Weighted convergence and regional clusters across China*

M. J. Herrerías† **Vicente Orts**† **Emili Tortosa-Ausina**‡

November 11, 2010

Abstract

We analyse per capita GDP convergence among 28 Chinese provinces between 1952 and 2005 using the distribution dynamics approach. Compared with previous studies, we provide a more complete view by including some additional information such as the asymptotic half-life of convergence, mobility indices and the continuous version of the ergodic distributions. In addition, we also extend the analysis to evaluate whether patterns could differ if weighted by either the population living in each province or their economic sizes, together with the existence and magnitude of spatial spillovers. The unweighted, unconditional analysis corroborates and supplements previous findings, especially those indicating that convergence patterns differ strongly under either pre- or post-reform trends. Both the weighted and space-conditioned analyses indicate that convergence could be much faster when these factors are introduced in the analysis. Implications are especially relevant when weighting by population, since results indicate that the number of people escaping from relative poverty would be much higher than the figure predicted by the unweighted analysis.

Key words and phrases: China, convergence, distribution dynamics, provinces, weights

JEL Classification: C16, O18, O47, R11

Communications to: M.J. Herrerías, Departament d'Economia, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló, Spain. Tel.: +34 964388615, fax: +34 964728591, e-mail: `herreria@uji.es`

*M.J. Herrerías and Vicente Orts gratefully acknowledge the financial support of the Ministerio de Ciencia e Innovación (ECO2008-06057/ECON) and Generalitat Valenciana (BFPI06/442 and PROMETEO/2009/068). Emili Tortosa-Ausina acknowledges the financial support of Fundació Caixa Castelló-Bancaixa (P1.1B2008-46), Ministerio de Ciencia e Innovación (ECO2008-03813/ECON and ECO2008-05908-C02-01/ECON) and Generalitat Valenciana (PROMETEO/2009/066). We also thank Qiao Yongyuan for his helpful assistance on data issues, as well as to three anonymous referees for helpful comments. The usual disclaimer applies.

†Universitat Jaume I.

‡Universitat Jaume I and Ivie.

1. Introduction

Income disparity across Chinese regions is a major concern among policymakers, and also an interesting case of study from an academic point of view. Growing inequality increases redistributive tax pressures, which deters investment incentives and can also lead to a more unstable socio-political environment for economic activities (Alesina and Perotti, 1993; Alesina and Rodrik, 1994). Given its potential to thwart both economic growth and stabilisation, inequality and poverty reduction across regions is one of the fundamental problems that the Chinese government must solve and, accordingly, several initiatives have been undertaken to promote income distribution across Chinese regions. One of the most prominent ones took place in the early 2000s, when the Chinese government launched the *Great Western Development Program* (GWDP) with the aim of investing more in the western regions, where economic development is lower. The purpose of this programme was to balance the degree of development across regions and to reduce poverty, but this expectation has not been fulfilled as of today.[1] Therefore, examining whether convergence is taking place across Chinese provinces is not only of great significance because of the sheer number of people whose welfare is involved (China represents about one-fifth of the world's population), but also because it makes it possible to evaluate the success of policies designed to alleviate the magnitude of the inequalities. In addition to this, China has an economy that is undergoing a transition from a centrally planned to a more market-oriented economy, and which has its own particular characteristics (see Herrerias and Orts, 2010). Therefore, as indicated by Sakamoto and Islam (2008), the Chinese case could help to assess whether switching from central planning to a market mechanism makes a difference with regard to convergence. This could be increased further to examine whether interregional differences in income levels tend to disappear or to increase over time as a result of this transformation.

As indicated by Islam (2003), there are different definitions of convergence, each of which is linked to growth theory in a different way. According to this author, not only are there different ways in which convergence can be understood but also different methodologies to evaluate it. The definitions would include convergence *within* an economy versus *across* economies, *growth rates* versus income *level* convergence, β-convergence versus σ-convergence, unconditional versus conditional convergence, global versus local or club-convergence, and deterministic versus stochastic convergence. The methods would be the informal cross-section approach, the for-

[1] A recent study by Ho and Li (2008) examines the time series properties of inequalities in China concluding that, although they actually increased between 1952 and 2000, policies can be effective to alleviate their severity.

mal cross-section approach, the panel approach, the time-series approach, and the distribution approach. This is perhaps the most comprehensive and up-to-date survey, but given the importance of the issue, other significant works have also been published such as Quah (1997b), De la Fuente (1997) Durlauf and Quah (1999), or Temple (1999).

From these surveys it can be seen that there is a substantial body of theoretical and empirical research focusing on the issue of country and regional convergence, and the Chinese case is by no means an exception. Most previous studies have examined convergence across Chinese provinces using parametric techniques which adopt either cross-section or panel data approaches (see, for example Rozelle, 1994; Jian et al., 1996; Chen and Fleisher, 1996; Raiser, 1998; Yao and Zhang, 2001a,b; Weeks and Yao, 2003; Wang, 2003; Pedroni and Yao, 2006).[2] However, the empirical evidence found in the literature on the subject is rather mixed. For example, Rozelle (1994) found divergence within the Jiangsu province for the 1984–1989 period and Jian et al. (1996) concluded that during the pre-reform period (prior to 1978) the Chinese provinces tended to diverge, while in the post-reform period (after 1978) until 1990, a tendency towards absolute convergence predominated. Nevertheless, this pattern changes to divergence in per capita incomes from 1990 to 1993. Using cross-section and panel data techniques, Chen and Fleisher (1996) provide evidence of divergence in the pre-reform period and convergence from 1978 to 1993.[3] In contrast, opposite results were found by Weeks and Yao (2003) using the GMM estimator. In addition, despite using different methodological approaches, Yao and Zhang (2001a,b) found that the Chinese regions did not converge in the reform period (1978–1995). Indeed, their results clearly indicated evidence of divergence in the different geo-economic clubs (coastal and non-coastal zones). Only when they controlled for regional effects and other determinants of growth, did they find conditional β-convergence using the panel data approach (Yao and Zhang, 2001b). More recently, in a study covering the 1952–1997 period, and using non-stationary panel techniques, Pedroni and Yao (2006) provided empirical support for the fact that the long-run tendency since the reforms has been for provincial-level incomes to continue to diverge. They added that this divergence cannot be attributed to the presence of separate, regional convergence clubs divided among common geographic sub-groupings such as the coastal versus interior provinces or preferential policies. This ambiguity in the results, of course, depends on the selection of the period under study,

[2]Studies such as Rozelle (1994), Jian et al. (1996), Gundlach (1997), and Yao and Zhang (2001a,b) did not consider the endogeneity or the dynamics of the models, and were thus in line with the critique stated by Caselli et al. (1996). Only a few studies try to deal with this problem, such as Weeks and Yao (2003), Wang (2003) and Ding et al. (2008) among others.

[3]Similar results were found by Raiser (1998) and Wang (2003) for the post-reform period.

the estimation method that is used and the variables that the researcher has considered. Furthermore, in studies that examine the differences in per capita income among clubs, the *a priori* selection of these small groups of regions could also affect the empirical analysis (Maasoumi and Wang, 2008).

In contrast, there is little empirical evidence of convergence across the Chinese provinces using the distribution dynamics model developed by Quah (1993a). Only two recent papers apply this new way of analysing economic convergence across the Chinese provinces: Bhalla et al. (2003) and Sakamoto and Islam (2008).The former investigated convergence patterns from 1952 to 1997 using per capita data among Chinese provinces. They concluded that there is evidence of convergence within the pre-defined geo-economic sub-regions, but no evidence was found of convergence between the sub-regions. In particular, they argued that the gap between the eastern and the central regions was small in the pre-reform period, but widened rapidly in the reform period. The same pattern occurs with eastern and western provinces, but with a more significant fluctuation over time. These results imply a strong divergence between these two pairs of regions. More recently, the latter authors—Sakamoto and Islam (2008)—found similar results. Indeed, their findings indicated that the distribution of per capita income across Chinese provinces over time has attained a bimodal characteristic with two opposing tendencies in the two sub-periods considered (1952–1978 and 1978–2003). During the pre-reform period the dynamics of the distribution indicated that there were more provinces piling at lower values of per capita income, whereas during the post-reform period the dynamics of the distribution moved in the opposite direction, namely, there were more provinces moving towards higher income groups. In spite of these results, Sakamoto and Islam (2008) argued that the distribution dynamics of the reform period do not seem to have led to a stable pattern yet, thus making prediction difficult, and hence it remains an open issue to be analysed further.

This paper examines the complexity of the convergence process in per capita GDP across the 28 Chinese provinces over the period 1952–2005, which means that our proposal therefore stands with those using the distribution approach developed by Quah (1993a,b). Unlike previous studies that apply either σ- or β-convergence in cross section or panel data techniques (which sometimes require strong assumptions) we allow data to reveal the nature of the relationship of interest by using nonparametric techniques. It is mainly a data driven approach, and we do not impose any assumption or restriction on the specification of the density of the distribution. So, in its initial steps, our investigation differs only slightly from that conducted by Sakamoto and Islam (2008).

However, we introduce a series of variations with respect to both Bhalla et al.'s (2003) and Sakamoto and Islam's (2008) proposals. Some of the differences we introduce have to do with the fact that the analyses of Chinese income distribution and convergence have dealt with the behaviour of incomes in terms of *provinces*—i.e., regional convergence. However, as indicated by Jones (1997), while this is a common way to view and analyse data, it can be highly misleading: should provincial borders be drawn differently, conclusions might vary remarkably. Alternatively, we could weight each province by its population (although other weighting schemes are possible) so that the unit of observation was then a *person* instead of a province. As indicated by Sala-i-Martin (2006), the unweighted approach is not useful if one is concerned about human welfare, since different provinces have varying population sizes. In this regard, the most important fact to note is that, for instance, by 2005 the population living in Sichuan was around 20 times larger than the population living in Qinghai or in Ningxia. Disparities were even higher by 1952—the population living in Sichuan was more than 40 times larger than that living in Ningxia. Therefore, the experience of the most populated provinces largely determines what happens to the "average" person in China.

By weighting by population, some researchers have drawn different conclusions to those reached via unweighted analyses. For instance, Jones (1997) showed that the emergence of a bimodal distribution disappeared once each country data point was weighted by population, whereas Schultz (1998) found that, when one uses population-weights, it is no longer true that incomes tend to diverge. Given the disparities in terms of both population and GDP across Chinese provinces, one may expect some interesting conclusions also to emerge in the case of China when comparing our results to the unweighted analysis by Sakamoto and Islam (2008).

The distribution analysis approach is also attractive because of its ability to disentangle the existence of spatial spillover effects, in a similar fashion to Quah (1996c). Following this author's approach, we measure whether these spillovers could exist or not by evaluating the magnitude of the contiguity effect across Chinese provinces. The rationale for this lies in the fact that the economic development of a particular region could be strongly related to that of its neighbouring provinces. The issue is particularly relevant in the case of China, whose government considered that by developing the coastal regions, the central and western provinces would also boost their development via (spatial) spillover effects. However, empirical evidence evaluating these policies is still scarce. Only Brun et al. (2002) have conducted research on the issue, their findings indicating a relative failure of the growth of the coastal regions from 1981 to 1998 to trigger development in the western provinces. Therefore, according to these authors, it would be wrong to expect spillover effects to be enough to reduce disparities between Chinese provinces, at least in the short run.

The rest of the paper is organised as follows. Section 2 describes the main trends in provincial distribution of per capita GDP in the Chinese economy during the period under consideration. Section 3 deals with the technical aspects of the distribution analysis model. In Section 4 we analyse the results of applying the model to per capita income data for 28 Chinese provinces. Finally, we present some concluding remarks in Section 5.

2. Emerging patterns in provincial distribution of Chinese per capita GDP: 1952–2005

A comprehensive description of the evolution of the Chinese economy is beyond the scope of this section and the reader is directed to other studies, such as Lardy (1992), Chai (1998) or Bramall (2000), for more details. Here, we briefly summarise the most important trends of the per capita GDP and population across Chinese provinces during the period under consideration (1952–2005). Table 1 shows the per capita GDP and population of the 28 Chinese provinces in 1952, 1978 and 2005, the growth rates of both magnitudes for the whole period (1952–2005) and for the two sub-periods considered (1952–1978 and 1978–2005), as well as the corresponding coefficient of variation for all magnitudes reported. As can be seen, there are substantial differences in per capita GDP among provinces in each year, as well as for each province between 1952 and 2005. In 1952 the average of provincial per capita GDP in China was of 134.60 Yuan and its coefficient of variation was 0.60. Between 1952 and 2005, the per capita GDP of Chinese provinces, measured in constant 1952 prices, was growing at a cumulative average growth rate of 6.41%, the result being that the average provincial per capita GDP achieved 4,382.60 Yuan in 2005. However, by this year the coefficient of variation increased to 1.08. This rapid growth and the increase in regional disparities have two very different steps in time: prior to the economic reforms, i.e. before 1978, and the post-reform period.

The pre-reform period (1952–1978) was characterised by the central planning of the allocation of economic resources and an unstable political environment. China experienced many booms and boosts, like the Great Leap Forward (1958–1961) or the famine caused by failures in the agricultural sector following the unstable economic and political environment that accompanied the Cultural Revolution (1966–1976). Nevertheless, in spite of these turbulences the average provincial per capita GDP grew at a rate of 3.73%, although regional inequalities increased significantly in that period. During those years the coefficient of variation of provincial per capita GDP increased from the aforementioned 0.60 to 1.27. The existence of barriers across regions (Rozelle, 1994) probably accounts for the divergence rather than convergence in per capita GDP that took place in the pre-reform period. These barriers are to be understood in the broad sense of the term and are notably related with the mobility of workers, the unequal specialisation in the different economic activities of provinces, the promotion of investment in heavy industry rather than in agriculture or the centralised fiscal system (Wei, 1996).

