Remoteness, Urbanization and India's Unbalanced Growth.

Samarjit Das
Indian Statistical Institute, Kolkata*

Chetan Ghate
Indian Statistical Institute, Delhi†

Peter E. Robertson‡
The University of Western Australia

August 9, 2014

*Economic Research Unit, Indian Statistical Institute, 203 B.T.Road, Kolkata-700108, India. E-mail: samarjit@isical.ac.in

†Economics and Planning Unit, Indian Statistical Institute - Delhi Center, 7 SJS Sanswal Marg, New Delhi - 110016, India. Email: cghate@isid.ac.in

‡Corresponding Author: Peter Robertson, The Business School, University of Western Australia, M251 UWA, Crawley, Perth, W.A, Australia. Email:peter.robertson@uwa.edu.au.

Abstract

The unbalanced nature of India's growth has caused considerable concern but little is known about its causes. We use a new data set of district level income and socio-economic data to explore the determinants of transitional growth at the district level. We find that there is absolute divergence across districts but weak conditional convergence once we allow for district characteristics, particularly urbanization and the distance from a major urban agglomeration. State-level effects have also significantly contributed to India's unbalanced growth. This results suggest that while geography is important, policy differences may also account for much of India's uneven growth growth.

Keywords: Convergence, Unbalanced Growth, India, Gravity Models, New Economic Geography, Urbanization.

JEL: O4, O5.

Acknowledgements

We thank Anu Rammohan, Ken Clements, Leandro Magnusson, M.H. Suryanarayana, Rod Tyers, participants at the Australian Development Economics Workshop, Monash University 2012, the 8th ISI Conference on Economic Growth and Development, ISI Delhi, 2012, and the University of Western Australia Economics Seminar Programme. We are grateful to the Editor, Arun Agrawal, and two referees for insightful comments. We also would like to thank Shradha Mohta and Theo Backhouse for excellent research assistance. Robertson gratefully acknowledges funding from the Australian Research Council (DP110101316). Ghate and Robertson gratefully acknowledge funding from the Australia-India Institute. Ghate acknowledges hospitality from ICRIER during 2012-2013, where part of this project was completed.

1 Introduction

India's tentative economic miracle faces many hurdles, but one of the chief difficulties is the unbalanced nature of growth, (Bardhan 2010). The resulting income disparities have stimulated considerable debate over how the gains from growth in India are being shared and may impede the political case for economic reform.[1] Unbalanced growth may reflect policy failures but may also be an efficient outcome of a dynamic growing economy, (World-Bank 2009). Thus it is important to gain an understanding of the causes of India's unbalanced growth in order to understand the trade-offs facing policy makers.

Evidence of India's unbalanced growth is apparent from the numerous studies that find richer states are growing faster, so that state average incomes are diverging (Cashin and Sahay 1996, Rao and Sen 1997, Rao, Shand and Kalirajan 1999, Trivedi 2003, Bandopadhyay 2004, Ghate 2008, Kar, Jha and Kateja 2011, Das 2012, Ghate and Wright 2012, Bandopadhyay 2012).[2] This pattern of divergence might be regarded as curious given that there are no political barriers to migration, approximately free trade, and a common set of federal institutions.

As emphasized by the new economic geography (NEG) literature, however, growth might be unbalanced for several reasons. These include trade and migration costs, and economies of scale associated with urbanization. Thus trying to offset unbalanced growth with a policy response could be costly for overall growth.[3]

[1] See for example Bhagwati and Panagariya (2013) and Dreze and Sen (2013).

[2] Complementing these state level studies is the literature on rising inequality at the individual or household level, and differences in wages across skill levels (Datt and Ravallion 2002, Mishra and Kumar 2005, Chamarbagwala 2008, Chaudhuri and Ravallion 2007, Dev and Ravi 2007, Cain, Hasan, Magsombol and Tandon 2010).

[3] See for example World-Bank (2009) for a general summary of this literature. With respect to India specifically, Desmet, Ghani, O'Connell and Rossi-Hansberg (2013) find that India's spatial pattern of growth displays a much higher than usual difference in growth rates across different sized urban areas.

Nevertheless regional growth disparities may also arise as a result of government policies such as the supply of public infrastructure and the quality of governance. It has been widely argued that India faces a severe shortage of public infrastructure resulting in differences between not only rural and urban areas, but also across rural areas and across cities (Basu and Maertens 2009, Sachs 2009, Lall, Wang and Deichmann 2010, Ghani, Goswami and Kerr 2012, Cain, Hasan and Mitra 2012). According to Crost and Kambhampati (2010), this differential supply of public infrastructure also applies to schooling infrastructure.[4] Likewise India's states have had different market reform programmes (Cain et al. 2012).[5]

One way to gain a better sense of the sources of the imbalance is to look at the growth experience across India within states, that is, at the district level. The aim of this paper is, therefore, to use newly available data on India's 575 districts to gain a better understanding of the causes of India's unbalanced growth. In particular we wish to see whether the pattern of divergence across states is similar within states, and, if so, how geographical factors, infrastructure, and other possible factors affect these district level differences.

We proceed, first, with a descriptive analysis of growth rates and income levels at the district level, between 2000-01 and 2007-08. This preliminary analysis shows a strong imbalance in growth rates across districts, suggesting that the growth in inequality across India runs much deeper than just differences across states.

Second we consider the causes of regional growth explicitly and, in particular, the role of geography, infrastructure and literacy rates emphasized in the new economic geography (NEG) literature. To achieve this we combine our data on per capita incomes with

---

[4]This last point may be important if, as suggested by some, that there has been a sharp increase in the returns to schooling following reforms (Cain et al. 2010, Azam 2012).

[5]Krishna and Sethupathy (2012) argue that the evidence of links between inequality and reforms in India are fairly weak.

district level social and economic characteristics for each district including literacy and infrastructure and spatial variables. Of particular interest is the role of the spatial distribution of markets faced by each district that capture the districts remotest or access to markets in terms of trade, migration and other linkages.

We find that urbanization, irrigation, electricity provision and state dummy variables are all highly significant factors in explaining differences in transitional growth rates and income levels across Indian districts. Interestingly we find no evidence that literacy and road quality have any impact on these district growth rates or income levels.

In terms of spatial factors we find very strong evidence that being close to a major city is a significant factor, but that being close to a large number of different markets is not important. We argue that this result is consistent with a setting where trade is largely in primary goods and there is relatively free mobility of labor and other factors across borders.

The results, therefore, point to a variety of factors as being important. They confirm that geography is important with significant benefits from urbanization and being close to cities. Nevertheless, even after for controlling for these factors, the results suggest that there remains scope to promote more balanced growth through policy reform.

2 Preliminary Statistical Analysis

2.1 District GDP data

To investigate the pattern of growth across India we use two new data sets of district level incomes and social and economic characteristics – respectively the *Indicus* "Development Landscape" and "District GDP" data-sets. The data consist of 575 district level observations of district income for two years, 2001 and 2008.[6]

The availability of district level income data provides the opportunity to observe regional disparities in India at a much finer level than previous studies based on Indian states. This is also advantageous insofar as there is likely to be a larger degree of heterogeneity in income levels, growth rates and other characteristics such as urbanization or literacy, compared to state level data.

We begin with a preliminary exploration of the data by considering different indicators of convergence and how the shape of the distribution of district incomes has changed over time. First, Table 1 shows the wide disparity in income levels across states. There is a 9.8 fold difference in 2007-2008 per-capita incomes between the richest state *Goa*, and the poorest state *Bihar*. This is larger than the real income gap between the GDP per capita of the USA and Angola, and only slightly smaller than the real income gap between the USA and India.[7]

At the district level, however, that gap is much larger. The range in per capita incomes in 2008 is from a minimum of RS. (m) 3858 in the *Sheohar* district (*Bihar*) to a maximum of RS. (m) 139868 in *Jamnagar* (*Gujarat*). This implies an income ratio of 36, which is equivalent, for example, to the ratio between the USA and Rwanda according to the Penn World Tables.

(Insert Table 1 about here)

The district data are shown visually in Figure 1. It can be seen that there are generally lower incomes in central districts as well as in the eastern states. Likewise the wealthy western corridor running from the north of Delhi down the west coast of India through

[6]This data has attracted some debate. See Himanshu (2009) but also, importantly, the reply by Bhandari (2009).

[7]This comparison is based on the Penn World Tables PPP values, that report Angola with a relative per capita GDP of 11.51 and India 7.21 in 2008.

Western *Maharashtra*, *Karnataka*, *Goa* and *Kerala* is easily observed. Figure 1 is thus suggestive of a strong geographic pattern in the differences in per-capita incomes across India.

(Insert Figure 1 about here)

The fact that the within-India differences are comparable to cross-country per-capita differences is remarkable given that there are no political barriers to migration, approximately free trade, and a common set of federal institutions, policies and governance. That such differences could persist over time is in stark contradiction to the standard competitive model that motivates the extensive literature on absolute β–convergence across regions.[8] In contrast, it points to the potential relevance of trade barriers, transport costs and agglomeration effects as emphasized in the NEG literature.

2.2 Absolute Convergence Across Districts

A simple starting point from which to analyse differences in transitional growth rates across districts is to employ the standard concept of absolute β–convergence (Baumol 1986, Sala-i Martin 1997, Durlauf et al. 2005). This is given by the coefficient β from (1):

$$y_{i,t} - y_{i,0} = \beta \, y_{i,0} \; + \; \varepsilon_i \tag{1}$$

where $y_{i,t}$ is the natural log of per-capita income at time t in region i and $y_{i,0}$ is initial per-capita income.[9] The left hand side of (1) represents the transitional growth rate over the period $(0, t)$. The results of estimating (1) across Indian districts are given in Table

[8]For example the hypothesis of absolute β–convergence has found widespread support in other countries (Sala-i Martin 1996, Durlauf, Johnson and Temple 2005).

[9]We report β for all states except *Goa*, *Pondichery* and *Chandigarh* where the number of districts is 2 or 1.

1. It can be seen that across India there is strong evidence of a small rate of divergence with $\beta = 0.007$, which is statistically significant at the 1 % level. Hence, on average, richer districts have been growing slightly faster than poorer districts.

Table 1 also shows the results of estimating (1) for each state separately. Thus we ask whether there is convergence across districts within each state. In four states, *Assam*, *Chhattisgarh*, *Kerala* and *Rajasthan*, there is significant absolute β–convergence of district level incomes. However there is also significant within-state divergence in three states – *Haryana*, *Orissa* and *Uttar Pradesh (UP)*.[10] Nevertheless for the vast majority of states the estimated β–convergence coefficient is insignificantly different from zero. Thus there is little evidence of strong convergence, either across the country as a whole or within individual states.

