Households or locations? Cities, catchment areas and prosperity in India

Yue Li and Martin Rama
World Bank
July 13, 2015
Motivation and approach
(Some) cities are drivers of prosperity in India

• Because most of the poor live in rural areas, and poverty rates are lower in cities, development programs have traditionally focused on rural areas, emphasizing skills and transfers.

• However, as the country urbanizes the distinction between rural and urban areas becomes blurred. The functioning of cities, and distance to them, matters for both urban and rural prosperity.

• The goal of this paper is to reassess the contribution the spatial dimension makes to prosperity, and develop a metric to identify what makes for top, average and bottom locations.
Poverty analysis *versus* urban economics

- **Poverty analysis** had traditionally focused on household characteristics (health, education) and government programs (social protection) as determinants of consumption.

- **Poverty maps** put more emphasis on the spatial dimension, mainly captured through location characteristics (local infrastructure and service delivery) but take them as residuals to household characteristics.

  (e.g. Demombynes et al. 2002, Elbers, Lanjouw and Lanjouw 2003, Hentschel, Lanjouw et al. 2000)

- **Urban economics** focuses on local externalities (agglomeration effects and congestion effects) as determinants of city productivity and earnings as well as city size.

  (e.g. Glaeser and Mare 2001, Wheaton and Lewis 2002, Moretti 2004a, Combes, Duranton, and Gobillon 2008, Combes et al. 2010)
Poverty analysis *meets* urban economics

Poverty analysis (as in poverty maps):

\[
\ln\left( \text{Real expenditure per capita} \right)_{il} = \alpha + \beta \cdot \left( \text{Household characteristics} \right)_{il} + \gamma \cdot \left( \text{Location characteristics} \right)_{il} + \mu_{l} + \epsilon_{il}
\]

Urban economics (as in the geography of jobs):

\[
\ln(\text{Nominal wage})_{jc} = \alpha + \beta \cdot \left( \text{worker characteristics} \right)_{j} + \gamma \cdot \left( \text{Agglomeration related location characteristics} \right)_{jc} + \epsilon_{jc}
\]

Our approach:

\[
\ln\left( \text{Nominal expenditure per capita} \right)_{il} = \alpha + \beta \cdot \left( \text{Household characteristics} \right)_{il} + \gamma_{l} + \epsilon_{il}
\]

Location effects, $\gamma_{l}$, to estimate spatial variations of productivity.

The empirical strategy
Defining locations

• Data are from the Household Consumption module of the National Sample Survey 2011-12, which uses stratified multi-staged sampling.

• Following Chatterjee, Murgai and Rama (2015), we classify first stage units (villages and urban frame survey blocks) into four groups:
  – Small rural (with less than 5,000 inhabitants)
  – Large rural (more than 5,000)
  – Small urban (with less than one million inhabitants)
  – Large urban (more than one million)

• The final sample includes 1,406 places from 599 districts of 31 states or UTs: 579 are for small rural, 221 for large rural, 581 for small urban, and 25 for large urban.
Matching spatial and household survey data

- The shape file of administrative boundaries are created based on the Administrative Atlas of India 2011 at the district level.
- The Atlas contains 640 polygons corresponding to the 640 districts defined by the Census 2011.
- To match the NSS 2011-12, Delhi districts are merged into one polygon, and Mumbai and Mumbai suburban into another.
- The 599 districts were matched with the districts defined by Census 2011.
- We do not conduct analysis for districts formed after the sampling frame of NSS 2011-12 was defined.
Measuring the distance between places

• Uncovering spatial correlations between location effects $\gamma_j$ requires a measure of the distance between places.

• The distance between places within the same district (of different population groups) is set equal to zero.

• The distance between places in different districts is set equal to the pairwise distance between the corresponding districts.

• For every pair of districts, distance is defined as the length of the shortest surface-level curve between their centroids.
The characteristics of the places

• Information on places is drawn from work in progress on the Spatial Database for South Asia, which covers four administrative levels (down to ward or village).