In contrast, in the post-reform period (1978–2005), the average growth rate of per capita GDP of the Chinese provinces increased to 8.97%, while the provincial inequality in per capita GDP, as measured by the coefficient of variation, declined from 1.27 to 1.08. This period is characterised by the economic reforms initiated in the late 1970s, including the progressive adoption of market-oriented and open-door strategies for development (that culminated in 2001 with its adhesion to the WTO), and which were gradually transforming the Chinese economy towards a more market-oriented, decentralised and open economy.

To sum up, these stylised facts reflected the most widely extended result in empirical studies, i.e. the absence of convergence or even divergence during the pre-reform period and slight convergence in the post-reform period.[4]

However, a closer look at the data reveals that, even within each sub-period, the dynamics of the first moments of the distribution of per capita GDP were very volatile, especially the variance throughout the first sub-period, but also in the second sub-period.[5] Although in the period 1978–2005, the variability of the variance of the distribution was much smaller, some changes in its trend can also be seen. Between 1978 and the late 1980s and early 1990s, the variance in distribution drops regularly, then the trend changes and increases again almost to the end of the sample (2004). At the end of the sample a new change in the trend of the variance appears indicating a new decrease in the inter-provincial inequality. Table 1 shows that the performances of the provinces in each economic zone (east, central and west) were also very different to each other throughout the period under consideration. Thus, while the provinces in the eastern zone reproduce the aggregate changes in average growth and inequality of per capita GDP on a different scale, in the central and western zones the dynamics varies markedly between them as well as between the two sub-periods. In the pre-reform period the average growth rate of the western provinces was superior to that of the provinces in

[4]See, for example, Bhalla et al. (2003).

[5]We do not report the coefficient of variation of the all distributions year by year, but they are available upon request.

the central zone (3.84% and 2.66% respectively), this difference being reversed in the post-reform period (8.03 and 8.83%). Furthermore, the dispersion of the provincial growth rates slowed down significantly in the post-reform period in all the zones. At the same time, while the inequality among provinces in the central zone declined in the pre-reform period, they did not change significantly throughout the whole of the post-reform period (the coefficient of variation was of 0.42, 0.37 and 0.38 in 1952, 1978 and 2005 respectively). The western provinces, however, showed only a slight increase in dispersion between 1952 and 1978, which was reversed during the post-reform (the coefficients of variation were of 0.37, 0.43 and 0.37 respectively). The picture is complex, and when we look at the data in more detail, the more differences we find in the performance of different provinces and zones. In fact, as stated by Quah (1996b,c,d, 1997a), looking only at the first and second moments of the distribution is likely to be uninformative when dealing with multimodal distributions, as could be the case, and therefore it is better to analyse the entire distribution of provincial per capita GDP and its dynamics.

3. Per capita income convergence as distribution dynamics

As indicated by Islam (2003), "research on convergence has indeed proceeded in many directions using many different definitions and methodologies". The same source further claims that "it is therefore not unreasonable to feel somewhat dazed by the variety of results and conclusions encountered in the literature". However, as also indicated by this author, a close review reveals that "at a broad level there is considerable agreement among the results" (Islam, 2003, p.341). Some of the definitions and methodologies available for approaching the convergence debate have been briefly reviewed in the introduction. In this Section we will provide the details of both the definition of convergence and the methods with which to measure it that are used in this paper.

Regarding the different ways in which convergence has been understood, our definition matches several of those proposed by Islam (2003). Specifically, it refers to convergence *within* an economy (as opposed to convergence *across* economies), and to *GDP per capita*-convergence (as opposed to *TFP* convergence). Initially we will analyse the issue of *unconditional* (absolute) convergence, although we will also partly examine the issue of *conditional* convergence, since this will assess whether physical contiguity and the fact that different provinces have different weights—in terms of both population and/or GDP—might affect the convergence process.

In a context of generalised growth, such as that of all Chinese provinces during the analysed period (1952–2005), some authors (Eaton and Eckstein, 1997) refer to convergence or divergence patterns as: (i) "parallel growth", which in our setting would occur when per capita GDP grows at similar rates for provinces with different per capita GDP levels; (ii) "divergent growth", which takes place when richer provinces grow faster than poorer ones, and (iii) "convergent growth", which refers to the growth of poorer provinces in relation to richer ones. In our particular setting, in which high growth rates have been generalised throughout the period (see Table 1) we will refer indistinctly to "convergent growth" and "convergence", and to "divergent growth" and "divergence", whereas "parallel growth" will imply that the per capita GDP distributions remain constant.

Regarding the use of different methodologies for examining convergence, we chose the distribution approach, which stresses the importance of analysing distribution dynamics in order to understand the mechanics of economic development. Choosing these methods enables us to provide a natural complement for one of the most recent contributions on the convergence debate in China, namely, the paper by Sakamoto and Islam (2008).

These views are strongly supported not only by Quah (1993a, 1996b,d) but also by many others who advocate the analysis of the dynamics of the entire cross-sectional distribution of per capita income (or labour productivity). The reasons for this lie in the fact that uncovering *all* the information on the dynamics using only summary statistics is a questionable procedure. Accordingly, empirical studies have shown consistent evidence of a cross-country income distribution displaying bimodality with a marked thinning in the middle. This transformation implies that, for instance, while by the 1960s many countries belonged to the middle income group, by the 1990s the world polarised into two groups, namely the rich and the poor, a phenomenon which Quah (1996d) refers to as "twin-peak" or "two-club" convergence. However, as indicated by Cetorelli (2002), there is also a positive probability of an economy moving from one group to the other, i.e. the bimodal distribution could be ergodic. One can therefore observe previously poor economies that grow rapidly and move to join the rich club; reversals of fortune, where fast growth is only temporary and may be followed by abrupt halts and decumulation; or economic disasters involving previously rich economies regressing to lower levels of income (Cetorelli, 2002).

The instruments provided by Quah (1993a,b), along with some others borrowed from the literature on income inequality (Shorrocks, 1978), are of remarkable interest for analysing provincial per capita income dynamics in China. Quah's critique on previous approaches to examining convergence (essentially those based on analysing β- and σ-convergence) points out that conclusions are based on only two summary statistics of the entire distribution. However, both the mean and the standard deviation give an interesting but incomplete picture of the entire distribution of per capita income, for it conceals some significant features such as the existence of multiple modes. This and related phenomena would be overlooked unless an analysis taking into account different groups of provinces were performed; however, focusing on the *entire* distribution is even better than carrying out the analysis for different groups of provinces.[6]

3.1. Unweighted analysis

3.1.1. Intra-distribution mobility and ergodic distributions

Our variable of interest is the normalised logarithm of per capita GDP over the period 1952–2005, i.e. divided by the mean for the 28 provinces.[7] We consider this type of normalisation because of the informativeness of its interpretation: the closer a value is to unity, the closer it will be to the national average. Therefore, the more values there are close to unity, the higher the convergence to this national average will be. Our selected variable is the same as the one chosen by Sakamoto and Islam (2008), but it is normalised in a slightly different way, since they use the log of normalised (divided by the mean) per capita GDP. We consider that our normalisation has the interesting feature of being more directly interpretable, since it allows us to measure, for instance, whether a particular province has twice or half the national average. While the normalisation selected by Sakamoto and Islam (2008) provides similar information, it is not as direct. Like Sakamoto and Islam (2008), we denote this variable by x_{it}, so that $x_{it} = \ln y_{it}/\ln \bar{y}_t$, where y_{it} is the per capita GDP of province i in year t, and $\bar{y}_t$ is the cross-sectional average of y_{it}.

In this context, probability mass concentrating tightly around unity would indicate that

[6] In contrast to the popularity of both the distribution dynamics and the β- and σ-convergence approaches, other proposals such as the Bayesian procedures introduced by Canova and Marcet (1995) have been applied less frequently (see, for instance Rodríguez-López et al., 2009).

[7] For the sake of simplicity, we use the concept of provinces throughout this paper. However, in China there are 23 provinces, 5 autonomous regions, 4 municipalities, and 2 special administration regions (SAR). Tibet and Hainan were excluded due to the lack of data. Moreover, this paper focuses on Mainland China, and consequently we have also excluded Taiwan, Hong Kong and Macao. In addition, Chongqing is included as part of Sichuan province, given that it was part of Sichuan until 1997. Although there is a debate in the literature about the quality of Chinese statistics, a number of researchers do offer evidence to show that they in fact meet the standards required to examine the long-run trends (see, for instance Holz, 2005; Chow, 2006; Bai et al., 2006). We use one of the latest revised compilations edited by the National Bureau of Statistics of China (NBS) in 2010, namely, "China Compendium of Statistics, 1949–2008". This source provides us with information that is sufficiently homogenous, both across Chinese provinces and over time, to perform this study properly.

convergence to this (national) average is taking place. Obviously, if probability mass concentrated around two different values, then it would indicate convergence to two different *clubs* (Ben-David, 1994).

Therefore, in our setting, $x_{i,t}$ refers to province i's normalised per capita GDP in period t, whereas $F_t(x)$ refers to the cumulative distribution of $x_{i,t}$ across provinces. Corresponding to $F_t(x)$ we can define a probability measure $\lambda_t((-\infty, x]) = F_t(x)$, $\forall x \in \mathbb{R}$, λ_t being the probability density function for each indicator across provinces in period t. Therefore, the model analyses the dynamics of λ_t, i.e., the dynamics of the cross-sectional distribution of per capita GDP,[8] for which we consider a stochastic difference equation $\lambda_t = P^*(\lambda_{t-1}, u_t)$, integer t, taking into account that $\{u_t : \text{integer } t\}$ is the sequence of disturbances of the entire distribution, and P^* is the operator mapping disturbances and probability measures into probability measures. In other words, the P^* operator reveals information on how the distribution of per capita GDP at time $t-1$ transforms into a different distribution at time t.

Following Redding (2002), we may assume that the stochastic difference equation is of first order and that operator P^* is time invariant. Thus, by setting null values to disturbances and iterating for $\lambda_t = P^*(\lambda_{t-1}, u_t)$ we obtain the future evolution of the distribution, namely, $\lambda_{t+\tau} = (P^*)^\tau \lambda_t$, and by discretising the set of possible values of x into a finite number of grids (what some authors call "states", "classes" or simply "intervals") e_k, $k \in \{1, \ldots, K\}$, P^* becomes a transition probability matrix, that is to say:

$$\lambda_{t+1} = P^* \cdot \lambda_t \tag{1}$$

In this transition probability matrix, λ_t turns into a $K \times 1$ vector of probabilities that the per capita GDP of a given province is located on a given grid at time t. Transition probability matrices make it possible to measure the probability of a given province moving to a higher (or lower) position on the grid. Calculation the transition probability matrices starts by discretising the set of observations into the selected states e_k.

After classifying each province-year observation into one of the K states, we build up a 5×5 matrix whose p_{kl} entries indicate the probability that a province that is initially in state k will transit to state l during the period or periods considered (T). Each row of the matrix constitutes a vector of transition probabilities, which add up to unity. The interpretation of the different figures in each matrix is straightforward. In the case of the limits between states, those for

[8]From now on, when talking about per capita GDP we will be referring implicitly to normalised log of per capita GDP.

which $e_k = (0.25, 0.50)$ would include provinces whose per capita GDP ranged between one quarter and half the national average. The different entries in the matrices indicate the probability of a given province transiting out from its initial state to other states during the period or periods considered. The boundaries between grid cells are chosen so that all province-year observations are divided approximately equally among the cells, each cell corresponding to approximately one fifth of the distribution of the selected variable across provinces and time. Interpretation is straightforward: observations in the first state refer to the poorest provinces. This way of constructing is common practice (see, for instance Redding, 2002; Lamo, 2000). Some other contributors have considered different criteria such as selecting the limits between states arbitrarily—although reasonably (Kremer et al., 2001; Quah, 1993a).[9]

To compute each transition matrix we count the number of transitions out of and into each cell, i.e., for each p_{kl} cell:

$$p_{kl} = \frac{1}{T-1} \sum_{t=1}^{T-1} \frac{n_{kl}^t}{n_k^t} \tag{2}$$

where T is the number of years or periods, n_{kl}^t is the number of provinces moving during one period from class k to class l, and n_k^t is the total number of provinces starting the period in class k.

By operating with the information offered by the transition probability matrix we can also characterise the hypothetical long-term ergodic or stationary distribution. The variety of resulting scenarios might be remarkable, including distributions with the probability mass concentrated mainly in the central classes (indicative of convergence to the mean if these central states contained the unity), polarised distributions ("twin peaks") indicating that the poorest and richest are becoming increasingly more distant from each other, or one with the probability mass distributed in the extreme classes (tails) of the distribution. Therefore, the ergodic distributions make it possible to determine the predominating long-run tendency for provincial per capita GDP in China.

3.1.2. Transition path analysis and mobility indices

We can also evaluate the speed with which the ergodic distribution, or steady-state, is approached by means of the concept of asymptotic half-life of the chain, $H - L$, which is how long it takes to cover half the distance from the stationary distribution. Following Shorrocks

[9] Alternatively, it is possible to dodge the discretisation problem by considering stochastic kernels (Quah, 1996c), although these present some difficulties for estimating the ergodic, or stationary distribution.