Next we consider σ–convergence, which is defined as a decline in the variance of district level per-capita log incomes across time. Table 2 shows the variance of district log per-capita incomes in the two periods, 2001 and 2008. It can be seen that there was a 30.7% increase in the variance of log per-capita incomes across districts – from 0.27 to 0.35. Thus there has also been σ–divergence.

(Insert Table 2 about here)

Table 2 reports a simple variance decomposition using log per-capita incomes.[11] Here, *within-state* variance, ν^W, refers to deviations of district log per-capita incomes, y_{ij}, from their state level mean log per-capita income, $\bar{y}_j$, $y_{ij} - \bar{y}_j$, and *between-state* variance, ν^B, refers to deviations of state level mean log per-capita incomes $\bar{y}_i$ from the country-wide mean log per-capita income, $\bar{y}$, $\bar{y}_i - \bar{y}$. By definition, the total India-wide variance of per-capita incomes across all districts, ν^T, is equal to the sum of the within-state variance

[10]Moreover both *UP* and *Orissa* are among the poorest states with the largest primary sector income shares, above 30%.

[11]Details of this simple decomposition are given in the appendix.

and between state variance, $\nu^T = \nu^W + \nu^B$. This variance decomposition shows that there has been a similar increase in σ–divergence both within states and between states.

(Insert Figure 2 about here)

Further evidence on the pattern of Indian growth can be obtained by examining other aspects of the change in the distribution of district incomes. To that end Figure 2 plots the kernel density estimate of the probability density function (PDF) for district log incomes for 2001 and 2008.

It shows the shift in mean income; a fall in peakedness (kurtosis) with a slight increase in concentration on the left tail (skewness). Likewise the cumulative distribution function (CDF) in Figure 3 shows that each district was better off in 2008 as compared to 2001. Together these visual images suggest while the income distribution has widened at the upper tail, incomes have increased at each point on the distribution. There is significant churning within the distribution, and only 16 districts (out of 575) remain in the same position on the distribution between 2001 and 2008. Overall however Kendall's rank correlation *tau* statistic is 0.8, suggesting a high correlation of rankings between the two periods.

Thus, though there is some evidence of convergence within a few states, among most states there is no correlation between initial income and growth. Examining the country as a whole, there is evidence of β– and σ–divergence, reflecting faster growth in higher income districts with most districts experiencing growth across the entire distribution.[12]

[12]This is consistent with evidence such as Dev and Ravi (2007) and Cain et al. (2010) who use household expenditure survey data to show that inequality rose over the sub-period 1993-2004, though absolute income levels were generally also rising. Hence the pattern across households, states and districts since 2000 appears to be similar, with growth occurring in all districts but greater gains for districts in the upper end of the distribution.

3 Transitional Growth Across Districts

The preceding model of absolute β–convergence explicitly assumes that all regions within a country have the same steady state income level (Barro and Sala-i Martin 1991, Durlauf et al. 2005, Barro and Sala-i Martin 2005). This can be justified, for example, by the factor price equalization theorem, which states that free-trade and identical technologies will result in a convergence of incomes across regions. Moreover, factor mobility will result in absolute convergence, even in the absence of identical technologies.

Nevertheless even within a country the assumptions that regions will converge to the same long run per capita income level seems fragile. In particular the NEG literature, following Krugman (1991) and Krugman and Venables (1995), has emphasized the importance of barriers to trade and factor migration and agglomeration effects due to external economies. Thus, even in a regional context, there may be significant obstacles to convergence and hence long run differences in per capita incomes.

There are two natural starting points for thinking about such barriers. The first is geographic barriers such as natural barriers, trade costs, transport and migration costs and agglomeration effects (Krugman (1991)). Head and Mayer (2004) however also point to the importance of human capital, knowledge externalities, and endowments.

The second broad set of explanation lies under the general heading of institutions and policy. State governments have considerable influence on market regulation (Besley and Burgess 2004, Acharya, Baghai and Subramanian 2010). Also there are observable differences in the provision of public provision of infrastructure. Both infrastructure provision and policy differences have featured in existing discussions of India's unbalanced growth (Ghate and Wright 2012, Crost and Kambhampati 2010, Lall et al. 2010, Cain et al. 2012, Desmet et al. 2013).

3.1 Proximity to Different Markets

The standard approach to allowing for these spatial considerations is based on trade in varieties of manufactured goods, with Dixit-Stiglitz preferences, and iceberg transport costs (Fujita, Krugman and Venables (2001)). Under these assumptions it is straightforward to show that the volume of trade will depend on a weighted average of all the trade costs to all markets. This is typically approximated in empirical work by *Market Access* defined as the GDP weighted average distance to all external markets: $MA = \sum_{j \in N} w_j \, d_{i,j}$ where: $d_{i,j}$ is the distance between district i and j; Y_j is income in region j and Y is aggregate GDP (all-India), $Y = \sum_{j \in N} Y_j$; $w_j = Y_j / Y$; and N is the total number of districts.[13]

There are two limitations of the Dixit-Stiglitz setting for our purposes. The first is that it abstracts from factor mobility, particularly migration. Within India, migration across districts to the cities is likely to be one of the main engines of convergence. As explained by Hering and Poncet (2010), in a Dixit-Stiglitz setting migration reduces the impact of trade on incomes so that market access becomes a less significant determinant of incomes.

A second limitation is that the bulk of inter-state trade India is in agricultural goods (Behera 2006). Primary goods trade is typically characterized by homogeneity and competitive markets in the NEG and new trade theory, so that the importance of being close to a variety of different markets is typically assumed to apply only to non-primary goods trade (Head and Mayer 2004, Redding and Venables 2004).

Thus with trade in relatively homogeneous primary goods, various import and export markets become perfect substitutes. Likewise, since there are no legal barriers to migration, for alternative city designations, employment opportunities are perfect substitutes.

[13] This definition is used in gravity models of trade as well as 'wage equation' models that attempt to explain differences in incomes across regions as a result of trade barriers (Head and Mayer (2004) and Redding and Venables (2004)).

In this setting trade transport and migration costs will not necessarily depend on the accessibility of a number of different markets. Rather it will depend only on the markets with lowest transport trade and migration costs.

3.2 Urbanization and Urban Agglomerations (UAs)

A key insight from the NEG literature is that the combination of increasing returns and factor mobility results in a spatial concentration of economic activity, or agglomerations (Helpman and Krugman (1985), Krugman (1991), Krugman and Venables (1995), Fujita et al. (2001)). It is argued that agglomerations reflect the existence of increasing returns combined with migration and factor movements. Consequently much of the growth process, such as technology adoption and capital accumulation, occurs in cities (Ciccone and Hall (1996) and Glaeser, Kallal, Scheinkman and Shleifer (1992)).

Similarly there is a growing literature on the urbanization-growth nexus in India ((Calì and Menon 2012, Desmet et al. 2013)). For instance, Desmet et al. (2013) show that high density service clusters in India exhibit increasing concentration suggesting that they continue to benefit from agglomeration economies. They also provide evidence of the lack of agglomeration economies in medium size locations in India. Finally, they show that a lack of infrastructure and poor policy choices have held back the growth of medium density locations.[14]

The likely role of urbanization, and UAs specifically, in explaining the growth of districts in India is therefore two-fold. First, more urbanized districts will be able to benefit themselves from increasing returns and hence may be wealthier or experience faster growth. Second, however, UAs will be able to offer higher prices for exports and higher wages for migrants due to increasing returns. Hence UAs may also benefit neighboring districts

[14]This pattern also suggests a higher than normal congestion in such places, (Desmet et al. 2013). Likewise the World-Bank (2009) argues that the general pattern of growth across the world is one of increasing urbanization but eventual stability of relative spatial concentration.

through these forward linkages.

3.3 Proximity to Urban Agglomerations

The importance of UAs along with: (i) the presence of internal migration, and; (ii) production and trade focused mainly of homogeneous primary goods, suggest that we should expect each district to generate most of its economic linkages with a nearby city and few linkages with other more distant districts or cities. A key hypothesis we wish to explore is whether transport and migration costs between a district and the closest UA can influence a district's growth rate.

To consider this issue we define the variable *Minimum Distance*, D_i as the distance between district i and the closest UA. This definition is straightforward if we can succinctly define the UAs in India. In practice, however, a UA is not a well defined empirical concept and requires some subjectivity. As shown in Table 3, India has three mega-cities with populations above 10 million, *Delhi*, *Mumbai*, and *Kolkata*. Of these, *Delhi* and *Mumbai* have extended urban agglomerations – defined as areas of unbroken urbanization – that exceed 20 million. Nevertheless even the smaller cities, *Bangalore*, *Hyderabad* and *Ahmedabad*, have populations of over six million and overall, there are 10 Indian cities with urban agglomerations over three million.

(Insert Table 3 about here)

We begin therefore by initially defining the UAs as the seven largest Indian cities. This includes all cities that have populations over seven million. Hence we define $D_i = \min d_{i,j},\ j \in M$, where M={*Delhi, Mumbai, Kolkata, Chennai, Bangalore, Hyderabad, Ahmedabad*}. Then, as a robustness check we also consider alternative definitions ranging from the six to 10 largest cities listed in Table 3.[15]

[15] As we shall see, the results are very robust to these alternative definitions.

As a visual reference Figure 4 shows a map with the values of D_i for each district in India, based on the seven largest UAs. The map shows a band of relatively remote districts between *Delhi* and *Hyderabad* through *Madhya Pradesh* and *Chhattisgarh*. The remaining remote districts are located in the geographic extremities, especially the far north of *Jammu and Kashmir*, the eastern most districts of *Gujarat* and the far western districts. It can also be seen that there are clusters of less remote districts along the western corridor from *Delhi* to *Bangalore* and *Chennai*.[16]

4 The Empirical Model

Our aim is to describe the association between the spatial factors discussed above, other socio-economic factors, and growth rates across districts. The proceeding discussion suggests the following descriptive model,

$$\ln y_i(t) - \ln y_i(0) = \alpha_0 + \alpha_1 \ \ln y_i(0) \ + \ \alpha_2 \ \ln D_i + \eta \, \mathbf{X}_i \ + \ \epsilon_i \tag{2}$$

where D_i is *Minimum Distance*, $\mathbf{X}_i$ is a vector of other characteristics of region i including *Market Access* (MA), state dummy variables, and ϵ_i is a district specific random shock reflecting, for example, institutions, climate and endowments.

