Courant Research Centre
'Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis'

Georg-August-Universität Göttingen
(founded in 1737)

Discussion Papers

No. 55

Distribution Dynamics of Regional GDP per Employee in Unified Germany

Sebastian Vollmer, Hajo Holzmann, Florian Ketterer, Stephan Klasen

first version: December 2010, revised: September 2011

Platz der Göttinger Sieben 3 · 37073 Goettingen · Germany
Phone: +49-(0)551-3914066 · Fax: +49-(0)551-3914059

Email: crc-peg@uni-goettingen.de Web: http://www.uni-goettingen.de/crc-peg

Distribution Dynamics of Regional GDP per Employee in Unified Germany

Sebastian Vollmer*
Hajo Holzmann†
Florian Ketterer†
Stephan Klasen‡

This version: September 2011

Abstract

We investigate to what extent convergence in production levels per worker has been achieved in Germany since unification. To this end, we model the distribution of GDP per employee across German districts using two-component normal mixtures. While in the first year after unification, the two component distributions were clearly separated and bimodal, corresponding to the East and West German districts, respectively, in the following years they started to merge showing only one mode. Still, using the recently developed EM-Test for homogeneity in normal mixtures, the hypothesis of just a single normal component for the whole distribution is clearly rejected for all years. A Posterior analysis shows that about a third of the East German districts were assigned to the richer component in 2006, thus catching up to levels of the West. The growth rate of a mover district is about one percentage point higher than the growth rate of a non-mover district which had the same initial level of GDP per employee.

JEL classification: O47, R11.
Keywords: Regional convergence, distribution dynamics, mixture models, Germany, unification.

Acknowledgements. We are grateful to two anonymous referees for valuable suggestions that considerably improved this paper. We also gratefully acknowledge financial support from the German Research Foundation (HO 3260/3-1, KL 1260/9-1 and VO 1592/3-1), the German Academic Exchange Service (Vollmer), the Claussen-Simon Stiftung (Holzmann) and the Landesstiftung Baden-Württemberg (Holzmann).

*University of Hannover & Harvard University. Corresponding author: svollmer@hsph.harvard.edu.
†University of Marburg.
‡University of Göttingen.

1 Introduction

About twenty years have passed since the unification of the two German states. Many predictions on the economic integration and convergence of the two German states had been made at the time of the unification. Sinn and Sinn (1991) and Akerlof et al. (1991) correctly predicted a massive output collapse after unification, linked to the exchange rate used for monetary union in 1990 which rendered East German moribund industry entirely uncompetitive; both papers suggested that unless corrective measures were taken (and they were not), it would take a long time for East German to recover from this output shock.

Comparable living standards across regions are a fundamental objective of both German and European Union regional policy. Thus, fiscal transfers to the East of Germany continue to be sizeable. While most of the transfers are effectively tied to higher unemployment and higher poverty in the East and hence are used to raise consumption levels, some of these transfers are used to promote production there by funding investments in infrastructure, industrial policies, and the like. These policies should ideally promote spatial convergence of production levels. On the other hand, new economic geography models would suggest that agglomeration tendencies in advanced economies might make it quite difficult for East Germany to attract and retain advanced industries and highly skilled workers which would work against spatial convergence; to the extent East Germany succeeds, it might again be spatially concentrated. Given these possibly opposing forces, it is important to empirically test whether or not convergence of GDP per employee can be observed.

We investigate to what extent convergence in production levels per worker has been achieved in Germany since unification. To this end, we apply a distribution dynamics approach to the distribution of GDP per employee across German districts. This approach to convergence analysis has been introduced to the literature by Quah (1993, 1996). He interpreted the emergence of a bimodal cross-country distribution of GDP per capita as polarization of distribution into a rich and a poor convergence club and coined the term "twin peaks". Bianchi (1997) was the first to empirically confirm the statistical significance of the second peak using a nonparametric procedure by Silverman (1981). Colavecchio et al. (2010) apply Silverman's test to the regional distribution of GDP per capita in Germany for the period of 1992–2001 and conclude that the distribution is bimodal or even trimodal. Jüssen (2008) applies Silverman's test to the regional distribution of GDP per employee for the period of 1992–2004 and finds that an initially bimodal distribution turns into a unimodal distribution around 2002. From this he concludes convergence and does not further investigate how distributional dynamics have developed within the East.

But studying the number of components that make up the distribution may be a superior approach to investigate convergence (and convergence clubs) for economic and technical reasons. Economically, components in a distribution have a stronger economic interpretation as they identify relevant sub-groups in a heterogeneous distribution; for convergence, the existence of these sub-groups and their development over time is the key. Technically, Vollmer et al. (2010) pointed out that it may be misleading to look at the number of peaks of a distribution if convergence clubs or sub-distributions are the true purpose of the analysis. They show that simple rescaling of the data (e.g. taking logs) produces a statistically significant third peak in the cross-country distribution of GDP per capita. Countries which were previously assigned to Quah's poor convergence club are considered middle-income on the log-scale, which introduces an arbitrary element in these analyses. Vollmer et al. (2010) model the cross-country distribution of GDP per capita with mixture models instead, where the distinct components correspond to subgroups (i.e. convergence clubs) in a heterogeneous population (the income distribution) in a natural fashion. As a further technical advantage, this approach is invariant to strictly monotonic transformation of the data and is thus robust towards this shortcoming of the twin peaks approach. Paap and van Dijk (1998) have pioneered the modeling of the cross-country distribution of GDP per capita with mixture models. Recent developments in the methodology of likelihood ratio tests that were not available to Paap and van Dijk (1998) allow us to determine the number and type of components with rigorous statistical testing. We adopt this approach to study the regional distribution of GDP per employee in post-unification Germany for the period of 1992–2006.

There is also a rich theoretical literature on the German integration process, much of which is inspired by the literature on new economic geography. Funke and Strulik (2000) set up a two-region endogenous growth model to discuss convergence of East and West Germany. They predict that East Germany will close 80 percent of the gap to West Germany between 20 and 30 years after unification. Uhlig (2006, 2008) develops a labor search model that allows for migration and network externalities. The model can result in two equilibria, a good equilibrium representing West Germany and a bad equilibrium representing East Germany (in terms of networking, labor productivity, unemployment and migration). Burda (2006, 2008) sets up a neoclassical model in which adjustment costs and initial conditions determine dynamics and the regional distribution of production factors. In each of these cases, it is, however, possible that regionally concentrated growth modes could develop in East Germany.

In a recent theoretical paper, that is particularly related to our empirical analysis, Schäfer and Steger (2010) set up a dynamic macroeconomic model of a small open economy with factor mobility and aggregate increasing returns to scale (representing East Germany) which features multiple equilibria as well as indeterminacy. They extend Krugman

(1991), who has shown that both "history" (initial conditions) as well as "expectations" (confidence) are potentially important mechanisms for equilibrium selection. They conclude that "the long term success of a specific region (or economy) results from the interaction between economic fundamentals, economic confidence, and public policy". It is precisely this interaction which might prevent convergence in many parts of East Germany while allowing for some regions to move up.

In our empirical analysis we indeed find that the regional distribution of GDP per employee in Germany is well described by a mixture of two normal distributions. In 1992, the two component distributions were clearly separated, corresponding to the East and West German districts, respectively. In the following years the two components started to merge, leading to a single mode but continuing to consist of two separate component distributions. A posterior analysis shows that 35 East German districts (out of 102) were assigned to the richer component in 2006, thus catching up to levels of the West (while only six districts from the West fell back to the poorer component). Interestingly, whether the East German districts move to the richer component or stay in the poor component does not depend on their initial level of GDP per employee.

