Bonn Econ Discussion Papers

Discussion Paper 1/2005

Price Convergence across Regions in India

by

Samarjit Das, Kaushik Bhattacharya

Dezember 2004

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is sponsored by the

Deutsche Post World Net

MAIL EXPRESS LOGISTICS FINANCE

Price Convergence across Regions in India

Samarjit Das and Kaushik Bhattacharya[*]

University of Bonn

December 31, 2004

Abstract

The paper attempts to examine whether there is price convergence across various regions in India. Our results indicate significant presence of cross-sectional dependence in prices in India, rendering some of the standard panel unit root tests inapplicable. Using various panel unit root tests that are robust to cross-sectional dependence, it is found that relative price levels among various regions in India mean-revert. We decompose each series into a set of common factors and idiosyncratic components. The decomposition enables us to test stationarity and estimate half-lives of the common factors and the idiosyncratic components separately. Both these components in case of India are found to be stationary. Idiosyncratic price shocks, however, are found to be more persistent as compared to the common factor. The results also indicate that transportation costs proxied by distance can explain a part of the variation in prices between two locations in India.

Key Words: Cross co-integration, Cross-sectional dependence, Panel unit root tests, Common factor, Price convergence

JEL Classification: C23 and E31

Corresponding Author: Samarjit Das, University of Bonn, Institute of Econometrics & O. R., Adenauerallee 24-42, 53113 Bonn, Germany
E-mail: samarjit@uni-bonn.de

[*]The authors would like to thank Dibyendu Bhaumik for arranging the data for this study.

1 Introduction

The issue of price convergence across regions within a single economy has received increasing attention in recent literature. A high dispersion of inflation across regions and its persistence over time may have serious implications on the regional wage rates and the standard of living. It also poses concern regarding allocation of resources. The existence of large systematic price divergence despite a common currency and no implicit or explicit restrictions on factor mobility may, therefore, indicate market segmentation, and its eradication is a challenge to policymakers.

The studies on regional price convergence provide a benchmark by testing the *law of one price* under more controlled condition, as problems due to fluctuations in exchange rate or factor market rigidities are eliminated. Price convergence in these studies could be tested under alternative frameworks. Earlier attempts were either through cointegrated vector autoregression (VAR) models or univariate unit root tests of relative prices across regions. However, estimating a co-integrating VAR model is difficult if the number of cross-sectional units are large. So far as univariate unit root tests are concerned, it is well known that for small series length and near unit root situations, these tests suffer from power deficiency. Increasing the time series length may, in fact, compound the problem as chances of structural break increase, leading to further loss of power. Recent literature, therefore, suggests to test stationarity *jointly* by conducting panel unit root tests on the regional price relatives. Recent applications on regional price convergence that adopt this framework include Parsley and Wei (1996) and Cecchetti *et al* (1998, 2002) for the US, Ceglowski (2003) for Canada, Engel and Rogers (1996) for both US and Canada, Menna (2001) for Italy and Fan and Wei (2003) for China.

In this study, we examine price convergence across regions in India. Common sense suggests that due to a single currency, near-free factor movements and policies adopted by the central authority, prices in different regions in India would be contemporaneously correlated. At the same time, it is also plausible that due to its large size, different agro-climatic and economic conditions and federal structure of governance, prices would also be affected by local shocks. We stress that a recognition of this dichotomy has important implications on panel unit root testing and hence, on the existing findings on regional price convergence. Recent studies like O'Connell (1998) and Breitung and Das (2003) have highlighted that, in the presence of contemporaneous correlation, standard panel unit root tests like Maddala and Wu (1999), Levin *et al* (2002) and Im *et al* (2003) may suffer from oversize problem. Another problem in these tests is the possibility of the presence of cross-co-integration - when individual series are non-stationary but have a common trend. In these cases, Bannerjee *et al* (2004) have shown that these tests are oversized.

A major implication of incorporating contemporaneous cross-sectional dependence in a panel framework is that any series in that framework may be decomposed into a number of common factors, besides the series specific idiosyncratic term. Econometrically, this decomposition serves three purposes. First, it provides a framework to test for non-stationarity separately for common factors and idiosyncratic components. Separate tests for them are meaningful because the common factors and idiosyncratic components may have different orders of integration, and it is well known that if a series is the sum of a non-stationary and a stationary series, it is difficult to establish the presence of unit root in the summed one (Schwert, 1989; Pantula, 1991; Bai and Ng, 2004). In such situations, it is perhaps desirable to test unit root hypothesis separately (Bai and Ng, 2004). Non-stationarities in the common factors, in fact, provide a parsimonious way of identifying common trends in the series without specifying a co-integrating VAR model. If, in addition, the idiosyncratic terms are found to be stationary, the series are then interpreted as cross-co-integrated (Bannerjee *et al*, 2004). Second, factor models capture the extent of dependence across cross-sectional units and therefore one should use tests which are robust to such dependence. Third, the decomposition also enables one to examine the speeds of convergence of the common factors and the series specific local shocks separately. This decomposition is meaningful for policy purpose because, if the idiosyncratic shocks are dominant, primary responsibility of control of inflation rests with the local government and often should involve the management of the local supplies alone. In contrast, predominance of common shocks reflects that inflation should be a concern for the federal or the central authorities.

To the best of our knowledge, no empirical work on regional price convergence has so far focussed on the decomposition of data into factors and idiosyncratic components, exploited the strength of separate unit root tests, and finally conducted the tests on convergence hypothesis. Existing studies have also not taken into account the possibilities of cross-co-integration and its econometric implications. In this paper, we attempt to examine to what extent prices in different places in India are affected due to common or local shocks, using panel unit root tests under both 'weak' and 'strong' cross-sectional dependence as formalised in Breitung and Das (2004). We also attempt to analyse to what extent transportation costs, (proxied by distance or its monotonic functions in this study) could explain deviations in inflation across regions in India. The time frame considered in this study is from January, 1995 to June, 2004, i.e., 114 months. It may be noted that during the early 1990s, the Indian economy was going through important structural changes. Choice of this period is, therefore, conscious, as it limits the possibilities of structural breaks in the series.

The plan of the paper is as follows: Section 2 presents a brief discussion on the methodologies. Section 3 describes and carries out a brief preliminery analysis of the data. Section 4 presents the empirical results on decompositions of prices into factors and idiosyncratic components. Section 5 attempts to find out the impact of distance on price deviations among pairs of regions. Finally, Section 6 summarizes the findings with some concluding observations.

