

DISCUSSION PAPER SERIES

No. 4689

EMPLOYMENT CONCENTRATION ACROSS US COUNTIES

Klaus Desmet and Marcel Fafchamps

INTERNATIONAL TRADE



Centre for **E**conomic **P**olicy **R**esearch

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP4689.asp

EMPLOYMENT CONCENTRATION ACROSS US COUNTIES

Klaus Desmet, Universidad Carlos III de Madrid and CEPR
Marcel Fafchamps, University of Oxford

Discussion Paper No. 4689
October 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL TRADE**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Klaus Desmet and Marcel Fafchamps

ABSTRACT

Employment Concentration Across US Counties*

This Paper examines the spatial distribution of jobs across US counties 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 US 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 and R12

Keywords: economic geography, ergodic distribution, spatial distribution of employment and US counties

Klaus Desmet
Departamento de Economía
Universidad Carlos III
28903 Getafe, (Madrid)
SPAIN
Tel: (34 91) 624 9845
Fax: (34 91) 624 9875
Email: desmet@eco.uc3m.es

Marcel Fafchamps
Department of Economics
University of Oxford
Manor Road
Oxford
OX1 3UL
Tel: (44 1865) 281446
Fax: (44 1865) 281447
Email: marcel.fafchamps@economics.ox.ac.uk

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=144849

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=116121

*We thank participants at the CEPR conference on Integration and Technological Change (Paris, June 2004). Financial aid from the Spanish Ministry of Science and Technology (BEC2002-03715), the Ramón and Cajal Program, and the Comunidad de Madrid (06/0096/2003) is gratefully acknowledged.

Submitted 04 August 2004

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, Krugman and Venables, 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.

This paper describes the geographical evolution of jobs in the U.S. between 1972 and 1992, 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, Levy and Young (2003) find 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 can address in our paper.¹ This seems similar to studying the distribution of population, a topic that has drawn a lot of attention, especially from researchers trying to document Zipf’s law for cities (see, e.g., Eaton and Eckstein, 1997). However, looking at employment, rather

¹If 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.

than population, has the advantage of allowing 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. This is probably because some of the more striking trends – such as suburbanization – have been affecting metropolitan areas. As a result, counties are somewhat understudied. However, as pointed out by Beeson, DeJong and Troesken (2001), limiting the analysis to urban areas introduces a selection bias, since cities are those areas which experienced high growth in the past. By studying counties, we do not have this problem, as we are looking at the entire distribution, and not just at its upper tail.

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, Romer and Weil, 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; Quah, 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 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 place 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), standard β -convergence tests using U.S. *county* data suggest the contrary, with jobs becoming more concentrated over recent decades (Desmet and Fafchamps, 2004). 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 Carlino and Chatterjee (2001) and Desmet and Fafchamps (2004).

The increased concentration evident in total employment stands in contrast to

²Moreover, 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.

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, Ellison and Glaeser, 2002; Kim, 1995).³ Our data confirm this empirical regularity. Although manufacturing cannot account for the spatial dynamics of aggregate employment, services can. Two of the three main service sectors — ‘retail’ and ‘finance, insurance and real estate’ — 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 ‘finance, insurance and real estate’, 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 manufacturing jobs moving to less dense areas, is not occurring. The reason is the rise of the service industry, particularly in ‘retail’ and in ‘finance, insurance, and real estate’ (Kolko, 1999).

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.

³Note that this is not the same as the other well documented phenomenon of suburbanization.

2.1 Unconditional convergence in a deterministic model

One approach to this question is that taken by the growth convergence literature (Durlauf and Quah, 1999). This 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, Romer and Weil, 1992; Quah, 1993):

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

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

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

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

$$\begin{aligned} Y^* &= e^{\frac{\alpha}{\beta}} \\ Y_0 &= 0 \end{aligned}$$

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 equation (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 equation (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:

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

where $\phi(\cdot)$ is an arbitrary smooth function. Equation (3) can be estimated by standard non-parametric techniques. If function $\phi(\cdot)$ cuts the 45 degree line more than once, the

process driving Y_t^i has multiple equilibria. Each point at which $\phi(\cdot)$ cuts the 45 degree 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 equation (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_t^i 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 $y_t^i \equiv \log Y_t^i$, equation (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_{it} 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_t^i is stochastic, a more adequate representation of its evolution over time is:

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

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. Equation (6) is itself a

deterministic law of motion. Provided that certain conditions are satisfied (Stokey, Lucas and Prescott, 1989; Luenberger, 1979), this system has a steady state or time-invariant distribution $f(y_{t+1})$ to which it converges.⁴ This time-invariant distribution is called the ergodic distribution. It is the distribution $f(y_{t+1})$ that satisfies:⁵

$$f(y_{t+1}) = \int_{-\infty}^{\infty} g(y_{t+1}|y_t)f(y_t)dy_t \quad (7)$$

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 (7) for a constant $f(\cdot)$. 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 $g(y_{t+1}|y_t)$

⁴In empirical applications, the most important issue that arises with respect to the existence of a non-degenerate ergodic distribution is that of de-trending. This is discussed in detail below.

⁵There might be multiple solutions to equation (7) and thus multiple ergodic distributions. Multiple solutions do not arise in our empirical analysis and are ignored here.

is obtained by dividing one by the other. For the third step, it is difficult to work with equation (7) directly. The standard approach in practice is to discretize the space of possible values y into N discrete cells Γ_k , with $k = \{1, \dots, N\}$. 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:

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

The ergodic distribution then a set of interval probabilities p_k that solves:⁶

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

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

⁶In matrix form we have:

$$\begin{aligned} p &= Ap \\ (I - A)p &= 0 \end{aligned}$$

It looks like the above system only has a solution of the form $p = 0$, 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:

$$\sum_i p_i = 1$$

We obtain a system of the form:

$$\begin{aligned} \begin{bmatrix} 1 - a_{11} & \dots & -a_{1N} \\ \dots & 1 - a_{ii} & \dots \\ -a_{N-1,1} & \dots & -a_{N-1,N} \\ 1 & \dots & 1 \end{bmatrix} [p] &= \begin{bmatrix} 0 \\ \dots \\ 0 \\ 1 \end{bmatrix} \\ Bp &= b \end{aligned}$$

The modified system can be solved by inverting:

$$p = B^{-1}b$$

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

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.⁷

An important practical detail is to make the data stationary. Suppose y_t^i has a trend component so that it is growing over time. If $f_t(y_t)$, and $f_{t+1,t}(y_{t+1}, y_t)$ are estimated directly from the data, the resulting ergodic distribution will be a mass point at the highest value for y_t . This is normal: if y_t^i grows without bounds, all realizations of y_t^i will eventually exceed the highest value of y_t^i at time t . The data must therefore be detrended first before undertaking steps (i) to (iii). The ergodic distribution then represents how the distribution of y_t^i evolves over time around its trend.

The shape of the ergodic distribution is potentially affected by whether the trend is linear or exponential. With a linear trend, the data generating process can be written:

$$\begin{aligned} y_t^i &= z_t^i \\ y_{t+s}^i &= \beta s + z_{t+s}^i \end{aligned} \tag{8}$$

where βs is a common time trend. This means that irrespective of their initial value, all y_t^i 's grow by the same quantity.⁸ Taking expectations we have:

$$\begin{aligned} E[y_t^i] &= E[z_t^i] \\ E[y_{t+s}^i] &= \beta s + E[z_{t+s}^i] \end{aligned}$$

Normalization means setting $E[z_t^i] = E[z_{t+s}^i]$. This yields:

$$\beta = \frac{E[y_{t+s}^i] - E[y_t^i]}{s}$$

or a sample equivalent. The corrected data for the second period is thus:

$$\begin{aligned} z_{t+s}^i &= y_{t+s}^i - \beta s \\ &= y_{t+s}^i - (E[y_{t+s}^i] - E[y_t^i]) \end{aligned}$$

⁷Quah (1996a, 1996b, 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.

⁸If $y_t^i = \log(x_t^i)$, this means that all x_t^i 's grow at the same rate.

If instead of a linear trend, we assume an exponential trend, the data generating process becomes:

$$\begin{aligned} y_t^i &= z_t^i \\ y_{t+s}^i &= z_{t+s}^i e^{\beta s} \end{aligned} \tag{9}$$

Growth regression analysis, for instance, works in logs and thus typically assumes an exponential trend. Again we want $E[z_t^i] = E[z_{t+s}^i]$. Straightforward algebra yields:

$$\begin{aligned} e^{\beta s} &= \frac{E[y_{t+s}^i]}{E[y_t^i]} \\ z_{t+s}^i &= y_{t+s}^i \frac{E[y_t^i]}{E[y_{t+s}^i]} \end{aligned}$$

Which approach is valid depends on the underlying data generation process. It is easy to characterize the bias introduced by choosing the wrong model. Suppose the true model is (8). Further suppose that there is positive growth, i.e., $(E[y_{t+s}^i] - E[y_t^i]) > 0$. Correcting using (9) yields the misspecified variable m_{t+s}^i :

$$\begin{aligned} m_{t+s}^i &= y_{t+s}^i \frac{E[y_t^i]}{E[y_{t+s}^i]} \\ m_{t+s}^i &> z_{t+s}^i \text{ if } y_{t+s}^i \text{ is small} \\ m_{t+s}^i &< z_{t+s}^i \text{ if } y_{t+s}^i \text{ is large} \end{aligned}$$

Whenever y_{t+s}^i and y_t^i are strongly correlated and y_{t+s}^i is close to y_t^i , using (9) incorrectly results in an overestimate of convergence probabilities, i.e., the transition up for low initial states and the transition down for high initial states. The opposite is also true: incorrectly applying (8) when (9) is true increases the variance of the ergodic distribution and thus underestimates convergence.