The inclusion of initial per capita GDP, $\ln y_i(0)$, as an explanatory variable follows the growth literature (Durlauf and Quah 1999). As shown in Appendix 2, this specification can be derived from a partial adjustment model where it is assumed that the long run steady state income level of a region is influenced by the value of D_i, as well as all the conditioning variables in $\mathbf{X}_i$.[17] Thus the coefficient α_1 can be interpreted as the

[16] This picture of a western corridor of relative urbanization is even stronger if we consider the ten largest UAs.

[17] This characterizes differences in short-run growth rates as observations along a transition path

conditional convergence coefficient. It captures the notion that a larger income gap between the i^{th} district and the UA in the initial time period will imply a faster growth rate for a given set of long run conditioning variables, *Minimum Distance*, *Market Access* and other elements of $\mathbf{X}_i$.

Finally note that a larger *Minimum Distance* is expected to negatively affect district transitional growth rates. It can also be shown that that α_2/α_1 can be interpreted as the elasticity of long run income with respect to *Minimum Distance*, (see Appendix 2).

4.1 Data

To construct *Market Access* and *Minimum Distance* we require data on the distance between various districts, $d_{i,j}$. For *Market Access* we obtain the $d_{i,j}$ from the latitude and longitude coordinates of each of district's headquarters. We then use these coordinates to construct a 575×575 matrix of district to district distances.[18] Likewise for *Minimum Distance* we use the minimum distance from one district headquarters to another.[19]

As noted above *Per-capita GDP* is the logarithm of district per-capita GDP in 2001. The other elements of the conditioning vector $\mathbf{X}_i$ are district and non-district socio-economic indicators. Specifically, the variables used are defined as follows: *Literacy* is

between an initial income level and a target long run steady state level or long run equilibrium "target". It is used for example by Krugman (1993) in a regional context but is commonly used as a motivation for cross-country empurple studies (Durlauf and Quah 1999). Though it is not assumed that each region has reached its long run or steady state income level, it is nevertheless shown that these long run values, in conjunction with initial income, $y_i(0)$, will determine the regions speed of convergence along its transitional growth path.

[18]The coordinates are obtained from http://www.gps-coordinates.net/ and converted to radians. The distances in kilometers are then calculated using the Haversine formula. The 575×575 matrix of district to district distances is available from the authors on request.

[19]We obtain data from *Google Maps* and a variety of other sources including Indian state tourism data.

the logarithm of the total literacy rate per hundred people; *Electricity* is the logarithm of the percentage of households with an electricity connection; *Commercial Banks* is the logarithm of the number of commercial banks per thousand people; *Urbanization* is the logarithm of the percentage of urban households in a given district, i.e. it is a measure of initial urban population; *Urbanization Squared* is the squared value of *Urbanization*; *Market Access*, as defined above, is the weighted average of trade costs to all markets; *Irrigated Land* is the logarithm of net irrigated land area (per million people) divided by the district population; *Pucca Road* is the logarithm of the percentage of households connected by pucca (hard) roads; *Metro Electricity* is the logarithm of the percentage of urban households with an electricity connection in the closest UA; *Metro Urbanization* is the logarithm of the percentage of urban households in the closest UA district; and *Metro Literacy* is the logarithm of the total literacy rate per hundred persons in the closest metropolitan districts. All variables refer to the initial (2000-01) level.

Finally we also include we also include state dummy variables, given by *State*. Summary statistics for the key variables of interest are given in Table 4.

(Insert Table 4 about here)

5 Results

In what follows we estimate equation (2) using our cross-section of Indian districts.[20] The results for our baseline model, equation (2), are given in Table 5a. The regression results in columns (1)-(3), (6) and (7) include the variable *Minimum Distance*. In column

[20]A visual inspection of the data suggests the presence of heteroscedasticity and the Breusch-Pagan (BP) test for heteroscedasticity on preliminary OLS results confirms this. As the form of heteroscedasticity is unknown, the application of GLS is not feasible. The implication of heteroscedasticity is that OLS will result in biased standard errors and tests based on these standard errors will be invalid. In what follows we therefore use White's (1982) robust standard errors to obtain valid inferences, even though efficiency is sacrificed.

(4) however we replace *Minimum Distance* by *Market Access*, and in column (5) we exclude both of these spatial variables. All regressions include the variable *Per-capita GDP*; otherwise the regressions differ by the number of additional explanatory variables included.

(Insert Tables 5a and 5b about here)

It can be seen that the sign of the convergence coefficient, β, is significant and negative across all models except in column (1). Specifically allowing for various observable characteristics across districts results in a finding of conditional convergence. For example in column (2) we find that once we control for differences in *Minimum Distance*, *Urbanization* and different States, districts that were initially poorer grew faster. This suggests that there is conditional convergence across districts where each district is converging to a particular level of long run wealth determined by *Minimum Distance*, *Urbanization* and *State*. The significance of the convergence coefficient, β, thus provides strong support for our basic partial adjustment model.

As expected we find that an increase in *Minimum Distance* reduces the transitional growth rate. This result is very robust across all the regressions and the coefficient is very stable with an elasticity of transitional growth with respect to D_i of -0.004. We discuss the implications of the size of this coefficient below.

Second, for a given initial income level, urban areas should also have faster transitional growth. This is verified in column (2) with *Urbanization* being positive and significant.

The final explanatory variable in column (2) is the vector of state dummy variables, *State*. It can be seen that the F-test of joint significance of the state dummy variables is highly significant across all of the various models. Since states have some autonomy with respect to laws and taxation, the *State* dummy variables reflect differences in institutions and governance. However state variables may also capture differences in climate, endowments and geography.

In column (3) we introduce a range of other possible explanatory variables. It can be seen that *Electricity* and *Irrigated land* are significant at the 1% level. Since electricity supply is government controlled this result suggests that public infrastructure is also important in understanding differences across districts.

We find that *Irrigated land* is also a significant explanatory variable and has a negative sign. This suggests that more irrigated land per capita is associated with slower transitional growth. There are a number of possible explanations for this. One possibility is that irrigated land is associated with rural districts that have high levels of home production so that market income understates actual income.[21] The other explanatory variables, *Commercial banks* and *Pucca Roads* are found to be insignificant. It can be seen that with the addition of these variables *Minimum Distance* remains significant at the 5% level.

In column (4) we drop *Minimum Distance* and include *Market Access*. As described above *Market Access* is the standard spatial variable used to describe the impact of trade costs on international trade flows. It can be seen however that *Market Access* is not significant. As discussed above, because there is migration across district borders and most internal trade in India is in primary goods, this result is not unexpected.

Next in column (5) we consider the effect of excluding both spatial variables. It can be seen that there is little change in the coefficients of the remaining parameters suggesting that there is little bias associated with this omission.

In column (6) we consider whether there is evidence of nonlinear effects of urbanization ((Krugman and Venables 1995, Bloom, Canning and Fink 2008, Glaeser 2011, Calì and Menon 2012, Desmet et al. 2013)). Thus in column (6) we consider an interaction term between *Urbanization* (in 2001) and initial income as well as *Urbanization squared* (in

[21]It is also possible that irrigated land simply captures more rural agricultural districts thus having the opposite sign to *Urbanization*.

2001). It can be seen however that these additional terms are insignificant. The main effect of including these variables is to increase the point estimate of the coefficient on per–capita GDP, which is simply due to the inclusion of per capita GDP in the interaction term with *Urbanization*.

Finally in column (7) we considers all of the variables including *Market Access*. It shows that the results are quite stable with little change in the sign or significance of any variables.

5.1 Characteristics of Urban Agglomerations

The results suggest that we can attribute some of the regional disparities in growth rates to differences in district characteristics, including the *Minimum Distance* to a UA. But in our discussion of *Minimum Distance* we implicitly assumed that all UAs are equivalent. Nevertheless UAs may differ in important ways and this may affect the potential growth of neighboring districts.

In Table 5b we therefore consider additional explanatory variables that relate to each district's closest UA. One way to interpret this is that the different UAs may have different long run incomes (see Appendix 2 for an analytical discussion of this point).

It can be seen that the additional explanatory variables are all insignificant with the exception of the literacy rate in the UA (*Metro Literacy)*. This is positive and significant across all models suggesting that districts are advantaged by being closer to cities where literacy is higher. This result is consistent with arguments in the NEG literature that emphasize the role of external economies that give rise to UAs and the possibility of complementarities between these external economies and human capital, (Head and Mayer 2004).

Including the additional UA district characteristics however has little effect on the estimated coefficients of the other variables which again tend to be very stable across all specifications in Tables 5a and 5b. Hence the conclusions on the low rate of convergence are robust to the inclusion of the UA district characteristics. Likewise *Minimum Distance* continues to be statistically significant in all specifications of the models.

5.2 Convergence

We have found that while there is absolute divergence across districts, there is also conditional convergence once we control for a few characteristics including *Minimum Distance*. Nevertheless though we have found strong evidence of conditional convergence, the estimated value of β = -0.85% to -1.3% is much slower than the values found in the growth literature across a wide array of counties.[22] The estimate of -1.3% (Column 3, Table 5b), for example, implies that the gap between each district's current income level, and its long run or steady state income level, is halved only every 62 years. At this rate, at the end of a decade, a per capita income gap between two districts would still be 90 percent of the gap that existed at the start of the decade. Thus the forces of convergence, or "trickle down", appear to be very weak across Indian districts.

For models where we include the interaction term, *Initial per-capita Income* $\times$ *Urbanization*, the joint significance test for per-capita income evaluated at the mean level of *Urbanization* is highly significant. This suggests that the strength of convergence is smaller in more urbanized districts.[23] The low convergence rate in more urbanized districts is consistent with less evidence of diminishing returns in more urbanized areas. It suggests that the pattern of absolute divergence and weak conditional convergence is mainly driven by growth in the more urbanized districts.

[22]For example it is is roughly half of Barro's " iron law of convergence", (Sala-i Martin 1996, Sala-i Martin 1997, Barro 2012). Nevertheless it should be noted that the conditional convergence model used here is quite different from the standard cross-country model.

[23]For the most rural area (least urbanized) the convergence coefficient estimate is -2.4%. For the most urbanized district, the convergence coefficient is very small (0.6%). The convergence effect estimated at the mean value *Urbanization* is -1.3%, similar to the preceding results.

5.3 The Impact of Proximity to Urban Agglomerations

To what extent does *Minimum Distance* matter for understanding differences in growth and incomes across India? The coefficient α_2 gives the impact on the transitional growth rate and ranges from approximately -0.003 to -0.005. This value is the partial effect of a one-percentage point change in the minimum distance to one of the seven UAs, on the growth rate, $\alpha_2 \equiv \partial \ln(y(t)/y(0))/\partial \ln D$. The estimate of α_2 implies that a district that is twice as remote will have a transitional growth rate that is 0.20 to 0.35 percentage points lower than the closer district.[24]

The most remote district in our data is *Tamenglong*, in Manipur, which is a mountainous district near the Burmese border and is 2531 kilometers from *Kolkata*, the nearest UA. At the other end of the spectrum the district *South 24 Parganas* is only 7.9 kilometers from *Kolkata*. For this maximum distance the more remote district would have a transitional growth rate that is 1.7 to 2.9 percentage points lower. Thus *Minimum Distance* variable has an economically important effect on observed transitional growth rates for the very remote regions.