2 Econometric Methodology

Though panel unit root tests are expected to carry more power than univariate time series tests, they should be used with caution. The standard panel unit root tests like Maddala and Wu (1999), Levin *et al* (2002) and Im *et al* (2003) are based on the restrictive assumption that cross-sectional units are independent. However, it is generally perceived that cross-sectional units are contemporaneously related. Various Monte-Carlo studies show that under cross-sectional dependence, such panel tests suffer from severe size distortion and lead to high probability of rejection of the null hypothesis of unit root (O'Connell, 1998; Breitung and Das, 2003). So, a necessary precondition of applying panel unit root test is to test for cross-sectional dependence.

2.1 Test for Cross-Sectional Dependence

As there is no *apriori* knowledge of spatial or weighting matrix, LM kind of test as proposed by Breusch and Pagan (1980) may be more appropriate in the panel unit root context. LM test is used to test for cross-sectional dependence in regression framework where the number of equations (N) is finite but time dimension (T) is infinite. However, simple modification of the original LM test provides normal distribution under very large N as opposed to chi-square distribution of LM test for finite N. LM test is defined as

$$LM = T \sum_{1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho_{ij}}^2,$$

where $\hat{\rho_{ij}}$, is the sample estimate of the pair-wise correlation of the residuals. Under independence, Breusch and Pagan (1980) showed that $T\hat{\rho_{ij}}^2 \Rightarrow \chi^2_1$and hence $LM \Rightarrow \chi^2_{\frac{N(N-1)}{2}}$ for large T, but for finite small N. However, it can be easily shown that under independence

$$MLM = \sqrt{\frac{1}{N(N-1)}} \sum_{1}^{N-1} \sum_{j=i+1}^{N} (T\hat{\rho_{ij}}^2 - 1),$$

follows $N(0,1)$ as $T \rightarrow \infty$, followed by $N \rightarrow \infty$.

2.2 Panel Unit Root Tests under Cross-Sectional Dependence

Cross-sectional dependence may be "weak" or "strong", depending upon whether for any number of cross-sectional dimension N, all the eigenvalues of the error covariance matrix are bounded ("weak"), or not ("strong"). Under strong form of dependence, where some of the eigenvalues of the error covariance matrix tend to infinity as the cross-section dimension (N) increases, each series may be decomposed into two components– the first part consists of a few factors that are common to all series, and the second, the idiosyncratic component (Forni *et al*, 2000).

Under weak cross-sectional dependence, we will confine ourselves to four statistics discussed in Chang (2002), Chang (2004) and Breitung and Das (2003). We briefly describe these test procedures. Consider the data generating process:

$$\Delta y_{it} \quad = \quad \mu_i + \phi y_{i,t-1} + \sum_{j=1}^{p_i} \nu_{ij} \Delta y_{i,t-j} + \varepsilon_{it} \qquad (1)$$

here the starting values $y_{i0} \ldots y_{i,-p_i}$ are set equal to zero. As short run dynamics is generally expected to be present in the data we have incorporated $\sum_{j=1}^{p_i} \nu_{ij} \Delta y_{i,t-j}$ to take care of autocorrelation in the data. Individual specific intercepts μ_i have also been considered keeping in mind that series mean are generally not zeroes. The error vector $\varepsilon_t = [\varepsilon_{1t}, \ldots, \varepsilon_{Nt}]'$ is i.i.d. with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = \Omega$, where Ω not necessarily a diagonal matrix.
The null hypothesis is

$$H_0 : \ \phi = 0 \ ,$$

that is, all time series are random walks. Under the alternative it is assumed that the time series are stationary with $\phi < 0$. [1]

For the IV method as suggested by Chang (2002), we consider the following nonlinear instruments

$$F(y_{i,t-1}) = y_{i,t-1} e^{-c_i |y_{i,t-1}|},$$

for $y_{i,t-1}$ in equation (1), where $c_i = KT^{-0.5} s_i$ and s_i is the estimated standard deviation of Δy_{it}. K, the truncation parameter, is taken as 3, as suggested by Chang (2002). So equation (1) looks like

$$\Delta y_{it} = \mu_i + \phi F(y_{i,t-1}) + \sum_{j=1}^{p_i} \nu_{ij} \Delta y_{i,t-j} + \varepsilon_{it}$$

Chang(2002) showed that the usual individual OLS t-statistics are asymptotically independent even in the presence of cross sectional dependence. So the actual test statistic is simply defined as a standardized sum of individual IV t ratios and follows $N(0,1)$.

[1]Some tests are applicable with hetergeneous and more general alternatives.

Another method developed by Chang(2004) is the bootstrap method which takes care of the oversize problem involved in standard OLS or GLS t statistics. In this method, one estimates equation (1), resamples the residual vectors and uses them to generate *pseudo* observations. OLS or GLS t statistics are calculated based on this pseudo observations. Collection of these t statistics form the bootstrap distribution under H_0.

Another simple method that also works when N is greater than T is the one proposed by Breitung and Das (2003). As OLS standard error is biased under cross-sectional dependence, Breitung and Das (2003) used modified standard error that is robust to cross-sectional dependence and showed that under H_0 the test statistic t_{rob} follows $N(0,1)$. However, to apply this method, one should 'pre-whiten' any serial correlation in the data.

We call these four statistics, IV (t_{iv}), OLS based bootstrap t (t^*_{ols}), GLS based bootstrap t (t^*_{gls}) and the robust test (t_{rob}) respectively.

2.3 Tests Using Common Factors

The standard form of the common factor model as in Breitung and Das (2004) is:

$$y_{it} = \mu_i + \gamma_i' f_t + u_{it} \ , \tag{2}$$

$$f_t = \psi f_{t-1} + \sum_{j=1}^{q} \eta_j \Delta f_{t-j} + v_t \ , \tag{3}$$

$$u_{it} = \theta u_{i,t-1} + \sum_{j=1}^{p_i} \eta_{i,j} \Delta u_{i,t-j} + \varepsilon_{it} \ , \tag{4}$$

where f_t is unobservable and random $k \times 1$ vector of common factors with γ_i as a non-random factor loading and u_{it} is an idiosyncratic error component that may be 'locally' cross-sectionally correlated . Equations and parameters are self explanatory.