In our empirical analysis, we try both approaches. As it turns out, both yield results that are qualitatively very similar. Consequently, we only present the results based on an exponential trend.

2.4 Conditional convergence

Until now we have implicitly assumed that all y_t^i have the same distribution, 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 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:

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

with $E[Xz] = 0$ and:

$$f_{t+1}(z_{t+1}) = \int_{-\infty}^{\infty} g(z_{t+1}|z_t)f_t(z_t)dz_t \quad (10)$$

with corresponding ergodic equation

$$f(z_{t+1}) = \int_{-\infty}^{\infty} g(z_{t+1}|z_t)f(z_t)dz_t \quad (11)$$

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

$$\begin{aligned} y_t^i &= X^i\theta + z_t^i \\ &\dots \\ y_{t+s}^i &= X^i\theta + z_{t+s}^i \end{aligned} \quad (12)$$

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

$$\begin{aligned} \widehat{z}_t^i &= y_t^i - X^i\widehat{\theta} \\ &\dots \\ \widehat{z}_{t+s}^i &= y_{t+s}^i - X^i\widehat{\theta} \end{aligned} \quad (13)$$

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

The long-term distribution of y_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:

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

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

$$\begin{aligned} f_Y(y) &= \sum_w f_Z(y-w)f_W(w) \\ &= \sum_{X\widehat{\theta}} f_Z(y-X\widehat{\theta})f_W(X\widehat{\theta}) \end{aligned} \tag{14}$$

where $f_Z(\cdot)$ is the ergodic distribution function of z^i . Applying this formula to the data yields the conditional ergodic distribution of y_t^i .⁹

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 (10) represents a first-order stochastic process, there exists an equivalent representation of (10) of the form:

$$z_{t+1}^i = \phi(z_t^i) + e_{t+1}^i \tag{15}$$

⁹To compute the long-term probability of a particular value of $y = Y$, we proceed as follows. Say we have 3000 values of $X^i\widehat{\theta}$, each with frequency 1.

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

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

where $\phi(\cdot)$ is an arbitrary smooth function.¹⁰ If $\phi(z_t^i) = \rho z_t^i$, we can write:

$$\begin{aligned} y_{t+1}^i &= X^i \theta + z_{t+1}^i \\ &= X^i \theta + \rho z_t^i + e_{t+1}^i \\ &= X^i \theta + \rho (y_t^i - X^i \theta) + e_{t+1}^i \end{aligned}$$

which can be rewritten:

$$y_{t+1}^i - y_t^i = (1 - \rho)X^i \theta + (\rho - 1)y_t^i + e_{t+1}^i \quad (16)$$

If y_t^i stands for log GDP per head, then the deterministic version of equation (16) is the standard conditional convergence model (Barro, 1991; Mankiw, Romer and Weil, 1992). What we estimate in this paper is a generalized version of model (16) where we replace the fixed parameter ρ with a smooth function $\phi(\cdot)$ to yield:

$$y_{t+1}^i = X^i \theta + \phi(y_t^i - X^i \theta) + e_{t+1}^i \quad (17)$$

In the deterministic version of (17) the shape of function $\phi(\cdot)$ captures the way in which y_{t+1}^i converges to its steady state $X^i \theta$. Equation (17) 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 $\phi(\cdot)$ by itself tells us little about σ -convergence. We therefore also compute conditional ergodic distributions.

Function $\phi(\cdot)$ can be estimated by replacing, in equation (17), $y_t^i - X^i \theta$ with \hat{z}_t^i (or a consistent estimate of it). After replacement, this boils down to applying a standard kernel regression to:

$$\hat{z}_{t+1}^i = \phi(\hat{z}_t^i) + e_{t+1}^i \quad (18)$$

In the unconditional case, we simply replace \hat{z}_t^i and \hat{z}_{t+1}^i with y_t^i and y_{t+1}^i .

¹⁰For our illustration, it is enough to assume that the errors e_{t+1}^i are not autocorrelated, but they need not be homoskedastic.

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 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 1972 and 1992 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 3092 counties. Sectoral employment data are missing for some counties, either because they are unavailable or because they are not disclosed.¹¹

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 square kilometers, corresponding to an average diameter of approximately 50 kilometers (30 miles).¹² 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

¹¹For 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.

¹²This approximation obviously underestimates the actual diameter, since counties are not perfect circles. It is nevertheless useful as a ballpark figure.

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, *Table 1* reports the standard deviation of log employment at the county level in 1972 and 1992. A decreasing standard deviation reflects employment becoming more equally spread across counties — this is what Sala-i-Martin (1996) calls σ -convergence —; an increasing standard deviation points to employment becoming more concentrated in space. As can be seen, for total employment the tendency has been towards more concentration (divergence). At the 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).

Table 2 reports the results of a standard linear regression of employment growth on initial log employment — the standard test of unconditional β -convergence. A positive coefficient on initial employment points to concentration (divergence), whereas a negative coefficient indicates deconcentration (convergence). Our findings from *Table 1* are confirmed. We find concentration of employment at the aggregate level and in most service sectors, and deconcentration in most other sectors. This suggests that services are driving aggregate employment dynamics. This should not come as a surprise, given the weight of services in the economy: in 1972 ‘retail’, ‘FIRE’ and ‘other services’ already made up 42% of total employment; by 1992 this share had grown to 54%.

The increasing concentration of aggregate county employment stands in contrast with the observed employment deconcentration across metropolitan areas (Carlino and Chatterjee, 2001). 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 86,000 workers in 1972, and the remaining 2875 counties. Although the overlap is not perfect, 198 out of the 200 largest counties are classified as metropolitan.¹³ 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 only holds for a limited group of high employment metropolitan counties. For the rest of the distribution, concentration seems to be the norm.

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. The unconditional estimating equation is of the form:

$$y_{1992}^i = \phi(y_{1972}^i) + e_{t+1}^i \quad (19)$$

where y is log employment. The data have been normalized, assuming an exponential trend.¹⁴ To facilitate interpretation, we have plotted the employment *growth* as a function of initial log employment.¹⁵ In this case, a negative slope indicates deconcentration (convergence) and a positive slope indicates concentration (divergence). Assuming the underlying model is deterministic, 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.

¹³This classification is based on the rural-urban codes for 1974, as established by the Office of Management and Budget (OMB).

¹⁴We also experimented with a linear trend; this did not significantly affect the quantitative results though.

¹⁵This amounts to subtracting y_{1972}^i from both sides of equation (19).

Figure 1 plots the results for total employment and by sector. The graphs include the 95% confidence interval. We start by looking at the picture for total employment (‘Total’). The curve cuts the horizontal axis three times: in a deterministic model, the two equilibria at the extremes are stable and the middle equilibrium unstable. 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. This result illustrates how misleading standard linear regressions can be when testing β -convergence. Kernel regressions represent a significant improvement over standard linear tests of convergence.

It turns out that exactly 50% of counties,¹⁶ i.e., all counties with less than 7664 jobs in 1972, are predicted to slowly empty out; the remaining 50% of counties are predicted to gain jobs and to end up in the high steady state. We only observe convergence in the lower and the upper end of the distribution, where the slope is negative. Regarding the upper tail, the 9% largest counties – corresponding to those with more than 57,611 jobs in 1972 – are predicted to exhibit convergence. This suggests that employment is moving out of large metro counties.¹⁷ These results are by and large consistent with the results from the previous sub-section.

Turning to individual sectors, we see that for most of the non-service sectors – such as ‘manufacturing’ and ‘farming’ – the slope is monotone and decreasing. This suggests deconcentration (convergence). The deterministic steady state is where the curve cuts the horizontal axis. In contrast, for two out of the three service sectors – ‘retail’ and ‘FIRE’ – the picture resembles that of aggregate employment. In ‘retail’, for instance, there is a

¹⁶To be precise, 1529 out of 3069 counties.

¹⁷According to the 1974 classification of the Office of Management and Budget (OMB), of the 9% largest counties, 263 out of 271 are metropolitan.

steady state with low retail employment and a steady state with high retail employment, with the middle group disappearing. More specifically, the deterministic model predicts that 48% of counties will end up in the low steady state (with 123 retail jobs, using 1972 figures) and the remaining 52% will end up in the high steady state (with 16,874 retail jobs, using 1972 figures).

4.3 Unconditional ergodic distributions

The kernel regressions are based on a deterministic underlying model, with all counties ending up in a small number of steady states. Predictions about the future spatial concentration of jobs are therefore extremely crude. To generate more realistic predictions, we turn to a stochastic model in which the long-term behavior of the system is characterized not by one or two steady states but by an entire distribution – the ergodic distribution.

Figure 2 plots the ergodic distributions for all sectors, and compares them to the distributions of employment in 1972 and 1992. The method used to derive the ergodic distribution was described in the methodology section. If the ergodic distribution is tighter than the distribution in 1972 and 1992, this suggests deconcentration: 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 1972 and 1992, 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.

These predictions about the long-term distribution of employment across counties assume that transition probabilities will remain constant. This is not quite realistic, as future changes in demand and technology are bound to affect those probabilities. Ergodic distributions are therefore best understood as a way of visualizing current trends. By showing how the world would like if current trends were to continue into the indefinite future, we improve our understanding of what those trends are.

Results shown in *Figure 2* fall basically into two categories: ‘total’, ‘retail’, ‘FIRE’ and ‘military’ employment exhibit hat shapes, while all other sectors exhibit (slightly skewed) bell shapes. For the case of ‘FIRE’ we even get ‘twin peaks’ – a bimodal distribution. Results for two sectors — ‘agricultural services’ and ‘mining’ — stand apart from the rest, because they are dominated by a very large number of counties with zero or undisclosed employment.