2 Data & Methods

2.1 Data

Germany is structured into 16 states (Länder, NUTS-1) and 429 districts (Kreise and kreisfreie Städte, NUTS-3), 326 of these districts are located in former West Germany, 102 districts are located in former East Germany, Berlin is neither assigned to the East nor the West.[1] We use data from the German Federal Statistical Office (Statistisches Bundesamt) on nominal GDP per employee and prices to obtain our variable of interest, namely real GDP per employee for all districts.[2] The data are available for the years 1992 and 1994–2006. We exclude the 28 richest West German districts in 1992 from our analysis, which were selected as follows. From fits of three-components normal mixtures, we chose those districts which were assigned to the (small) richest component in more than half of the years under investigation. For more details see Section 3.3. These districts, which typically include the central cities of major industrial centers in Western Germany where production is heavily concentrated and which pull in workers from a wide

[1] These figures predate a reform in Saxony which reduced the number of districts from 29 to 13. Berlin is just one district.

[2] Unfortunately, price data are only available at the state level. Thus, we have to assume that price levels are relatively similar within a state. While there might be level differences within a state, it is not unreasonable to believe that trends of prices within a state, which is most relevant for our analysis, are rather similar.

surrounding, stay way ahead of the rest of Germany's distribution of GDP per employee, and do not affect the main part of the distribution between east and west. Hence, these districts do not have any relevance for our research question, namely the catch-up process of East German districts to Western standards after unification.[3] We further verify these assertions in Section 3.3. Figure 1 shows the development of real mean GDP/employee and the standard deviation for East and West Germany over time. Apparently we observe convergence, with GDP per employee in the East initially growing much faster than in the West and thus catching up. Since 1998 the gap between East and West has shrunk at a much lower rate than before. The overall standard deviation of GDP per employee decreased strongly in the first few years and stayed more or less constant since then. However, the standard deviation in the East increased since 1999 when the speed of convergence had already slowed down. Note that an assessment of gross national income per capita would lead to quite different conclusions. On the one hand, due to much higher unemployment in the East and a lower share of working age people, the difference in income per capita between the East and West would be much larger than GDP per employee. On the other hand, the sizable transfer payments from West to East ensure that the incomes in the East are much higher than their output levels, sharply reducing the difference between East and West; as our focus is on convergence of labor productivity (rather than living standards) here, we focus on GDP/worker rather than GNI/capita.

2.2 Two-component normal mixtures

A natural way to model a heterogeneous population such as Germany's distribution of GDP per employee after unification is by finite mixture models. In a two-component normal mixture, the observations have density

$$f(x;\alpha,\mu_1,\mu_2,\sigma_1,\sigma_2) = (1-\alpha)\phi(x;\mu_1,\sigma_1) + \alpha\phi(x;\mu_2,\sigma_2), \tag{1}$$

with $0 \leq \alpha \leq 1$ and

$$\phi(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

[3] Specifically, we exclude Stuttgart, Landeshauptstadt, Kreisfreie Stadt; Böblingen, Landkreis; Heilbronn, Kreisfreie Stadt; Mannheim, Universitätsstadt, Kreisfreie Stadt; Ingolstadt, Kreisfreie Stadt; München, Landeshauptstadt, Kreisfreie Stadt; Altötting, Landkreis; Freising, Landkreis; München, Landkreis; Pfaffenhofen a.d.Ilm, Landkreis; Starnberg, Landkreis; Erlangen, Kreisfreie Stadt; Fürth, Kreisfreie Stadt; Hamburg; Frankfurt am Main, Kreisfreie Stadt; Offenbach am Main, Kreisfreie Stadt; Wiesbaden, Landeshauptstadt, Kreisfreie Stadt; Groß-Gerau, Landkreis; Hochtaunuskreis; Main-Taunus-Kreis; Wolfsburg, Kreisfreie Stadt; Wilhelmshaven, Kreisfreie Stadt; Düsseldorf, Kreisfreie Stadt; Rhein-Kreis Neuss, Kreis; Köln, Kreisfreie Stadt; Leverkusen, Kreisfreie Stadt; Rhein-Erft-Kreis; Ludwigshafen am Rhein, Kreisfreie Stadt from our analysis.

We assume without loss of generality that $\mu_1 \leq \mu_2$. $\phi(x; \mu_1, \sigma_1)$ and $\phi(x; \mu_2, \sigma_2)$ correspond to the distributions of the two sub-populations, and α and $1-\alpha$ are interpreted as their relative sizes.

We fit two-component normal mixtures to the log data. Note that it is essential to set up a joint model for the two populations, since we want to investigate convergence within the complete distribution of GDP per employee in Germany.

In order to check that the parametric components are well-specified, we investigate the log-data in 1992 for East and West separately, since in this year the distributions were clearly separated. We apply Shapiro-Wilk's (SW) and Anderson-Darling's (AD) tests to check whether normality can be rejected, yielding SW p-value East: 0.87; AD p-value East: 0.91; SW p-value West: 0.82; AD p-value West: 0.60. Hence we conclude that a mixture of two normal distributions fits the log-data well. Note that this also implies that several components in the distribution will not arise due to lack of fit of the normal distribution, so that the components, if detected in the following years, have their natural interpretation as subgroups.

The parameters α, μ_1, μ_2, σ_1 and σ_2 are estimated from the data by maximum likelihood. We allow for unequal variances σ_1^2 and σ_2^2, because a likelihood ratio test shows that the simplifying assumption of equal variances does not hold for all years.

Let $X_1, \ldots, X_n$ denote independent, identically distributed observations with densities (1). The log-likelihood

$$\mathcal{L}_n(\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2) = \sum_{i=1}^{n} \log \big((1-\alpha)\phi(X_i; \mu_1, \sigma_1) + \alpha\phi(X_i; \mu_2, \sigma_2) \big)$$

in finite normal mixtures with different variances is unbounded, since for any given n, $\mathcal{L}_n(\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2) \to \infty$, if $X_1 = \mu_1$ and $\sigma_1 \to 0$, holding the other parameters fixed. Thus, a global maximizer of the likelihood function does not exist. There are some formal ways around this problem, e.g. choose the largest local maximum or restrict the possible variances by restrictions of the form $\sigma_2^2 \leq c\sigma_1^2$ and $\sigma_1^2 \leq c\sigma_2^2$ for some $c > 1$ (cf. Hathaway 1985), which again leads to the existence of a global maximum. We found that using reasonable starting values (which are easy to obtain in our problem by considering East and West German districts separately), maximization algorithms such as EM or quasi Newton found stable local maxima of the log-likelihood function.