(1978), the half-life is defined as:

$$H - L = -\frac{\ln 2}{\ln |\lambda_2|} \tag{3}$$

where $|\lambda_2|$ is the second largest eigenvalue (after 1) of the transition probability matrix. It ranges between infinity (when the second eigenvalue is equal to 1 and the stationary distribution does not exist) and 0 (when $\lambda_2 = 0$ and the system has already reached its stationary equilibrium (Magrini, 1999)).

We also consider the mobility indices proposed by the literature on economic inequality (Shorrocks, 1978; Geweke et al., 1986). As suggested by Quah (1996a), analogously to the measures of income inequality designed to collapse the information contained in an entire distribution in a single scalar, a mobility index summarises the mobility information in a transition probability matrix into one number. We consider the proposals by Shorrocks (1978) and Geweke et al. (1986), summarised by Quah (1996a). In their proposals, the mobility index (μ_1) evaluates the trace of the transition probability matrix, disclosing information on the relative magnitude of diagonal and off-diagonal terms. It is identical to the inverse of the harmonic mean of expected durations of remaining in a certain state and, following Quah (1996a), its particular expression is:

$$\mu_1(P^*) = \frac{K - \mathrm{tr}(P^*)}{K - 1} = \frac{\sum_j (1 - p_{jj})}{K - 1} \tag{4}$$

where K is the number of classes, and p_{jj} is the j-diagonal entry of matrix P^*, which represents the probability of remaining in state j. Large values of μ_1 indicate less persistence (or more mobility) in P^*.

3.1.3. The evolution of the external shape of the distributions

It is also important to provide information on both the initial and final distributions for the variable of interest, in order to gain further insights into how distributions have evolved. Therefore, we provide four sets of additional results for all the indicators, namely, transition probability matrices, ergodic distributions, initial distributions, and final distributions.

However, in their present form, the three sets of distributions share a common disadvantage, namely, they are discrete and probability is spread out across only one set of states. Although we have provided reasons why such a disadvantage may not be as restrictive as some authors suggest, we try to be as informative as possible by also providing the continuous counterpart to this discrete estimation, namely, the non-parametric estimation of density functions via kernel smoothing. This is the first step in Quah's model of distribution dynamics, and it provides remarkable insights about the convergence process. If the probability mass became tighter, it would indicate convergence, whereas if it became flatter, it would be indicative of divergence. As can be easily inferred, multiple scenarios may result. Therefore, this literature usually considers a kernel estimator for each indicator:

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\Big(\frac{\|x - X_i\|_x}{h}\Big) \tag{5}$$

where x is the point of evaluation, X is the indicator of interest, N is the number of observations (Chinese provinces in our case), h is the bandwidth, $\| \cdot \|_x$ is a distance metric on the space of X, and $K(x)$ is a kernel function (see Härdle and Linton, 1994). These kernel estimators are generally required to hold that:

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \int_{\mathbb{R}} xK(x)dx = 0, \quad \sigma_K^2 = \int_{\mathbb{R}} x^2 K(x)dx < \infty \tag{6}$$

There are several choices for $K(x)$, which may be defined in terms of univariate and unimodal probability density functions. For the sake of simplicity, we consider a Gaussian kernel:

$$K(x) = (1/\sqrt{2\pi})e^{-\frac{1}{2}x^2} \tag{7}$$

As indicated by Loader (1999), the kernel density estimate has been widely studied, and there are several monographs dealing with the issue in a competent manner (see, for instance Silverman, 1986; Wand and Jones, 1995; Li and Racine, 2007). However, because it is based on a local constant approximation, it suffers from problems such as trimming of peaks, or problems in the tails, since increasing bandwidths for data sparsity can lead to severe bias. These problems can be alleviated using the local likelihood variant of density estimation, as indicated by Loader (1996) or Hjort and Jones (1996).[10] In addition, these methods performed much better in our specific setting, in which both weighted densities and ergodic densities have to be estimated.

Following Loader (1996), for density estimation, the appropriate local likelihood criterion is:

$$\sum_{i=1}^{N} \omega_i(x) \ln(f(X_i)) - N \int W\Big(\frac{u - x}{h}\Big) f(u)du \tag{8}$$

where W indicates that we are considering a locally weighted least squares criterion for each

[10] For instance, Loader (1996) have investigated these issues more thoroughly, comparing the relative efficiencies of kernel and local log-polynomial methods.

fitting point x, $\omega_i(x)$ are the localisation weights, and the log-link is used, i.e., $\ln(f(x))$ is modelled by local polynomials. The term on the right is the added penalty term. See Loader (1999) for details.

The continuous version of the ergodic distributions is more difficult to estimate. In this case, related literature is scarce. Some studies provide estimations for ergodic densities (see Johnson, 2000, 2005). However, no studies provide, simultaneously, results for ergodic distributions yielded by transition probability matrices *and* ergodic densities. In order to obtain a fully compatible view between the results of the transition probability matrices and their continuous counterpart, we generated ergodic densities considering the information in the (discretised) ergodic distributions (1×20). Specifically, we generated normal distributions for each of the twenty states which probability is spread out over, with a number of observations proportional to each state's share of ergodic probability. This generates a pseudo-histogram in which we do not have bars, but normal distributions. Then we proceed in exactly the same way as when smoothing both initial and final distributions, i.e. by considering local likelihood density estimation methods to smooth the observations in each of these twenty states. This algorithm yields ergodic densities which are fully consistent with the ergodic distributions computed from transition probability matrices. The continuous state approach naturally complements the view provided by discrete ergodic distributions, which tend to summarise too much information in a few states. Although the information provided by ergodic densities is essentially the same, we remove the arbitrariness implied by selecting a grid.

3.2. Weighted analysis

According to the methods described above, transitions are estimated by counting the number of provinces moving from one class to another. However, as indicated in the introduction, using provinces as units of analysis would not be useful if we were concerned with human welfare because different provinces have different population sizes. Therefore, the unweighted analysis does not help to answer questions such as "How many people in China live in poverty?" or "How have poverty rates changed over the last few decades?" Therefore, it is also relevant to estimate *weighted* transition probability matrices, for which different weighting schemes are feasible, and are not limited to just population. The underlying idea is that the impact on Chinese per capita GDP will be greater if a larger country transits out than if a small province does so. Therefore, we count the transitions of provinces, but in this case each province is represented by its entire share of Chinese population (in the case of population-weighted transition probability matrices), so that the unit of observation is now a *person* instead of a province, i.e. we count the number of people moving between states. This issue is often ignored, although exceptions do exist, such as Kremer et al. (2001) or Jones (1997).

In the case of the weighted analysis, Equation (2) has to be slightly modified to take into account that the number of people involved in a given transition (if we weighted by population) depends on the particular province that is moving from one class to another. The modified expression corresponding to each cell in the weighted transition probability matrix follows:

$$p_{kl}^{\omega} = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{i=1}^{n_{kl}} \frac{W_{ikl}^{t}}{W_{ik}^{t}} \tag{9}$$

where W_{ikl}^{t} is the population (or GDP, depending on the weighting scheme) corresponding to province i, which is moving from class k to class l in period t, and W_{ik}^{t} is the population (or GDP) corresponding to province i, which starts the period in class k.

Weighting densities also requires slight modifications. As indicated previously, few studies have considered this, despite its potential relevance in some specific contexts. Following Goerlich (2003) or Tortosa-Ausina et al. (2005), expression (5) would be slightly modified, thus becoming:

$$\hat{f}_{\omega}(x) = \frac{1}{h} \sum_{i=1}^{N} \omega_i K\Big(\frac{\|x - X_i\|_x}{h}\Big) \tag{10}$$

where ω_i is the share of either China GDP or China population (depending on the type of weighting we consider) corresponding to province i. In our case, in which we use local likelihood density estimation, the weights can be entered directly into Equation (8).

3.3. Conditioning on neighbour-relative information

Furthermore, the techniques employed enable us to analyse the importance of spatial factors in explaining regional convergence or divergence. In particular, following Quah (1996c), we can analyse the role played by surrounding regions in explaining the dynamics of regional distribution of per capita GDP (conditional distribution dynamics). The specific hypothesis to be tested is whether Chinese provinces might be converging with their neighbours, i.e. the provinces around them. Quah (1996c, 1997a) provides reasons as to why such a convergence pattern could exist, along with methods to evaluate conditional convergence with our instruments. As indicated by Quah (1996c), and most of the literature on spatial economics, geographical location and spatial interactions between regions *matter*. Increasing returns to scale, together with enhancing market access, and probably a combination of labour migration across regions and vertical linkages between industries explain the cumulative process of regional growth, which endogenously turns into a polarisation of the spatial distribution of per capita GDP.[11] Additionally, the existence of localisation and urbanisation economies, or knowledge spillovers, reinforces the capacity of areas surrounding more highly developed regions to grow. Not only geographical location but also proximity matters for growth.[12] As can be seen in Table 1, the Chinese provinces with higher growth rates of per capita GDP between 1978 and 2005 (over 9%) are all next to each other and stretch almost continuously from Liaoning in the northeast to Guangdong in the southeast and to Sichuan in the west. Liaoning, Hebei and Shandong are provinces located around Beijing and Tianjin; Zhejiang is located between Shanghai and Fujian; while Guangdong and Fujian are located next to Hong Kong, Taiwan and Shanghai. This set of coastal provinces also stretches westward through Anhui, Henan, Hubei and Sichuan. Proximity, or even neighbourhood, could become a key factor in its growth.

In order to elucidate the existence and magnitude of these spatial spillovers, we conducted an analysis which hinges on the comparison of two normalised per capita GDP series: (i) *state-relative* GDP per capita, where we normalise each province's per capita GDP by the per capita GDP in China (which are the data used to conduct the analysis in the previous subsections); and (ii) *neighbour-relative* per capita GDP, where we normalise each province's per capita GDP by the average per capita GDP of the surrounding, physically contiguous provinces, excluding the province itself. As indicated by Quah (1996c), it is convenient to consider these two relative GDP per capita series as the parts unexplained by nation-state factors and physical-location factors, respectively. Formally, the expression corresponding to the neighbour-relative per capita GDP series is

$$x_i^{NR} = \frac{\ln y_i}{\ln \frac{1}{NE-1}(\sum_{j \in NE \setminus i} y_j)} \tag{11}$$

where NE is the number of (physically-contiguous) neighbours each i province has, and nr is the super-index indicating that we are referring to the neighbour-relative per capita GDP series. By default, the expression of x_i introduced on page 9 referred to the state-relative per capita GDP series. Therefore, in the case of Beijing (for instance), the sum in the denominator

[11]From the pioneering work of Marshall (1890) to the more recent developments of the "new economic geography" Krugman (1991, 1993), economists have emphasized a combination of these forces to explain the strong localization of economic activity. A recent study by Moreno-Monroy (2010) highlights the relevance of New Economic Geography models to analyze wage disparities across Chinese cities.

[12]Several studies, using different methodologies, have examined convergence using spatial econometrics' approaches. See, for instance, Montouri (1999), Bosker (2009), or Checherita (2009), among others.

would include the GDP of Tianjin and Hebei, but it would exclude Beijing itself, in order not to bias the average of the neighbours' GDP.

The same analyses as those presented in subsections 3.1.1–3.1.3 to this new series of neighbour-relative per capita GDP, focusing on the comparison with the state-relative per capita GDP series. Interpretations are also straightforward: the closer the values of the neighbour-relative series are to unity, the lower the inequalities among neighbour provinces will be and, therefore, the higher the magnitude of the spillover effects will also be. From this, it can easily be inferred that comparing these two series would be equivalent to analysing *unconditional* versus *conditional* convergence.

4. Results

Results concerning both transition probability matrices and ergodic distributions are reported in tables 2 to 5. They constitute a total of 12 panels in which different sorts of related information are reported. The four different tables present results for the different sub-periods considered, i.e. the first panel in each table provides results for the entire period 1952–2005, whereas the second and third panels provide results for the periods 1952–1978 and 1978–2005, respectively. In addition, the last three rows in each panel display information on the initial, final and ergodic distribution of the variable under analysis. Furthermore, apart from the analysis of the unweighted distribution of per capita GDP, the aforementioned additional conditioning schemes are also reported (GDP-weighted, population-weighted and physically-contiguous conditioned). The notes at the end of each table offer a brief explanation on how to interpret different pieces of information reported therein.

4.1. Unweighted analysis

Table 2 reports on unweighted transition probability matrices for all the periods considered. In each of the matrices contained in the table, the upper limits, or cut-off points, have been set to the same values in order to facilitate comparisons. The criterion to be used to specify the grid is the one usually found in the literature (see, for instance Lamo, 2000), i.e. considering all observations for the entire period 1952–2005 (28 observations per year and 54 years, which totals 1,512 observations); these are divided into five equally-sized intervals, yielding an (almost) uniform distribution over the observed sample. The numbers in brackets on the left are the numbers of observations beginning from a particular state. They total 1,372 instead of

1,512 because five-year transitions are considered, and thus the last five years are excluded, i.e., $1,372 = 1,512 - 28$ provinces $\times$ 5 years. The limits of the grid (or cut-off points that demarcate the intervals) are displayed in the first row of each panel. Their interpretations are straightforward: the upper limit for the first state is 0.915, indicating that approximately one fifth of the total number of observations lie below that threshold—i.e. below 91.5% of the average. At the other extreme, the upper-state has observations lying above 1.053 (105.3%) of the average. Note that the average is unity, since our data have been normalised by the mean: the closer a value is to unity, the closer it is to the average for its particular year.