It is also useful to think of the impact of *Minimum Distance* in terms of income levels. That is, given these differences in transitional growth rates, what would be the implication in the long term for inequality across regions? As discussed in Appendix 2, the empirical model (2) has the form of a standard partial adjustment model. This allows us to interpret the ratio of coefficients as the elasticity of *Minimum Distance* with respect to long run equilibrium differences in per capita incomes, $\gamma = -\alpha_2/\alpha_1$.

The value of γ is reported for each model in Tables 5a and 5b, along with a joint significance test. It can be seen that the estimates of γ are significant at the 1% level across

[24]This follows since Since $\ln 2 = 0.69$. The distance to a UA in the sample is 532km with a standard deviation of just under 400km. So doubling the distance is approximately equal to increasing the distance by one standard deviation from the mean.

all models with a value ranging from approximately -0.25 to -0.57.

Suppose we consider the most conservative estimate of the gravity parameter of $\gamma = -0.25$ This means, for example, that if a more isolated district, i, is twice the distance from the UA than a closer district, j, $D_i/D_j = 2$, then this implies that the more remote district will have a steady state income level that is approximately 84% of the closer district.[25] The negative effect of distance on transitional growth is supported by other papers in the literature, such as (Calì and Menon 2012), who highlight the positive spill-overs from urban growth to nearby rural areas.

However at the maximum difference in remoteness, of 320:1, we would expect the more remote district to have a per-capita income level of only 24% of the closer district. Thus the distance coefficient suggests quite a large impact on income levels for very remote districts but relatively modest effects for districts that are within a range of twice or half the average distance.

In sum, the significance of the variables in Tables 5a and 5b sheds some light on the observed pattern of divergence across India. First we have found that even in the absence of policy failures, we should find some significant variation in growth rates, and long run incomes, across districts due to spatial factors. This supports our conjecture that transport costs impose important regional constraints on development. The results suggest that this effect is particularly important in understanding the reasons for low growth rates in very remote districts, but only has a modest impact on most districts.

Likewise we find that differences in urbanization rates also explain much of the variation in district incomes in line with the NEG theory that emphasizes the role of external economies that give rise to UAs.

[25] That is $y_i^*/y_j^* = (D_i/D_j)^\gamma = 0.84$. Alternatively if a more remote district had a long run equilibrium income level that is approximately 50% of the closer districts, it would be 16 times further from the center.

In addition our analysis points to public infrastructure, as indicated by electrification, as being an important determinants of long run district income levels. This is a useful starting point in considering potential policy responses to address the unbalanced nature of India's growth. It suggests that absolute convergence will depend on increasing equality in public investment. Divergence in growth rates may be mitigated through improving infrastructure investment, in low growth regions. This potential policy implication is underscored by the fact that the state dummy variables are also highly significant. There are significant policy differences at the state level, particularly with respect to labor laws, (Besley and Burgess 2004, Acharya et al. 2010).

6 Robustness

6.1 Parameter stability

As a robustness test we then extend our definition of a metropolitan district, or UA, to include the 10 largest UAs in India by population as in Table 3.[26] The overall conclusion is also robust to those UAs with very little change in significance of the key variables or the estimated size of the coefficients. Second we consider whether our distance variable is stable across different data sets. To do this we divide the whole data set into several subgroups, and then examine the stability of the model parameters. To this end, we re-estimate equation (2) and (8) but drop several districts. Specifically, we first drop all north-east districts, then all districts from *Bihar* and *Maharashtra*. We also drop other observations; the various alternatives we consider are listed in Table 6a and 6b.

(Insert Tables 6a and 6b about here)

[26]Because of space constraints, we do not include these results, but they are available from the authors on request.

A stability test is then conducted by using interaction dummy variables, where the dummy variable takes the value one for included districts and takes the value zero for excluded districts. Then we examine whether such interaction dummies are significant or not based on a F-test. The results are depicted in Tables 6a and 6b.

All the parameters, including the distance variable are found to be very stable across the data subsets, as shown in Tables 6a and 6b where the estimated p-values for the F-tests (given in parenthesis) are significantly larger than 0.05. Thus we do not reject the null hypothesis of constant coefficients. Hence this test indicates there is no evidence that the parameters change across the subsets of the data districts.[27]

6.2 Endogeneity

We also consider the potential for the explanatory variables to be endogenous, leading the OLS estimates to be biased and inconsistent. To investigate this, we first apply the Hausman test by comparing 2SLS and the OLS estimates.[28] The Hausman tests are negative for all these cases, which is not unexpected since, as discussed above, there is evidence that our data are strongly heteroscedastic, invalidating the use of the Hausman test.

We therefore compare the equality of two parameter vectors (OLS and 2SLS) in a SUR setting.[29] Tables 7a and 7b provide results of this endogeneity test for three variables:

[27]We examine parameter stability for the genuine regressors excluding the intercept and the state dummy variables. Note also that it is important that these subsets of the full data set are selected in a random fashion. For example creating subsets of the data based on different income groups would introduce a sample selection problem.

[28]For 2SLS, the identifying variables we use are the percentage of household with telephones, percentage of people below the poverty line and female literacy rates

[29]Stata provides an indirect test for endogeneity in a seeming unrelated regression (SUR) framework. The *Suest* command in Stata compares the equality of two parameter vectors (OLS and 2SLS) in a SUR setting. The test statistic follows a χ^2 distribution with the number of model parameters as the

District Literacy, *District Pucca Roads*, and *District Urbanization*, which are included as controls in Tables 5a and 5b. We have also tested the exogeneity status of the Metro variables included in Tables 5a and 5b. The SUR based tests provide some support for exogeneity.[30]

(Insert Tables 7a and 7b about here)

Endogeneity can also occur through the variable *Market Access* because of the presence of the district GDP weights $w_j = Y_j/Y$ ($d_{i,j}$, which is the distance in kilometers in a straight line between district i and j is always exogenous). This is confirmed by the Hausman test.

We first note that the variable *Market Access* is always insignificant no matter what specification we use in the presence of the variable, *Minimum Distance.* For instance, in column (7) in Table 5a, *Market Access* is insignificant even at the 10% level. Controlling for UA characteristics (in Table 5b), *Market Access* continues to be insignificant.

While these results suggest that endogeneity due to the introduction of *Market Access* may not be a problem, we re-run the model proxying for district GDP weights $w_j = Y_j/Y$ by 2001 district population weights. We find that *Market Access* continues to be insignificant when included along with *Minimum Distance* and the other variables (i.e., irrepective of whether we control just for district characteristics, or district and metro characteristics). Next, we use the IV approach to confirm that endogeneity is not a problem. We use the variable, *Minimum Distance* and other variables listed in Tables 7a to instrument for *Market Access.* Under IV, we find that instrument for *Market Access*

degrees of freedom

[30]Specifically they do not reject exogeneity. There is one exception however: we find that the variable, Pucca Road, is endogenous, although insignificant. To correct for this, we use an instrumental variable (IV) approach. Under IV, we find that the coefficient estimate for Pucca Road continues to be insignificant suggesting that endogeneity through this variable is not a concern.

is always insignificant in the presence of the variable *Minimum Distance* not only when we control for district characteristics but also district and metro characteristics.[31]

7 Conclusion

Though India's growth has been unbalanced, the causes of this pattern of divergence are not well understood. We therefore examine the evidence for convergence of per-capita incomes at the district level using a new data set of district per-capita incomes and socio-economic characteristics. We find little evidence of convergence either within States or across all districts as a whole. Rather there is absolute divergence of income levels across districts.

We therefore attempt to explain differences in transitional growth across districts with reference to district characteristics and initial district per capita incomes. We argue that an important spatial variable to consider in the case of India, is the the the district's proximity to a major urban agglomeration. This follows from the NEG literature which emphasizes the importance of urban agglomerations and increasing returns, and the fact that, for migrants, the largest cities are likely to be close substitutes in terms of employment opportunities.

We find that urbanization and electrification are significantly associated with higher transitional growth rates across all our models. Thus the results support Desmet et al. (2013) who argue that frictions, policies, and a general lack of infrastructure in medium-density cities is preventing the spread of growth in India. Likewise we find that the state dummy variables are jointly significant. This supports studies that have emphasized the role of different degrees of regulation across states (Besley and Burgess 2004, Acharya et al. 2010).

[31] These results are available from the authors on request.

We also find that the proximity of a district to major UA is also a highly significant explanatory variable across our various models. Thus the model is a capable of explaining much lower growth rates in very remote districts. Notably, however, we also find that the more conventional market access variable used in the NEG literature - which gives more weight to the number of different markets - is not significant in any of our models. This makes sense in the Indian regional context where manufactured goods trade that depends on varieties, is very small.

Thus geographical factors, particularly urbanization and proximity to a large urban agglomeration, are found to be very important. Nevertheless we also find evidence that some factors associated with the policy setting, such as electrification and differences across States, are also important in understanding the differences in growth across India's districts.