These results by and large confirm our findings in the kernel regressions. Total employment is becoming more 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’ and ‘FIRE’, and to a much smaller extent — the sector is small — in ‘military’.

All other sectors are predicted to either remain at their current level of concentration or to become less concentrated. We indeed see that, for most of these sectors, the ergodic distribution is tighter than the current distribution. Furthermore, we also see that, for ‘farming’, ‘manufacturing’, ‘transportation and utilities’, and ‘wholesale’, the ergodic distributions shift to the left compared to past distributions. This suggests fewer counties with large employment levels and more counties with low or intermediate employment levels. For these sectors, convergence is predicted to take place at levels of county employment lower than current levels. This is consistent with a spreading of these activities over space. In contrast, concentration in ‘construction’, ‘federal civilian’ employment, ‘state and local’ administration, and ‘other services’ is predicted to remain basically constant. In the case of ‘federal employment’ and ‘other services’ there remains a fairly thick upper tail though, suggesting the continued presence of a small number of counties with large employment in those sectors.

If we compare *Figure 2* with *Figure 1*, 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. Finally, when the kernel regression has a flat – albeit negative – segment above the steady state, as in the case of federal employment and other services, we observe that the ergodic distribution has a fat upper tail. The size of this upper tail cannot, however, be assessed solely on the basis of the shape of the kernel regression. We therefore see that, in all cases, 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.

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 (12). The X^i characteristics include a variety of geographical features for which we have data.¹⁸ We then obtain the \widehat{z}_t^i using (13). These \widehat{z}_t^i are then used to calculate a new set of kernel regressions and a new set of ergodic distributions using formula (14).

Figure 3 plots the outcome of the kernel regressions of the \widehat{z}_t^i obtained using (18).

¹⁸The complete list was given in the data section (Section 3).

This kernel regression assumes that the time-varying component of y_t^i is deterministic and examines whether it has a steady state. As explained in Section 2, this approach is equivalent to the standard conditional β -convergence approach, except that it allows for nonlinearities.

Comparing *Figure 3* with *Figure 1*, we again get divergence in the the middle and convergence in the lower tail for ‘total’ employment, though there is less evidence of convergence in the upper tail. Turning to individual sectors, our previous results are confirmed. The non-service sectors exhibit deconcentration (convergence), whereas two out of the three service sectors – ‘retail’ and ‘FIRE’ – and to some extent the military 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 *Figure 4*. The basic pattern is similar to that depicted in *Figure 2*: ‘total’ employment and ‘retail’ exhibit a hat shaped ergodic distribution, suggesting increasing concentration. But once we condition on geographical characteristics, ‘FIRE’ and the ‘military’ no longer have a hat shape, although their ergodic distributions continue to display a higher variance and fatter upper tails than past distributions. For ‘state and local’ administration, employment is expected to become more concentrated over time once we condition on geographical factors. Taken together, these results suggest increasing concentration in some, but not all, service sectors.

The results for non-service sectors are also broadly similar to *Figure 2*, although in several instances the shift to the left, indicating deconcentration, is less noticeable once we condition on geographical characteristics. This is especially true for ‘farming’. This hardly comes as a surprise given that geographical characteristics are thought to play an important role in the location of agricultural activity. The tendency towards less concentration for ‘manufacturing’ is confirmed, although the conditional ergodic distribution

exhibits a larger variance than the unconditional one, again suggesting that geographical factors play a role in the positioning of manufacturing enterprises.

5 Conclusion

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 heavily 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 9% largest counties and the 15% smallest counties exhibit deconcentration; the remaining 76% exhibit concentration. This result is consistent with deconcentration across urban areas (Carlino and Chatterjee, 2001) and concentration across U.S. counties (Desmet and Fafchamps, 2004). Whether the overall picture is one of concentration or deconcentration is not entirely obvious. In terms of number of counties, concentration holds the upper hand. However, in terms of number of people, deconcentration dominates, since the 9% highest employment counties accounted for two thirds of total employment in 1972.

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, a number of service activities are becoming more concentrated, in particular ‘retail’ trade and ‘finance, insurance and real estate’. 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.

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 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 counterfactual 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, Onatski and Stock, 2000). Developing such tests is left for future research.

References

- [1] Beeson, P., DeJong, D., and Troesken, W., 2001. Population Growth in U.S. Counties, 1840-1990, *Regional Science and Urban Economics*, 31(6), 669-699.
- [2] Barro, R.J., 1991. Economic Growth in a Cross-Section of Countries, *Quarterly Journal of Economics*, 106(2), 407-443.

- [3] Blanchard, O. and Katz, L., 1991. Regional Evolutions, *Brookings Papers on Economic Activity*, 1, 1-75.
- [4] Carlino, G.A. and Chatterjee, S., 2001. Aggregate Metropolitan Employment Growth and the Deconcentration of Metropolitan Employment, *Journal of Monetary Economics*, 48, 549-583.
- [5] Desmet, K. and Fafchamps, M., 2004. Changes in the Spatial Concentration of Employment across U.S. Counties: A Sectoral Analysis 1972-2000, *Journal of Economic Geography*, forthcoming.
- [6] Dumais, G., Ellison, G. and Glaeser, E., 2002. Geographic Concentration as a Dynamic Process, *Review of Economics and Statistics*, 84(2), 193-204.
- [7] Durlauf, S.N. and Quah, D.T., 1999. The New Empirics of Economic Growth, in: J.B. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, volume 1A, Chapter 4, North Holland Elsevier Science: Amsterdam.
- [8] Eaton, J. and Eckstein, Z., 1997. Cities and Growth: Theory and Evidence from France and Japan, *Regional Science and Urban Economics*, 27, 443-474.
- [9] Fujita, M., Krugman, P. and Venables, A.J., 1999. *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press: Cambridge, Mass.
- [10] Henderson, J.V., 1988. *Urban Development: Theory, Fact and Illusion*, Oxford University Press: New York.
- [11] Higgins, M., Levy, D. and Young, A., 2003. Growth and Convergence across the U.S.: Evidence from County-Level Data, Emory Economics Working Paper 0306.
- [12] Kim, S., 1995. Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987, *Quarterly Journal of Economics*, 110, 881-908.
- [13] Kolko, J., 1999. Can I Get Some Service Here?, mimeo, Harvard University.

- [14] Kremer, M., Onatski, A. and J. Stock, 2000. Searching for Prosperity, Department of Economics, Harvard university, mimeo.
- [15] Luenberger, D.G., 1979. Introduction to Dynamic Systems, Wiley: New York.
- [16] Mankiw, N.G., Romer, D. and Weil, D.N., 1992. A Contribution to the Empirics of Economic Growth, Quarterly Journal of Economics, 107, 407-437.
- [17] Mood, A.M., Graybill, F.A., and Boes, D.C., 1974. Introduction to the Theory of Statistics, McGraw Hill.
- [18] Quah, D.T., 1993. Galton's Fallacy and Tests of the Convergence Hypothesis, Scandinavian Journal of Economics, 95(4), 427-443.
- [19] Quah, D.T., 1996a. Empirics for Economic Growth and Convergence, European Economic Review, 40, 1353-75.
- [20] Quah, D.T., 1996b. Convergence Empirics Across Economies with (Some) Capital Mobility, Journal of Economic Growth, 1, 95-124.
- [21] Quah, D.T., 1997. Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs, Journal of Economic Growth, 2, 27-59.
- [22] Sala-i-Martin. X., 1996. Regional Cohesion: Evidence and Theories of Regional Growth and Convergence, European Economic Review, 40, 1325-52.
- [23] Stokey, N.L. Lucas, R.E. and Prescott, E.C., 1989. Recursive Methods in Economic Dynamics, Harvard University Press: Cambridge, Mass.

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

Sector	standard deviation (logs) 1972	standard deviation (logs) 1992
Total	1.34	1.43
Farming	.94	.84
Agricultural Services	1.32	1.31
Mining	1.67	1.78
Construction	1.51	1.58
Manufacturing	1.97	1.88
Transportation/Utilities	1.54	1.57
Wholesale	1.73	1.67
Retail	1.42	1.58
FIRE	1.54	1.64
Other Services	1.50	1.62
Federal Civilian	1.57	1.60
Military	1.49	1.50
State/Local	1.31	1.36

Source: REIS, Bureau of Economic Analysis

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

Dependent variable: annual growth rate in sectoral employment 1972-92							
	Total	Farming	Ag serv	Mining	Constr	Manuf	Trans/util
const	-.0045894 (2.346)	.031724 (18.468)	.0795073 (37.606)	.0319508 (12.668)	.0232263 (11.562)	.054432 (26.750)	.0356632 (18.522)
emp72	.0022102 (10.385)	-.006113 (24.146)	-.0072239 (15.188)	-.0054879 (9.313)	-.0009096 (2.805)	-.0060793 (21.744)	-.0030251 (9.506)
R^2	0.0399	0.1595	0.0855	0.0368	0.0027	0.1372	0.0305
	Wholesale	Retail	FIRE	Other serv	Fed civ	Milit	State/Loc
const	.0554855 (28.480)	-.0042489 (2.294)	.0031527 (1.950)	.0172376 (8.878)	.009566 (7.904)	.0058711 (5.114)	.0127573 (8.322)
emp72	-.0043428 (12.543)	.0032981 (12.969)	.0019356 (7.445)	.0021457 (8.147)	-.0007685 (3.222)	-.0011588 (5.269)	.0005706 (2.660)
R^2	0.0528	0.0521	0.0184	0.0221	0.0034	0.0090	0.0023

Absolute values of t-statistics in brackets.