Note that such formulation allows for serial correlation in the data and also for deterministic terms like individual intercepts. If either ψ or θ is exactly 1, the data must necessarily be non-stationary. In the special case of $\psi = 1$, but $\theta < 1$, the series are represented by a common trend and are called *cross-co-integrated* (Banerjee *et al*, 2004; Breitung and Das, 2004). In this case, common factors (i.e., common trends) are the binding force for the series to move together and convergence hypothesis can be accepted. So under factor models, convergence may take place in two ways. First, when both factors and idiosyncratic errors are stationary. Second, when there is cross-co-integration.

The first problem in factor models is to determine the number of factors (k). Factors are estimated using principal components. In this paper, we follow the procedures developed by Bai and Ng (2002), e.g., we use three PC_p criteria and three IC_p criteria.

After estimating factors, one can apply standard univariate unit root test to each factors to gauge stationarity/nonstationarity of factors or do a joint test as suggested by

Bai and Ng (2004). Each factors as estimated by principal component follows standard Dickey-Fuller test under the null hypothesis of unit root (Bai and Ng, 2004). Similarly, we can test for stationarity of idiosyncratic errors using panel approach. Moon and Perron (2004) developed panel unit root test based on idiosyncratic errors after removing the factors from the data. Breitung and Das (2004) showed that the robust test as in Subsection 2.2 follows standard Dickey-Fuller test under the null hypothesis of unit root when cross-sectional dependence is strong, i.e., when there is a factor.

3 Data and Descriptive Analysis

Our empirical study on the Indian economy is carried out with the data on monthly consumer price indices for industrial workers (CPIIW) in India. The data on CPIIW are collected by the Indian Labour Bureau (ILB) from 76 different cities/towns or regions, which appear to be more or less uniformly distributed across 24 States or Union Territories (UTs) in India. The large number of centres available in India enables one to observe the spatial distribution of the prices clearly. As distances between pairs of centres vary over a wide range, from about 10 kilometres between Kolkata and Howrah to more than 2000 kilometres between Ahmedabad and Guwahati, one has more control on variables like distance that are used to explain variations in regional price relatives.

Appendix A presents the names of these centres along with the States (UTs) in which they are located, the bracketed numbers in the first column being the number of such centres within the State (UT). It is observed that in many large States, there are more than one cities/towns from which these data are collected. Since in India, States are major units within which a large part of fiscal policy would be common, the data provide us an opportunity to examine price convergence in more detail than are generally available in the literature. To apprise the closeness of the centres with other centres, we indicate the geographical location of each centre in Appendix A. The first and the second terms in the bracket after each centre in the second column reflect the latitude and the longitude corresponding to that centre respectively. [2]

The various items covered in CPIIW can be classified into five major groups, viz. (i) food, beverages and tobacco, (ii) fuel and light, (iii) clothing and footwear, (iv) housing and (v) miscellaneous. Among these groups, the first three could be interpreted as tradables and the last two non-tradables. It may be noted that due to heterogeneous living conditions, the weights for these items in the regional CPIIW may vary across centres. The weights in CPIIW for the above five groups at All India level are 60.15, 6.28, 8.54, 8.67 and 16.36 per cent respectively, indicating that approximately one fourth of the total

[2]The latitudes and the longitudes have been collected from the website http://www.indiapress.org and are expressed in the units of degrees and minutes.

weights in CPIIW at the all-India level is bourne by the non-tradables.[3]

The extent of long-run cross-sectional variation in regional inflation in India is presented in Figure 1. In Figure 1, we plot the estimated kernel density of the annual average deviations of the regional inflation rates from the All-India average between January 1995 to June 2004. Figure 1 reveals considerable differences in the rates of inflation experienced by different regions in India, the range being close to 3.0 percentage points. It may be noted that in case of US, such differentials for ten year spans are 1.13 percentage points on average, though in earlier periods a differential of 1.55 percentage points have also been observed (Cecchetti *et al*, 1998). Figure 1, therefore, suggests that local price shocks could be highly persistent.

The descriptive statistics relating to annual inflation rates for different years are presented in Table 1. In contrast to Figure 1, Table 1 thus provides the short-run features corresponding to regional price movements. The first feature that we observe from Table 1 is the dip in inflation rate in the later years. From the year 2000 onwards, the annual average rate of inflation based on CPIIW in India is consistently less than 4.0 per cent. This difference is due to the changes in the general macroeconomic environment in India.

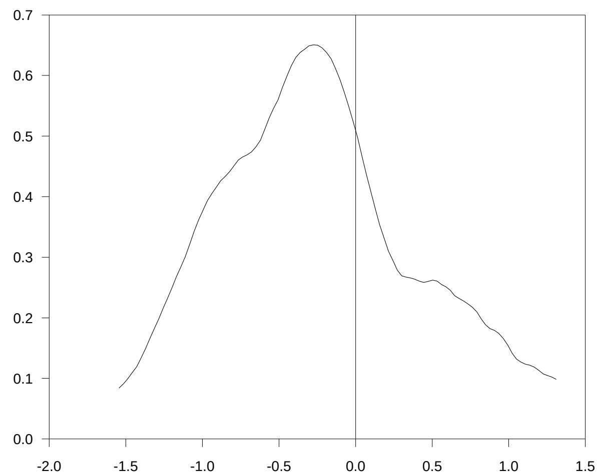

Figure 1: Kernel Densities of the Annual Average Deviation of Regional Inflation Rates from the All-India Average (Between January, 1995 to June, 2004)

For our analysis, however, an examination of the dispersion among regional inflation rates would be more important. Table 1 provides some interesting insights on these variations. It reveals that the range over which regional inflation rates vary could be as high as about 20.0 percentage points in a single year, as in 1998. The standard deviations for different years are, however, found to be stable. Except during the year 1998 in which

[3]Further details of the coverages and the weighting scheme are available in the ILB website (http://labourbureau.nic.in).