References

Acharya, Viral V., Ramin P. Baghai, and Krishnamurthy V. Subramanian (2010) 'Labor laws and innovation.' Technical Report 16484, NBER Working Paper

Azam, Mehtabul (2012) 'Changes in wage structure in urban india, 1983–2004: A quantile regression decomposition.' *World Development* 40(6), 1135–1150

Bandopadhyay, S. (2004) 'Twin peaks-distribution dynamics of economic growth across indian states.' In *Growth, inequality and poverty: prospects for pro-poor growth,* ed. A. Shorrocks and R. van der Hoeven (New York: Oxford University Press)

—— (2012) 'Convergence club empirics: Evidence from indian states.' *Research in Economic Inequality (forthcoming)*

Bardhan, Pranab (2010) *Awakening Giants: Feet of Clay* (Princeton N.J.: Princeton University Press)

Barro, R.J., and X Sala-i Martin (1991) 'Convergence Across States and Regions.' *Brookings Papers on Economic Activity* 1(1), 107–182

Barro, Robert J. (2012) 'Convergence and modernization revisted.' *Paper Presented at the Nobel Symposium on Growth and Development, Stockholm, September 3-5, 2012*

Barro, Robert J., and J. Sala-i Martin (2005) *Economic Growth* (Singapore: McGraw Hill)

Basu, Kaushik, and Annemie Maertens (2009) 'The growth of industry and services in south asia and its impact on employment.' In 'Accelerating Growth and Job Creation in South Asia' (New Delhi: Oxford University Press) pp. 81–140

Baumol, William J. (1986) 'Productivity growth, convergence, and welfare: What the long run data show.' *American Economic Review* 76(5), 1027–1085

Behera, T. (2006) 'India's internal trade: A review of interstate movement of major commodities.' *Vaanijya* pp. 43–63

Besley, Timothy, and Robin Burgess (2004) 'Can labor regulation hinder economic performance? evidence from india.' *The Quarterly Journal of Economics* 100(6), 91–134

Bhagwati, Jagdish, and Arvind Panagariya (2013) *Why Growth Matters: How Economic Growth in India Reduced Poverty and the Lessons for Other Developing Countries* (Washington DC: PublicAffairs, Council of Foreign Relations)

Bhandari, Laveesh (2009) 'Socio-economic performance of constituencies: A response.' *Economic and Political Weekly* 44(40), 61–63

Bloom, David E, David Canning, and Günther Fink (2008) 'Urbanization and the wealth of nations.' *Science* 319(5864), 772–775

Cain, J, Rana Hasan, and Devashish Mitra (2012) 'Trade liberalization and poverty reduction: New evidence from indian states.' In *India's Reforms: How They Produced Inclusive Growth,* ed. Jagdish Bhagwati and Arvind Panagariya (N.Y.: Oxford University Press) chapter 4, pp. 91–169

Cain, J Salcedo, Rana Hasan, Rhoda Magsombol, and Ajay Tandon (2010) 'Accounting for inequality in india: Evidence from household expenditures.' *World Development* 38(3), 282–297

Calì, Massimiliano, and Carlo Menon (2012) 'Does urbanization affect rural poverty? evidence from indian districts.' *The World Bank Economic Review* p. lhs019

Cashin, Paul, and Ratna Sahay (1996) 'Internal migration, center-state grants, and economic growth in the states of india.' *International Monetary Fund Staff Papers* pp. 123–171

Chamarbagwala, Rubiana (2008) 'Economic liberalization and wage inequality in india.' *World Development* 34(12), 1997–2015

Chaudhuri, Shubham, and Martin Ravallion (2007) 'Partially awakened giants: Uneven growth in china and india.' In *Dancing With Giants: China, India, And The Global Economy,* ed. L. A. Winters and Shahid Yusuf (Washington State DC: The World Bank)

Ciccone, Antonio, and Robert E Hall (1996) 'Productivity and the density of economic activity.' Technical Report, National Bureau of Economic Research

Crost, Benjamin, and Uma S Kambhampati (2010) 'Political market characteristics and the provision of educational infrastructure in north india.' *World Development* 38(2), 195–204

Das, Samarjit (2012) 'The convergence debate and econometric approaches: Evidence from india.' In *The Oxford Handbook of the Indian Economy,* ed. C. Ghate (New York: Oxford University Press) pp. 766–782

Datt, Gaurav, and Martin Ravallion (2002) *Is India's Economic Growth Leaving the Poor Behing?,* vol. 2846 (World Bank-free PDF)

Desmet, Klaus, Ejaz Ghani, Stephen O'Connell, and Esteban Rossi-Hansberg (2013) 'The spatial development of india.' *Journal of Regional Science*

Dev, S Mahendra, and C Ravi (2007) 'Poverty and inequality: all-india and states, 1983-2005.' *Economic and Political Weekly* pp. 509–521

Dreze, Jean, and Amartya Sen (2013) *An Uncertain Glory: India and its Contradictions* (Allen Lane)

Durlauf, S. N., P. A. Johnson, and J. R. W. Temple (2005) 'Growth econometrics.' In *Handbook of Economic Growth,* ed. P. Aghion and Durlauf S. N., vol. 1A (North-Holland: Amsterdam) pp. 555–677

Durlauf, Steven N, and Danny T Quah (1999) 'The new empirics of economic growth.' *Handbook of macroeconomics* 1, 235–308

Fujita, Masahisa, Paul R Krugman, and Anthony J Venables (2001) *The spatial economy: Cities, regions, and international trade* (MIT press)

Ghani, Ejaz, Arti Grover Goswami, and William R Kerr (2012) 'Is india's manufacturing sector moving away from cities?' Technical Report, National Bureau of Economic Research

Ghate, Chetan (2008) 'Understanding divergence in india: A political economy approach.' *Journal of Economic Policy Reform* 11(1), 1–9

Ghate, Chetan, and Stephen Wright (2012) 'The v-factor: Distribution, timing and correlates of the great indian growth turnaround.' *Journal of Development Economics* 99(1), 58–67

Glaeser, Edward (2011) 'Cities, productivity, and quality of life.' *Science* 333(6042), 592–594

Glaeser, Edward L, Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer (1992) 'Growth in cities.' *Journal of Political Economy* 100(6), 1126–52

Government of India (2013) 'Press information bureau.' http://pibmumbai.gov.in/scripts/detail.asp?releaseId=E2011IS3

Head, Keith, and Thierry Mayer (2004) 'The empirics of agglomeration and trade.' *Handbook of regional and urban economics* 4, 2609–2669

Helpman, E., and Paul R Krugman (1985) *Market Structure and Foreign Trade* (Cambridge MA: MIT press)

Hering, Laura, and Sandra Poncet (2010) 'Income per capita inequality in china: the role of economic geography and spatial interactions.' *The World Economy* 33(5), 655–679

Himanshu (2009) 'Electoral politics and the manipulation of statistics.' *Economic and Political Weekly* 44(19), 31–35

Kar, Sabyasachi, Debajit Jha, and Alpana Kateja (2011) 'Club-convergence and polarization of states: A nonparametric analysis of post-reform india.' *Indian Growth and Development Review* 4(1), 53 – 72

Krishna, Pravin, and Guru Sethupathy (2012) 'Trade and inequality in india.' In *India's Reforms: How They Produced Inclusive Growth,* ed. Jagdish Bhagwati and Arvind Panagariya (N.Y.: Oxford University Press) chapter 6, pp. 247–278

Krugman, Paul (1993) 'On the number and location of cities.' *European Economic Review* 37(2), 293–298

Krugman, Paul, and Anthony J Venables (1995) 'Globalization and the inequality of nations.' *The Quarterly Journal of Economics* 110(4), 857–880

Krugman, Paul R. (1991) 'Increasing returns and economic geography.' *Journal of Political Economy* 99(1), 483–499

Lall, Somik V, Hyoung Gun Wang, and Uwe Deichmann (2010) 'Infrastructure and city competitiveness in india.' Technical Report, Working paper//World Institute for Development Economics Research

Mishra, Prachi, and Utsav Kumar (2005) 'Trade liberalization and wage inequality: Evidence from india.' *IMF Working Paper WP0520*

Rao, G., R. Shand, and K. Kalirajan (1999) 'Convergence of incomes across indian states: A divergent view.' *Economic and Political Weekly* 17(13), 769–778

Rao, M Govinda, and Kunal Sen (1997) 'Internal migration, center-state grants, and economic growth in the states of india: A comment on Cashin and Sahay.' *Staff Papers-International Monetary Fund* 44(2), 283–288

Redding, Stephen, and Anthony J Venables (2004) 'Economic geography and international inequality.' *Journal of international Economics* 62(1), 53–82

Sachs, Jeffrey D (2009) 'South asia story of development opportunities and risks.' In *Accelerating Growth and Job Creation in South Asia,* ed. Ejaz Ghani and Ahmed Sadiq (New Delhi: Oxford University Press) pp. 42–49

Sala-i Martin, Xavier (1997) 'I just ran two million regressions.' *American Economic Review* 87(2), 178–83

Sala-i Martin, Xavier X. (1996) 'Regional cohesion: Evidence and theories of regional growth and convergence.' *European Economic Review* 40(6), 1325–1352

Trivedi, K. (2003) 'Regional convergence and catch-up in india between 1960 and 1992.' *Mimeo, Oxford University Economics Department*

World-Bank (2009) *World Development Report: Reshaping economic geography* (Washington DC: World Bank)

Table 1: State data and β–Convergence Coefficients

State	Pop (Millions)	Per-Capita GDP Rs 000's 2007-08	Share Primary Sector %	β	p-value
All India	1,137.1	38	21	0.0107***	-0.0019
Andhra Pradesh	82.2	38	29	-0.0032	-0.0069
Arunachal Pradesh	1.2	34	26	-0.0134	-0.0345
Assam	29.3	24	35	-0.0332***	-0.0091
Bihar	95.6	11	25	-0.0068	-0.0138
Chhattisgarh	23.2	33	24	0.0188**	-0.008
Goa	1.5	108	14	na	na
Gujarat	55.9	52	19	0.0012	-0.0057
Haryana	23.8	62	21	0.0333*	-0.0114
Himachal Pradesh	6.5	49	22	0.0081	-0.0308
Jammu & Kashmir	11.0	29	27	0.0047	-0.0098
Jharkhand	30.2	23	22	0.0304	-0.0179
Karnataka	56.7	38	19	0.0102	-0.0091
Kerala	33.8	48	17	-0.0391*	-0.0206
Madhya Pradesh	69.0	20	33	-0.0005	-0.0096
Maharashtra	107.1	53	13	0.0119*	-0.0065
Manipur	2.4	24	26	-0.0009	-0.0184
Meghalaya	2.5	30	27	0.0102	-0.0164
Mizoram	1.0	34	15	0.0176	-0.013
Nagaland	2.2	33	34	-0.0157	-0.0305
Orissa	39.7	26	31	0.0492***	-0.0085
Punjab	26.4	52	31	-0.0054	-0.0298
Rajasthan	64.1	26	28	-0.0338***	-0.0123
Sikkim	0.6	40	18	na	na
Tamil Nadu	66.0	44	14	0.0089	-0.0092
Tripura	3.5	33	24	na	na
Uttar Pradesh	189.3	18	31	0.0133***	-0.0046
Uttaranchal	9.4	36	20	0.008	-0.0147
West Bengal	86.4	35	23	0.0033	-0.065

Note 1: *, **, *** denotes 10, 5 and 1 percent levels of significance respectively.
Note 2: Robust (White) standard errors are used.
Note 3: Union Territories are Excluded.
Note 4: n.a. indicates not enough observations available to estimate β

Table 2: Decomposition of σ–Convergence

	Variance	Between State Variance	Within State Variance	Skewness	Kurtosis	Gini
2001	0.27	0.15	0.12	0.15	3.09	0.0307
2008	0.35	0.20	0.15	0.16	2.88	0.0322
Change	0.08	0.05	0.03			

Table 3: Metropolitan Districts

Extended Urban Agglomeration	Population 2011 (Millions)
Delhi	21,753,486
Greater Mumbai	20,748,395
Kolkata	14,617,882
Chennai	8,917,749
Bangalore	8,728,906
Hyderabad	7,749,334
Ahmedabad	6,352,254
Pune	5,049,968
Surat	4,585,367
Jaipur	3,073,350

Source: Government of India (2013)

Table 4: Descriptive Statistics

	Mean	Variance	Minimum	Maximum	Skewness
Initial Per-capita GDP	9.583	0.274	8.243	11.313	0.148
Minimum Distance	6.004	0.671	2.067	8.018	-1.091
Market Access	7.039	0.036	6.776	7.796	.7557
Literacy	4.131	0.046	3.408	4.570	-0.750
Electricity (%)	3.776	0.578	1.131	4.588	-1.212
Commercial Banks	-9.698	0.175	-11.194	-8.227	0.500
Urbanization	2.870	0.565	0.279	4.605	-0.199
Irrigated Land	-3.253	1.163	-7.782	-1.139	-0.980
Pucca Road	3.968	0.617	-1.204	4.605	-3.063
Metro Electricity	4.557	0.000	4.543	4.583	1.367
Metro Urbanization	4.579	0.001	4.479	4.605	-1.638
Metro Literacy	4.412	0.001	4.367	4.459	0.064

Note: See the text for a description of all variables.