In order to formally investigate whether the two components in Germany's distribution of GDP per employee finally merged, one can test in model (1) whether it effectively consists of just a single component. This amounts to testing the hypothesis

$$H_0 : \alpha(1-\alpha) = 0 \text{ or } (\mu_1, \sigma_1) = (\mu_2, \sigma_2)$$

against the full model (1). This turns out to be a quite difficult parametric testing problem, see Chen and Chen (2003) for some history. In the following we present a novel approach, the EM-test by Chen and Li (2009) for normal mixtures in mean and variance parameters, which overcomes many drawbacks of the simple likelihood ratio test for the same problem. The test by Chen and Li (2009) is based on a penalized log-likelihood function

$$pl_n(\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2) = \mathcal{L}_n(\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2) + p(\alpha) + p_n(\sigma_1) + p_n(\sigma_2),$$

Here, $p : [0, 1] \to \mathbb{R}$ is a continuous function that is maximized at $\alpha = 0.5$ and tends to negative infinity as α goes to 0 or 1 and $p_n : [0, \infty) \to \mathbb{R}$ is bounded, when σ is large, but tends to negative infinity when σ goes to 0. The test statistic is computed as follows:

Step 0 Choose a set of initial α values, say $\alpha_1, \alpha_2, \ldots, \alpha_J$ and a positive integer K. Compute

$$(\widehat{\mu}_0, \widehat{\sigma}_0) = \arg\max_{\mu,\sigma} pl_n(0.5, \mu, \mu, \sigma, \sigma).$$

Let $j = 1, k = 0$.
Step 1 Let $\alpha_j^{(k)} = \alpha_j$.
Step 2 Compute

$$(\mu_{j1}^{(k)}, \mu_{j2}^{(k)}, \sigma_{j1}^{(k)}, \sigma_{j2}^{(k)}) = \arg\max_{\mu_1,\mu_2,\sigma_1,\sigma_2} pl_n(\alpha_j^{(k)}, \mu_1, \mu_2, \sigma_1, \sigma_2)$$

Step 3 For $i = 1, 2, \ldots, n$, compute the weights

$$w_{ij}^{(k)} = \frac{\alpha_j^{(k)}\phi(X_i; \mu_{j2}^{(k)}, \sigma_{j2}^{(k)})}{(1-\alpha_j^{(k)})\phi(X_i; \mu_{j1}^{(k)}, \sigma_{j1}^{(k)}) + \alpha_j^{(k)}\phi(X_i; \mu_{j2}^{(k)}, \sigma_{j2}^{(k)})}.$$

and then use the M-step to update the parameters

$$\alpha_j^{(k+1)} = \arg\max_{\alpha} \left((n - \sum_{i=1}^{n} w_{ij}^{(k)}) \log(1-\alpha) + \sum_{i=1}^{n} w_{ij}^{(k)} \log(\alpha) + p(\alpha) \right)$$

and

$$\begin{aligned}(\mu_{j1}^{(k+1)}, \mu_{j2}^{(k+1)}, \sigma_{j1}^{(k+1)}, \sigma_{j2}^{(k+1)}) &= \arg\max_{\mu_1,\mu_2,\sigma_1,\sigma_2} \Big(\sum_{i=1}^{n} (1 - w_{ij}^{(k)}) \log\big(\phi(X_i; \mu_1, \sigma_1)\big) + p_n(\sigma_1) \\ &\quad + \sum_{i=1}^{n} w_{ij}^{(k)} \log\big(\phi(X_i; \mu_2, \sigma_2)\big) + p_n(\sigma_2) \Big).\end{aligned}$$

Repeat Step 3 until $k+1=K$.

Step 4 Let $j=j+1$, $k=0$ and go to Step 1, until $j=J$.

Step 5 Calculate

$$EM_n^{(K)} = \max\{M_n^{(K)}(\alpha_j), j=1,2,\ldots,J\}$$

where

$$M_n^{(K)}(\alpha_j) = 2\{pl_n(\alpha_j^{(K)}, \mu_{j1}^{(K)}, \mu_{j2}^{(K)}, \sigma_{j1}^{(K)}, \sigma_{j2}^{(K)}) - pl_n(0.5, \widehat{\mu}_0, \widehat{\mu}_0, \widehat{\sigma}_0, \widehat{\sigma}_0)\}$$

Chen and Li (2009) show that under the null hypothesis H_0, if one of the α_j's is equal to 0.5, then as $n \to \infty$,

$$EM_n^{(K)} \xrightarrow{d} \chi_2^2.$$

As parameters of the EM-test, following the recommendations in Chen and Li (2009) we choose $p(\alpha) = \log(1-|1-2\alpha|)$ and $p_n(\sigma) = -0.25\{s_n/\sigma^2 + \log(\sigma^2/s_n)\}$ where $s_n = \sum_{i=1}^n (X_i - \bar{X})^2/n$. Further we choose $\{\alpha_1, \alpha_2, \alpha_3\} = \{0.1, 0.3, 0.5\}$ and $K=3$.

While the test results of the EM-test are quite robust with respect to the choice of the set of initial values for α, the corresponding EM-estimates somewhat depend on this choice. Therefore, we decided to fit a two-component normal mixture with distinct means and variances via maximum likelihood and use these estimates for the a posteriori analysis.

One advantage of modeling (log) GDP per employee is that we can relate the estimated density to sub-populations. That means that we can use mixture models for a discriminant analysis, see e.g. Fraley and Raftery (2002). Once we have fitted a two-component normal mixture

$$f(x; \widehat{\alpha}, \widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1, \widehat{\sigma}_2) = (1-\widehat{\alpha})\phi(x; \widehat{\mu}_1, \widehat{\sigma}_1) + \widehat{\alpha}\phi(x; \widehat{\mu}_2, \widehat{\sigma}_2)$$

to the data (here $\widehat{\alpha}, \widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1$ and $\widehat{\sigma}_2$ denote the ML estimates, i.e. the parameters maximizing $\mathcal{L}_n(\alpha, \mu_1, \mu_2, \sigma_1, \sigma_2)$), each observation X_i can be assigned the posterior probabilities

$$p(1; X_i) = \frac{(1-\widehat{\alpha})\phi(X_i; \widehat{\mu}_1, \widehat{\sigma}_1)}{f(X_i; \widehat{\alpha}, \widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1, \widehat{\sigma}_2)}, \qquad p(2; X_i) = 1 - p(1; X_i),$$

which give the probability that X_i belongs to the corresponding component in the mixture model. One may then assign X_i to one of the components by using the maximum a-posterior estimate (MPE), which assigns the $j \in \{1,2\}$ to district X_i for which $p(j; X_i)$ is maximal.

3 Results

3.1 The Regional Distribution of GDP per Employee

Figure 2 shows the fitted normal mixture and a kernel density estimate of the regional distribution of GDP per employee for the first and last year of our analysis, 1992 and 2006, for which we chose the bandwidth according to Silverman's rule of thumb (Silverman, 1986). Figures for all other years can be found in the appendix (cf. Figures 5 and 6). It is apparent that we observe two quite distinct components in 1992, which have moved considerably together by 2006. The pictures show two modes and two components in 1992, in 2006 there is only a single mode and it is not obvious whether the population consists of two sub-populations or just a single one. Therefore, we test both for the number of modes and for the number of components for all years. Silverman (1981) introduced a nonparametric test for the hypothesis that a density function has k modes against the alternative that it has more than k modes, Bianchi (1997) was the first to apply this test to income distributions. We refer to these two papers for a detailed description of the test. For our data, Silverman's test rejects unimodality in favor of bimodality from 1992 to 2000, see column $p_{\text{Silverman}}(1\text{vs}2)$ of Table 1 for the corresponding p-values. In 2001 the hypothesis of unimodality can only be rejected at the 10 percent level, from 2002 on it cannot be rejected anymore. We confirm these results using the recently introduced likelihood ratio test for bimodality in two-component normal mixture models in mean and variance by Holzmann and Vollmer (2008). The corresponding p-values are shown in column p_{HV}. The hypothesis of two modes cannot be rejected in favor of even more modes in any year by Silverman's test, see column $p_{\text{Silverman}}(2\text{vs}3)$ of Table 1.