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

Dependent variable: annual growth rate in sectoral employment 1972-92							
	Total	Farming	Ag serv	Mining	Constr	Manuf	Trans/util
const	.0654953 (3.960)	-.0020957 (0.538)	.0979573 (9.623)	.0842865 (7.272)	.081878 (4.892)	.1346491 (8.551)	.0934903 (6.686)
emp72	-.0037547 (2.785)	-.0010493 (1.901)	-.0071837 (4.790)	-.0107865 (5.568)	-.0075326 (4.195)	-.0130225 (8.736)	-.0084311 (5.605)
R^2	0.0377	0.0179	0.1083	0.1409	0.0820	.2782	0.1375
	Wholesale	Retail	FIRE	Other serv	Fed civ	Milit	State/Loc
const	.1189618 (8.686)	.0997601 (6.481)	.0465353 (3.841)	.0795902 (6.284)	.0362754 (4.601)	.0362712 (4.557)	.0364982 (3.155)
emp72	-.0104643 (7.031)	-.0074854 (5.065)	-.0023342 (1.855)	-.0035878 (3.012)	-.003329 (3.621)	-.004913 (5.167)	-.0020426 (1.777)
R^2	0.1998	0.1147	0.0171	0.0483	0.0621	0.1188	0.0157

Absolute values of t-statistics in brackets.

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

Dependent variable: annual growth rate in sectoral employment 1972-92							
	Total	Farming	Ag serv	Mining	Constr	Manuf	Trans/util
const	-.0116011 (4.686)	.0403319 (21.371)	.0898732 (37.196)	.0330733 (12.556)	.0198499 (8.166)	.0569191 (25.047)	.0414091 (17.571)
emp72	.0030247 (10.924)	-.0074054 (26.547)	-.0101888 (17.670)	-.006341 (9.845)	-.0002885 (0.701)	-.0065219 (20.040)	-.0041513 (10.105)
R^2	0.0399	0.1967	0.1202	0.0440	0.0002	0.1265	0.0368
	Wholesale	Retail	FIRE	Other serv	Fed civ	Milit	State/Loc
const	.0597689 (25.129)	-.0156651 (6.931)	.0045739 (2.254)	.0176736 (7.262)	.0123751 (8.304)	.0043236 (3.144)	.0089734 (4.688)
emp72	-.0052954 (11.703)	.0050313 (15.613)	.0016601 (4.827)	.0020689 (6.015)	-.0014565 (4.630)	-.0008204 (2.939)	.0011516 (4.150)
R^2	0.0496	0.0785	0.084	0.0130	0.0074	0.0030	0.0060

Absolute values of t-statistics in brackets.

Figure 1: Unconditional kernel regressions

Employment growth (1972-1992) on log of initial employment (1972)
Upper and lower curve represent 95% confidence interval

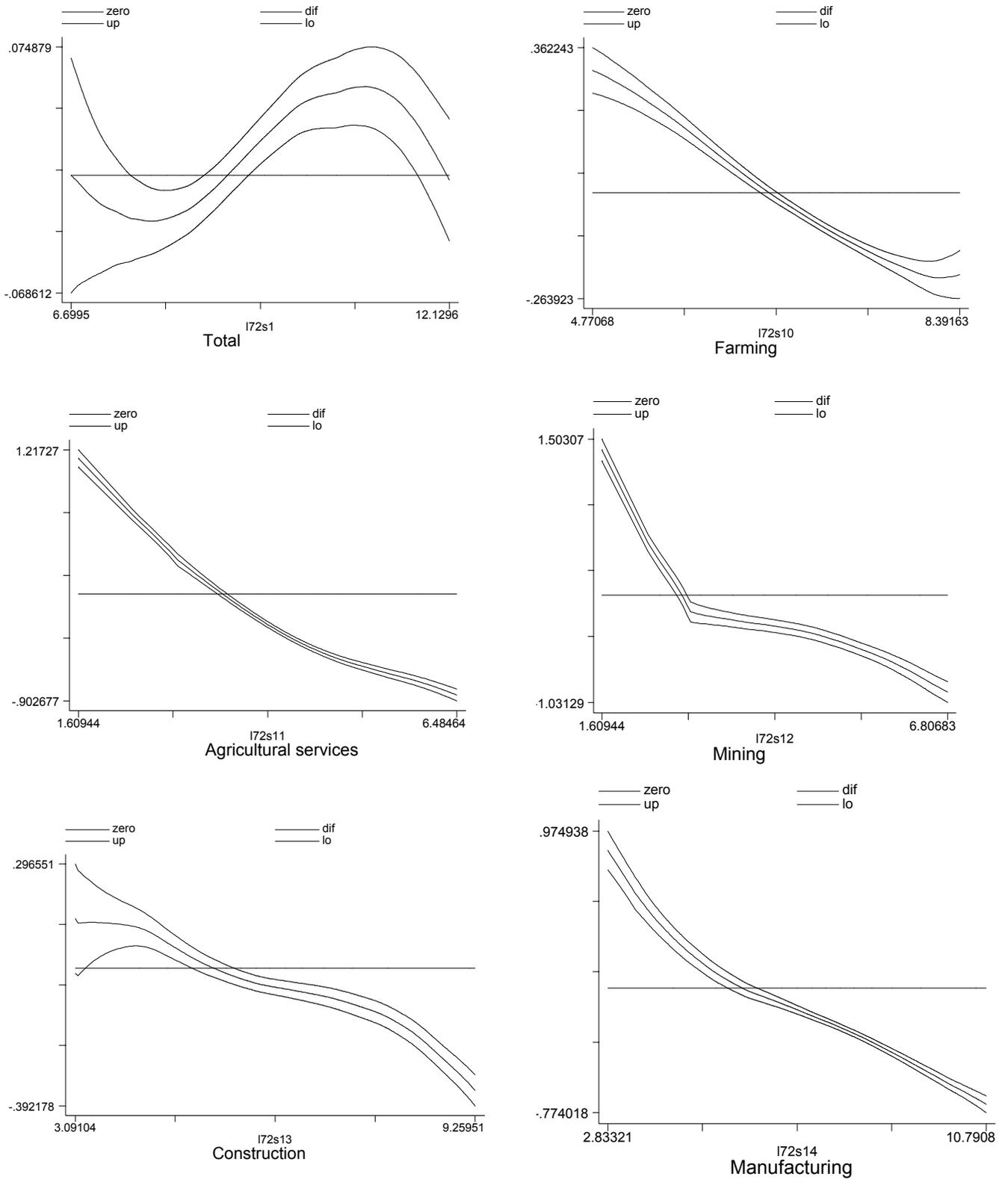


Figure 1: unconditional kernel regressions (cont'd)

Employment growth (1972-1992) on log of initial employment (1972)
Upper and lower curve represent 95% confidence interval