• The Database combines the Census of India, other NSS modules and rounds, the Economic Census, administrative records, remote sensing data and open source data.

• It provides a wide range of socioeconomic indicators, include but not limited to urban extent, employment, infrastructure, education, health, and environment.
Spatial correlations between residuals are small
Spatial correlations between location effects are significant.
Location matters
The more granularity, the more locations matter...

<table>
<thead>
<tr>
<th>OLS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>District</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Place</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditure</td>
<td>0.388</td>
<td>0.388</td>
<td>0.388</td>
<td>0.388</td>
</tr>
<tr>
<td>Total</td>
<td>0.200</td>
<td>0.217</td>
<td>0.236</td>
<td>0.241</td>
</tr>
<tr>
<td>Household</td>
<td>0.200</td>
<td>0.176</td>
<td>0.165</td>
<td>0.153</td>
</tr>
<tr>
<td>Location</td>
<td>0.020</td>
<td>0.041</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>0.021</td>
<td>0.030</td>
<td>0.038</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditure 100.00 100.00 100.00 100.00</td>
</tr>
<tr>
<td>Total 51.50 55.97 60.77 62.15</td>
</tr>
<tr>
<td>Household 51.50 45.44 42.49 39.50</td>
</tr>
<tr>
<td>Location 0.00 5.12 10.57 12.85</td>
</tr>
<tr>
<td>Interaction 0.00 5.41 7.70 9.79</td>
</tr>
</tbody>
</table>

Growing granularity

A greater share explained by locations and interaction

Higher explanatory power
Defining location more narrowly changes household effects

With location effects (absolute values)

Without location effects

OLS
In particular, caste and religion matter less.
Cities and catchment areas
A rural-urban gradation of location effects

Two-sample t tests and Kolmogorov-Smirnov tests show that the means and the distributions of location fixed effects are significantly different between small and large rural areas and between small and large urban areas, but not between large rural and small urban areas.
Spatial correlations between location effects are significant

- OLS, correlations w.r.t. small rural
- OLS, correlations w.r.t. small urban
- OLS, correlations w.r.t. large rural
Top locations are spatially correlated

Delhi-Faridabad

But Delhi’s surroundings are more productive than Bangalore’s

And as a result Delhi’s catchment area is much bigger than Bangalore’s

Bangalore

Bangalore/large urban is more productive than Delhi/large urban

nominal consumption based, OLS
Four “tiers” of locations, across all of India

- **Top locations**: Top 100 locations, urban or rural.
- **Catchment areas**: contiguous to a top location or another catchment and > 1 sd. above the mean.
- **Average locations**: above -1 sd. below the mean but neither top nor catchment.
- **Bottom locations**: below -1 sd. below the mean.
Three steps to identify clusters

Clusters of the cores
- Places with the 100 highest location effects are the cores.
- Contiguous top 100 places form clusters of the cores.

Catchment areas
- Places with location effects greater than one standard deviation above the mean are potential catchment areas.
  - Among the potential ones, those contiguous with the cluster of the cores are catchment areas.
  - Among potential ones, those contiguous with the catchment areas identified above are also catchment areas.

Final clusters of the top locations and their catchment areas
- Contiguous cores and their contiguous catchment areas form the final clusters of top location and their catchment areas.
17 clusters of top locations and catchment areas across India

And the worst locations tend to be contiguous too
Bottom small rural places are NOT on the Ganga

Bottom small urban places ARE on the Ganga

Color intensity indicates size of location effect
Location and inclusion
Clusters include administratively rural places

- Total
- Bottom locations
- Average locations
- Catchment areas of top locations
- Top locations

Legend:
- Blue: Rural population share
- Orange: Urban population share
Population shares are similar in large rural and in small urban areas.
In clusters, location effects are similar for rural and urban places.
Higher location effects do not always imply greater inequality