The contents of each matrix in Table 2 have some commonalities with the concept of β-convergence, since they provide information about intra-distribution mobility, or churning (Quah, 1996c). Each cell in the matrices must be interpreted as the probability of remaining in that particular state after five years (remember that we compute 5-year transitions). For instance, the upper-left entry of the matrix in Table 2a would indicate that the probability of the observations in the lowest relative per capita GDP state (below 0.915) remaining in that state was 83%, whereas the remainder moved up to richer states. Persistence was even higher at the other end of the distribution, as revealed by the lower right cell in the matrix, which indicates that 91% of the observations in the richest state remain in the same class after five years, on average. The other values in the main diagonal show a higher degree of mobility. For instance, entry p_{22} would indicate that, after 5 years, only 66% of observations remain in the same state of relative wealth, whereas 14% move down to lower per capita GDP states and the remaining 20% move up to higher per capita GDP states. In general, values in the main diagonal closer to 1 indicate more persistence, whereas values closer to zero indicate higher mobility.

In the matrices examined in Table 2, values on the main diagonal average 0.732, 0.684 and 0.812 for the periods 1952–2005, 1952–1978 and 1978-2005, respectively. This information is rich, but it would be richer still if additional ways of evaluating persistence/mobility such as the mobility indices presented in Equation (4) were considered. The results for these indices are shown in Table 6 and, in general, they corroborate what the averages for the diagonal entries revealed, i.e. the sub-period 1952–1978 shows much higher mobility than that of 1978–2005 (0.713 vs. 0.573) and, for the whole period, total mobility lies somewhere in-between (0.640). However, as will be shown below, it is not only the intensity of mobility that differs across periods but, more importantly, its sign. Mobility leads to probability mass polarising at both lower and upper states in the first sub-period, whereas probability mass concentrates in states 4 and 5 under 1978–2005 trends.

The last three rows in each table support this claim. They contain information on the initial (1952), final (2005) and ergodic (steady-state) distributions for the three periods considered. Table 2a indicates that the initial and final distributions differ, but not strongly. What is more revealing is that, under current trends, although the ergodic distribution will not be uniform, convergence to a given interval will be modest, with state 4 absorbing 27% of the probability mass. However, we must bear in mind that in our particular setting under current trends may be a misleading statement, since trends differed remarkably before and after the reform. The ergodic distributions in Table 2b (pre-reform) and Table 2c (post-reform) differ markedly, not only compared to the ergodic distribution in Table 2a but more notably with respect to each other. For the pre-reform period (1952–1978) the ergodic distribution polarises towards both the left and right tails of the distribution, whereas for the post-reform period (1978–2005) the distribution becomes left skewed (negative skew). This would imply that the effects of the reform were positive for convergence among provinces and they are likely to continue over time, indicating that, under 1978–2005 trends, the two states of highest relative per capita GDP will contain 79% of the provinces.

It is important not only to compute the values of the steady-state distribution but also to analyse the speed at which it is approached. As indicated in previous sections, this can be evaluated via the concept of the asymptotic half-life of the chain, i.e. how long it takes to cover half the distance from the ergodic distribution (Magrini, 1999). Therefore, computing Equation (3) leads to the results in Table 7. The steady-state reached considering only 1978–2005 information (Table 2c) is not only more favourable than that obtained using 1952–1978 information (Table 2b), but it will take much less to reach it, in fact, virtually one third of the time. This result emerges in spite of the higher intra-distribution mobility found for the pre-reform period, as revealed by Table 6. Therefore, the future predicted using only 1978–2005 information is far more promising, and it will take less time to reach it, probably because of the entry p_{55} in Table 2c, which indicates that, on average, 99% of the richest countries remain rich after five years.

Bulli (2001), Johnson (2000, 2005) and many others have pointed out that it may be problematic to consider a discrete approach in which probability is split in some states whose limits are somehow arbitrary. Sakamoto and Islam (2008) partly circumvent this criticism and add some additional robustness to their analysis by considering different grids (5 and 7 grids), the results being similar for the different choices. We believe it is more interesting to consider a *fully* continuous counterpart to the initial and final distributions in Table 2, but continuous counterparts to the steady-state distributions reported are also taken into account.

Figure 1a displays the continuous counterparts to the discrete initial (1952) and final (2005) distributions in the tables corresponding to the unweighted analysis in Table 2a. Although the densities basically corroborate the results of the discrete analysis, we can perceive more clearly that, although moderate convergence has taken place (the 2005 density is slightly tighter), we can also see that by 1978 the distribution became slightly bimodal (in the vicinity of 1.2). Figure 5a displays the continuous counterpart to the steady-state distributions in tables 2b and 2c. Although results are generally corroborated (see, for instance, the emerging bi-modality under 1952–1978 trends), some subtleties that the 5-grid analysis could not show are perceived. These are basically related to the type of multi-modality that will prevail (much tighter under 1978–2005 trends). Taking into account the pre-reform information, the ergodic density (solid line in Figure 5a) indicates that some very rich provinces (upper tail of the distribution) will coexist with some others (fewer) that are very poor (lower tail of the distribution). This extreme behaviour will fade away if only post-reform information (dashed line in Figure 5a) is considered, although we can still distinguish two bigger modes, known as twin peaks, to use Quah's (1996d) term for them.

Therefore, the results obtained by Sakamoto and Islam (2008) are generally corroborated, but we have complemented them in several ways. Although their way of normalising differs, they use a slightly shorter time period (1952–2003) and they add some robustness to the analysis by considering a different number of grids, we find the same broad results, i.e. divergence before the reform and strong convergence afterwards. However, the mobility indices, transition path analysis and continuous approach to the steady-state distributions all enrich the analysis.

4.2. Weighted analysis

The analysis performed in the previous section is relevant, but it might be judged as being partly biased because the same importance is attached to all provinces. This is especially the case if we are concerned about human welfare, because the different provinces have different population sizes. As indicated in the introduction, the unweighted analysis could be highly misleading if we drew national borders differently, as this would affect the shape of the densities. It may be more natural to attach a weight to the observations, where the weights reflect the contribution of each observation in the sample. As indicated in previous sections, we will consider different weighting schemes, i.e. population (N) and economic size (GDP). In the case of countries, both variables are very unevenly distributed. This is especially blatant in the case of population, for which India and China, two of the poorest countries in terms of per capita GDP, account for more than one third of the world's total population, whereas some of the richest countries, such as Iceland or Luxembourg, account for only 0.01% of the world population (Goerlich, 2003). In our particular case, it does not seem fair either to treat all Chinese provinces equally in the estimation. As can be seen on the right-hand side of Table 1, there is a significant dispersion in the population of the different provinces. More important still, although the growth rates of provincial population have gone in the direction of reducing the dispersion of the population among provinces, their dispersion has been modest (higher in the post-reform than in the pre-reform—coefficients of variation were 0.35 and 0.29 respectively), and consequently they only reduced the dispersion of provincial population moderately throughout the considered period (coefficients of variation were 0.77, 0.66 and 0.63 in 1952, 1978 and 2005 respectively). These differences in the distribution of provincial population have relevant implications when we are looking at per capita distribution of GDP from an individual or personal welfare perspective instead of from a provincial point of view. For example, by 2005, as indicated in Table 1, the population of Sichuan was 110,010,000 (larger than any European country), whereas that of Qinghai was 5,430,000. Therefore, the welfare implications of Sichuan converging with the rest of the provinces are not the same as if Qinghai converged, because of the number of people involved.

Results are shown in the GDP-weighted and population-weighted panels in Tables 3 and 4 respectively. The mobility indices, transition path analysis and continuous analysis are reported in the same tables and figures as those corresponding to the unweighted analysis. Both tables 3 and 4 offer new perspectives on the evaluation of convergence. Although the unweighted analysis did not clearly predict convergence or divergence (in accordance with the ergodic distribution), in terms of probability mass strongly concentrated in a given state, the weighted analyses yielded different results. Under the population-weighted scenario (Table 4), according to which we evaluate transitions of people moving across classes, considering the entire period (1952–2005), the steady-state distribution has more than half of the probability mass (65%) in the two upper states. This indicates that a large part of the *population* will escape from poverty in the long run. However, similarly to what we obtained for the unweighted analysis, the tendencies differ remarkably between the pre- and post-reform sub-periods, and it is the effect of the second sub-period which drives the convergence pattern most. As shown in Table 4b, although the predicted pattern using the 1952–1978 information was convergence, most of the population was being driven deep down into poverty, since the probability mass is overwhelmingly accumulated (73%) in the two lowest relative per capita GDP states. This result is shared when weighting by economic size (Table 3b), i.e. the largest provinces in terms of GDP were becoming relatively poorer. In contrast, the post-reform period shows opposite patterns. As Table 4c reveals, in the hypothetical long run (i.e. under 1978–2005 trends) all the population (100%) will reach the two highest per capita GDP states.[13] This occurs because of a lower right entry in the matrix in Table 4c, where not a single observation leaves this class to reach poorer classes. The same result is found when weighting by GDP (100% probability in the two wealthiest states), thus also indicating that large provinces in terms of GDP are also the ones that are escaping from poverty.

Although the general tendency when weighting by GDP or by population is similar for both sub-periods, there are some differences when evaluating the implied mobility in each matrix (Table 6) or the half-life time of convergence (Table 7). Regarding the former, results are very similar to the unweighted case. In the case of the half-life of convergence (Table 7), when weighting by both GDP or by population convergence is faster in the post-reform sub-period than in the pre-reform period. For the pre-reform period, convergence is a bit faster when weighting by population than when either weighting by GDP or for the unweighted case.

Finally, the continuous analysis in figures 1 and 5 further corroborates how relevant it is to perform the weighted analysis. Figure 2 and Figure 3 report information already displayed in Figure 1 in a different way so as to facilitate visualisation of the patterns. As shown in Figure 1b and Figure 1c and, particularly, in figures 2 and 3, the evolution of the shape of the *weighted* densities differs compared to the one shown in Figure 1a, especially in 2005. Weighting by GDP makes the density shift rightwards (Figure 2c), although some additional bumps also emerge, thus indicating that some important shares of GDP will remain in poor provinces. The result of weighting by population is more striking, since it indicates that by 2005 a large share of the population was reaching higher GDP levels, but a larger share was also trailing behind, as indicated by a marked bimodality. This is what Quah (1996d) refers to as "twin peaks". However, in the steady-state (figures 5b and 5c), and confirming what we found via the discrete analysis of the transition matrices, much of this bimodality will fade away, and the distributions will be basically skewed when using 1978–2005 information, which contrasts sharply with the bimodality found for the unweighted case (Figure 5a).

[13] Although there is a 1.00 in the lower right entry of the matrix, this only because of rounding.

In synthesis, uneven distribution of per capita GDP across Chinese provinces becomes less strong when weighted by GDP or population, that is, in terms of average personal welfare, and when using the post-reform information the implicit steady-state distributions will be skewed rightwards and reflect an improvement in the symptoms of convergence. Nevertheless, some peaks persist on the upper tails of the distribution and it will also take a long time to reach the steady state. These stylised facts, together with the variability and changes in the trend of the variance of the distribution mentioned above, are quite consistent with the timing of the reforms, the unbalanced regional implications of these reforms and with the changes in emphasis in the main policy objectives during the period.

In the first phase of the economic reforms, but before economic liberalisation, the strategy was concentrated on the rural areas. The commune system was removed in favour of the Household Responsibility System, where workers were allowed to operate on their own, although with some restrictions.[14] After decollectivisation, the Chinese government promoted economic policies addressed at diversifying agriculture, especially by enhancing the rural industries and the township and village enterprises (TVEs). In fact, the promotion of TVEs was the most important way of transferring excess rural labour into industrial production, given the strong restrictions on interprovincial migration (Fujita and Hu, 2001). As a result, rural industrial output increased sharply in this period. However, the effectiveness of TVEs also raised some doubts owing to the fact that they often operated according to non-economic criteria in the early years of the reforms. Some regions improved in this phase, especially those oriented towards industry, but the income differentials persisted among provinces because of the barriers that existed across provinces (Rozelle, 1994).

In the 1980s, the second phase of the reforms was characterised by the gradual opening up of the Chinese economy, the increased presence of the non-state sector (collective and private sectors) and a fiscal reform that endowed the provinces with more fiscal power (Wei, 1996). At first the open-door policy was especially favourable for the coastal areas (open cities and Special Economic Zones —SEZs—). Thus the geographic and economic policy factors allowed trade and FDI to become concentrated in the coastal areas.[15] At the same time, this period was distinguished by a major liberalisation and decentralisation of the economy compared with the previous stage. For example, price liberalisation accelerated the entry of non-state enterprises, and the profit-oriented incentive schemes in state industry led to a rapid increase in industrial

[14] Further details on rural reforms and agricultural growth can be found in Lin (1992).

[15] Although FDI was allowed in 1979, the effects on output are more significant in the 1980s and 1990s.

output and gains in productivity by the mid-1980s. As a result, the non-state sector gradually became more important in the economic development of China.[16] Although the interprovincial mobility of workers was still costly, there was an increase in migrational movements from rural areas to urban and coastal areas. On the other hand, the fiscal decentralisation of 1980 played a key role in improving the autonomy of local governments, but generated a significant budget deficit. Consequently, the fiscal system was reformed in 1985. The immediate effect of this reform was a reduction in the central government's ability to redistribute revenues among regions which, together with the economic developments that favoured coastal provinces, increased symptoms of divergence and led to a new fiscal reform in 1994. The main feature of this reform was the separation of the national tax service from the local tax service, with an additional mixed category that was shared between central and local government, without negotiation and applied to all the provinces with the aim of reducing the income gap across provinces.