As argued in the introduction, the more important feature is the number of components that generate the distribution rather than the number of modes, since components correpond to underlying convergence clubs in a heterogeneous distribution. To test for the number of components here, we will apply the EM test as discussed above. When testing for two components in a normal mixture model, the EM test clearly finds two different components for all years (no matter which level of confidence we apply, the first three digits of the p-values are always zero, see column p_{EM} in Table 1).

The interpretation of this finding is, that levels of GDP per employee in East and West have moved close enough together so that the two underlying distributions do no longer result in separate modes. However, the complete distribution of GDP per employee continues to be best described by a mixture of two separate underlying distributions. Jüssen (2008) exclusively focuses on the number of modes in the distribution and therefore concludes convergence at this point. His study on the number of modes misses the point that the distribution is still generated by two components which cannot be revealed with

Silverman's test. In addition, looking at the dynamics within these two components turns out to be fruitful, to which we turn now.

3.2 Convergence

The posterior analysis confirms the visual first impression that in 1992 all East German districts belonged to the first (poorer) component and all West German districts belonged to the second (richer) component (based on the MPE). Over the years, 35 East German districts moved up from the first to the second component, and six West German districts, namely Aurich, Bentheim, Friesland, Wittmund (all in Lower Saxony), Cochem-Zell (in Rhineland-Palatinate), and Bottrop (in Northrhine Westphalia) fell back from the second to the first component. Figure 4 shows a map of Germany where all Eastern districts which moved up to the second component are colored black and the six Western districts which moved down to the first component are colored gray. Table 2 lists all Eastern districts that moved up to the second component with growth rates and ranks in 1992 and 2006 (within the East). The map shows some interesting patterns of the movement between components. First, the six declining districts in Western Germany are mostly rural districts in rather remote areas where income levels had already been rather low in 1992.[4] There is also a clear regional pattern to the districts in Eastern Germany that have moved up. In particular, three types of districts have moved up. First, seven districts bordering the (former) border with West Germany have moved up. Second, four districts surrounding Berlin have also moved up. Third, a regional cluster of some 15 districts in Thuringia and Saxony have moved up; most of these are close to economically dynamic cities such as Leipzig, Dresden, or Jena. Thus we find clear geographical patterns of districts that moved up that appear to be linked to proximity to Western growth areas as well as emerging Eastern ones.

To investigate this selective convergence process further, we perform additional empirical analyses. Following the classical Barro and Sala-I-Martin (1991, 1992) framework, we find β-convergence among all districts, but also within East and West respectively. This finding is visualized in Figure 3 (left). On the other hand, Figure 1 shows that the standard deviation of GDP per employee across East German districts increases; in other words we do not find σ-convergence in the East. However, we observe σ-convergence within the first component. As reported in Table 1, the σ-parameter of the first component decreases over time (which is due to the fact that some districts move up to the second component).

As one can see in Figure 3 (left), the districts in the East were growing faster than

[4] The exception is Bottrop which is the middle of the Ruhr industrial area that suffered from serious economic decline of the coal and steel industries already since the 1980s.

the districts in the West and would eventually catch up to the West if this development continued. The timing of the catch up (or in the words of our empirical model: movement from the first to the second component) is in theory determined by two factors: The initial level (GDP per employee in 1992) and the speed of convergence (annual growth rate relative to the growth of the second component). However, a closer look at the East suggests that initial levels seemed not to matter. As one can see in Figure 3 (right), at each initial level of GDP per employee there are districts that move from the first to the second component and districts that stay in the first component. Let us illustrate this for two extreme examples: Märkisch-Oderland was the second richest district in the East in 1992, but it had not moved to the second component by 2006. This implies that all 35 districts, which moved to the second component, (and also some others) were ranked higher than Märkisch-Oderland in 2006. On the other extreme, Wartburgkreis was the sixth poorest district in 1992, but it nevertheless managed to catch-up to the West, in fact it had the highest average annual growth rate of all districts, both East and West.

In Table 3 we show the results of simple β-convergence regressions for the East. The dependent variable is the average annual growth rate of GDP per employee between 1992 and 2006. The main independent variable is the initial level of GDP per employee in 1992. As we have already mentioned before, the coefficient of this variable is always negative and highly significant. This means that we observe β-convergence within East Germany. The second column shows the same regression with a "mover" dummy variable that is coded one if the district moves from the first to the second component between 1992 and 2006, and it is coded zero otherwise. The "mover" dummy is highly significant and has a coefficient of 0.012. This variable is clearly endogenous, but it nevertheless gives us the interesting descriptive observation that a district from the East which moved from the first to the second component between 1992 and 2006 had an average annual growth rate that was a little bit more than one percentage point higher than a district that stayed in the first component (and had the same initial level of GDP per employee). In the last column we include an interaction term between the mover dummy and initial level of GDP per employee, which turns out to be insignificant. Thus, there is a level difference of about 1 percentage point in growth between movers and non-movers with the same initial level of GDP per employee, which is independent of initial GDP per employee. This finding (in hand with the visual observations mentioned above) suggests that there are two distinct convergence clubs for GDP per employee in the East of Germany. As we suggest above, proximity to growth nodes in the West or emerging ones in the East appear to be important factors affecting the membership in those two clubs.

3.3 Robustness analysis

In this section we discuss to what extent the exclusion of the 28 richest West German districts affects our results. To this end we fit a three-component normal mixtures to the data to explicitly capture the "very high income" group which had been excluded before. Hence, the observations have density

$$f(x;\psi) = \alpha_1\phi(x;\mu_1,\sigma_1) + \alpha_2\phi(x;\mu_2,\sigma_2) + (1-\alpha_1-\alpha_2)\phi(x;\mu_3,\sigma_3)$$

with $\alpha_j \geq 0$ and $\alpha_1 + \alpha_2 \leq 1$ and $\psi = (\alpha_1, \alpha_2, \mu_1, \mu_2, \mu_3, \sigma_1, \sigma_2, \sigma_3)$. We assume $\mu_1 \leq \mu_2 \leq \mu_3$. We fitted the models using maximum likelihood, i.e. for given observations $X_1, \ldots, X_n$ we searched for the parameter $\widehat{\psi}$ maximizing $\mathcal{L}(\psi) = \sum_{i=1}^{n} \log\big(f(X_i;\psi)\big)$. In Figure 7 we show the fits for the years 1992 and 2006. Given the fitted three-component normal mixture model, we assigned each observation X_i the posterior probabilities

$$p(1;X_i) = \frac{\widehat{\alpha}_1\phi(X_i;\widehat{\mu}_1,\widehat{\sigma}_1)}{f(X_i;\widehat{\psi})},\ p(2;X_i) = \frac{\widehat{\alpha}_2\phi(X_i;\widehat{\mu}_2,\widehat{\sigma}_2)}{f(X_i;\widehat{\psi})},\ p(3;X_i) = 1 - p(1;X_i) - p(2;X_i).$$

We assigned each observation X_i to component j, $j = 1, 2, 3$, according to MPE.

In the above analysis, we have excluded those districts which were assigned more than half of the years to the third component, i.e. the component with the highest mean level in the GDP per employee. If we keep the three-component model and perform the a-posteriori analysis once again, we find that in 1992 the Eastern districts are assigned to the lowest component and the Western districts are assigned to either the second or the third component. In 2006, 37 instead of 35 East German districts are assigned to the second component (same as before plus Meißen, Landkreis and Altmarkkreis Salzwedel). Both districts were also close to the boundary between the first and second component in the two-component model (but not yet above it). It is thus safe to conclude that the analysis of the convergence process is not affected by the exclusion of the 28 richest districts and the reduction to the simpler two-component model.