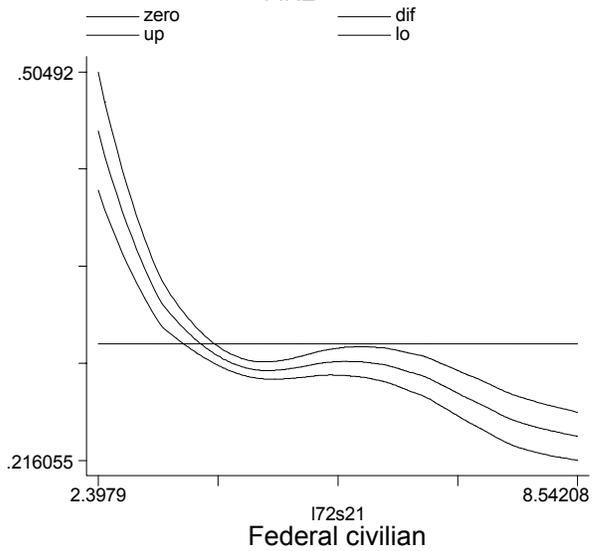
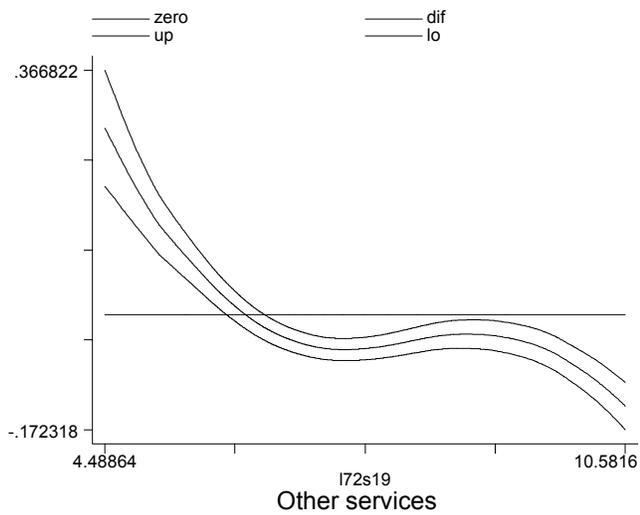
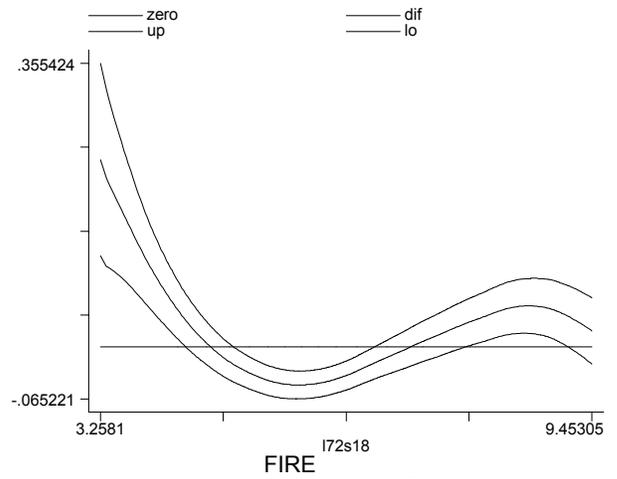
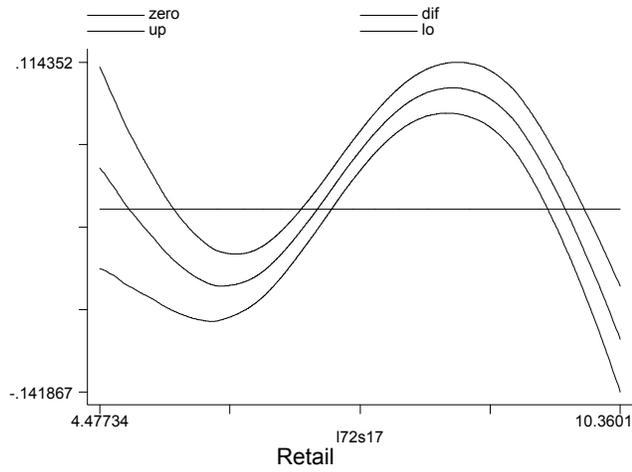
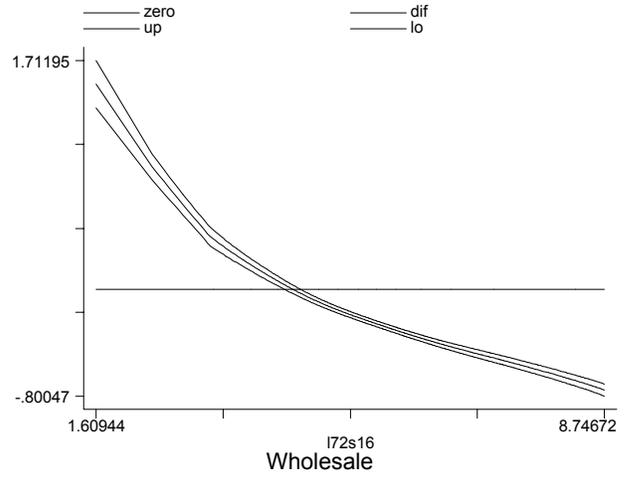
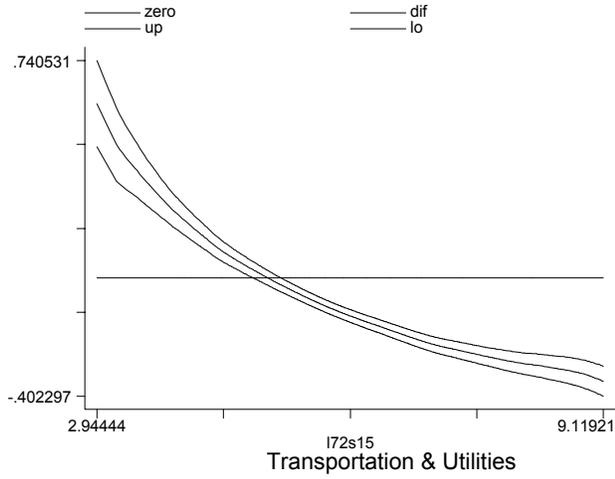


Figure 1: unconditional kernel regressions (cont'd)

Employment growth (1972-1992) on log of initial employment (1972)
Upper and lower curve represent 95% confidence interval

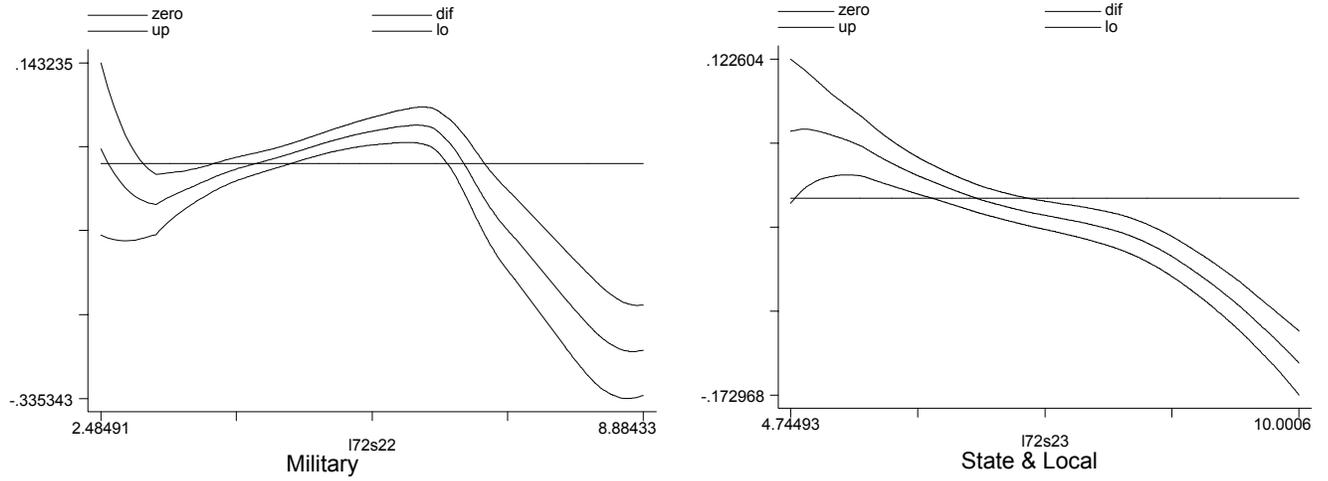


Figure 2: Unconditional ergodic distributions

Ergodic distributions are given by the full curves. For comparison purposes, we also plot the distributions in 1972 (the dashed curves) and the distributions in 1992 (the dotted curves).

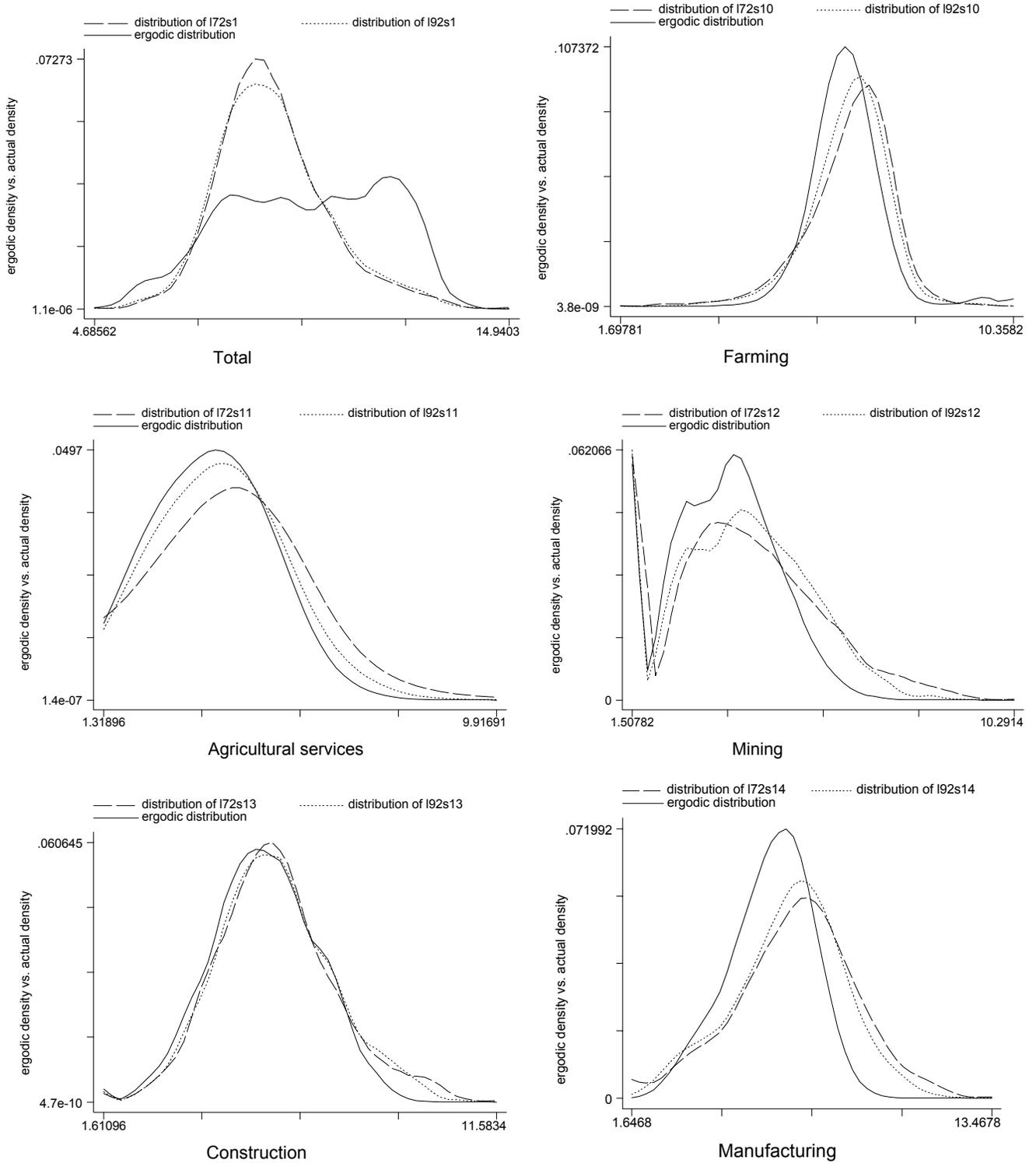


Figure 2: Unconditional ergodic distributions (cont'd)

Ergodic distributions are given by the full curves. For comparison purposes, we also plot the distributions in 1972 (the dashed curves) and the distributions in 1992 (the dotted curves).