In 1995, the Chinese government recognised that:

> "Since the adoption of reforms and open-door policies, we have encouraged some regions to develop faster and get richer, and we have advocated that the richer should act as a model for and help the poor. Each region has had immense economic development and the people's standard of living has had great improvement. But for some reasons, regional economic inequalities have widened somewhat" (People's Daily Overseas Edition, October 5, 1995, p. 4.)

Thus, the strategy was changing in favour of promoting a more evenly balanced regional development, in an attempt to reduce the tendencies towards uneven regional development. This strategy became obvious in the Eighth Five-Year Plan and, more especially, in the Ninth Five-Year Plan (1996–2000). The Chinese government launched a strategy to promote the development of the central and western regions that relied, at least partly, on the spillovers generated by the more developed coastal provinces.

4.3. Conditioning: spatial analysis

The transition probability matrices in Tables 5 show neighbour-relative counterparts to the transition probability analyses carried out for weighted and unweighted state-relative series

[16]Further details about the effects of the reform on the performance of the Chinese State Enterprises can be found in (Li, 1997).

(tables 2, 3 and 4). Likewise, the neighbour-relative analysis indicates that conclusions differ notably prior to and after the reform, i.e. they hinge critically on whether we base the future projections (ergodic distributions) on 1952–1978 or 1978–2005 information.

If the entire period 1952–2005 (Table 5a) is considered, the diagonal entries average 0.72, which is only slightly lower than the 0.73 corresponding to the state-relative series (Table 2a). The mobility indices in Table 6 corroborate this finding, since $\mu_1 = 0.640$ in the case of the state-relative series, and $\mu_1 = 0.676$ when conditioning by neighbours' information. Under these trends, the (slightly) higher mobility would lead to an apparently multi-modal ergodic distribution, but it is difficult to discern tendencies. The analysis for the different sub-periods shows, once again, different patterns. The ergodic distribution corresponding to the 1952–1978 trends indicates that multi-modality will prevail in the future (Table 5b), with probability mass thinning in the middle state. Multi-modality vanishes if we focus on 1978–2005 trends (Table 5c) and the pattern indicates strong convergence, with the probability mass concentrated in the middle class. In both cases, but especially for 1978–2005 trends, the ergodic distribution differs remarkably when compared to state-relative information (tables 2c, 3c and 4c). However, as indicated by the asymptotic half-life of convergence in Table 7, the ergodic distribution will be achieved much faster when conditioning by neighbouring information, i.e. *conditional* convergence will be faster than *unconditional* convergence.

Figure 1d and Figure 4 also show the impact of conditioning on neighbouring-province information. Although the information contained in Figure 4 was already reported in Figure 1a and Figure 1d, the way it is presented allows a clearer understanding of the effect of spatial conditioning. Both Figure 1d and Figure 4 show tighter distributions for neighbour-relative compared to state-relative per capita GDP series, for all years in the figures, especially in 1952 and 2005. This would indicate that each province's per capita GDP is closer to the average of its surrounding provinces than to the national average, thereby suggesting that spatial spillovers *do* matter. Yet some subtleties also exist. For instance, the uni-modal state-relative distribution of per capita GDP turns into a tighter but multi-modal distribution when conditioning by neighbouring information in 1978 (Figure 4b). This implies that, although the general tendency is towards convergence *within* spatial clusters, in certain years there are some provinces which outperform their neighbours, thus yielding a bimodal distribution. Therefore, spatial spillovers are relevant but *not* for everyone.

Figure 5d reports continuous counterparts for the ergodic distributions in tables 5b (solid line) and 5c (dashed line). The solid line in Figure 5d, corresponding to 1952–1978 trends, shows a big mode in the vicinity of 1, but also another one, much less marked, at the lower tail of the distribution. The dashed line indicates that the ergodic distribution that would prevail under 1978–2005 trends would be much tighter in the vicinity of 1, thus indicating that the members of spatial clusters' will be quite similar in terms of per capita GDP (conditional convergence). However, since several observations will lie in the lower tail of the distribution, in the hypothetical long-run scenario some provinces will continue to be outperformed by their neighbours, i.e. although there will still be inequalities that cannot be explained by physical-location factors, they will affect provinces differently.

These results, and especially the tendency towards the stratification of provinces in different clubs, are of no minor concern to authorities, and reveal that there is still some room for policies promoting convergence in per capita GDP among Chinese provinces, because the natural tendency towards spatial agglomeration seems to be persistent. Thus, together with the explicit regional policies and the use of other central government policies to re-balance regional development (central investment projects, endowment of infrastructures, credit policy, etc.), other measures are also needed to balance the tendency towards the localisation of economic activity induced by market forces. Improvements in the accessibility and the role of market mechanisms in the interior are needed, but increasing the role assigned to official interprovincial migrations is probably necessary too.

5. Conclusions

Nobody doubts that the acceleration of the economic reforms initiated at the end of the 1970s has encouraged economic growth over the last four decades. The open-door policy, with a strong drive towards industrialisation focused on foreign investment (especially in the coastal regions), along with a series of economic reforms oriented more towards the market probably explained this exceptional performance during the 1980s and 1990s. In 1995, however, the Chinese government recognised that the income gap between western and central regions and the coastal areas was increasing, thus making it necessary to implement pro-active policies to reduce these inequalities. The stimulus package that was carried out was focused mainly on the development of inland provinces through the promotion of investment as a way to reduce those imbalances.

In this scenario, a plethora of research studies have examined not only the aggregate growth of the country but also other related questions, such as whether differences in per capita GDP across provinces exist, along with the evolution of disparities over time. This ample body of literature analysing convergence across Chinese provinces continuues to grow, and includes such relevant topics as those examined by country and regional convergence studies. Some papers have analysed provincial convergence following the early proposals of Barro and Sala-i-Martin (1992), i.e. by examining β- and σ-convergence, together with some of the posterior refinements of these techniques. Some others (fewer) have leaned towards the distribution dynamics' model initially proposed by Quah (1993a,b). Our article follows this second line of research. Recent contributions such as Bhalla et al. (2003) or Sakamoto and Islam (2008), have applied Quah's basic proposals to examine provincial convergence in per capita income. Our paper complements their methods and findings and extends them in several directions.

Similarly to Sakamoto and Islam (2008), the ergodic distributions obtained using either pre-reform or post-reform information are quite different, a positively skewed (and slightly bi-modal) distribution being produced for 1952–1978 and a negatively skewed (and bi-modal) one for the period 1978–2005. Therefore, it would be corroborated that the post-reform policies have led most provinces to escape from relatively low per capita income levels. However, this analysis has some limitations such as the need to specify a *discrete* grid with a limited number of states. Few contributions try to fix this by considering a *continuous* state space approach (Johnson, 2000). We follow Johnson's (2000) approach to provide continuous counterparts to the ergodic distributions yielded by transition probability matrices, which offered more detailed results. Under both pre- and post-reform information, the hypothetical long-run scenario shows multi-modality. For the 1952–1978 information, the distances separating the biggest modes are quite important, with predominance of one large mode comprising most of the provinces with incomes close to the average, and a small group of provinces that are becoming very rich. However, using 1978–2005 information these two modes become more balanced, with one of them above the national average and the other one in the vicinity of 1. We can also corroborate Sakamoto and Islam's claim that "the dynamics of the post-reform period do not yet seem to have settled into a stable pattern". In our case, although the analysis of the asymptotic half-life of convergence indicates that it will take much longer to reach the steady state under 1952–1978 trends, it will also take a long time to reach the steady-state under 1978–2005 trends. Under this scenario, although most provinces will escape from relative poverty, it will also take a long because of the complexity of intra-distribution dynamics.

We extend the analysis to control for some relevant characteristics of Chinese provinces. Specifically, although *unweighted* analyses of country/regional convergence are commonplace, *weighted* analyses are far less widely extended. However, in many circumstances and especially if we focus on human welfare (Sala-i-Martin, 2006), a weighted analysis might be more relevant than its unweighted counterpart. Several weighting schemes are possible, but because of their significance we considered the population and economic size (GDP) of each province. As strongly stressed throughout the article, since both population and GDP differences across Chinese provinces are outstanding, controlling for these differences might substantially alter the results—which in fact turned out to be the case.

For the entire period 1952–2005, we find that under the population-weighted scenario, the steady-state distribution has a substantial part of the probability mass (65%) in the two upper states, thereby indicating that much of the population will escape from poverty in the long run. As expected, the tendencies differ remarkably between the pre- and the post-reform periods, and it is the effect of the second sub-period (1978-2005) which, for the most part, drives the convergence pattern to top states. Specifically, for the pre-reform period, although the predicted pattern was convergence, most of the population was driven deep down into relative poverty, since probability mass is overwhelmingly accumulated (83%) in the lowest relative per capita income states. This result is shared when weighting by economic size. However, the continuous ergodic distributions also indicate that some provinces will still be much richer than the rest, as indicated by the existence of a slight bump well above unity. In contrast, the post-reform period, shows opposite patterns. In the hypothetical long run (under 1978-2005 trends) most of the population (97%) will reach the highest per capita income state. An analogous result is found when weighting by GDP (98% probability in the wealthiest state), thus indicating that large provinces in terms of GDP are also the ones escaping from poverty. Moreover, when weighting by GDP or by population, convergence is faster in the post-reform period. Thus, the marked bimodality yielded by the unweighted analysis turns into a tighter pattern of convergence when weighting by the population of each province or economic size, as suggested by Sala-i-Martin (2006).

Finally, our study also analysed whether spatial spillovers exist. Although a more thorough analysis would be welcome, the techniques we use can be easily adapted to provide some insights into the magnitude of these effects. This can be thought of as conditional convergence analysis, in which a province is compared only with its contiguous provinces, and therefore it could converge towards its neighbours' average (conditional, club or cluster convergence) instead of towards the national average (unconditional convergence). Compared to the unweighted, unconditional analysis, the long-run scenario will be multi-modal under pre-reform trends—in fact, we could even talk of club *divergence*. However, under post-reform trends, cluster convergence will be much stronger, with probability mass concentrating tightly around unity as indicated by the ergodic density; provinces will converge strongly with their neighbours.

According to our results it seems that all the reforms have enabled poorer regions to gradually converge with the richer ones in the post-reform period, while in the pre-reform period no convergence was found in terms of poor provinces catching up with the richest. However, more economic reforms are needed in this regard to guarantee balance and steady economic growth, thereby improving the standards of living of the whole population.

References

Alesina, A. and Perotti, R. (1993). Income distribution, political instability, and investment. Working Paper 4486, NBER.

Alesina, A. F. and Rodrik, D. (1994). Distributive politics and economic growth. The Quarterly Journal of Economics, 109:465–490.

Bai, C. E., Hsieh, C., and Qian, Y. (2006). The return to capital in China. Technical Report 12755, NBER.

Barro, R. J. and Sala-i-Martin, X. (1992). Convergence. Journal of Political Economy, 100(2):223–251.

Ben-David, D. (1994). Convergence clubs and diverging economies. Discussion Paper 922, Centre for Economic Policy Research, London.

Bhalla, A., Yao, S., and Zhang, Z. (2003). Regional economic performance in China. The Economics of Transition, 11(1):25–39.

Bosker, M. (2009). The spatial evolution of regional GDP disparities in the 'old' and the 'new' Europe. Papers in Regional Science, 88(1):3–27.

Bramall, C. (2000). Sources of Chinese Economic Growth, 1978–1996. Studies on Contemporary China. Oxford University Press.

Brun, J. F., Combes, J. L., and Renard, M. F. (2002). Are there spillover effects between coastal and noncoastal regions in China? China Economic Review, 13(2-3):161–169.

Bulli, S. (2001). Distribution dynamics and cross-country convergence: a new approach. Scottish Journal of Political Economy, 48:226–243.

Canova, F. and Marcet, A. (1995). The poor stay poor: non-convergence across countries and regions. Discussion Paper 1265, CEPR, London.

Caselli, F., Esquivel, G., and Lefort, F. (1996). Re-opening the convergence debate: a new look at cross-country growth empirics. Journal of Economic Growth, 1(3):363–389.

Cetorelli, N. (2002). Could Prometheus be bound again? A contribution to the convergence controversy. Journal of Economic Dynamics and Control, 27(1):29–50.

Chai, J. C. (1998). China: Transition to a Market Economy. Studies on Contemporary China. Oxford University Press.

Checherita, C. D. (2009). Variations on economic convergence: The case of the United States. Papers in Regional Science, 88(2):259–278.

Chen, J. and Fleisher, B. (1996). Regional income inequality and economic growth in China. Journal of Comparative Economics, 22(2):141–164.

Chow, G. (2006). Are Chinese official statistics reliable? CESifo Economic Studies, 52(2):396–414.

De la Fuente, Á. (1997). The empirics of growth and convergence: A selective review. Journal of Economic Dynamics and Control, 21(1):23–73.

Ding, L., Haynes, K., and Liu, Y. (2008). Telecommunications infrastructure and regional income convergence in China: panel data approaches. The Annals of Regional Science, 42(4):843–861.

Durlauf, S. N. and Quah, D. T. (1999). The new empirics of economic growth. In Handbook of Macroeconomics, volume 1A, pages 231–304. North Holland, Amsterdam.

Eaton, J. and Eckstein, Z. (1997). Cities and growth: Theory and evidence from France and Japan. Regional Science and Urban Economics, 27(4-5):443–474.