Table 1: Descriptive Statistics Corresponding to Annual Rates of Inflation across Centres

	1996	1997	1998	1999	2000	2001	2002	2003	2004
Mean	9.2	6.7	13.4	5.1	3.4	3.5	3.9	3.7	3.3
Median	9.0	6.8	13.4	5.0	3.6	3.6	3.7	3.8	3.5
S. D.	2.2	2.2	3.7	2.4	2.4	2.4	2.1	1.6	2.1
Skewness	1.6	-0.1	-0.3	0.8	-0.2	0.4	0.6	0.1	-1.0
Kurtosis	6.5	2.0	0.8	1.4	-0.5	0.4	0.1	-0.6	2.0
Minimum	5.3	-1.4	0.9	0.6	-2.9	-1.7	-0.4	0.7	-4.4
Maximum	19.9	12.4	21.7	13.9	7.9	10.6	9.3	7.3	7.9

Note: For the year 2004, rate of inflation is computed based on average CPI's from January to June

it was high at 3.7, standard deviations are generally in the range of 2.1 to 2.4. Moments of higher order in Table 1 reveal that the skewness changes sign frequently, indicating near symmetry. Further, the distributions tend to be slightly leptokurtic for most of the years.

The kurtosis observed in the data could be due to two possibilities. First, it could be due to the existence of large local shocks in the prices in a few places. The second possibility is the existence of measurement errors in the concerned centres. It is, therefore, imperative to examine this aspect further. In Table 2, we present the regional inflation rates for the five bottom and top centres. It is expected that if the price rise (fall) is due to genuine economic factors, the rate of inflation in the neighbouring centres would reflect similar patterns. Thus, existence of centres that are geographically close in either the top or bottom 5 centres would be viewed as a cross-validation of the existence of local price shocks. In Table 2, there are several such clusters (i.e., centres that are either geographically close or belong to the same State), indicating that in many cases the slightly long tails observed in the distribution of regional inflation rates are due to genuine economic reasons. In a few cases (e.g., Rajkot in the year 1996 and 1997), however, the regional inflation rates appear to be outliers, i.e., the differences with neighbouring centres are high. Thus, the possibility of measurement errors in data in a few cases can not be completely ruled out.

4 Empirical Results

As contemporaneous cross-sectional dependence is a major problem with panel unit root tests, we first apply tests (LM and MLM) that could indicate its presence in the data. In this paper, we have worked with the log-ratio of regional prices to that of all India averages. However, to check robustness of the *numeraire*, we have considered Nagpur

Table 2: Centres that Experienced Lowest or Highest Inflation

Year	Centres with Lowest Inflation	Centres with Highest Inflation
1996	Saharanpur(5.3), Tripura (5.6), Ludhiana(5.7), Srinagar(5.7),Delhi(5.9)	Pondichery (12.5), Gudur (12.7), Mudakayam (12.9), Jamshedpur (14.4), Rajkot (19.9)
1997	Rajkot (-1.4),Tinsukia (3.4), Indore (3.5), Barbil (3.6), Solapur (4.0)	Pondichery (10.6), Mercara (10.6), Tiruchirapally (11.3), Goa (11.5), Rourkela (12.4)
1998	Quilon (0.9), Tiruchirapally (7.2), Coonoor (7.2), Bangalore (8.2), Salem (8.3)	Faridabad (18.9), Srinagar (19.3), Howrah (20.8), Varanasi (21.4), Jalpaiguri (21.7)
1999	Jaipur (0.6), Kethgudem (1.1), Pondichery (0.6), Noamundi (1.6), Rourkela (1.6)	Thiruvananthapuram (8.9), Monghyr (9.4), Howrah (9.7), Chandigarh (11.3), Srinagar (13.9)
2000	Kodarma (-2.9), Varanasi (-1.3), Lalbac-Silchar (-1.1), Tezpur (-0.9), Darjeeling (-0.3)	Delhi (7.1), Coimbatore (7.6), Nasik (7.6), Goa (7.8), Mumbai (7.9)
2001	Mariani-Jorhat (-1.7), D. D. Tinsukia (-1.2), Mundakayam (-0.4), Mercara (-0.4), Monghyr (-0.1)	Durgapur (7.8), Bhopal (8.2), Srinagar (8.4), Kolkata (9.2), Haldia (9.3)
2002	Tezpur (-0.4), Noamundi (0.2), Mariani-Jorhat (0.2), Mercara (0.4), Lalbac-Silchar (0.8)	Kolkata (7.8), Guntur (7.9), Durgapur (8.8), Tiruchirapally (9.2), Haldia (9.3)
2003	Tripura (0.7), Vadodara (0.7), Jalpaiguri (0.9), Surat (1.0), Ranch-Hatia (1.2)	Bhilai (6.2), Pondichery (6.5), Guntur (6.6), Tiruchirapally (6.6), Quillon (7.3)
2004	Tiruchirapally (-4.4), Surat (-2.1), Kethgudem (-1.5), Warangal (-0.9), Rajkot (-0.1)	Jharia (6.3), Faridabad (6.4), Raniganj (6.7), Solapur (6.7), Lalbac-Silchar (7.9)

Note: For the year 2004, rate of inflation is computed based on average CPI's from January to June prices as *numeraire* as well.[4] For empirical analysis, the data for different centres have also been pre-whitened to remove traces of serial correlation present in it. To do that, an AR(12) filter has been applied to all series after some preliminary analysis and appears to be reasonable, given that we have monthly data.

Results on both LM and MLM tests in Table 3 clearly establish the presence of contemporaneous cross-sectional dependence. This conclusion does not change whether we consider centres as units, aggregate the centres to have State specific CPIIWs or separately apply the tests to States that have at least 4 centres. In fact, the results remain same when (i) Nagpur is specified as the *numeraire* or (ii) the tests are applied to smaller group of panels (with $N = 5, 10, 15, 20$) where the units are randomly chosen. This finding suggests use of panel unit root tests which are robust to cross-sectional dependence. Therefore, in this study we have not considered tests that are based on cross-sectional independence, like Maddala and Wu (1999), Levin *et al* (2002)and Im *et al* (2003).

Table 3 also summarizes the results of panel unit roots tests under cross-sectional dependence. These tests implicitly assume that cross-sectional dependence is of arbitrary form and weak. For all tests, individual specific intercepts are incorporated. For robust test, first observation has been subtracted from all observations to take into account individual specific intercepts (Breitung and Das, 2003). For bootstrap based tests, we have considered 5000 bootstrap replication. The 'truncation' parameter K for IV method of has been taken as 3 following Chang (2002), without going into the debate about its appropriateness.[5]

For centrewise data, Table 3 suggests that unit root null hypothesis is rejected by all tests. In other words, all tests suggest that the regional prices (relative to a common *numeraire*)are jointly stationary, implying that shocks to regional relative prices do not drive them away from the average All India prices. In otherwords, all tests strongly support price convergence. This finding does not change when centre-specific CPIIW's are aggregated to State level, (i.e., when States are considered as cross-sectional units with $N = 24$, instead of regions with $N = 76$).