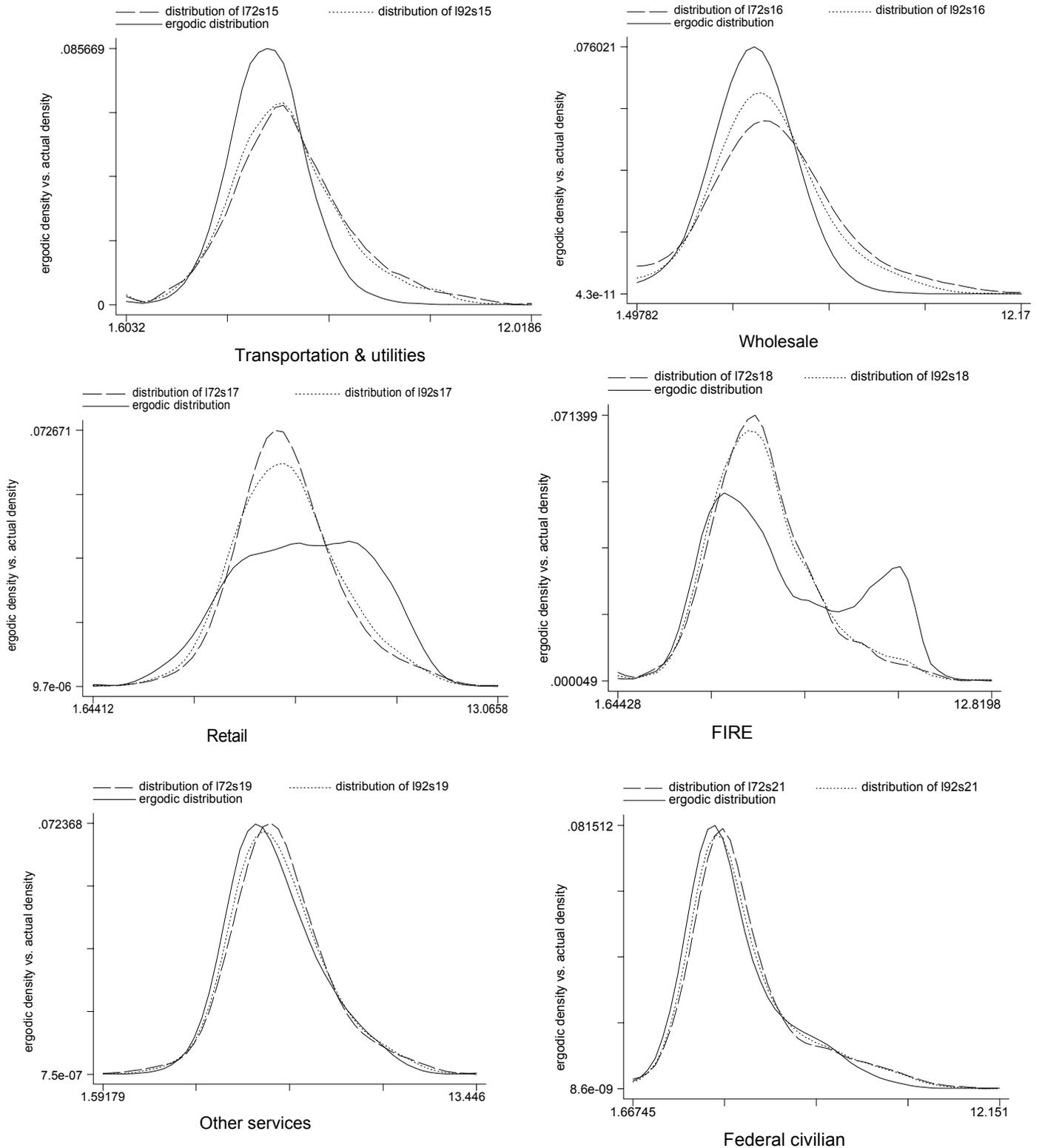


Figure 2: Unconditional ergodic distributions (cont'd)

Ergodic distributions are given by the full curves. For comparison purposes, we also plot the distributions in 1972 (the dashed curves) and the distributions in 1992 (the dotted curves).

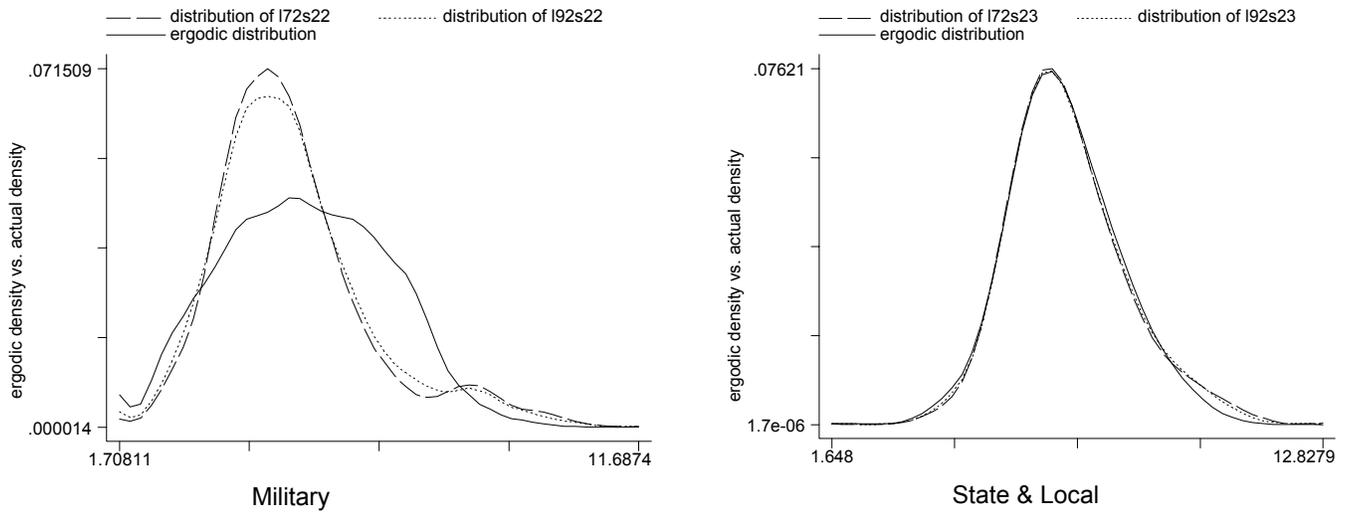


Figure 3: conditional kernel regressions

Employment growth (1972-1992) on log initial employment (1972)
Upper and lower curve represent 95% confidence intervals