Fujita, M. and Hu, D. (2001). Regional disparity in China 1985–1994: The effects of globalization and economic liberalization. The Annals of Regional Science, 35(1):3–37.

Geweke, J., Marshall, R. C., and Zarkin, G. A. (1986). Mobility indices in continuous time Markov chains. Econometrica, 54(6):1407–1423.

Goerlich, F. J. (2003). Weighed, kernel density estimators and convergence. Empirical Economics, 28(2):335–351.

Gundlach, E. (1997). Regional convergence of output per worker in China: a neoclassical interpretation. Asian Economic Journal, 11(4):423–442.

Härdle, W. and Linton, O. (1994). Applied nonparametric methods. In Engle, R. and McFadden, D., editors, Handbook of Econometrics, volume 4. North Holland, Amsterdam.

Herrerias, M. J. and Orts, V. (2010). Is the export-led growth hypothesis enough to account for China's growth? China & World Economy, 18(4):34–51.

Hjort, N. L. and Jones, M. C. (1996). Locally parametric nonparametric density estimation. The Annals of Statistics, 24(4):1619–1647.

Ho, C. Y. and Li, D. (2008). Rising regional inequality in China: Policy regimes and structural changes. Papers in Regional Science, 87(2):245–259.

Holz, C. A. (2005). China governance project: The institutional arrangements for the production statistics. Statistics Working Papers 2005/1, OECD, Paris.

Islam, N. (2003). What have we learnt from the convergence debate? Journal of Economic Surveys, 17(3):309–362.

Jian, T., Sachs, J., and Warner, A. (1996). Trends in regional inequality in China. China Economic Review, 7(1):1–21.

Johnson, P. A. (2000). A nonparametric analysis of income convergence across the US states. Economics Letters, 69:219–223.

Johnson, P. A. (2005). A continuous state space approach to "Convergence by parts". Economics Letters, 86:317–321.

Jones, C. I. (1997). On the evolution of the world income distribution. Journal of Economic Perspectives, 11(3):19–36.

Kremer, M., Onatski, A., and Stock, J. (2001). Searching for prosperity. Carnegie-Rochester Conference Series on Public Policy, 55:275–303.

Krugman, P. R. (1991). Increasing returns and economic geography. Journal of Political Economy, 99(3):483.

Krugman, P. R. (1993). Geography and Trade. MIT Press.

Lamo, A. (2000). On convergence empirics: Some evidence for Spanish regions. Investigaciones Económicas, 24:681–707.

Lardy, N. R. (1992). Foreign Trade and Economic Reform in China, 1978–1990. Cambridge University Press.

Li, Q. and Racine, J. S. (2007). Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton and Oxford.

Li, W. (1997). The impact of economic reform on the performance of Chinese state enterprises, 1980–1989. Journal of Political Economy, 105(5):1080–1106.

Lin, J. Y. (1992). Rural reforms and agricultural growth in China. The American Economic Review, 82(1):34–51.

Loader, C. R. (1996). Local likelihood density estimation. The Annals of Statistics, 24(4):1602–1618.

Loader, C. R. (1999). Local Regression and Likelihood. Springer Verlag, New York.

Maasoumi, E. and Wang, L. (2008). Economic reform, growth and convergence in China. Econometrics Journal, 11(1):128–154.

Magrini, S. (1999). The evolution of income disparities among the regions of the European Union. Regional Science and Urban Economics, 29(2):257–281.

Marshall, A. (1890). Principles of Economics. Macmillan and Co.

Montouri, S. J. R. B. D. (1999). US regional income convergence: a spatial econometric perspective. Regional Studies, 33(2):143–156.

Moreno-Monroy, A. I. (2010). Market access and the heterogeneous effect of shocks on wages: Evidence from Chinese cities. Papers in Regional Science, forthcoming.

Pedroni, P. and Yao, J. Y. (2006). Regional income divergence in China. Journal of Asian Economics, 17(2):294–315.

Quah, D. T. (1993a). Empirical cross-section dynamics in economic growth. European Economic Review, 37:426–434.

Quah, D. T. (1993b). Galton's fallacy and tests of the convergence hypothesis. Scandinavian Journal of Economics, 95(4):427–443.

Quah, D. T. (1996a). Aggregate and regional disaggregate fluctuations. Empirical Economics, 21:137–159.

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

Quah, D. T. (1996c). Regional convergence clusters across Europe. European Economic Review, 40:951–958.

Quah, D. T. (1996d). Twin Peaks: Growth and convergence in models of distribution dynamics. Economic Journal, 106(437):1045–1055.

Quah, D. T. (1997a). Empirics for growth and distribution: Stratification, polarization and convergence clubs. Journal of Economic Growth, 2(1):27–59.

Quah, D. T. (1997b). Regional cohesion from local isolated actions. II. Conditioning. Discussion Paper 379, CEPR.

Raiser, M. (1998). Subsidising inequality: Economic reforms, fiscal transfers and convergence across Chinese provinces. Journal of Development Studies, 34(3):1–26.

Redding, S. (2002). Specialization dynamics. Journal of International Economics, 58(2):299–334.

Rodríguez-López, J., Martínez-López, D., and Romero-Ávila (2009). Persistence of inequalities across the Spanish regions. Papers in Regional Science, 88(4):841–862.

Rozelle, S. (1994). Rural industrialization and increasing inequality: Emerging patterns in China's reforming economy. Journal of Comparative Economics, 19(3):362 – 391.

Sakamoto, H. and Islam, N. (2008). Convergence across Chinese provinces: An analysis using Markov transition matrix. China Economic Review, 19(1):66–79.

Sala-i-Martin, X. (2006). The world distribution of income: Falling poverty and... convergence, period. Quarterly Journal of Economics, 121(2):351–398.

Schultz, T. P. (1998). Inequality in the distribution of personal income in the world: How it is changing and why. Journal of Population Economics, 11(3):307–344.

Shorrocks, A. F. (1978). The measurement of mobility. Econometrica, 46(5):1013–1024.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Temple, J. (1999). The new growth evidence. Journal of Economic Literature, 37:112–156.

Tortosa-Ausina, E., Pérez, F., Mas, M., and Goerlich, F. J. (2005). Growth and convergence profiles in Spanish provinces. Journal of Regional Science, 45:147–182.

Wand, M. P. and Jones, M. C. (1995). Kernel Smoothing. Chapman and Hall, London.

Wang, Z. (2003). Regional Divergence of per capita GDP in China: 1991-1999. Journal of Chinese Economic and Business Studies, 2:39–53.

Weeks, M. and Yao, J. Y. (2003). Provincial conditional income convergence in China, 1953–1997: A panel data approach. Econometric Reviews, 22(1):59–77.

Wei, Y. (1996). Fiscal systems and uneven regional development in China, 1978–1991. Geoforum, 27(3):329–344.

Yao, S. and Zhang, Z. (2001a). On regional inequality and diverging clubs: a case study of contemporary China. Journal of Comparative Economics, 29(3):466–484.

Yao, S. and Zhang, Z. (2001b). Regional growth in China under economic reforms. Journal of Development Studies, 38(2):167–186.

Table 1: Descriptive statistics for Chinese provinces, per capita income (GDP/N) and population (N), 1952 to 2005

Province	Y/N^a			Y/N annual growth rate (%)			N^b			N annual growth rates (%)		
East	**1952**	**1978**	**2005**	**1952-78**	**1978-05**	**1952-05**	**1952**	**1978**	**2005**	**1952–78**	**1978–05**	**1952–05**
Shanghai	430	2,942	24,374	7.68	8.15	7.92	573	1,098	1,360	2.54	0.80	1.65
Beijing	165	1,441	11,809	8.69	8.10	8.39	490	872	1,538	2.24	2.13	2.18
Tianjin	298	1,086	10,679	5.10	8.83	6.99	439	724	939	1.94	0.97	1.44
Liaoning	218	815	7,072	5.20	8.33	6.79	1,932	3,394	4,221	2.19	0.81	1.49
Jiangsu	131	310	5,777	3.36	11.45	7.41	3,739	5,834	7,475	1.73	0.92	1.32
Zhejiang	112	274	5,948	3.49	12.08	7.78	2,213	3,751	4,898	2.05	0.99	1.51
Guangdong	101	217	3,851	2.99	11.24	7.11	2,910	5,064	9,194	2.15	2.23	2.19
Shangdong	91	286	4,711	4.51	10.93	7.73	4,827	7,160	9,248	1.53	0.95	1.23
Fujian	102	234	3,976	3.25	11.06	7.16	1,270	2,446	3,535	2.55	1.37	1.95
Guangxi	67	202	1,602	4.33	7.98	6.17	1,943	3,402	4,660	2.18	1.17	1.66
Hebei	125	341	3,788	3.94	9.33	6.65	3,272	5,057	6,851	1.69	1.13	1.40
Mean	167.27	740.66	7,598.79	4.78	9.77	7.28	2,146.18	3,527.50	4,901.73	2.07	1.23	1.64
Median	125.00	309.55	5,776.58	4.33	9.33	7.16	1,943.00	3,402.00	4,660.00	2.15	0.99	1.51
Coefficient of variation	0.65	1.13	0.83	0.39	0.16	0.09	0.67	0.60	0.61	0.16	0.41	0.20
Central												
Heilongjiang	238	399	2,622	2.01	7.22	4.63	1,111	3,130	3,820	4.07	0.74	2.36
Jilin	153	324	3,112	2.93	8.73	5.85	1,065	2,149	2,669	2.74	0.81	1.75
Hubei	90	215	2,312	3.40	9.20	6.31	2,751	4,575	6,031	1.98	1.03	1.49
Shanxi	116	322	2,917	4.00	8.50	6.27	1,395	2,424	3,355	2.15	1.21	1.67
Hunan	86	211	1,748	3.51	8.15	5.85	3,271	5,166	6,732	1.77	0.99	1.37
Anhui	78	120	1,246	1.68	9.05	5.37	2,966	4,713	6,516	1.80	1.21	1.50
Jiangxi	114	179	1,642	1.75	8.56	5.16	1,656	3,183	4,307	2.55	1.13	1.82
Henan	83	158	1,919	2.50	9.69	6.10	4,371	7,067	9,768	1.87	1.21	1.53
Inner Mongolia	173	300	4,288	2.14	10.35	6.24	716	1,823	2,386	3.66	1.00	2.30
Mean	125.68	247.61	2,422.88	2.66	8.83	5.75	2,144.59	3,803.28	5,065.01	2.51	1.03	1.75
Median	114.00	215.05	2,312.20	2.50	8.73	5.85	1,655.69	3,182.82	4,307.00	2.15	1.03	1.67
Coefficient of variation	0.42	0.37	0.38	0.32	0.10	0.10	0.58	0.45	0.47	0.34	0.17	0.20
West												
Sichuan	67	158	1,698	3.37	9.20	6.30	6,405	9,707	11,010	1.61	0.47	1.03
Xinjiang	166	259	2,489	1.73	8.74	5.24	465	1,233	2,010	3.82	1.83	2.80
Qinghai	101	376	2,052	5.18	6.49	5.85	161	365	543	3.19	1.48	2.32
Ningxia	126	462	3,087	5.13	7.29	6.22	142	356	596	3.58	1.93	2.74
Gansu	125	322	2,541	3.70	7.96	5.85	1,065	1,870	2,594	2.19	1.22	1.69
Shaanxi	85	322	3,141	5.25	8.80	7.05	1,528	2,779	3,718	2.33	1.08	1.69
Yunnan	70	178	1,439	3.65	8.05	5.87	1,695	3,091	4,450	2.34	1.36	1.84
Guizhou	58	117	874	2.74	7.73	5.25	1,490	2,686	3,730	2.29	1.22	1.75
Mean	99.70	274.09	2,165.01	3.84	8.03	5.95	1,618.96	2,760.98	3,581.56	2.67	1.32	1.98
Median	93.00	290.40	2,270.38	3.67	8.00	5.86	1,277.30	2,278.23	3,156.18	2.33	1.29	1.79
Coefficient of variation	0.37	0.43	0.37	0.33	0.11	0.10	1.25	1.09	0.93	0.29	0.35	0.30
Total 28 provinces												
Mean	134.60	448.88	4,382.60	3.83	8.97	6.41	1,995.04	3,397.14	4,577.02	2.38	1.19	1.77
Median	113.00	293.18	3,002.24	3.50	8.74	6.26	1,591.85	3,110.54	4,020.50	2.19	1.13	1.68
Coefficient of variation	0.60	1.27	1.08	0.43	0.15	0.14	0.77	0.66	0.63	0.29	0.35	0.25

[a] In 1952 yuan/person.
[b] In 10,000 persons.