Table 5a: Transitional Growth Results

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Initial Per-capita GDP	0.0018	-0.0082**	-0.0129***	-0.0112***	-0.0112***	-0.0249**	-0.0238**
	(0.0030)	(0.0037)	(0.0042)	(0.0042)	(0.0042)	(0.0120)	(0.0119)
Mn. Distance to UA	-0.0055***	-0.0047***	-0.0040**			-0.0039**	-0.0039**
	(0.0015)	(0.0015)	(0.0016)			(0.0016)	(0.0016)
Market Access				0.0176			0.0160
				(0.0157)			(0.0156)
Urbanization		0.0103***	0.0067***	0.0069***	0.0066***	-0.0363	-0.0333
		(0.0021)	(0.0022)	(0.0022)	(0.0022)	(0.0313)	(0.0310)
Literacy			-0.0081	-0.0073	-0.0082	-0.0079	-0.0072
			(0.0087)	(0.0085)	(0.0086)	(0.0086)	(0.0086)
Electricity			0.0123***	0.0129***	0.0135***	0.0136***	0.0131***
			(0.0045)	(0.0045)	(0.0044)	(0.0047)	(0.0047)
Commercial Banks			0.0053	0.0025	0.0032	0.0042	0.0035
			(0.0049)	(0.0050)	(0.0048)	(0.0049)	(0.0050)
Irrigated Land			-0.0036***	-0.0037***	-0.0039***	-0.0030**	-0.0028**
			(0.0013)	(0.0012)	(0.0012)	(0.0013)	(0.0013)
Pucca Road			0.0020	0.0019	0.0020	0.0017	0.0016
			(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Urban. x Initial per-capita GDP						0.0040	0.0036
						(0.0038)	(0.0038)
Urbanization Squared						0.0009	0.0011
						(0.0020)	(0.0020)
Metro Electricity							
Metro Literacy							
Metro Urbanization							
Constant	0.0975***	0.1521***	0.1956**	-0.0001	0.1322*	0.3029**	0.1744
	(0.0358)	(0.0385)	(0.0861)	(0.1489)	(0.0820)	(0.1289)	(0.1785)
Gravity Parameter	3.0196	-0.5696***	-0.3087***			-0.1585***	-0.1657***
	(20.6305)	(0.0188)	(0.0136)			(0.0206)	(0.0214)
F Test *State*	24.23***	32.62***	10.30***	11.26***	11.28***	10.56***	10.44***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
F Test *Per-capita GDP*						12.29***	12.35***
						0.0005	0.0004
BP Test	0.47	2.40	6.00**	5.93**	6.79***	6.56**	5.85**
	0.4921	0.1210	0.0143	0.0149	0.0092	0.0104	0.0155
Observations	566	556	544	548	548	544	544
R-squared	0.3354	0.3841	0.4031	0.3994	0.3980	0.4066	0.4078

Note 1: *, **, *** denotes 10, 5 and 1 percent levels of significance respectively.
Note 2: Robust (White) standard errors are used.
Note 3: F Test *State* are joint tests for State dummy variables.
Note 4: F Test *Per-capita GDP* are joint tests for significance of *Per-capita GDP* +*Per-capita GDP* * *Urbanization*.

Table 5b: Transitional Growth Results including Metro Variables

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Initial Per-capita GDP	0.0014	-0.0085**	-0.0130***	-0.0116***	-0.0116***	-0.0250**	-0.0248**
	(0.0030)	(0.0037)	(0.0042)	(0.0042)	(0.0042)	(0.0122)	(0.0121)
Mn. Distance to UA	-0.0049***	-0.0039***	-0.0033**			-0.0032**	-0.0033**
	(0.0016)	(0.0015)	(0.0016)			(0.0016)	(0.0016)
Market Access				0.0016			0.0026
				(0.0168)			(0.0167)
Urbanization		0.0103***	0.0067***	0.0068***	0.0067***	-0.0356	-0.0351
		(0.0020)	(0.0022)	(0.0022)	(0.0021)	(0.0317)	(0.0315)
Literacy			-0.0057	-0.0057	-0.0058	-0.0056	-0.0055
			(0.0093)	(0.0092)	(0.0092)	(0.0093)	(0.0093)
Electricity			0.0121***	0.0130***	0.0130***	0.0135***	0.0134***
			(0.0045)	(0.0046)	(0.0045)	(0.0047)	(0.0048)
Commercial Banks			0.0038	0.0019	0.0019	0.0028	0.0027
			(0.0049)	(0.0049)	(0.0048)	(0.0049)	(0.0051)
Irrigated Land			-0.0036***	-0.0039***	-0.0039***	-0.0030**	-0.0030**
			(0.0012)	(0.0012)	(0.0012)	(0.0012)	(0.0012)
Pucca Road			0.0019	0.0019	0.0019	0.0016	0.0016
			(0.0025)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Urban. x Initial per-capita GDP						0.0040	0.0039
						(0.0039)	(0.0039)
Urbanization Squared						0.0008	0.0008
						(0.0020)	(0.0020)
Metro Electricity	0.2349	0.2586	0.2537	0.2505	0.2510	0.2397	0.2388
	(0.1602)	(0.1600)	(0.1731)	(0.1750)	(0.1742)	(0.1729)	(0.1738)
Metro Literacy	0.0966*	0.1084**	0.0998*	0.1226**	0.1242**	0.0990*	0.0964*
	(0.0500)	(0.500)	(0.0517)	(0.0529)	(0.0505)	(0.0517)	(0.0542)
Metro Urbanization	-0.0837*	-0.0704	-0.0730	-0.0740	-0.0750	-0.0722	-0.0706
	(0.0485)	(0.0473)	(0.0485)	(0.0497)	(0.0492)	(0.0483)	(0.0489)
Constant	-1.0146	-1.1826	-1.0914	-1.2384	-1.2384	-0.9202	-0.9326
	(0.7672)	(0.7711)	(0.8416)	(0.8402)	(0.7571)	(0.8576)	(0.8540)
Gravity Parameter	3.4018	-0.4596***	-0.2509***			-0.1294***	-0.1311***
	(0.2078)	(0.0180)	(0.0137)			(0.0208)	(0.0209)
F Test *State*	58.98***	27.04***	10.29***	11.54***	11.66***	10.70***	10.26***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
F Test *Per-capita GDP*						12.48***	12.46***
						0.0005	0.0004
BP Test	0.62	2.58	6.07**	6.26**	6.32**	6.55**	6.46**
	0.4297	0.1081	0.0137	0.0123	0.0119	0.0105	0.0110
Observations	566	556	544	548	548	544	544
R-squared	0.3436	0.3928	0.4109	0.4084	0.4084	0.4142	0.4142

Note 1: *, **, *** denotes 10, 5 and 1 percent levels of significance respectively.
Note 2: Robust (White) standard errors are used.
Note 3: F Test *State* are joint tests for State dummy variables.
Note 4: F Test *Per-capita GDP* are joint tests for significance of *Per-capita GDP* +*Per-capita GDP* * *Urbanization*.

Table 6a: Stability Test for models excluding metro variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
North East	0.18	0.35	0.35	0.55	0.40	0.45	0.57
	(0.84)	(0.79)	(0.95)	(0.85)	(0.90)	(0.89)	(0.85)
Maharastra	0.40	1.20	1.20	0.97	1.57	1.28	0.80
	(0.67)	(0.31)	(0.30)	(0.47)	(0.14)	(0.25)	(0.64)
Bihar	0.93	0.42	2.24**	1.75*	2.59**	2.19**	1.57
	(0.40)	(0.74)	(0.02)	(0.07)	(0.01)	(0.03)	(0.10)
North East and Bihar	1.07	0.72	0.72	0.56	0.82	0.68	0.51
	(0.35)	(0.54)	(0.68)	(0.85)	(0.57)	(0.71)	(0.90)
Maharastra and Bihar	0.81	1.23	1.49	1.46	1.26	2.03**	1.49
	(0.44)	(0.30)	(0.16)	(0.15)	(0.27)	(0.04)	(0.13)
Karnataka	0.81	1.44	1.50	1.20	0.98	1.27	1.43
	(0.44)	(0.23)	(0.16)	(0.29)	(0.45)	(0.26)	(0.15)

Note 1: P-values are given in the parenthesis.

Note 2: F-tests are joint tests for sate dummy variables.

Table 6b: Stability Test for models including metro variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
North East	0.51	0.75	0.51	0.59	0.44	0.46	0.74
	(0.73)	(0.56)	(0.87)	(0.84)	(0.90)	(0.90)	(0.71)
Maharastra	0.15	0.64	0.74	0.61	0.85	0.82	0.57
	(0.93)	(0.63)	(0.68)	(0.82)	(0.56)	(0.59)	(0.87)
Bihar	0.94	0.43	2.24**	1.75*	2.58**	2.26**	1.60*
	(0.39)	(0.73)	(0.02)	(0.07)	(0.01)	(0.02)	(0.10)
North East and Bihar	0.69	0.70	0.71	0.56	0.83	0.75	0.54
	(0.60)	(0.59)	(0.70)	(0.86)	(0.58)	(0.66)	(0.89)
Maharastra and Bihar	0.37	0.65	1.17	1.12	1.05	1.50	1.20
	(0.77)	(0.63)	(0.31)	(0.34)	(0.40)	(0.15)	(0.28)
Karnataka	1.54	1.60	1.53	1.25	1.29	1.43	1.34
	(0.19)	(0.16	(0.12	(0.24	(0.24	(0.16	(0.18)

Note 1: P-values are given in the parenthesis.