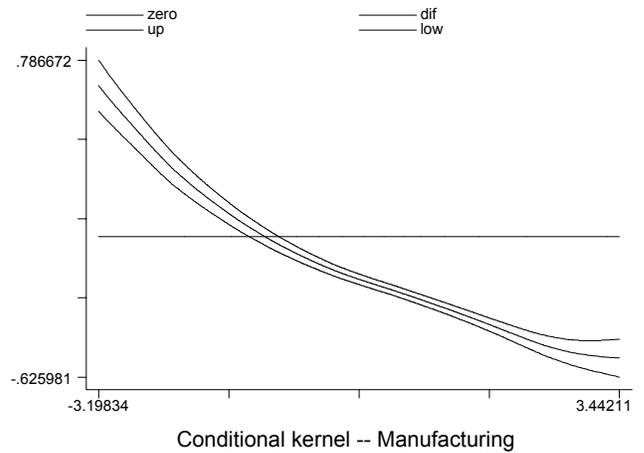
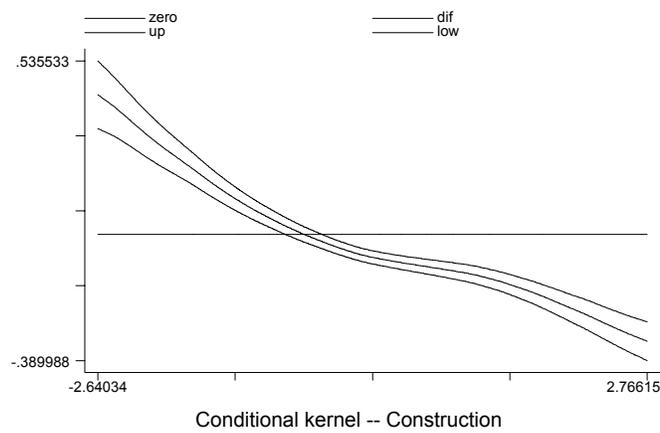
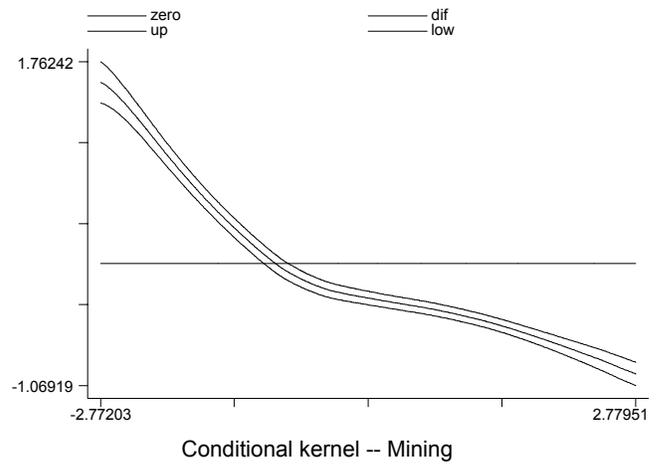
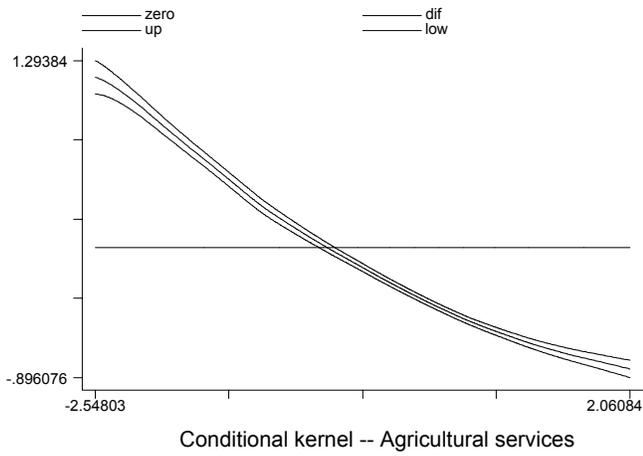
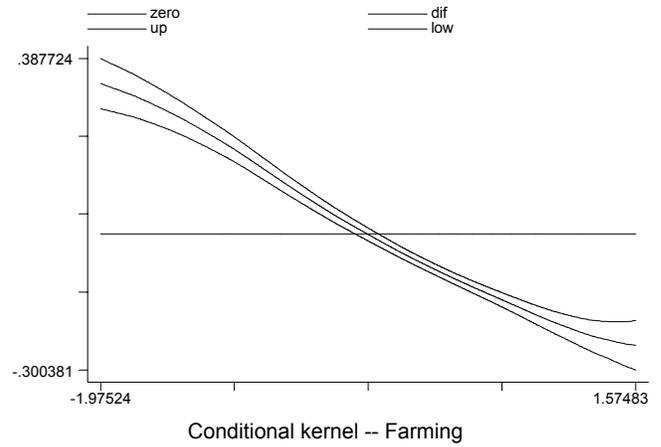
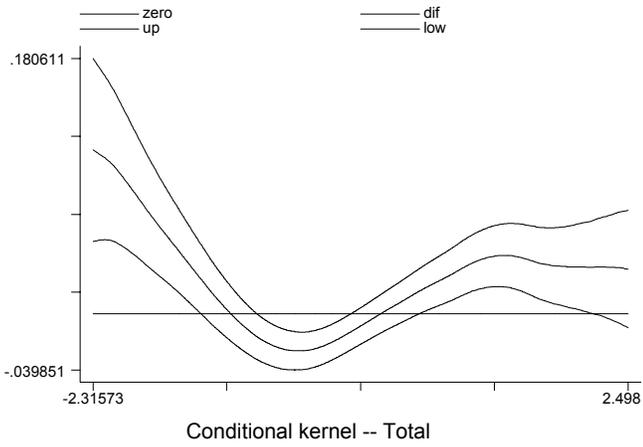


Figure 3: conditional kernel regressions (cont'd)

Employment growth (1972-1992) on log initial employment (1972)
Upper and lower curve represent 95% confidence intervals

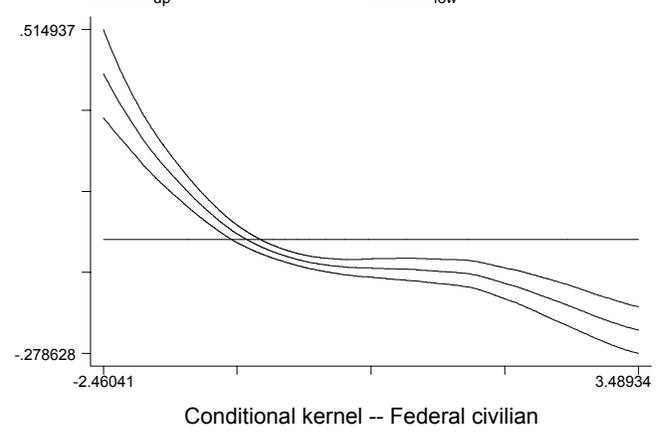
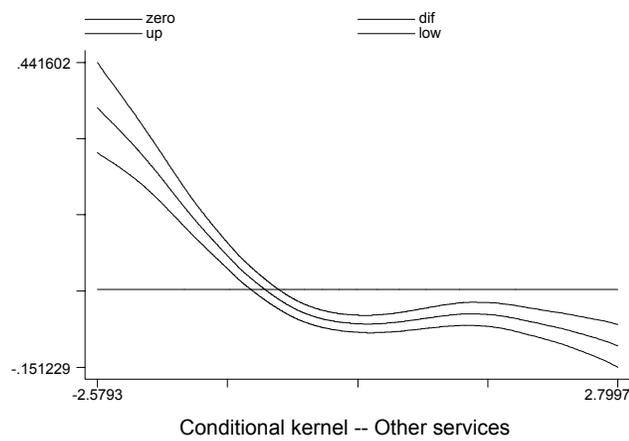
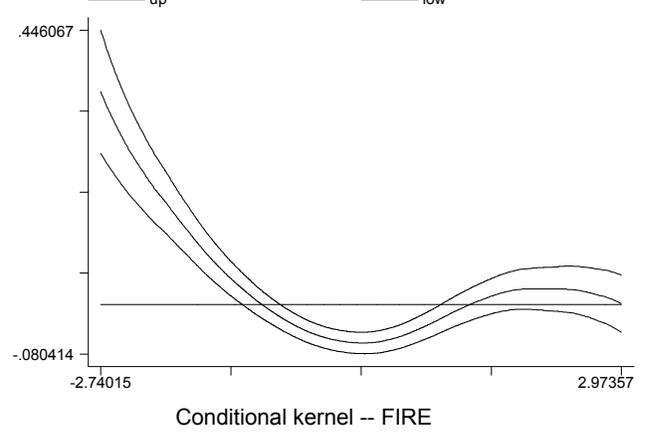
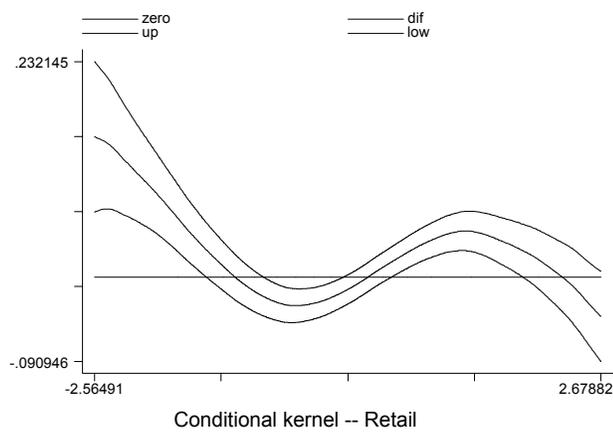
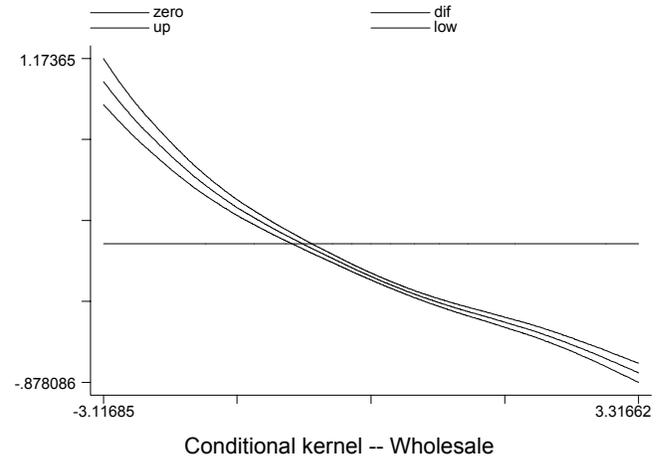
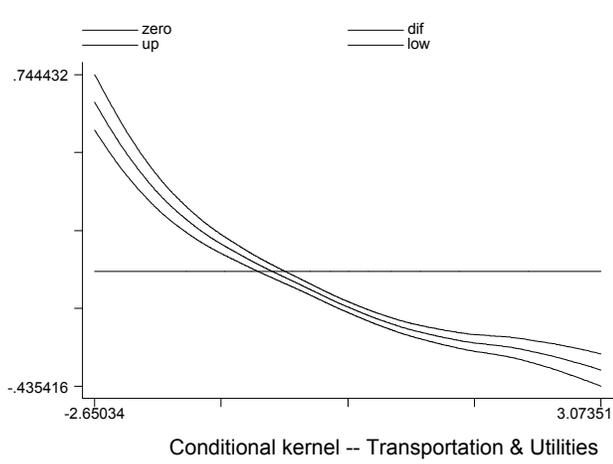


Figure 3: conditional kernel regressions (cont'd)

Employment growth (1972-1992) on log initial employment (1972)
Upper and lower curve represent 95% confidence intervals