When we carry out similar tests for States that have at least 4 centres to examine price convergence within a State, all States except Madhya Pradesh indicate price convergence unambiguously. The evidence for Madhya Pradesh is mixed. Though the precise reason of non-rejection of unit root by some tests for Madhya Pradesh is difficult to identify, it may be noted that Madhya Pradesh is one of the largest and backward States in India. It is plausible that the large size of this State coupled with poor management of local shocks have led to market segmentation within the State. Panel unit root tests actually test joint

[4]Nagpur is chosen as a *numeraire* because it is located near the centre of India.

[5]Im and Pesaran (2003) remark that the choice of K sometimes depends on the data generating process.

Table 3: Panel Unit Root Tests and Cross-Sectional Dependence Tests

Regions	t^*_{ols}	t^*_{gls}	t_{rob}	t_{iv}	LM	MLM	Estimated Half-Life
Andhra Pradesh (N=6)	-4.44 (-1.80)	-3.56 (-1.93)	-4.13 (-1.65)	-4.21 (-1.65)	67.33 (7.26)	9.55 (1.96)	11.00
Assam (N=5)	-2.38 (-2.08)	-2.70 (-1.60)	-1.98 (-1.65)	-2.72 (-1.65)	179.30 (3.94)	37.85 (1.96)	43.80
Gujarat (N=5)	-2.88 (-1.93)	-4.01 (-1.90)	-2.48 (-1.65)	-4.19 (-1.65)	76.40 (3.94)	14.84 (1.96)	12.46
Jharkhand (N=5)	-3.32 (-1.77)	-3.51 (-1.91)	-2.89 (-1.65)	-3.09 (-1.65)	54.21 (3.94)	9.88 (1.96)	14.21
Karnataka (N=4)	-3.04 (-1.97)	-2.35 (-1.84)	-2.97 (-1.65)	-2.25 (-1.65)	69.33 (1.63)	18.21 (1.96)	17.56
Kerala (N=4)	-4.50 (-1.91)	-4.73 (-1.75)	-3.95 (-1.65)	-4.79 (-1.65)	90.53 (1.63)	24.40 (1.96)	06.77
Madhya Pradesh (N=5)	-1.64 (-1.78)	-1.81 (-1.89)	-1.85 (-1.65)	-0.98 (-1.65)	47.81 (3.94)	5.99 (1.96)	—
Maharashtra (N=5)	-4.68 (-2.00)	-4.30 (-1.90)	-4.25 (-1.65)	- 4.30 (-1.65)	35.22 (3.94)	5.64 (1.96)	8.90
Tamil Nadu (N=6)	-2.83 (-1.80)	-3.38 (-1.93)	-2.07 (-1.65)	-3.87 (-1.65)	243.78 (7.26)	41.78 (1.96)	18.90
Uttar Pradesh (N=5)	-2.57 (-2.11)	-2.59 (-1.90)	-2.57 (-1.65)	-2.67 (-1.65)	247.25 (3.94)	53.05 (1.96)	25.80
West Bengal (N=8)	-2.67 (-2.10)	-3.30 (-1.97)	-2.01 (-1.65)	-2.58 (-1.65)	340.12 (16.92)	41.72 (1.96)	26.94
All States (N=24)	-5.23 (-2.08)	-4.24 (-1.81)	-4.12 (-1.65)	-6.11 (-1.65)	1402.07 (238.52)	47.92 (1.96)	18.85
All Centres (N=76)	-10.22 (-2.64)	-6.26 (-1.29)	-6.18 (-1.65)	-11.32 (-1.65)	9561.09 (1152.74)	91.09 (1.96)	19.83

Note: (i) t^*_{ols}, t_{rob}, t^*_{gls} and t_{iv}, denote the t-statistics corresponding to Bootstrap-OLS, Robust-OLS, Bootstrap-GLS and Chang's(2002) instrumental variable method, respectively. LM and MLM statistics defined for testing cross sectional dependence as defined earlier. 5% critical values are given in (.) below the test statistics. (ii) Half-lives are presented in months.

Table 4: Selection Criteria for the Number of Factors for Centres ($N = 76$)

R	PC_{p1}	PC_{p2}	PC_{p3}	IC_{p1}	IC_{p2}	IC_{p3}
1	16434.4	16606.9	16023.2	9.7	9.7	9.7
2	28058.3	28604.5	26756.3	10.3	10.3	10.2
3	70501.7	72420.7	65927.7	11.2	11.2	11.1
4	97887.8	101214.6	89958.4	11.5	11.6	11.4
5	89871.0	93460.7	81314.9	11.5	11.5	11.3
6	111952.6	117016.0	99883.8	11.7	11.8	11.6
7	101667.5	106745.7	89563.4	11.7	11.7	11.5
8	98313.3	103641.1	85614.3	11.7	11.7	11.4
9	87067.2	92119.3	75025.4	11.6	11.7	11.3
10	110570.3	117371.0	94360.6	11.8	12.0	11.6

Note: (i) R denotes the possible number of factors. (ii) PC_{pi}'s and IC_{pi}'s are selection criteria as suggested by Bai and Ng (2002)

stationarity. Therefore, it is not surprising that in a large panel, erratic price behaviour of a few centres do not change results of panel unit root tests significantly. Our results thus highlight that more examinations at micro-level are perhaps necessary to identify erratic price movements within specific areas in India. Interestingly, panel unit root tests like Levin *et al* (2002) and Im *et al* (2003), that ignore cross-sectional dependence, suggest price convergence in Madhya Pradesh, as probability of rejection of a unit root is high in these tests.[6]

As we have mentioned earlier that the possible presence of common factor may provide a different conclusion and, therefore, we need to to decompose data into factors and idiosyncratic components. As Bai and Ng (2002) have suggested, we have considered all six criteria to select optimal number of factors that may be present in data. We have searched for over 10 possible factors. The values of the criteria corresponding to centres ($N = 76$)are presented in Table 4.[7] Interestingly all six criteria uniformly suggest presence of only one factor, irrespective of whether the tests are applied to centres ($N = 76$) or aggregated to States ($N = 24$). As the number of factors is only one and the number of idiosyncratic error terms is relatively quite large, it is likely that idiosyncratic errors will dominate the test performance. Therefore, separate tests for the common factor as well as idiosyncratic errors are quite appropriate in this context. We present Moon-Perron (2003) (MP) test and direct Dickey-Fuller (DDF) test on the estimated factor and the Robust test as developed by Breitung and Das (2004) on the series as a whole.

