applied to the study of any distributional dynamics.

This paper leaves a number of other questions unanswered. First, it is unclear whether the forces identified here operate in a similar manner in other time periods and other parts of the world. Applying the same approach to other data sets is necessary before we can conclude that the process described here generalizes beyond the confines of this study. Second, the methodology presented here does not (yet) allow statistical inference in the normal sense. Statistical tests are reported for some of the statistics presented here, such as confidence intervals for kernel regressions. But we do not present a 'test' of (conditional or unconditional) convergence based on estimated ergodic distributions. In principle, such a test could be developed provided an intuitively satisfying counter-factual distribution could be devised. It should also be possible to use bootstrapping to test whether the mode of the ergodic distribution has shifted to the left or the right relative to the current distribution (Kremer et al., 2000). Developing such tests is left for future research.

Acknowledgements

We thank participants at the CEPR conference on Integration and Technological Change (Paris, June 2004). We also benefitted from the comments of two anonymous referees and the editor. Financial aid from the Spanish Ministry of Education (SEJ2005-05831), the Ramón y Cajal program, the Comunidad de Madrid (06/0096/2003), and the Fundación Ramón Areces is gratefully acknowledged.

References

Beale, C., 1977. The recent shift of United States population to nonmetropolitan areas, 1970 1975. International Regional Science Review 2, 113 122.

Beeson, P.E., DeJong, D.N., 2002. Divergence. Contributions to Macroeconomics 2 (1) Article 6.

Beeson, P., DeJong, D., Troesken, W., 2001. Population growth in U.S. counties, 1840 1990. Regional Science and Urban Economics 31 (6), 669 699.

Barro, R.J., 1991. Economic growth in a cross-section of countries. Quarterly Journal of Economics 106 (2), 407 443.

Blanchard, O., Katz, L., 1991. Regional evolutions. Brookings Papers on Economic Activity 1, 1 75.

Carlino, G., Chatterjee, S., 2002. Employment deconcentration: a new perspective on America's postwar urban evolution. Journal of Regional Science 42, 455 475.
Chatterjee, S., Carlino, G.A., 2001. Aggregate metropolitan employment growth and the deconcentration of metropolitan employment. Journal of Monetary Economics 48, 549 583.
Ciccone, A., 2002. Agglomeration effects in Europe. European Economic Review 46, 213 227.
Ciccone, A., Hall, R.E., 1996. Productivity and the density of economic activity. American Economic Review 86, 54 70.
Desmet, K., Fafchamps, M., 2005. Changes in the spatial concentration of employment across U.S. counties: a sectoral analysis 1972 2000. Journal of Economic Geography 5, 261 284.
Dumais, G., Ellison, G., Glaeser, E., 2002. Geographic concentration as a dynamic process. Review of Economics and Statistics 84 (2), 193 204.
Durlauf, S.N., Quah, D.T., 1999. The new empirics of economic growth. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, vol. 1A. North Holland Elsevier Science, Amsterdam. Chapter 4.
Eeckhout, J., 2004. Gibrat's law for (all) cities. American Economic Review 94, 1429 1451.
Fujita, M., Krugman, P., Venables, A.J., 1999. The Spatial Economy: Cities, Regions, and International Trade. MIT Press, Cambridge, Mass.
Harris, T.F., Ioannides, Y.M., 2000. Productivity and Metropolitan Density, unpublished manuscript.
Henderson, J.V., 1988. Urban Development: Theory, Fact and Illusion. Oxford University Press, New York.
Higgins, M., Levy, D., Young, A., 2003. Growth and convergence across the U.S.: evidence from county-level data. Emory Economics Working Paper, vol. 0306.
Kim, S., 1995. Expansion of markets and the geographic distribution of economic activities: the tends in U.S. regional manufacturing structure, 1860 1987. Quarterly Journal of Economics 110, 881 908.
Kolko, J., 1999. Can I Get Some Service Here?, mimeo, Harvard University.
Kremer, M., Onatski, A., J. Stock, 2000. Searching for Prosperity, Department of Economics, Harvard University, mimeo.
Luenberger, D.G., 1979. Introduction to Dynamic Systems. Wiley, New York.
Mankiw, N.G., Romer, D., Weil, D.N., 1992. A Contribution to the Empirics of Economic Growth. Quarterly Journal of Economics 107, 407 437.
Mood, A.M., Graybill, F.A., Boes, D.C., 1974. Introduction to the Theory of Statistics. McGraw Hill.
Quah, D.T., 1993. Galton's fallacy and tests of the convergence hypothesis. Scandinavian Journal of Economics 95 (4), 427 443.
Quah, D.T., 1996a. Empirics for economic growth and convergence. European Economic Review 40, 1353 1375.
Quah, D.T., 1996b. Convergence empirics across economies with (some) capital mobility. Journal of Economic Growth 1, 95 124.
Quah, D.T., 1997. Empirics for growth and distribution: stratification, polarization, and convergence clubs. Journal of Economic Growth 2, 27 59.
Sala-i-Martin, X., 1996. Regional cohesion: evidence and theories of regional growth and convergence. European Economic Review 40, 1325 1352.
Silverman, B.W., 1986. Density Estimation: For Statistics and Data Analysis. Chapman and Hall, London.
Stokey, N.L., Lucas, R.E., Prescott, E.C., 1989. Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge, Mass.
Sutton, J., 1997. Gibrat's Legacy. Journal of Economic Literature 35, 40 59.