Note 2: F-tests are joint tests for sate dummy variables.

Table 7a: Results for Endogeneity Test for models excluding metro variables

	(2)	(3)	(4)	(5)	(6)	(7)
Market Access	NA	NA	7.006***	NA	NA	NA
			(0.0081)			
District Literacy	619.547***	1.0245	1.0152	1.0557	0.9531	0.9577
	(0.0000)	(1.0000)	(1.0000)	(1.000)	(1.000)	(1.000)
District Pucca Road	856.3743***	120.0632***	68.3737***	119.4803***	91.4705***	65.7513***
	(0.0000)	(0.0000)	(0.0009)	(0.0000)	(0.000)	(0.0025)
District Urbanization	490.4656***	21.8968	9.089	21.4989	22.107	8.4231
	(0.0000)	(0.9299)	(1.0000)	(0.9202)	(0.9252)	(1.0000)

Note 1: *, **, *** denotes 10, 5 and 1 percent levels of significance respectively.

Note 2: All tests follow χ^2 with appropriate degrees of freedom equal to the number of model parameters.

Note 3: Endogeneity tests are performed by comparing OLS and 2SLS parameter estimates. This comparison is done in SUR framework. The Hausman test is not appropriate as data has heteroscedasticity.

Table 7b: Results for Endogeneity Test for models including metro variables

	(2)	(3)	(4)	(5)	(6)	(7)
Market Access	NA	NA	4.394**	NA	NA	NA
			(0.0361)			
District Literacy	301.7376***	0.5442	0.5591	0.5399	0.5445	0.569
	(0.0000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)
District Pucca Road	297.4828***	24.7409	64.2781***	23.4085	20.3217	59.8895**
	(0.0000)	(0.9385)	(0.0066)	(0.9477)	(0.9882)	(0.0224)
District Urbanization	283.7623***	20.43	6.807	20.1736	20.7042	7.0247
	(0.0000)	(0.9828)	(1.0000)	(0.9787)	(0.9806)	(1.0000)

Note 1: *, **, *** denotes 10, 5 and 1 percent levels of significance respectively.

Note 2: All tests follow χ^2 with appropriate degrees of freedom equal to the number of model parameters.

Note 3: Endogeneity tests are performed by comparing OLS and 2SLS parameter estimates. This comparison is done in SUR framework. The Hausman test is not appropriate as data has heteroscedasticity.

Appendix 1: Variance Decomposition

This appendix briefly describes our variance decomposition. Let y_{ij} be the underlying variable (say, per-capita logged income) of j^{th} district in i^{th} state, $j = 1, 2 \ldots, n_i, \quad i = 1, 2 \ldots K$. Let $N = \sum_{i=1}^{n_i}$, the total number of observations. Define $\bar{\bar{y}} = \frac{1}{N} \sum_{i=1}^{K} \sum_{j=1}^{n_i} y_{ij}$, the grand mean. Define $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}, i = 1, 2 \ldots K$, the within mean. We define following three quantities...

Total sum of square (TSS)= $\sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2$.

Within Sum of square (WSS)= $\sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$.

Between Sum of Square (BSS)= $\sum_{i=1}^{K} n_i (\bar{y}_i - \bar{\bar{y}})^2$.

Then

$$TSS = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{\bar{y}})^2 = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i + \bar{y}_i - \bar{\bar{y}})^2 = WSS + BSS.$$

Finally dividing each term by N gives the total, between and within-state variances, $\nu^T = TSS/N$, $\nu^W = WSS/N$ and $\nu^B = BSS/N$. Hence $\nu^T = \nu^W + \nu^B$.

Appendix 2: Derivation of Equation 2

The purpose of this appendix is to use a standard partial adjustment model to assist with interpreting the inefficiencies in the empirical model in equation (2). First consider a standard partial adjustment model given by

$$\ln \hat{y}_i(t) - \ln \hat{y}_i(0) = \beta(\ln \hat{y}_i^* - \ln \hat{y}_i(0)), \tag{3}$$

where $\hat{y}_i$ is district income per effective worker, $\hat{y}_i \equiv (y_i/A_i)$, A_i is a labor productivity term and y_i is income per worker in district i. The left hand side of (3) is the transitional growth rate of output per effective worker in region i. On the right hand side is the gap between current income per effective worker and the long run steady state value of output per effective worker $\hat{y}_i^*$. Thus the transitional growth rate of district i is assumed to depend on the gap between the current income initial levels of output per effective worker. In what follows the speed of adjustment, β will be a parameter to be estimated.

Next let y_i denote district i per capita income and y^* denote the steady state income per worker in a nearby UA. Then for district i consider a variable θ_i such that, in a steady-state equilibrium,

$$y_i^* = \theta_i \, y^* \tag{4}$$

The variable $\theta_i \leq 1$ thus measures the extent of all barriers to complete convergence, such as trade and transport costs, communications costs, road quality and other geographic barriers. If $\theta_i < 1$ district i will only achieve partial convergence to the metropolitan center or UA.

In terms of effective workers (4) implies $\hat{y}_i^* = \theta_i \, \hat{y}^*$.[32] Then using (3) the transitional

[32] We assume long run technology convergence so that $A_i^* = A^*$. Alternatively one could assume that technological gaps exist in the long run and that this difference is absorbed as an argument in the function θ_i.

growth process for some non-metropolitan district i, can be derived as

$$\ln y_i(t) - \ln y_i(0) = gt - \beta \ln y_i(0) + \ln A_i(0) + \beta \left(\ln \hat{y}^* + \ln \theta_i\right). \tag{5}$$

In equation (5) the transitional growth rate of district i depends on: (i) the initial per capita income of district i, $y_i(0)$; (ii) the level of labor productivity of district i, $A_i(0)$; (iii) the steady-state value of income per effective worker in the relevant UA, $\hat{y}^*$ and; (iv) the distance between district i and the UA, θ_i.

To operationalize (5) we need to specify an empirical counterpart to (4). The gravity literature suggests a simple inverse relationship such as $\theta_{i,j} = \theta \ D_i^{\gamma}$. Hence, using logarithms we have

$$\ln \theta_{i,j} = \ln \theta \ + \gamma \ \ln D_i + \eta \, \mathbf{X}_i \tag{6}$$

where $\gamma < 0$, is the distance elasticity, $\mathbf{X}_i$ is a vector of characteristics of region i and η is a vector of coefficients. [33]

From (5) and (6) we obtain an empirical model, which is (2) in the text.

$$\ln y_i(t) - \ln y_i(0) = \alpha_0 + \alpha_1 \ \ln y_i(0) \ + \ \alpha_2 \ \ln D_i + \eta \, \mathbf{X}_i \ + \ \epsilon_i \tag{7}$$

where $\alpha_1 \equiv -\beta$, $\alpha_2 = \beta\gamma$, $\alpha_0 = g + \beta \ln A_i(0) + \beta \ln \hat{y}^* + \theta$, and $\ln A_i(0) = \ln A + \epsilon_i$, where ϵ_i is a district specific random shock reflecting, for example, institutions, climate and endowments.

This shows that $\alpha_2/\alpha_1 = \gamma$ recovers the elasticity of long run income with respect to *Minimum Distance*, as claimed in the text. Specifically from (4) and (6) we have

$$\gamma = \frac{\partial \ln(y_i^*/y_j^*)}{\partial \ln(D_i/D_j)}$$

[33]This also requires the restriction that $D_i \geq 1$, which will be true in our data.

Thus if we consider two districts i and j with identical characteristics, except for their distance from the metropolitan district k, *Minimum Distance*, then the value of γ determines the difference in long run incomes in the long run equilibrium. As can also be seen from (2), the coefficient $\alpha_2 = \beta\gamma$ gives the impact on the transitional growth rate.

Finally, a further simple extension of (2) allows for the possibility that the UAs have different balanced path income levels. Specifically, suppose $\hat{y}_j^* = f(\mathbf{Z}_j)\ \hat{y}^*$, where $\mathbf{Z}_j$ is a vector of characteristics that affect the steady state income levels of UA district j. Then, assuming $f(\mathbf{Z})$ is log linear gives

$$\ln y_i(t) - \ln y_i(0) = \alpha_0 - \alpha_1 \ln y_i(0) \ + \ \alpha_2 \ \ln D_i + \eta \ \mathbf{X}_i + \delta \ \mathbf{Z}_j + \epsilon_i. \tag{8}$$

This then provides a basis for including the additional UA characteristics as discussed above in Section 5.3 and Table 5b.