Table 5 summarizes panel unit roots tests under common factor structure. All three

[6]The value of the test statistic for Levin *et al* (2002) is -1.79 and for Im *et al* (2003), -2.08.

[7]Results corresponding to States are similar and are not presented here.

Table 5: Results of Various Panel Unit Root Tests under Common Factor

Regions	DDF	t_{rob}	MP	Half-Life of Common Factor	Half-Life of Local Shock
Centre (N=76)	-2.54	-6.18	-10.08	8.14	22.89
State (N=24)	-2.16	-4.12	-6.48	9.20	16.85
Critical value	-1.945	-1.945	-1.645	—	—

Note: (i) DDF, t_{rob}, MP and denote the t-statistics corresponding to direct Dickey Fuller test on the estimated principal component, Robust-OLS, Moon-Perron (2002) method, respectively. For all tests the nominal size is 0.05, (ii) Half-lives are presented in months.

tests uniformly suggest price convergence across regions/States when the series are bifurcated into common factors and idiosyncratic components.

The extent of persistence of shocks may be observed from the estimated half-lives in Table 3 and Table 5.[8] Table 3 reveals that estimated half lives for different States vary over a wide range. One plausible explanation for this phenomenon is market segmentation. However, in our case, the price series in different regions also contain non-tradables in differential proportions. Studies have found that estimated half-lives of convergence in non-tradables tend to be more than that of tradables. For example, in the case of the US economy, Parsley and Wei (1996) have found that the median rates of convergence for perishables, non-perishables and services are four, five and fifteen quarters respectively. Similar findings have also been obtained by Menna (2001) for Italy. It is plausible that in case of India the differential impact of non-tradables have led to the varied range of estimated half-lives pertaining to the States.

In Table 5, while half lives of shocks to the common factor is estimated as 8.14 and 9.20 months respectively for regional and State-specific data, the corresponding figures are 22.89 and 16.85 months for the idiosyncratic components. One possible reason of such differences of estimated half-lives between the common factor and the idiosyncratic components may be due to the presence of non-tradables in aggregate CPIIW, which are expected to have more impact on idiosyncratic errors.[9] The difference in speeds of convergence of the common factor and the idiosyncratic errors may perhaps be explained in policy terms as well. A rise in the common factor leads to an all-around increase in prices in India and immediately raises policy concerns. In contrast, reaction times to local shocks might not be as fast due to lack of sufficient media attention and in case of India, may as well indicate market segmentation at the micro-level.

[8] All half-lives in this paper are adjusted for 'Nickel Bias' of first order assuming an AR(1) process.

[9] Lack of availability of detailed commoditywise regional data on prices constrained us to examine this in further detail.

It may be noted that we have not presented any result on half-lives of common factors and idiosyncratic components within specific States. This is because the number of cross-sectional units that are required to estimate factors is quite large and given that the highest number of centres within a State is only 8 (for West Bengal) in our case, such applications are not possible. We have, however, repeated the exercise with respect to Nagpur as the common *numeraire* instead of all India average and the overall results remain unchanged.

5 The Impact of Distance on Price Deviations

This section examines the role of distance as a possible determinant for any systematic price deviations between two cities or regions in India. Prices in two different regions could be different due to transportation costs. In the literature, Engel (1993), Engel and Rogers (1996), Parsley and Wei (1996), Cecchetti *et al* (1998) and Menna (2001) have all used distance to proxy for this variable. The impact of distance on regional prices is typically analysed by carrying out a few cross-sectioal regressions. The dependent variable in these regressions is generally a measure that reflects the extent of price convergence in a pair of cities or regions over time. These measures are then regressed on distance or some monotonic transformations of it.

These regressions are often carried out with respect to a specific city or region that acts as a *numeraire*. For example, Cecchetti *et al* (1998) in the case of the US economy and Menna (2001) in case of Italy, have analysed the impact of distance by respectively considering the prices in Chicago and Rome as the *numeraire*. The independent variable was either the logarithm of distance between city i and the numeraire city or some other monotonic functions of distance. Thus, data on N cities typically lead to $(N-1)$ observations in a cross-sectional regression. It may be noted that in this approach, the impact of distance on the prices in a city pair (i, j) is ignored, unless one of them is the numeraire city.

In this study, we have considered an alternative approach that does not require the specification of any *numeraire*. For specific measures of price divergence, we find their values for all possible pairs of regions and regress it on measures relating to distance between the pair. This approach increases the number of observations in the regression dramatically. In our case, the number of observations in each cross-sectional regression is ${}^{76}C_2$, i.e., 2850.

So far as the dependent variable is concerned, we consider two measures in this study. The first, INFCORR attempts to measure the correlations of monthly rate of inflations between two cities. The second, INFGAP, is the deviation between the annual average rate of inflation between two cities during January 1995 to June 2004. Thus, the first and the second variable respectively attempt to measure short and long-run variations in prices. As explanatory variables, we consider four alternatives, viz., distance, square of distance, log of distance and double log of distance respectively. It may be noted that Cecchetti *et al* (1998) have considered all these variables except the square of distance as explanatory variables, whereas square of distance has been used by Engel (1996). Approximations of distance between pairs of cities have been obtained from the latitudes and the longitudes presented in Appendix A. Although, latitudes and longitudes yield locations on a spehere, for simplicity, we have assumed them to be planar (x, y) positions (after transforming them in decimal numbers) and approximated distance between two cities in locations (x_1, y_1) and (x_2, y_2) by the formula $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

To find out whether State-specific fiscal and administrative policies have any impact, initially we also added a dummy variable called STATEDUMMY that took the value unity when both the centres in a pair were in the same State. However, in none of the regressions, the variable turned out to be significant. Therefore, it was dropped from subsequent specifications.