4 Discussion and Conclusions

We find that the regional distribution of GDP per employee in Germany is well described by a mixture of two normal distributions that is twin peaked in 1992. The two components move closer together to a single peaked distribution but continue to consist of two separate component distributions over the entire observation period. Our analysis is based on a parametric mixture model, which is obtained through rigorous testing of all assumptions involved.

In a posterior analysis we have identified the East German districts which converged to West German levels of GDP per employee since unification. While we do find β-convergence within the entire country and also within the East and the West respectively, movements from the first to the second component cannot be explained by initial levels of GDP per employee. The annual growth rate of a district that moves from the first to the second component is about one percentage point higher than the growth rate of a district that stays in the first component and had the same level of GDP per employee in 1992 as the district that moved up. We thus conclude that there are two different convergence regimes in the East which are independent of a districts' initial levels of GDP per employee.

Membership in the better convergence regime seems to be related to proximity to the (former) West and Berlin as well as proximity to the emerging growth nodes in Thuringia and Saxony. This largely confirms the key predictions of the new economic geography models discussed above. For one, due to rapid monetary and economic union which largely wiped out East German industry and led to substantial out-migration of skilled labor, East Germany started with distinct locational disadvantages which the models above suggest might have led to persistent bad equilibria. For many districts in East Germany, this still describes the situation as of 2006. At the same time, some East German districts, due to a combination of locational and public policy factors, were able to build up, or benefit from proximity to, agglomeration tendencies wither in the West, Berlin, or emerging growth poles in East Germany. The policy challenge for coming years will be to either support the creation of more such growth poles in the East or facilitate the linkage of more districts to existing ones.

References

Akerlof, G., A. Rose, J. Yellen, and H. Hessenius (1991) East Germany in from the Cold: The Economic Aftermath of Currency Union. *Brookings Papers on Economic Activity* **22**, 1–106.

Barro, R. and X. Sala-i-Martin (1991) Convergence across States and Regions. *Brookings Papers on Economic Activity* **22**, 107–182.

Barro, R. and X. Sala-i-Martin (1992) Convergence. *Journal of Political Economy* **100**, 223–251.

Bianchi, M. (1997) Testing for Convergence: Evidence from Non-Parametric Multimodality Tests. *Journal of Applied Econometrics* **12**, 393–409.

Burda, M. (2006) Factor Reallocation in Eastern Germany after Reunification. *American Economic Review Papers and Proceedings* **96**, 368-374.

Burda, M. (2008) What kind of shock was it? Regional Integration and Structural Change in Germany after Unification. *Journal of Comparative Economics* **36**, 557-567.

Chen, H. and J. Chen (2003). Tests for homogeneity in normal mixtures with presence of a structural parameter. *Statistica Sinica* **13**, 351–365.

Chen, J. and P. Li. (2009) Hypothesis test for Normal Mixture Models: the EM Approach. *Annals of Statistics* **37**, 2523–2542.

Colavecchio, R., D. Curran and M. Funke (2010) Drifting Together or Falling Apart? The Empirics of Regional Economic Growth in Post-Unification Germany. *Applied Economics* forthcoming.

Fraley, C. and A. Raftery (2002) Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* **97**, 611–631.

Funke, M. and H. Strulik (2000) Growth and Convergence in a Two-Region Model of Unified Germany. *German Economic Review* **1**, 363-384.

Hathaway, R.J. (1985) A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Annals of Statistics* **13**, 795–800.

Holzmann, H. and S. Vollmer (2008) A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU. *AStA Advances in Statistical Analysis* **92**, 57–69.

Jüssen, F. (2008) A distribution dynamics approach to regional GDP convergence in unified Germany. *Empirical Economics* **37**, 627–652.

Krugman, P. (1991) History versus Expectations. *Quarterly Journal of Economics* **106**, 651- 667.

Paap, R. and H.K. van Dijk (1998) Distribution and mobility of wealth of nations. *European Economic Review* **42**, 1269–1293.

Quah, D.T. (1993) Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics* **95**, 427–443.

Quah, D.T. (1996) Empirics for economic growth and convergence. *European Economic Review* **40**, 1353–1375.

Schäfer, A. and T. Steger (2010) History, Expectations, and Public Policy: Economic Development in Eastern Germany. Discussion paper. University of Leipzig.

Silverman, B. (1981) Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society Series B* **43**, 97–99.

Silverman B. (1986) *Density estimation for statistics and data analysis.* Chapman and Hall, New York.

Sinn, G. and H.-W. Sinn (1991) *Kaltstart. Volkswirtschaftliche Aspekte der Deutschen Vereinigung.* Mohr-Siebeck, Tübingen.

Vollmer, S., H. Holzmann and J. Weisbrod (2010) Peaks vs. Components. Discussion Paper. University of Hannover.

Uhlig, H. (2006) Regional Labor Markets, Network Externalities and Migration: The Case of German Reunification. *American Economic Review Papers and Proceedings* **96**, 383-387.

Uhlig, H. (2008) The slow decline of East Germany. *Journal of Comparative Economics* **36**, 517-541.

Table 1: ML estimates of a two component normal mixture, p-values of the EM test (p_{EM}), the bimodality test by Holzmann and Vollmer (2008) (p_{HV}) and p-values of Silverman's test.

Year	$\widehat{\alpha}$	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$	p_{EM}	p_{HV}	$p_{\mathrm{Silverman}}(1\mathrm{vs}2)$	$p_{\mathrm{Silverman}}(2\mathrm{vs}3)$
1992	0.745	4.420	4.674	0.0448	0.0383	< 0.01	< 0.01	< 0.01	0.79
1994	0.747	4.521	4.678	0.0362	0.0359	< 0.01	< 0.01	< 0.01	0.47
1995	0.745	4.540	4.681	0.0340	0.0359	< 0.01	< 0.01	< 0.01	0.42
1996	0.736	4.557	4.687	0.0301	0.0334	< 0.01	< 0.01	< 0.01	0.89
1997	0.731	4.572	4.693	0.0309	0.0325	< 0.01	< 0.01	< 0.01	0.97
1998	0.737	4.573	4.697	0.0318	0.0336	< 0.01	< 0.01	< 0.01	0.89
1999	0.728	4.584	4.701	0.0315	0.0313	< 0.01	< 0.01	< 0.01	0.22
2000	0.752	4.589	4.706	0.0298	0.0328	< 0.01	< 0.01	< 0.01	0.55
2001	0.775	4.598	4.705	0.0289	0.0359	< 0.01	0.07	0.05	0.70
2002	0.756	4.614	4.709	0.0300	0.0338	< 0.01	0.30	0.96	0.68
2003	0.724	4.626	4.714	0.0301	0.0341	< 0.01	0.35	0.76	0.72
2004	0.790	4.621	4.714	0.0259	0.0371	< 0.01	0.21	0.40	0.40
2005	0.841	4.617	4.715	0.0221	0.0426	< 0.01	0.12	0.15	0.72
2006	0.832	4.624	4.725	0.0234	0.0419	< 0.01	0.12	0.15	0.87