Table 2: Transition probability matrix and ergodic distribution, per capita income (GDP/N), unweighted, 5-year transitions, limits all years

	Upper limit, all years:				
(Number of observations)	0.915	0.965	0.997	1.053	Max.
(280)	0.83	0.15	0.01	0.01	0.00
(272)	0.14	0.66	0.19	0.01	0.00
(273)	0.01	0.15	0.58	0.26	0.00
(274)	0.01	0.01	0.22	0.68	0.07
(273)	0.00	0.00	0.01	0.08	0.91
Initial distribution (1952)	0.18	0.25	0.14	0.18	0.25
Final distribution (2005)	0.14	0.21	0.25	0.18	0.21
Ergodic distribution	0.16	0.19	0.22	0.27	0.17

a) 1952–2005

	Upper limit, all years:				
(Number of observations)	0.915	0.965	0.997	1.053	Max.
(122)	0.76	0.20	0.03	0.01	0.00
(151)	0.16	0.61	0.23	0.01	0.00
(106)	0.02	0.28	0.45	0.25	0.00
(95)	0.01	0.05	0.14	0.71	0.08
(142)	0.00	0.00	0.01	0.10	0.89
Initial distribution (1952)	0.18	0.25	0.14	0.18	0.25
Final distribution (1978)	0.21	0.18	0.18	0.25	0.18
Ergodic distribution	0.21	0.18	0.12	0.13	0.37

b) 1952–1978

	Upper limit, all years:				
(Number of observations)	0.915	0.965	0.997	1.053	Max.
(134)	0.87	0.12	0.00	0.01	0.00
(94)	0.10	0.80	0.10	0.00	0.00
(149)	0.00	0.06	0.68	0.26	0.00
(156)	0.00	0.00	0.19	0.72	0.08
(111)	0.00	0.00	0.00	0.01	0.99
Initial distribution (1978)	0.21	0.18	0.18	0.25	0.18
Final distribution (2005)	0.14	0.21	0.25	0.18	0.21
Ergodic distribution	0.02	0.06	0.13	0.27	0.52

c) 1978–2005

Notes: The variable of analysis is $x_{it} = \ln y_{it} / \ln \bar{y}_t$, where y_{it} is the per capita GDP of the province (in constant 1952 prices). The 5-year (or quinquennial) transition refers to the movement of x_{it} from one of the five states in period t to another (including staying in the same) state in period $t+5$. The transition matrices presented in this Table are estimated by averaging the observed 5-year transitions of *provinces* during the periods 1952–2005 (top panel), 1952–1978 (middle panel), and 1978–2005 (bottom panel). The transition matrices and ergodic distributions displayed in each panel are based on five states, whose upper limits (the "grid") are chosen to yield a virtually uniform distribution over the observed sample. In order to facilitate comparisons, these cut-off points were calculated using the entire 1952–2005 sample (totalling 28 provinces $\times$ 54 years $=$ 1512 observations), i.e., the top panel, and remained unchanged throughout the entire analysis. The numbers in parentheses on the left are the numbers of observations beginning from a particular state. The cells are arranged in ascending order, with the upper left cell in each matrix showing transitions from the poorest to the poorest. The way the variable of analysis x_{it} is computed allows an economically meaningful interpretation of each state to be made, i.e., observations in state one are those with GDP per capita lower than the 91.5% of the national average, as indicated by its cut-off point. The ergodic distributions are computed following Kremer et al. (2001).

Table 3: Transition probability matrix and ergodic distribution, per capita income (GDP/N), GDP-weighted, 5-year transitions, limits all years

	Upper limit, all years:				
(Share of GDP)	0.915	0.965	0.997	1.053	Max.
(0.16)	0.83	0.15	0.01	0.01	0.00
(0.18)	0.12	0.69	0.18	0.01	0.00
(0.17)	0.01	0.15	0.56	0.28	0.00
(0.18)	0.01	0.02	0.18	0.70	0.09
(0.31)	0.00	0.00	0.00	0.05	0.94
Initial distribution (1952)	0.16	0.29	0.12	0.20	0.24
Final distribution (2005)	0.06	0.16	0.12	0.26	0.40
Ergodic distribution	0.09	0.12	0.12	0.25	0.41

a) 1952–2005

	Upper limit, all years:				
(Share of GDP)	0.915	0.965	0.997	1.053	Max.
(0.18)	0.79	0.18	0.03	0.01	0.00
(0.25)	0.14	0.64	0.22	0.00	0.00
(0.17)	0.03	0.28	0.51	0.18	0.00
(0.09)	0.01	0.12	0.15	0.67	0.05
(0.31)	0.00	0.00	0.01	0.11	0.88
Initial distribution (1952)	0.16	0.29	0.12	0.20	0.24
Final distribution (1978)	0.16	0.16	0.20	0.19	0.29
Ergodic distribution	0.42	0.28	0.17	0.07	0.06

b) 1952–1978

	Upper limit, all years:				
(Share of GDP)	0.915	0.965	0.997	1.053	Max.
(0.15)	0.87	0.13	0.00	0.00	0.00
(0.12)	0.07	0.81	0.12	0.00	0.00
(0.16)	0.00	0.03	0.63	0.35	0.00
(0.26)	0.00	0.00	0.14	0.74	0.12
(0.31)	0.00	0.00	0.00	0.01	0.99
Initial distribution (1978)	0.16	0.16	0.20	0.19	0.29
Final distribution (2005)	0.06	0.16	0.12	0.26	0.40
Ergodic distribution	0.00	0.00	0.00	0.02	0.98

c) 1978–2005

Notes: Table 2's notes also apply here with the exception that the transition matrices are estimated by averaging the observed 5-year transitions of *GDP* (i.e., the GDP of each province that moves from one state to another) during the periods 1952–2005 (top panel), 1952–1978 (middle panel), and 1978–2005 (bottom panel). Therefore, the numbers in parentheses on the left are the percentage of GDP beginning from a particular state; these percentages were calculated taking into account the GDP of each province beginning from a particular state, and the sum of the numbers in parentheses in Table 2.a represents 100%.

Table 4: Transition probability matrix and ergodic distribution, per capita income (GDP/N), population-weighted, 5-year transitions, limits all years

	Upper limit, all years:				
(Share of population)	0.915	0.965	0.997	1.053	Max.
(0.30)	0.84	0.15	0.01	0.00	0.00
(0.25)	0.11	0.69	0.18	0.01	0.00
(0.19)	0.01	0.14	0.56	0.29	0.00
(0.16)	0.00	0.02	0.18	0.71	0.08
(0.10)	0.00	0.00	0.00	0.06	0.94
Initial distribution (1952)	0.26	0.36	0.12	0.17	0.10
Final distribution (2005)	0.15	0.29	0.15	0.25	0.16
Ergodic distribution	0.10	0.12	0.13	0.30	0.35

a) 1952–2005

	Upper limit, all years:				
(Share of population)	0.915	0.965	0.997	1.053	Max.
(0.31)	0.79	0.18	0.03	0.00	0.00
(0.31)	0.14	0.64	0.22	0.00	0.00
(0.19)	0.02	0.27	0.52	0.19	0.00
(0.08)	0.01	0.12	0.15	0.67	0.05
(0.11)	0.00	0.00	0.01	0.11	0.88
Initial distribution (1952)	0.26	0.35	0.12	0.17	0.10
Final distribution (1978)	0.32	0.22	0.21	0.19	0.07
Ergodic distribution	0.45	0.28	0.17	0.07	0.03

b) 1952–1978

	Upper limit, all years:				
(Share of population)	0.915	0.965	0.997	1.053	Max.
(0.30)	0.86	0.13	0.00	0.00	0.00
(0.18)	0.07	0.82	0.11	0.00	0.00
(0.18)	0.00	0.02	0.63	0.34	0.00
(0.24)	0.00	0.00	0.14	0.74	0.11
(0.10)	0.00	0.00	0.00	0.00	1.00
Initial distribution (1978)	0.32	0.22	0.21	0.19	0.07
Final distribution (2005)	0.15	0.30	0.15	0.24	0.16
Ergodic distribution	0.00	0.00	0.00	0.03	0.97

c) 1978–2005

Notes: Table 2's notes also apply here with the exception that the transition matrices are estimated by averaging the observed 5-year transitions of *people* (i.e., the population of each province that moves from one state to another) during the periods 1952–2005 (top panel), 1952–1978 (middle panel), and 1978–2005 (bottom panel). Therefore, the numbers in parentheses on the left are the percentage of population beginning from a particular state; these percentages were calculated taking into account the population of each province beginning from a particular state, and the sum of the numbers in parentheses in Table 2.a represents 100%.

Table 5: Transition probability matrix and ergodic distribution, per capita income (GDP/N), physically contiguous-conditioned, 5-year transitions, limits all years

	Upper limit, all years:				
(Number of observations)	0.915	0.965	0.997	1.053	Max.
(193)	0.73	0.25	0.02	0.00	0.00
(250)	0.18	0.66	0.14	0.02	0.00
(207)	0.02	0.12	0.57	0.29	0.00
(474)	0.00	0.01	0.18	0.74	0.06
(248)	0.00	0.00	0.00	0.08	0.91
Initial distribution (1952)	0.00	0.46	0.07	0.29	0.18
Final distribution (2005)	0.04	0.25	0.25	0.32	0.14
Ergodic distribution	0.13	0.17	0.19	0.38	0.13

a) 1952–2005

	Upper limit, all years:				
(Number of observations)	0.915	0.965	0.997	1.053	Max.
(94)	0.82	0.16	0.02	0.00	0.00
(99)	0.34	0.49	0.12	0.05	0.01
(85)	0.03	0.16	0.31	0.49	0.01
(207)	0.00	0.01	0.19	0.69	0.11
(131)	0.00	0.00	0.01	0.17	0.82
Initial distribution (1952)	0.00	0.46	0.07	0.29	0.18
Final distribution (1978)	0.25	0.11	0.04	0.46	0.11
Ergodic distribution	0.32	0.10	0.08	0.26	0.24

b) 1952–1978

	Upper limit, all years:				
(Number of observations)	0.915	0.965	0.997	1.053	Max.
(74)	0.69	0.30	0.01	0.00	0.00
(136)	0.05	0.81	0.13	0.00	0.00
(113)	0.01	0.07	0.78	0.13	0.00
(224)	0.00	0.01	0.17	0.78	0.03
(97)	0.00	0.00	0.00	0.07	0.93
Initial distribution (1978)	0.25	0.11	0.04	0.46	0.11
Final distribution (2005)	0.04	0.25	0.25	0.32	0.14
Ergodic distribution	0.04	0.21	0.45	0.27	0.02

c) 1978–2005

Notes: Table 2's notes also apply here with the exception that the variable of analysis is the neighbour-relative GDP per capita series of province i in period t, x_{it}^{NR}, as defined in Equation (11). The 5-year (or quinquennial) transition refers to the movement of x_{it}^{NR} from one of the five states in period t to another (including staying in the same) state in period $t+5$. Therefore, the transition matrices presented in this Table are estimated by averaging the observed 5-year transitions of *provinces* during the periods of 1952–2005 (top panel), 1952–1978 (middle panel), and 1978–2005 (bottom panel).

Table 6: Mobility indices (μ_1)[a]

Transition matrix	1952-1978	1978-2005	1952-2005
Unweighted	0.713	0.573	0.640
GDP-weighted	0.728	0.584	0.648
Population-weighted	0.729	0.580	0.645
Physically contiguous-conditioned	0.783	0.624	0.676

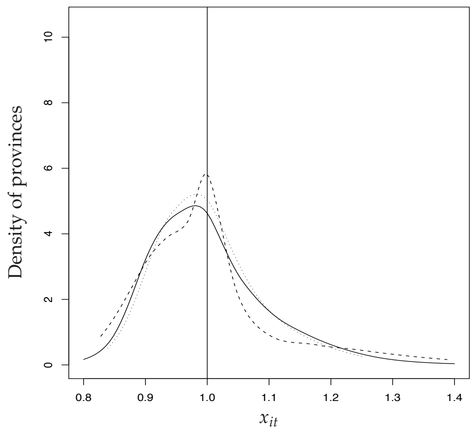
[a] The values refer to the μ_1 index, as defined in Equation (4), which summarises the mobility information in each transition probability matrix in one number so as to facilitate comparisons across them.

Table 7: Transition path analysis (asymptotic half-life of convergence, $H - L$)[a]

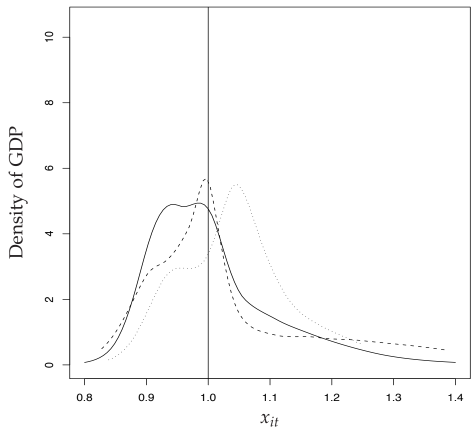
Transition matrix	1952–1978	1978–2005	1952–2005
Unweighted	69.490	25.680	20.969
GDP-weighted	73.116	20.470	25.692
Population-weighted	44.379	21.452	19.808
Physically contiguous-conditioned	17.095	9.250	9.661

[a] The values indicate the speed at which the ergodic or steady-state distribution is approached. Specifically, they refer to the concept of the asymptotic half-life of the chain, $H - L$, which is how long it takes to cover half the distance from the stationary distribution. Since we are using 5-year transitions, these numbers should be multiplied by 5 in order to have them in years.