The results of the different regressions are presented in Table 6. Table 6 reveals that the short-run movements in prices are more similar in nearby areas. In all the regressions in which INFCORR is a dependent variable, the explanatory variables are significant, with a negative sign. Among different regressions, the one with DISTANCE as an explanatory variable performs the best in terms of R^2.

The impact of distance on long-run price deviations, however, appears to be small. In all the regressions involving INFGAP, though explanatory variables are significant and display the expected sign, the fit in terms of R^2 is not very good. The equation with square of DISTANCE as the explanatory variable provides the best explanation, although the fit with DISTANCE as an explanatory variable comes close.

Taken together, the results indicate that transportation costs can perhaps explain a part of the variation in prices between two locations, although their overall explanatory power is small.

6 Conclusion

The paper attempted to examine whether there is any significant long-run price disparity across various regions in India. The results indicated significant presence of contemporaneous cross-sectional dependence in prices in India, rendering some of the earlier and more traditional panel unit root tests inapplicable. Using various panel unit root tests that were robust to cross-sectional dependence, evidences in favour of mean reversion of regional relative price levels were obtained. The evidence appeared to be similar when tests were restricted to specific parts of India, as well as when the number of cross-sectional units

Table 6: Regression Results of Impact of Distance on Regional Price Deviations

Dependent Variable	INTERCEPT	DISTANCE	Square of DISTANCE	Log of DISTANCE	Double-Log of DISTANCE	R^2
INFCORR	0.4550 (71.0)	-0.0101 (-19.1)				0.1138
INFCORR	0.4070 (92.0)		-0.0004 (-18.5)			0.1070
INFCORR	0.4963 (52.3)			-0.0687 (-16.7)		0.0888
INFCORR	0.4065 (75.3)				-0.0850 (-13.7)	0.0624
INFGAP	0.6912 (29.3)	0.0075 (3.88)				0.0049
INFGAP	0.7204 (44.4)		0.0004 (4.31)			0.0061
INFGAP	0.6507 (18.9)			0.0557 (3.72)		0.0045
INFGAP	0.7372 (38.0)				0.0524 (2.35)	0.0016

Note: The bracketed numbers are *t*-ratios.

were aggregated to reflect State specific measures. A decomposition of each series into a set of common factors and idiosyncratic component revealed the existence of only one common factor in case of India. The decomposition also enabled us to test stationarity and estimate half lives of the common factor and the idiosyncratic component separately. Both these components in case of India were found to be stationary. Local price shocks were, however, found to be more persistent as compared to the common factor. Further analysis indicated that proportional transportation costs could perhaps explain a part of the variation in prices between two locations, limiting the possibilities of unexploited arbitrage opportunities.

The paper ends with a few comments. It may be noted that we have worked with the aggregate CPIIW series that also contains non-tradables. As it is well known that prices of non-tradables tend to be more dispersed across regions, it is likely that a restriction to CPI on tradable commodities alone would yield stronger evidence in favour of price convergence. An interesting future research agenda would be to examine price convergence commoditywise. In particular, the decomposition of commoditywise regional prices in India into a set of common factors and local shocks and juxtapositions of their estimated half-lives would enrich the existing findings on regional price convergence.

References

Arellano, M. (1987): Computing Robust Standard Errors for Within-groups Estimators, *Oxford Bulletin of Economics and Statistics*, 49, 431-434.

Bai, J. and S Ng (2002): Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191-221.

Bai, J. and S Ng (2004): A Panic Attack on Unit Roots and Cointegration, *Econometrica*, 72, 1127-1177.

Banerjee, A., M. Marcellino and C. Osbat (2004): Testing for PPP: Should we use panel Methods? *Empirical Economics* , forthcoming.

Breitung, J. and S. Das (2003): Panel Unit Root Tests under Cross Sectional Dependence, Mimeo, University of Bonn.

Breitung, J. and S. Das (2004): Testing for Unit Roots in Panels with a Factor Structure, Mimeo, University of Bonn.

Breusch, T. S. and A. R. Pagan (1980): The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics, *Review of Economic Studies*, 47, 239-253.

Cecchetti S. G., N. C. Mark and R. J. Sonora (1998): Price Level Convergence among United States Cities: Lessons for the European Central Bank, *Working Paper No. 32, Oesterreichische Nationalbank*.

Cecchetti S. G., N. C. Mark and R. J. Sonora (2002): Price Index Convergence among United States Cities, *International Economic Review*, 43(4), 1081-1099.

Ceglowski J., (2003): The Law of One Price: International Evidence for Canada, *Canadian Journal of Economics*, 36(2), 373-400.

Chang, Y. (2002): Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, 110, 261-292.

Chang, Y. (2004): Bootstrap Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics*, 120, 263-293.

Engel, C. (1993): Real Exchange Rates and Relative Prices: An Empirical Investigation, *Journal of Monetary Economics*, 32 (August), 35-50.

Engel, C. and J. H. Rogers (1996): How Wide Is the Border? *American Economic Review*, 86 (December), 1112-1125.

Engel, C. and J. H. Rogers (2001): Violating the Law of One Price: Should We Make a Federal Case Out of It? *Journal of Money, Credit and Banking*, 33(1), 1-15.

Fan C. S. and X. Wei, (2003): The Law of One Price: Evidence from the Transitional Economy of China, *Mimeo*, Department of Economics, Lingnan University, China.

Forni, M., M. Hallin, M Lippi, and L. Reichlin (2000): The Generalized Dynamic Factor Model: Identification and Estimation, *Review of Economics and Statistics*, 82, 540-554.

Im, K. S. and M. H. Pesaran(2003): On the Panel Unit Root Tests Using Non-Linear Instrumental Variables, *Mimeo*, University of Cambridge.

Im, K. S., M. H. Pesaran and S. Shin (2003): Testing for Unit Roots in Heterogeneous Panels, *Journal of Econometrics*, 115, 53-74.

Levin, A., C. Lin and C. J. Chu (2002): Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties, *Journal of Econometrics*, 108, 1-24.

Maddala, G. S. and S. Wu (1999): A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, 61, 631-652.

Menna M. (2001): Price Level Convergence among Italian Cities: Any Role for the Harrod-Balassa-Samuelson Hypothesis? *Mimeo, University of Rome.*

Moon, R. and B. Perron (2004): Testing for Unit Root in Panels with Dynamic Factors, *Journal of Econometrics*, 122, 81-126.