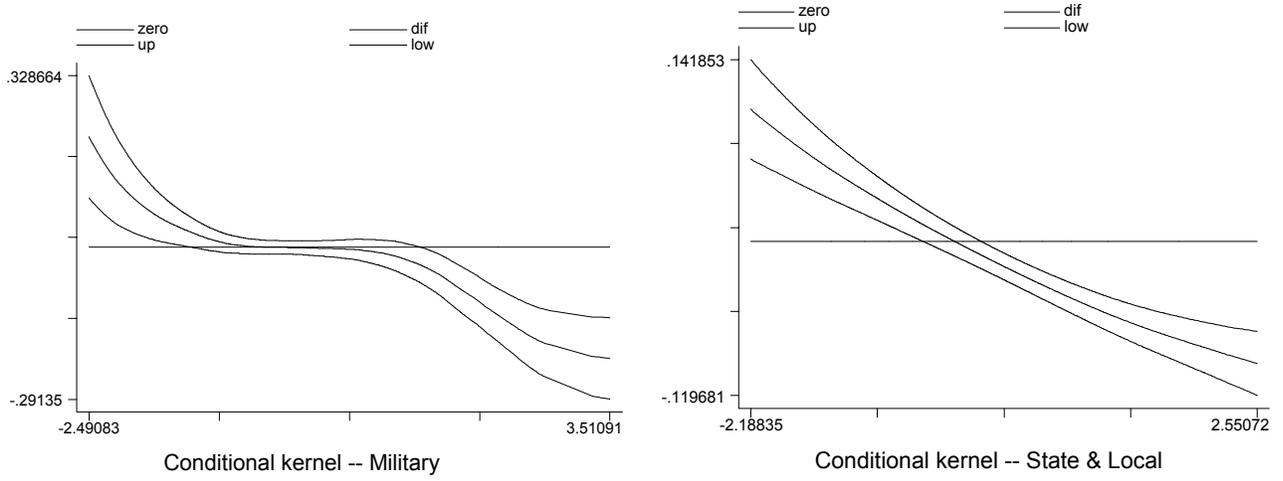


Figure 4: conditional ergodic distributions

Ergodic distributions are given by the full curves. For comparison purposes, we also plot the distributions in 1972 (the dashed curves) and the distributions in 1992 (the dotted curves).

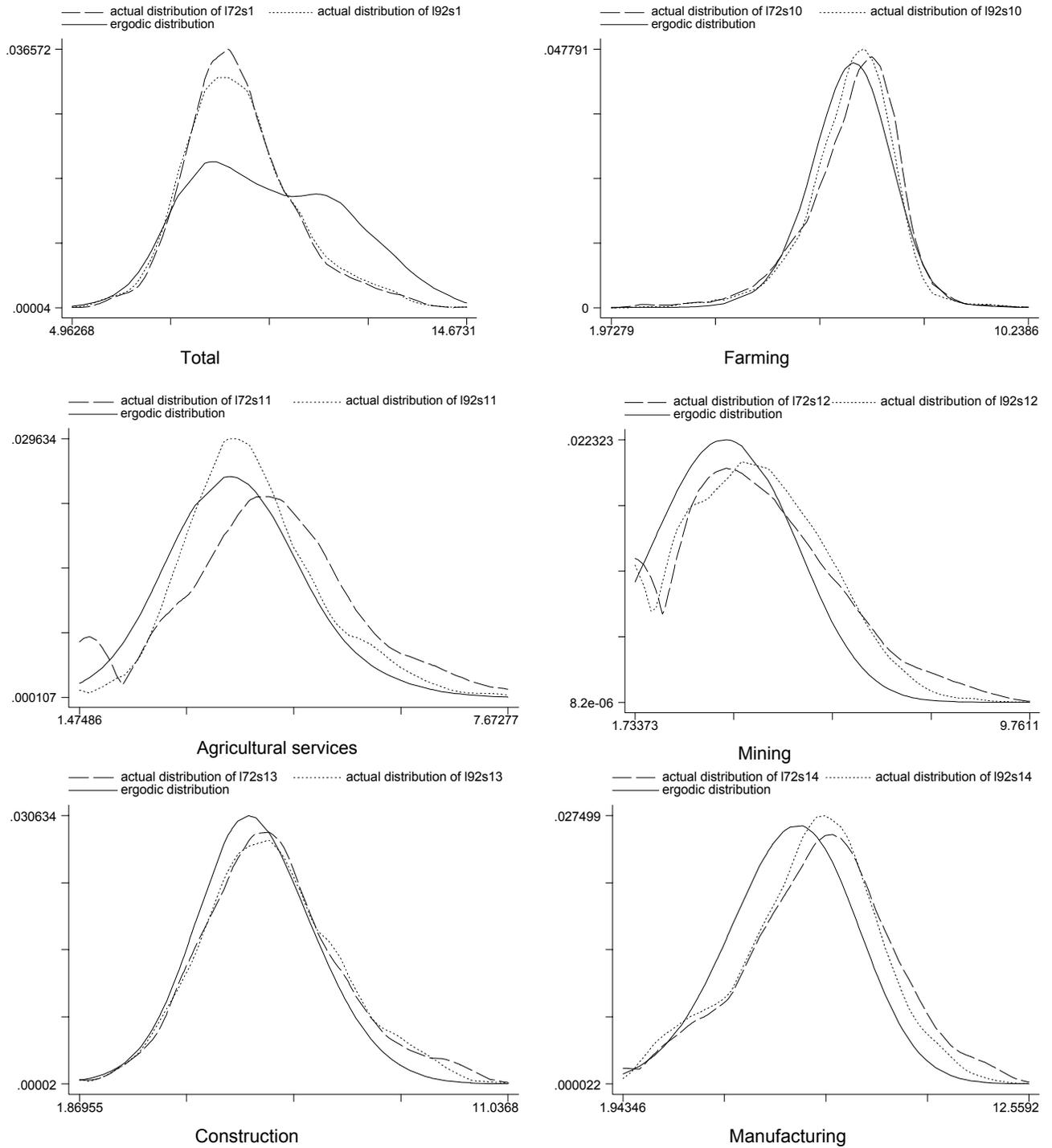


Figure 4: conditional ergodic distributions (cont'd)

Ergodic distributions are given by the full curves. For comparison purposes, we also plot the distributions in 1972 (the dashed curves) and the distributions in 1992 (the dotted curves).

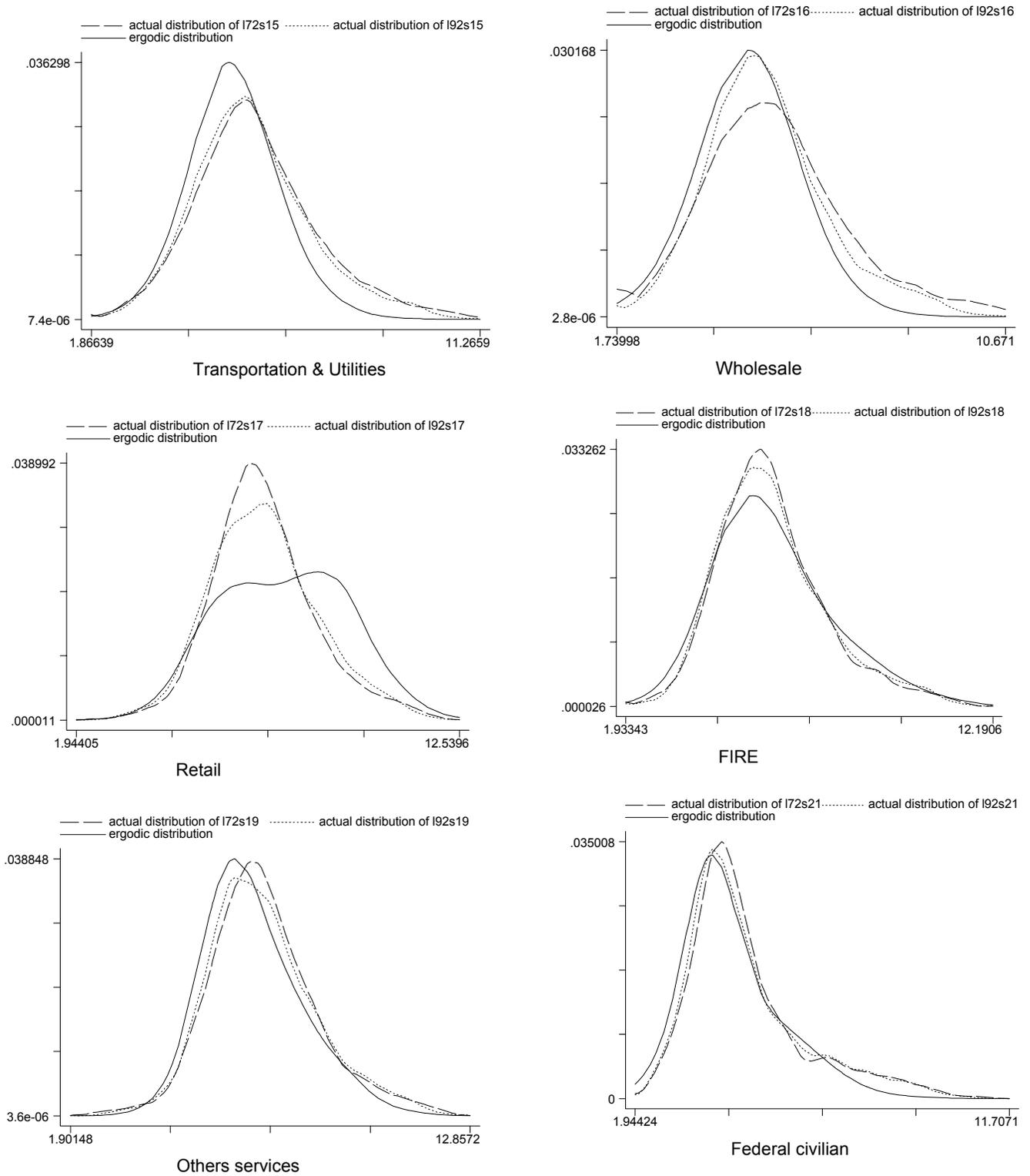


Figure 4: conditional ergodic distributions (cont'd)

Ergodic distributions are given by the full curves. For comparison purposes, we also plot the distributions in 1972 (the dashed curves) and the distributions in 1992 (the dotted curves).