Clusters and large urban

All places

- clusters of top locations and catchment areas
- average, large urban
- average, small urban, large rural and small rural
- bottom locations

OLS
There is a positive sorting of households across places

![Graph showing average expenditure versus location effects with lines indicating slopes of 1.03 and 1.63.](image)

- slope = 1.03
- slope = 1.63

Symbols:
- Red circles: top locations and catchment areas
- Purple squares: average, large urban
- Green asterisks: average, small urban, large rural and small rural
- Yellow triangles: bottom locations

OLS
Top locations and catchment areas are associated with factors that are believed to drive agglomeration economies and faster productivity growth

Catchment areas resemble top locations except in road and population density

<table>
<thead>
<tr>
<th></th>
<th>Bottom locations</th>
<th>Average locations</th>
<th>Catchment areas</th>
<th>Top locations</th>
<th>Unit of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of population density (people/sqkm)</td>
<td>5.761***</td>
<td>5.831***</td>
<td>6.007*</td>
<td>6.358</td>
<td>district</td>
</tr>
<tr>
<td>Employment structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of main workers (% of total workers)</td>
<td>67.241***</td>
<td>73.000***</td>
<td>79.799</td>
<td>82.098</td>
<td>district</td>
</tr>
<tr>
<td>Share of marginal workers (% of total workers)</td>
<td>32.759***</td>
<td>27.000***</td>
<td>20.201</td>
<td>17.902</td>
<td>district</td>
</tr>
<tr>
<td>Regular wage employment share (% of total labor force)</td>
<td>12.888***</td>
<td>20.683***</td>
<td>29.085</td>
<td>31.107</td>
<td>place</td>
</tr>
<tr>
<td>Economic sector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture sector employment share (% of total labor force)</td>
<td>47.847***</td>
<td>39.230***</td>
<td>28.375</td>
<td>28.896</td>
<td>place</td>
</tr>
<tr>
<td>Manufacturing sector employment share (% of total labor force)</td>
<td>7.881***</td>
<td>11.879***</td>
<td>16.995</td>
<td>15.694</td>
<td>place</td>
</tr>
<tr>
<td>Skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of people with secondary education (% total working age population)</td>
<td>25.117***</td>
<td>26.795***</td>
<td>33.885</td>
<td>33.277</td>
<td>place</td>
</tr>
<tr>
<td>Share of people with tertiary education (% total working age population)</td>
<td>8.168***</td>
<td>10.170***</td>
<td>12.538</td>
<td>14.368</td>
<td>place</td>
</tr>
<tr>
<td>Infrastructure and services</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road density (km/sqkm)</td>
<td>0.163***</td>
<td>0.284***</td>
<td>0.488**</td>
<td>0.751</td>
<td>district</td>
</tr>
<tr>
<td>Access to electricity (% of total households)</td>
<td>48.780***</td>
<td>64.151***</td>
<td>8.45</td>
<td>89.87</td>
<td>district</td>
</tr>
<tr>
<td>Access to cellphone (% of total households)</td>
<td>44.551***</td>
<td>56.547***</td>
<td>69.889</td>
<td>72.594</td>
<td>district</td>
</tr>
<tr>
<td>Access to banking services (% of total households)</td>
<td>51.288***</td>
<td>57.771***</td>
<td>66.775</td>
<td>68.118</td>
<td>district</td>
</tr>
<tr>
<td>Social and religious affiliations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of Scheduled Tribes (% total population)</td>
<td>20.844***</td>
<td>12.182</td>
<td>6.849</td>
<td>9.215</td>
<td>place</td>
</tr>
<tr>
<td>Share of Scheduled Castes (% total population)</td>
<td>16.197</td>
<td>16.396**</td>
<td>20.829</td>
<td>19.989</td>
<td>place</td>
</tr>
</tbody>
</table>

Note: Difference with top locations *significant at 0.1 level, ** significant at 0.05 level, *** significant at 0.01 level

Scheduled Tribes are concentrated in bottom locations