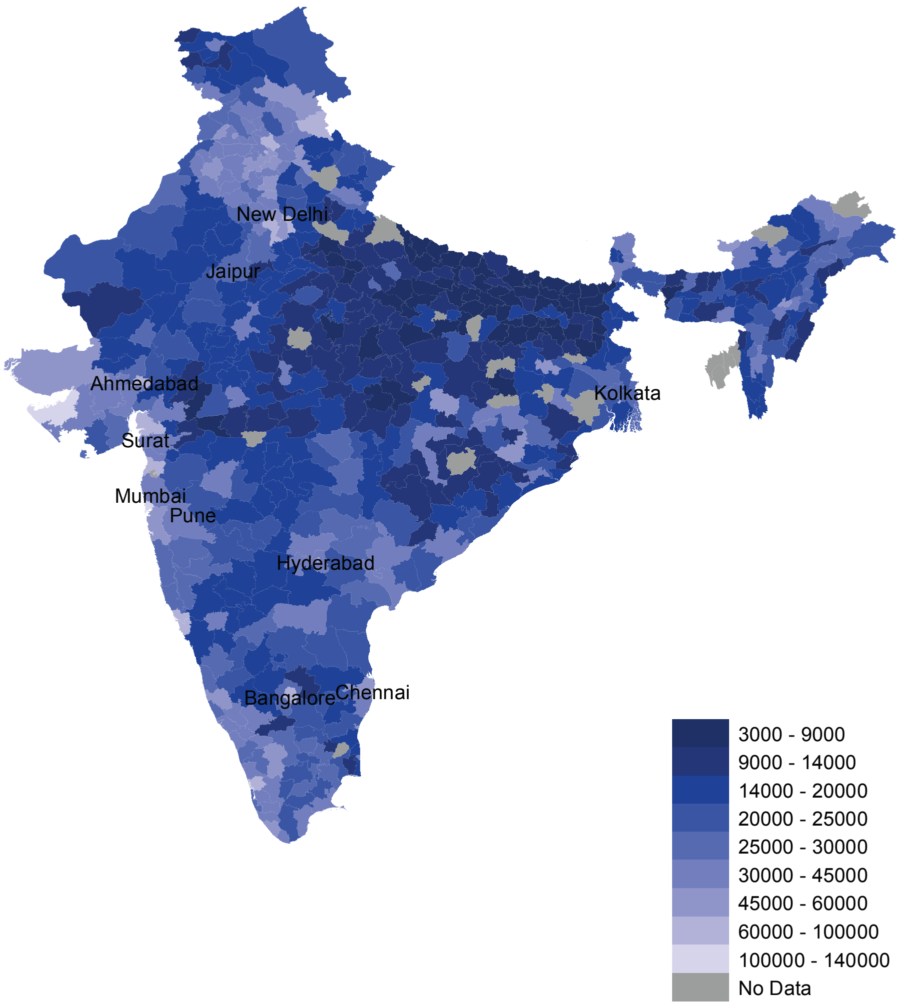


Figure 1: Per-Capita Income by District

Figure 2: Probability Density Function for Indian District Incomes

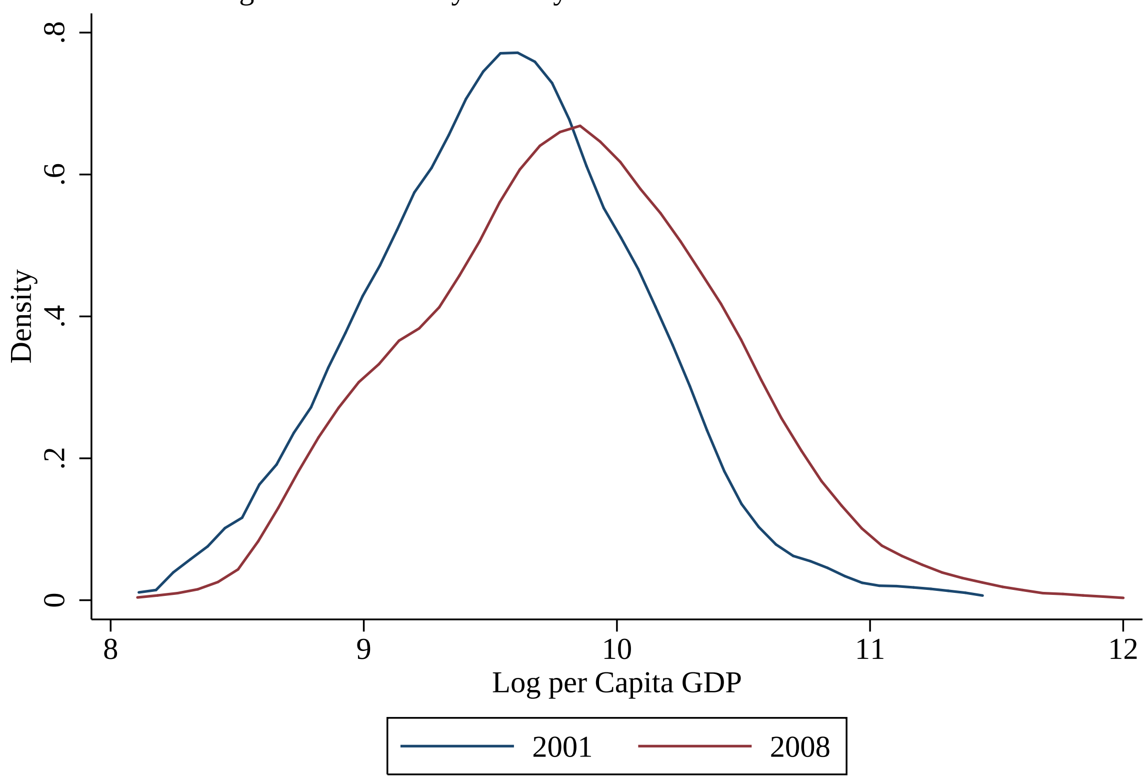
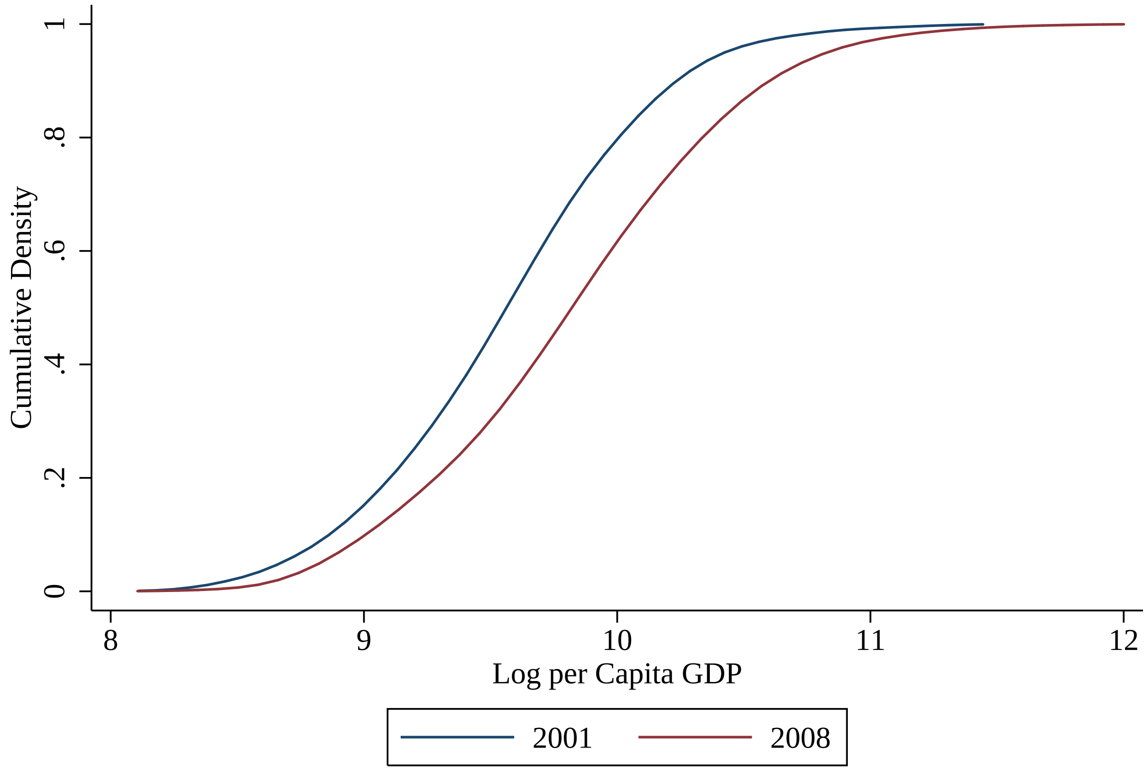


Figure 3: Cumulative Distribution Function for Indian District Incomes



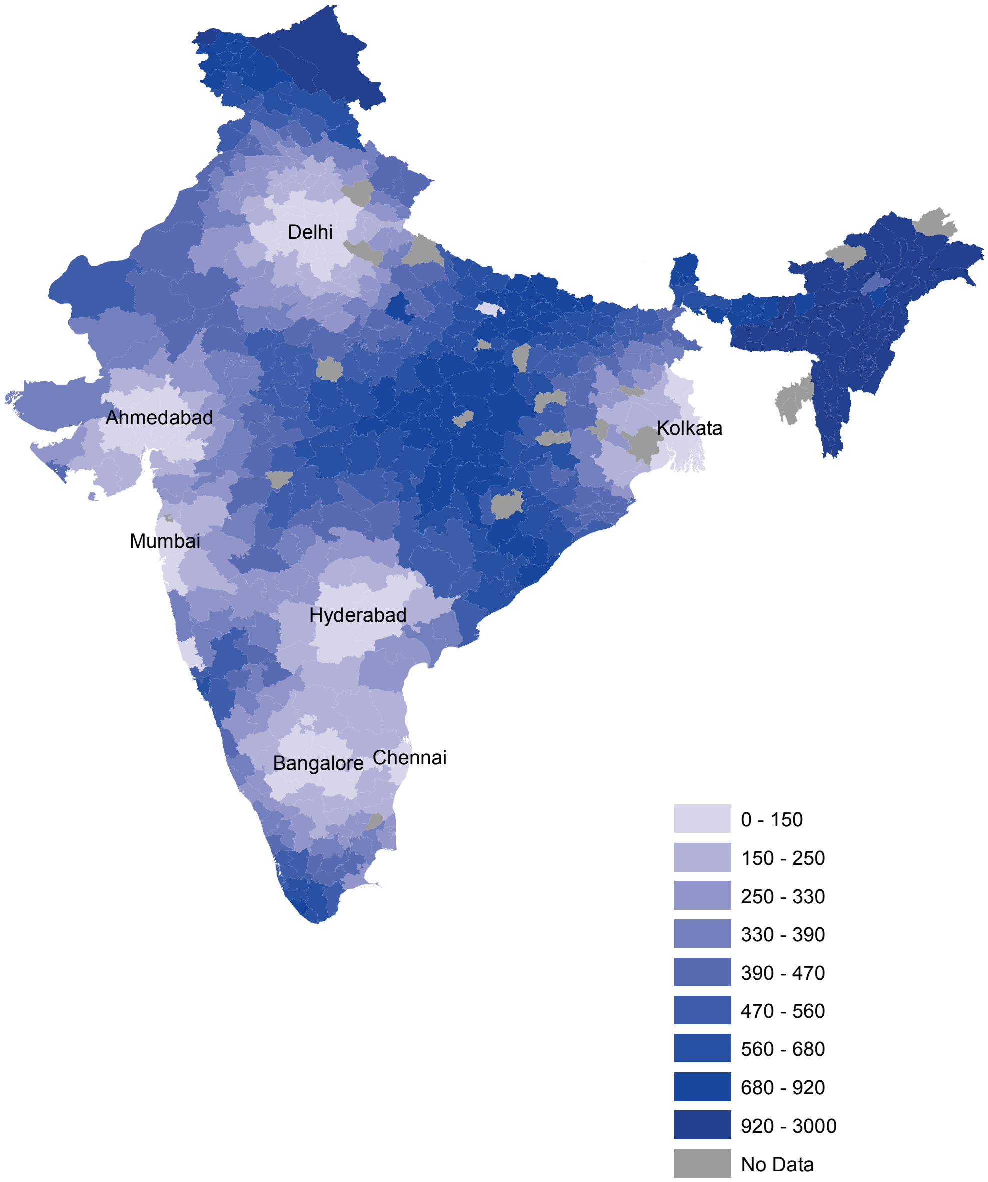


Figure 2: Minimum Distance to Seven Largest Metropolitan Centers