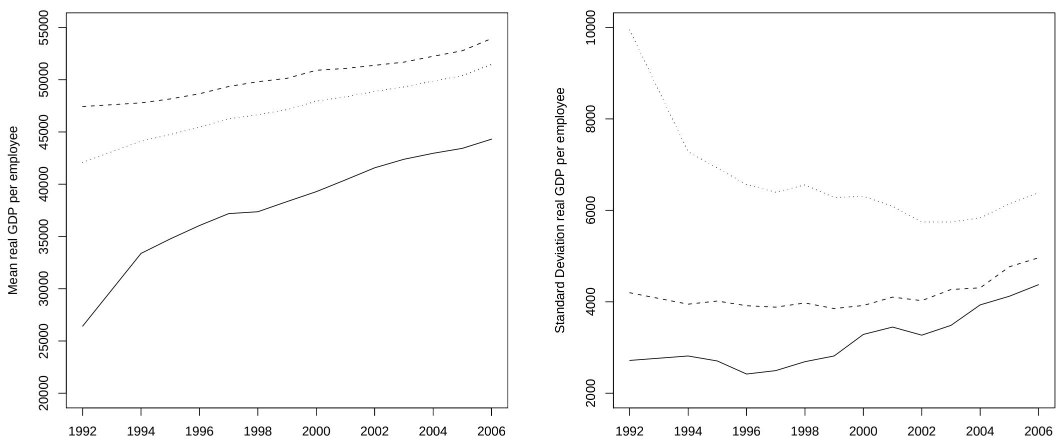

Figure 1: Mean and Standard Deviation of GDP per employee for all districts (dotted line), districts from the East (solid line) and districts from the West (dashed line).

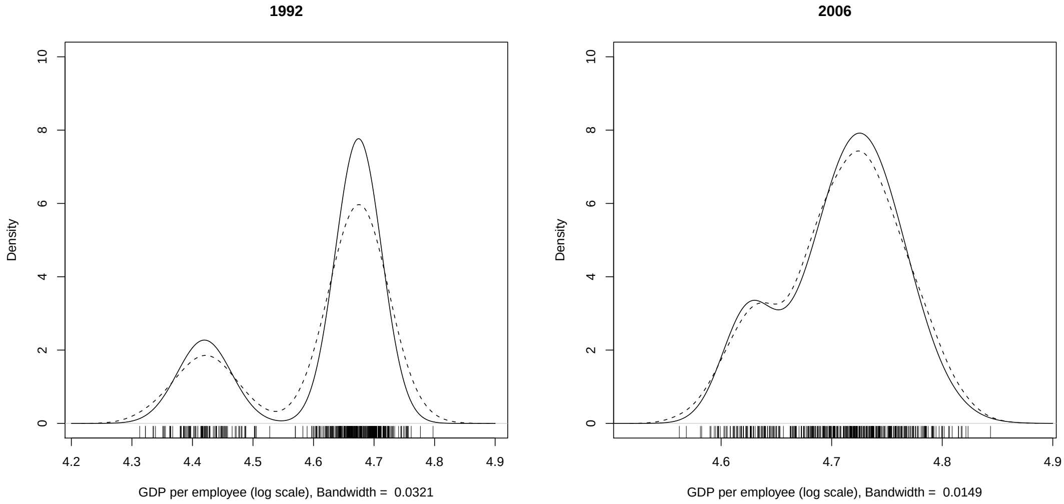

Figure 2: Distribution of GDP per employee in 1992 and 2006. A fitted two component normal mixture (solid line) compared to a kernel density estimator (dashed line).

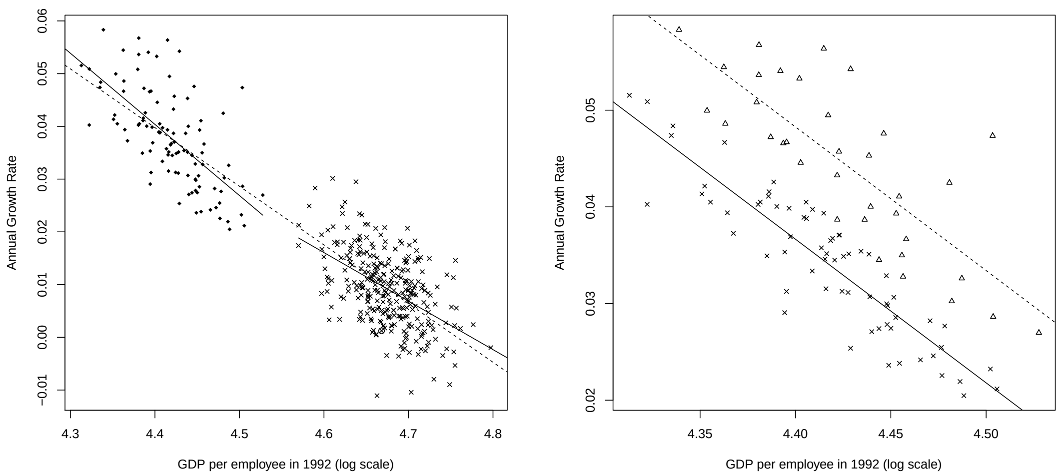

Figure 3: Left: β-convergence in Germany (dashed line) and East and West respectively (solid lines). Right: Two convergence clubs in the East, movers (triangles) and non-movers (crosses).

Table 2: East German districts that caught up to the West

County Name	Rank 1992	Rank 2006	Year(s) moved	Growth Rate
Brandenburg (out of 18)				
Dahme-Spreewald, Landkreis	4	1	1996-2004	4.74
Spree-Neiße, Landkreis	10	4	1995	4.25
Oberspreewald-Lausitz, Landkreis	21	11	2002-2005	4.11
Oder-Spree, Landkreis	48	13	1999	4.57
Ostprignitz-Ruppin, Landkreis	50	31	2006	3.87
Uckermark, Landkreis	55	9	2000	4.95
Teltow-Fläming, Landkreis	75	25	1999	4.66
Oberhavel, Landkreis	84	7	2001	5.68
Prignitz, Landkreis	90	33	2006	4.86
Mecklenburg-Vorpommern (out of 18)				
Rostock, Kreisfreie Stadt	7	18	1998-2004	3.26
Wismar, Kreisfreie Stadt	31	5	2001-2004	4.76
Güstrow, Landkreis	32	34	2001-2005	3.45
Demmin, Landkreis	68	6	2003	5.33
Saxony (out of 29)				
Niederschlesischer Oberlausitzkreis	1	14	1994-2004	2.7
Kamenz, Landkreis	3	23	2005	2.86
Chemnitzer Land, Landkreis	9	30	2004	3.02
Muldentalkreis	19	29	1994-2004	3.5
Dresden, Kreisfreie Stadt	35	21	2002	4.00
Mittweida, Landkreis	39	28	2005	3.87
Leipziger Land, Landkreis	51	20	2004	4.33
Döbeln, Landkreis	71	24	2004	4.67
Freiberg, Landkreis	85	12	2005	5.36
Riesa-Großenhain, Landkreis	92	16	2001-2005	5.45
Saxony-Anhalt (out of 14)				
Burgenlandkreis	17	22	2002-2004	3.67
Jerichower Land	22	15	2002-2005	3.93
Börde	37	10	2001	4.53
Saalekreis	59	3	2000	5.64
Harz	67	26	2005	4.46
Anhalt-Bitterfeld	87	19	2003	5.08
Thuringia (out of 23)				
Weimarer-Land, Kreis	18	32	2005	3.28
Sömmerda, Kreis	41	2	2000	5.42
Jena, Kreisfreie Stadt	76	8	2001-2003	5.41
Saalfeld-Rudolstadt, Kreis	79	27	2005	4.72
Ilm-Kreis	94	35	2006	5.00
Wartburgkreis	97	17	2005	5.83

Brandenburg, Mecklenburg-Vorpommern, Saxony, Saxony-Anhalt and Thuringia are East German Länder (NUTS-1).