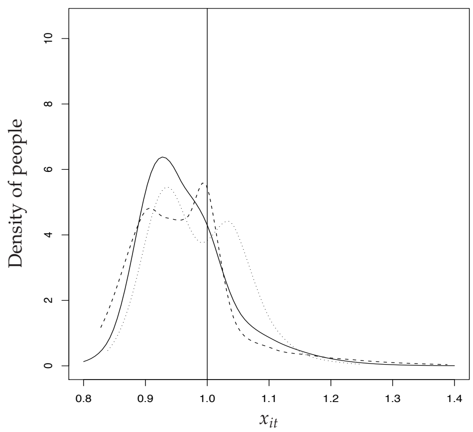
Figure 1: GDP/N, densities, 1952 vs. 1978 vs. 2005



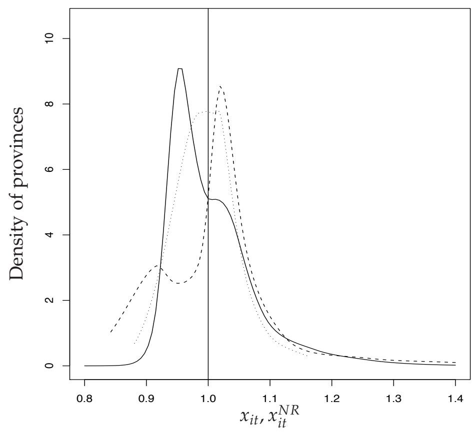
(a) Unweighted



(b) GDP-weighted



(c) Population-weighted

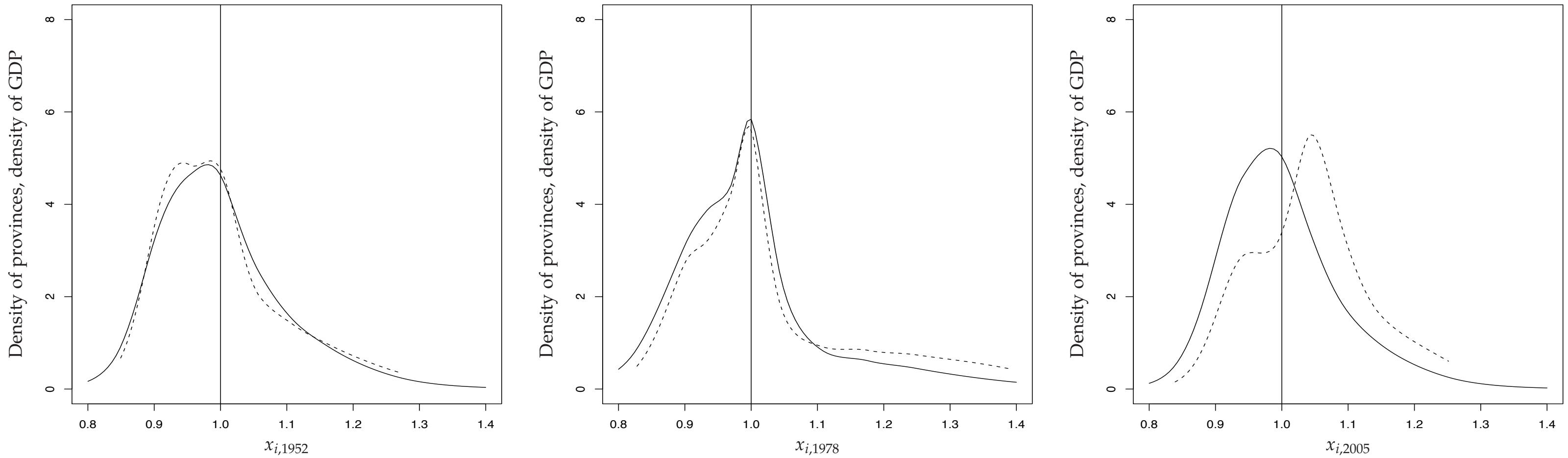


(d) Physically contiguous-conditioned

1952 —— 1978 ------ 2005 ········

Notes: All figures contain densities estimated using local likelihood density estimation. The vertical line represents the average, which is the unity because we have normalised the variable of analysis, $x_{it} = \ln y_{it} / \ln \bar{y}_t$, where y_{it} is the per capita GDP of the province (in constant 1952 prices).

Figure 2: GDP/N, densities, unweighted vs. GDP-weighted

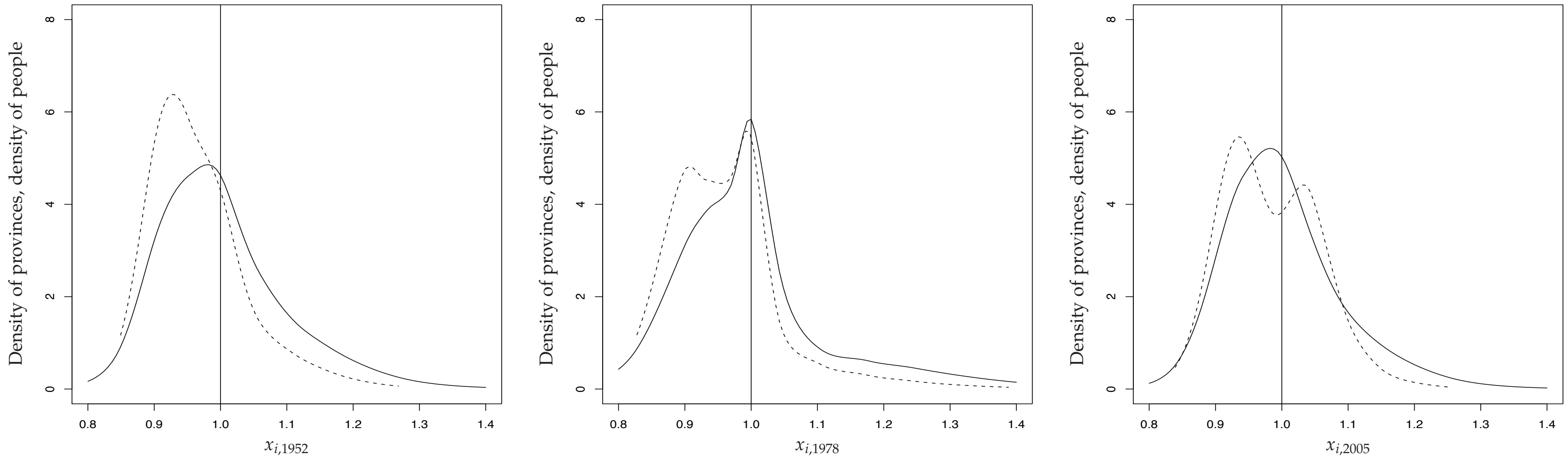


(a) 1952 **(b)** 1978 **(c)** 2005

Unweighted —— GDP-weighted ------

Notes: All figures contain densities estimated using local likelihood density estimation for the years 1952, 1978 and 2005. The vertical line represents the average, which is unity because we have normalised the variable of analysis, $x_{it} = \ln y_{it} / \ln \bar{y}_t$, where y_{it} is the per capita GDP of the province (in constant 1952 prices). The solid line is the unweighted density of the x_{it}, whereas the dashed line refers to the GDP-weighted density.

Figure 3: GDP/N, densities, unweighted vs. population-weighted

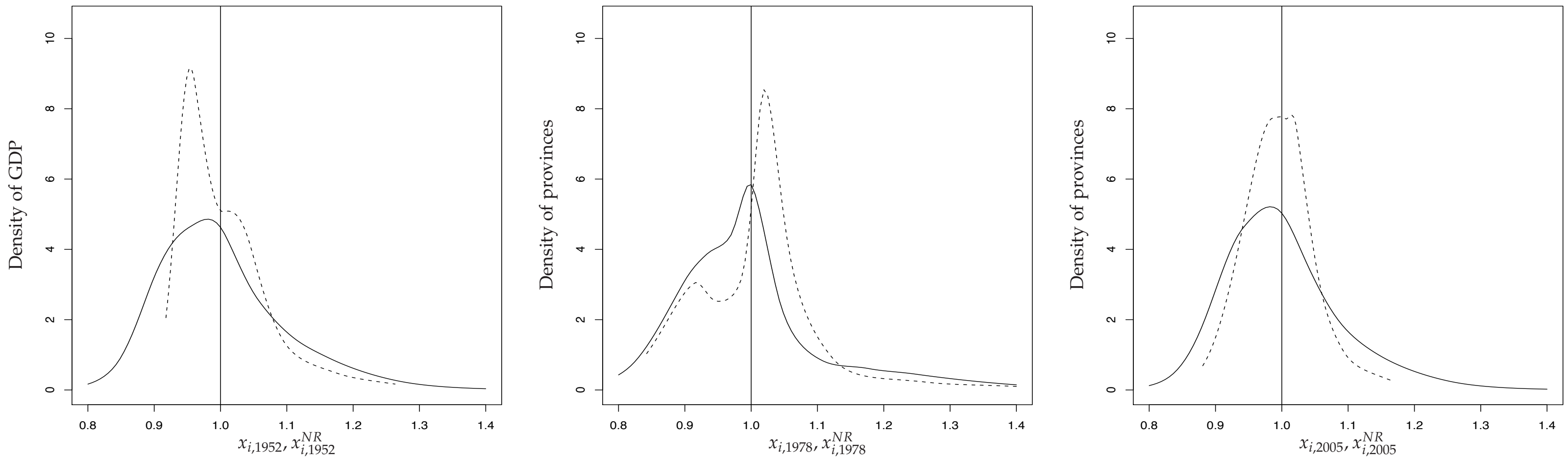


(a) 1952 **(b)** 1978 **(c)** 2005

Unweighted —— Population-weighted ------

Notes: All figures contain densities estimated using local likelihood density estimation for the years 1952, 1978 and 2005. The vertical line represents the average, which is unity because we have normalised the variable of analysis, $x_{it} = \ln y_{it} / \ln \bar{y}_t$, where y_{it} is the per capita GDP of the province (in constant 1952 prices). The solid line is the unweighted density of the x_{it}, whereas the dashed line refers to the population-weighted density.

Figure 4: GDP/N, densities, unweighted vs. physically contiguous-conditioned

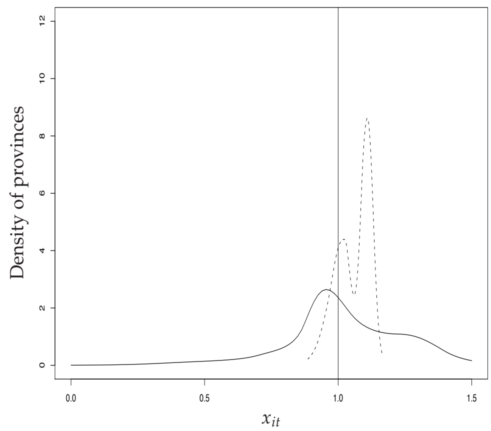


(a) 1952 **(b)** 1978 **(c)** 2005

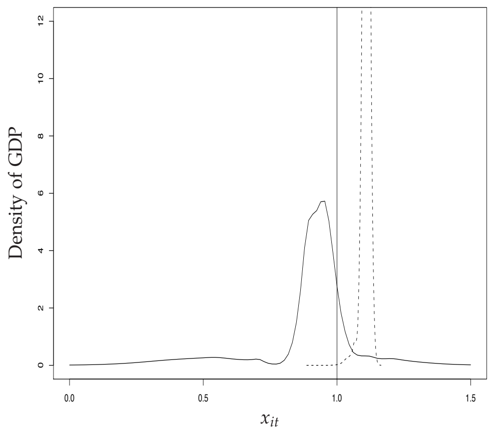
Unweighted —— Physically contiguous-conditioned ------

Notes: All figures contain densities estimated using local likelihood density estimation for the years 1952, 1978 and 2005. The vertical line represents the average, which is unity because we have normalised the variable of analysis, $x_{it} = \ln y_{it}/\ln \bar{y}_t$, where y_{it} is the per capita GDP of the province (in constant 1952 prices). The solid line is the unweighted density of the x_{it}, whereas the dashed line refers to the neighbour-relative GDP per capita series of province i in period t, x^{NR}_{it}, as defined in Equation (11).

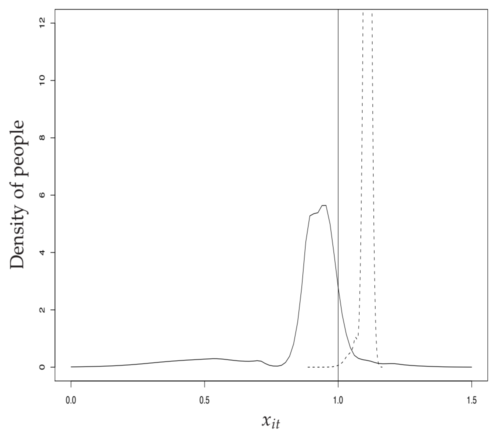
Figure 5: GDP/N, ergodic distributions, 1952–1978 vs. 1978–2005



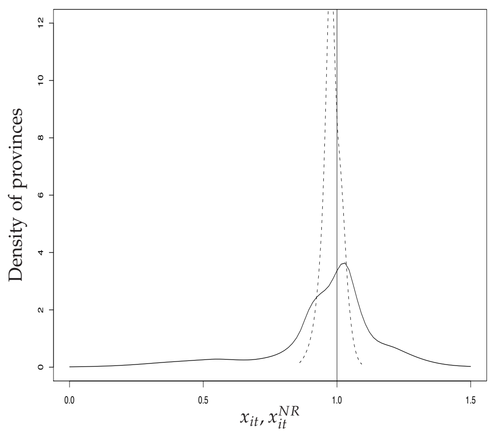
(a) Unweighted



(b) GDP-weighted



(c) Population-weighted



(d) Physically contiguous-conditioned

1952–1978 —— 1978–2005 ------

Notes: All figures contain ergodic densities estimated using local likelihood density estimation. The vertical line represents the average, which is unity because we have normalised the variable of analysis, $x_{it} = \ln y_{it} / \ln \bar{y}_t$, where y_{it} is the per capita GDP of the province (in constant 1952 prices). The solid line in each subfigure represents the ergodic densities under 1952–1978 trends, whereas the dashed lines are the ergodic densities under 1978–2005 trends. The scale of the vertical axes is not displayed in full in order to facilitate comparison of the densities.