O'Connell, P. (1998): The Overvaluation of Purchasing Power Parity, *Journal of International Economics*, 44, 1-19.

Pantula S. G. (1991): Asymptotic Distributions of Unit Root Tests when the Process is Nearly Stationary, *Journal of Business and Economic Statistics*, 9, 63-71.

Parsley D. and S. Wei (1996) : Convergence to the Law of One Price without Trade Barriers or Currency Fluctuations, *Quarterly Journal of Economics*, 111(4), 1211-1236.

Schwert, W. G. (1989): Tests for Unit Roots: A Monte Carlo Investigation, *Journal of Business and Economic Statistics*, 7, 147-160.

Taylor A. M. (2001): Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price, *Econometrica*, 69(2), 473-498.

Appendix A: Cities/Towns/Regions from which the Data on CPI are Collected in India

State	Cities / Towns/ Regions
Andhra Pradesh (6)	Gudur ($14^o08'$, $79^o51'$), Guntur ($16^o18'$, $80^o27'$), Hyderabad ($17^o23'$, $78^o29'$), Kethgudem ($17^o40'$, $80^o56'$), Visakhapatnam ($17^o42'$, $83^o18'$), Warangal ($18^o00'$, $79^o35'$)
Assam (5)	D. D. Tinsukia ($27^o30'$, $95^o22'$), Guwahati ($26^o11'$, $91^o44'$), Labac-Silchar ($24^o49'$, $92^o48'$), Mariani-Jorhat ($26^o45'$, $94^o13'$), Rangapara-Tezpur ($26^o38'$, $92^o48'$)
Bihar (1)	Monghyr ($25^o24'$, $86^o30'$)
Chhatisgarh (1)	Bhilai ($21^o13'$, $81^o26'$)
Gujarat (5)	Ahmedabad ($23^o02'$, $72^o37'$), Bhavnagar ($21^o48'$, $72^o06'$), Rajkot ($22^o18'$, $70^o47'$), Surat ($21^o10'$, $72^o50'$), Vadodara ($22^o18'$, $73^o12'$)
Hariana (2)	Faridabad ($88^o26'$, $77^o19'$), Yamunanagar ($30^o07'$, $77^o18'$)
Jammu and Kashmir (1)	Srinagar ($34^o05'$, $74^o49'$)
Jharkhand (5)	Jamshedpur ($22^o48'$, $86^o11'$), Jharia ($23^o45'$, $86^o24'$), Kodarma ($24^o28'$, $85^o36'$), Noamundi ($22^o09'$, $85^o32'$), Ranchi-Hatia ($23^o21'$, $85^o20'$)
Karnataka (4)	Bangalore ($12^o59'$, $77^o35'$), Belgaum ($15^o52'$, $74^o31'$), Hubli-Dharwar ($15^o21'$, $75^o10'$), Mercara ($12^o25'$, $75^o44'$)
Kerala (4)	Alwaye ($10^o07'$, $76^o21'$), Mundakayam ($9^o36'$, $76^o34'$), Quilon ($8^o53'$, $76^o36'$), Thiruvananthapuram ($8^o29'$, $76^o55'$)
Madhya Pradesh (5)	Balaghat ($21^o48'$, $80^o11'$), Bhopal ($23^o16'$, $77^o24'$), Chhindwara ($22^o04'$, $78^o56'$), Indore ($22^o43'$, $75^o50'$), Jabalpur ($23^o10'$, $79^o57'$)
Maharashtra (5)	Mumbai ($19^o00'$, $72^o48'$), Nagpur ($21^o09'$, $79^o06'$), Nasik ($19^o59'$, $73^o48'$), Pune ($18^o32'$, $73^o52'$), Solapur ($17^o42'$, $75^o48'$)
Orissa (2)	Barbil ($22^o06'$, $85^o20'$), Rourkela ($22^o13'$, $84^o53'$)
Punjab (2)	Amritsar ($31^o35'$, $74^o53'$), Ludhiana ($30^o54'$, $75^o51'$)
Rajasthan (3)	Ajmer ($26^o27'$, $74^o38'$), Bhilwara ($25^o21'$, $74^o38'$), Jaipur ($26^o55'$, $75^o49'$)
Tamil Nadu (6)	Chennai ($13^o05'$, $80^o17'$), Coimbatore ($11^o00'$, $76^o58'$), Coonor ($11^o21'$, $76^o49'$), Madurai ($9^o56'$, $78^o07'$), Salem ($11^o39'$, $78^o10'$), Tiruchirapally ($10^o49'$, $78^o41'$)
Uttar Pradesh (5)	Agra ($27^o11'$, $78^o10'$), Ghaziabad ($28^o40'$, $77^o26'$), Kanpur ($26^o28'$, $80^o21'$), Saharanpur ($29^o58'$, $77^o33'$), Varanasi ($25^o20'$, $83^o00'$)
West Bengal (8)	Asansol ($23^o41'$, $86^o59'$), Darjeeling ($27^o02'$, $88^o16'$), Durgapur ($23^o29'$, $87^o20'$), Haldia ($22^o06'$, $88^o06'$), Howrah ($22^o35'$, $88^o20'$), Jalpaiguri ($26^o31'$, $88^o44'$), Kolkata ($22^o34'$, $88^o21'$), Raniganj ($25^o52'$, $87^o52'$)
Chandigarh (1)	Chandigarh ($30^o44'$, $76^o55'$)
Delhi (1)	Delhi ($28^o39'$, $77^o13'$)
Pondichery (1)	Pondichery ($11^o56'$, $79^o53'$)
Himachal Pradesh (1)	The entire industrial belt of the State ($31^o06'$, $77^o10'$)
Tripura (1)	The entire industrial belt of the State ($23^o49'$, $91^o16'$)
Goa (1)	The entire industrial belt of the State ($15^o29'$, $73^o50'$)

Note: (i) As all regions in India are at the North of the equator and East of the Greenwich meridian, the N in latitude and the E in the longitude have not been mentioned after the respective numbers.(ii) For the States, Himachal Pradesh, Tripura and Goa, the latitudes and longitudes of the State capitals have been considered. Areas of these States are small compared to many other States in India. (iii) For Kethgudem and Mundakayam, locations of the nearest railway stations (Bhadrachalam and Kottayam respectively) have been considered.