Table 3: β-Convergence in East Germany

Dependent variable: Annual growth rate of GDP per employee			
log GDP per employee 1992	-0.135^{***}	-0.149^{***}	-0.149^{***}
	(0.015)	(0.009)	(0.011)
Mover		0.012^{***}	0.002
		(0.001)	(0.085)
Mover * log GDP per employee 1992			0.002
			(0.019)
Constant	0.634^{***}	0.691^{***}	0.694^{***}
	(0.067)	(0.040)	(0.050)

*** denotes statistical significance at the 1 percent level.

** denotes statistical significance at the 5 percent level.

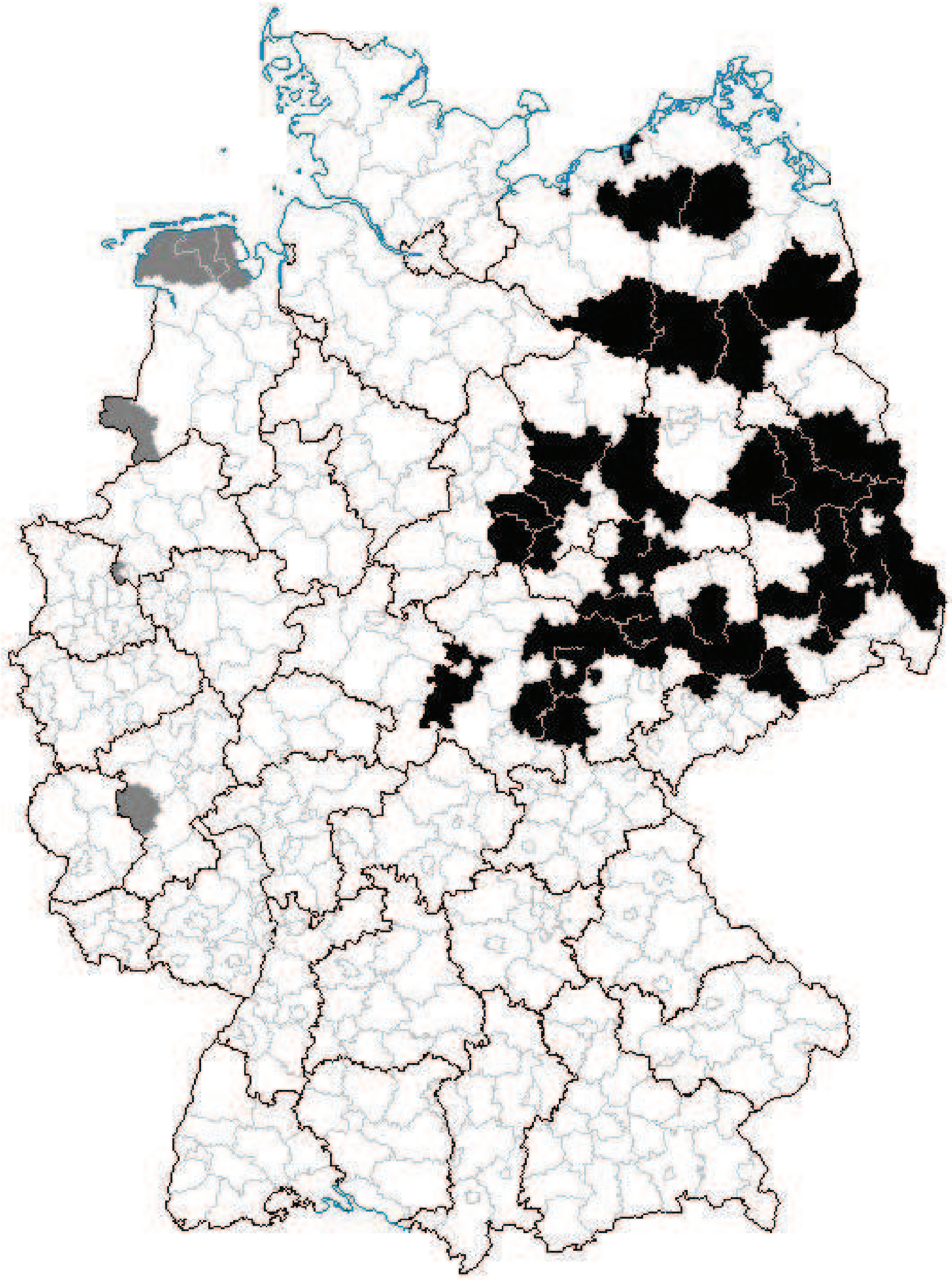

Figure 4: Map of Germany. East German districts that catch up to the West are colored black, and the six West German districts that fell back are colored gray.

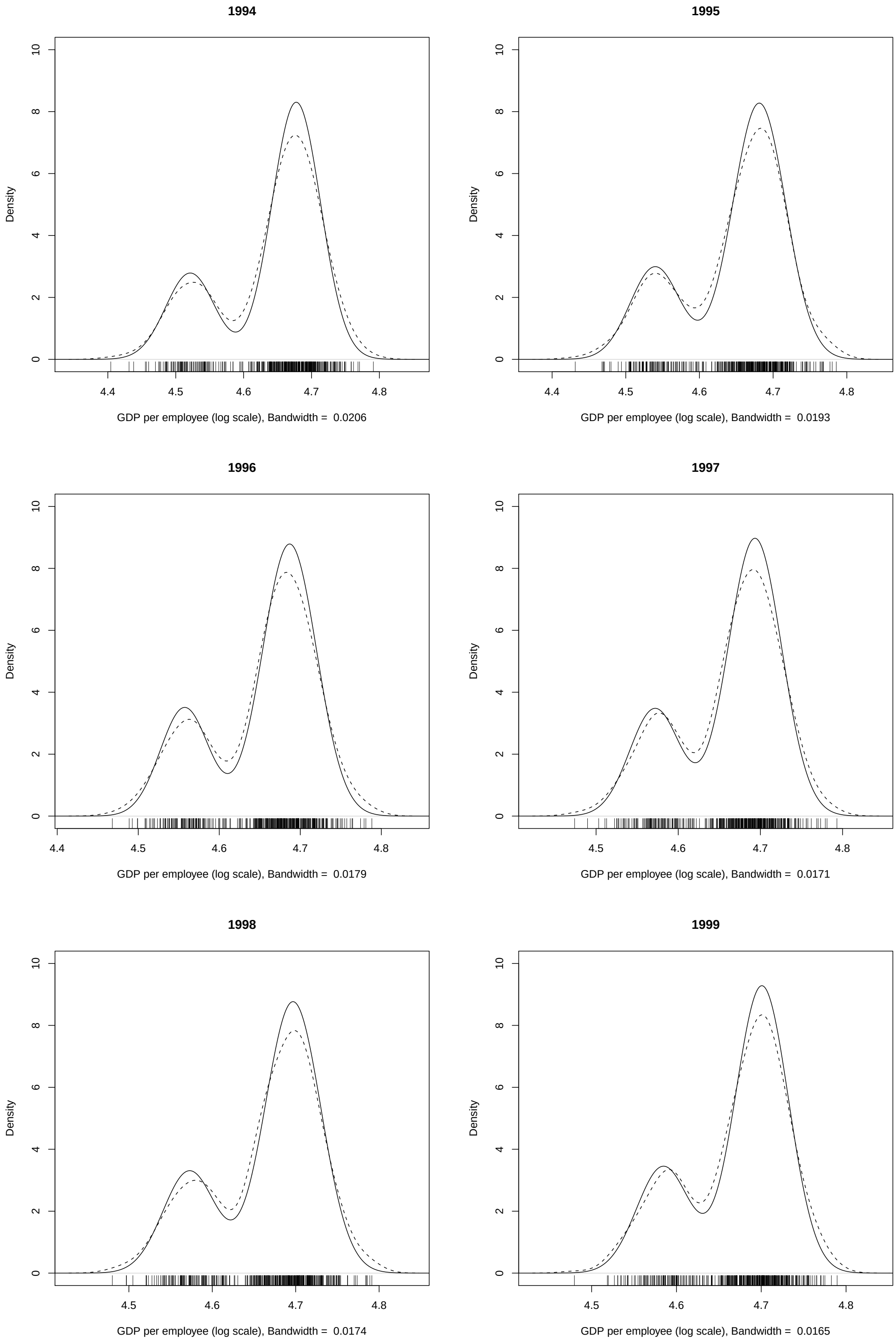

Figure 5: Distribution of GDP per employee between 1994 and 1999. A fitted two component normal mixture model (solid line) compared to a kernel density estimator (dashed line).

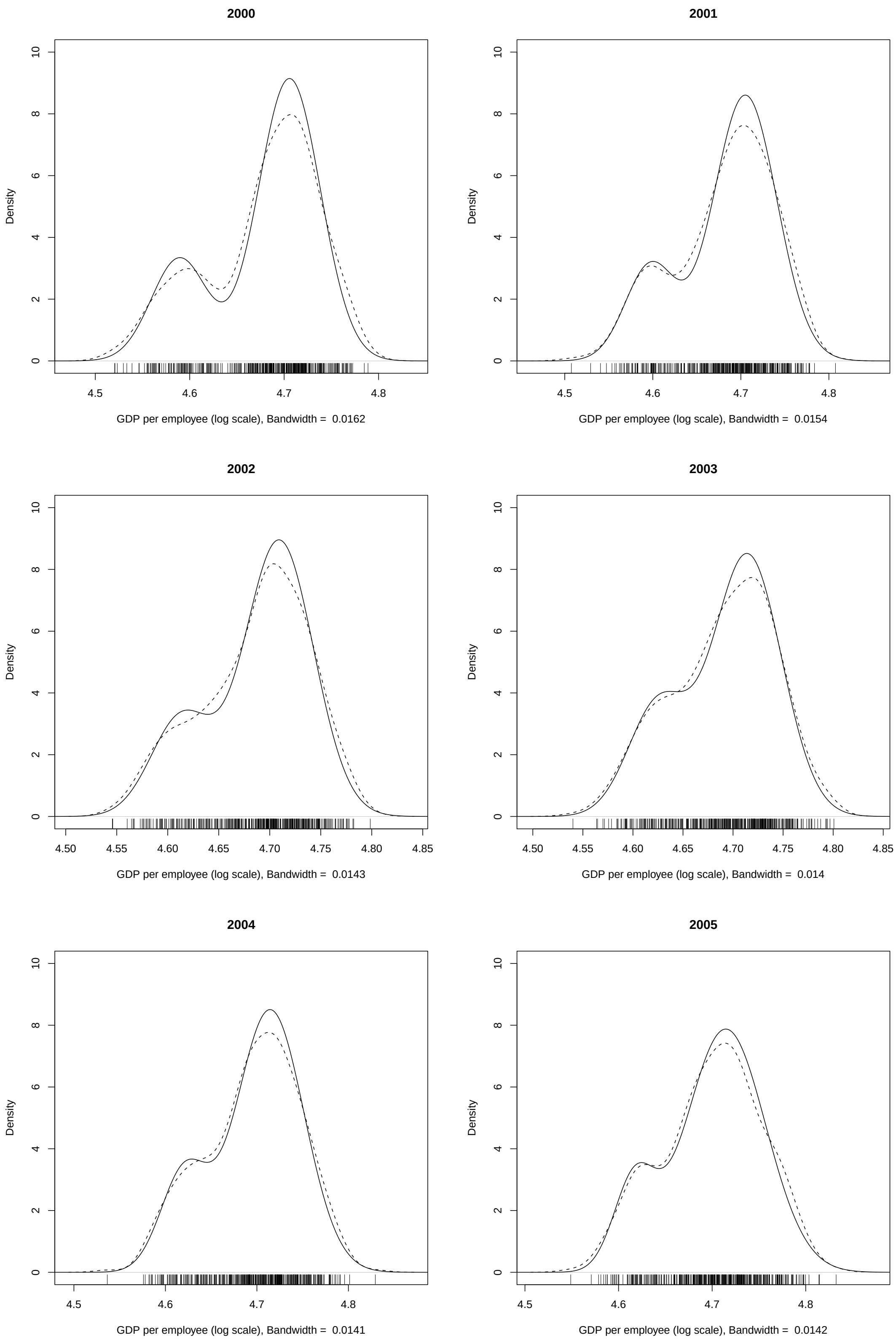

Figure 6: Distribution of GDP per employee between 2000 and 2005. A fitted two component normal mixture model (solid line) compared to a kernel density estimator (dashed line).

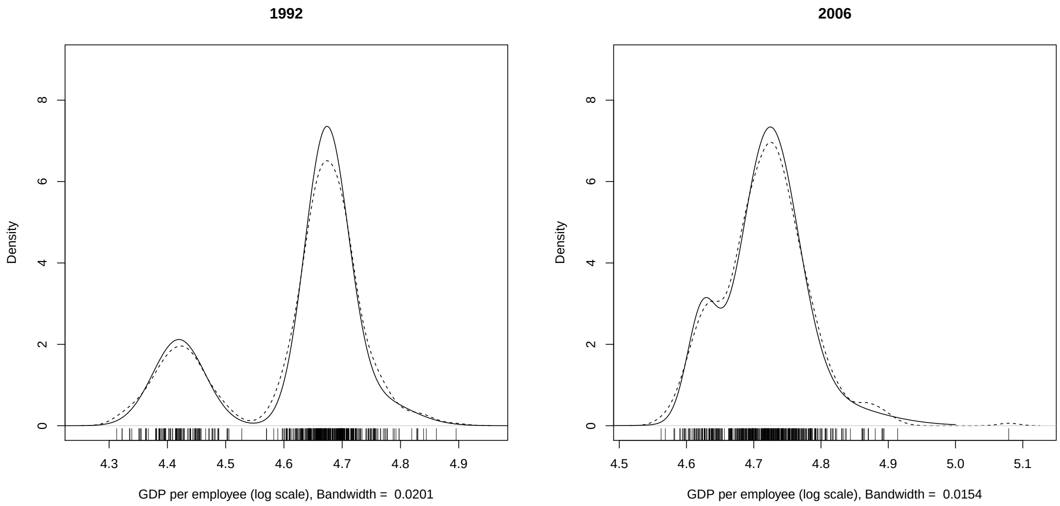

Figure 7: Distribution of GDP per employee in 1992 and 2006. A fitted three component normal mixture (solid line) compared to a kernel density estimator (dashed line).