WPS6919

Policy Research Working Paper 6919

Regional Diversity and Inclusive Growth in Indian Cities

Ejaz Ghani
William R. Kerr
Ishani Tewari

The World Bank
Poverty Reduction and Economic Management Network
Economic Policy and Debt Unit
June 2014

Policy Research Working Paper 6919

Abstract

This paper examines the employment growth of Indian districts from 2000 to 2010 in the manufacturing and services sectors. Specialization and diversity metrics that combine industries in both sectors are calculated and related to subsequent job growth. The analysis finds robust and consistent evidence that the diversity of industries in the district across the two sectors links to subsequent job growth. Somewhat surprisingly, this link finds its strongest expression outside typical stories about the role of diversity. For example, the growth is strongest in rural areas of districts and in districts with low population density. Diversity correlates with disproportionately higher employment growth in the informal sector and plays a role in generating employment in the district's smaller industries. These findings point toward the "inclusive" nature of diversity-driven growth and highlight a potentially important agenda item for policy makers concerned with inclusive development.

This paper is a product of the Economic Policy and Debt Unit, Poverty Reduction and Economic Management Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at EGhani@worldbank.org.

Regional Diversity and Inclusive Growth in Indian Cities

Ejaz Ghani, William R. Kerr, and Ishani Tewari

Keywords: India, specialization, diversity, growth, development, employment, urban, services, manufacturing.
JEL: D24, L10, L25, L26, L60, L80, O10, O40, R00, R10, R11, R12

Author institutions and contact details: Ghani: World Bank, Eghani@worldbank.org; Kerr: Harvard University, Bank of Finland, and NBER, wkerr@hbs.edu; Tewari: Yale School of Management, ishani.tewari@yale.edu.

Acknowledgments: We are grateful to Gilles Duranton and Stephen O'Connell for helpful suggestions/comments and data calculations. Funding for this project was provided by the World Bank..

1. Introduction

The continued development of India's cities and regions is a topic of exceptional importance, impacting the lives of over a billion people. Many academics and policy makers are looking for the right local recipes for success. A number of publications point out the challenges ahead (e.g., World Bank, 2013) and concern is often expressed about whether employment and job growth can keep pace with the anticipated demand. India is seeking to boost its growth from 5% to 10% and increase its rate of urbanization, which has lagged behind its peers. McKinsey Global Institute (2010) contends that cities in India and South Asia have the potential to generate 70% of the country's new jobs and GDP over the next 20 years, a process that could drive a four-fold increase in per capita incomes. While promising and greatly desired, such urbanization imposes unprecedented challenges, including needed infrastructure investments and policy reforms. Perhaps more fundamentally, we do not yet know what industrial structures of cities have the best connection to expansive, inclusive growth and the creation of new jobs for a broad base of workers.

This project considers a simple but important question for this end. To what extent does the local specialization or diversity of districts link to employment growth in India since 2000? This is in many respects a very old question dating back to Alfred Marshall and Jane Jacobs. Marshall (1890) established the field of agglomeration and the study of clusters by noting the many ways in which similar firms from the same industry can benefit from locating together. Glaeser et al. (1992) famously described how Marshall-Arrow-Romer (MAR) type externalities predict that spillovers come from within industries, especially when concentration is high, and can lead to local job growth. Porter (1990) also holds that spillovers emerge from within industries, but only in the presence of competition. Jacobs (1969), on the other hand, pushed back against these perspectives. She emphasized how knowledge flows across industries and that industrial variety and diversity is conducive to growth. Haussmann (2014) strongly advocates against local specialization.

While these issues have been widely discussed over the past two decades, our understanding of their practical application to an environment like India is very limited. Duranton and Puga (2000) began a line of research for advanced economies, but there has been very little work quantifying specialization and diversity for developing economies, and this has

rarely then been connected to employment growth relationships. In prior work, Ghani, Kerr and Tewari (2013) started down this path for Indian manufacturing. This earlier work measured specialization and diversity for the manufacturing sector and compared it to the United States. It found that specialization and diversity jointly enhanced manufacturing productivity—the subtle distinctions between the measures are described further below—and that initial manufacturing specialization in 1989 was a strong predictor of employment growth over the 1989 to 2010 period for Indian districts.[1]

This study makes two central improvements on the earlier work. First, and most important, it provides a more comprehensive view of the local industrial sectors in India by bringing together manufacturing and services data, which in the Indian context also includes industries often grouped with retail trade or transportation. Services have been very important for India's development, and notions of local specialization and diversity extend well beyond the manufacturing sector. The broader sector reach limits us to an analysis of employment growth since 2000, but this shorter time span is strongly compensated for by the more complete perspective on local specialization and diversity. In Sections 2 and 3 of this paper, we conduct many of the metrics and analyses carried out by Ghani, Kerr and Tewari (2013) to identify and convey how these patterns fit with those depicted earlier for the manufacturing sector.

We next use the developed metrics to consider employment growth at the district level across the 2000 to 2010 period. We conduct this analysis at many levels and with multiple decompositions in Section 4. We have two key findings. First, overall employment is strongly linked to broader district diversity. This pattern is distinct from our earlier paper that identifies manufacturing specialization in 1989-1994 as a key correlate of manufacturing growth across the 1989 to 2010 period, and we conduct several analyses to show their compatibility. This broader view of the local industrial base was important to discern the role of diversity and is highly robust. We also find that initial clusters of modern services experienced abnormally high

[1] Some earlier studies have provided evidence through productivity estimations and related empirical work. Indian studies include Lall et al. (2004), Lall and Mengistae (2005), Deichmann et al. (2008), and Fernandes and Sharma (2011). Duranton and Puga (2004) and Rosenthal and Strange (2004) provide broad background to the agglomeration literature. Acemoglu and Zilibotti (1997) and Imbs and Wacziarg (2003) are examples of work on national diversification, and Kaulich (2012) provides an extensive survey and empirical evidence.

employment growth in the post 2000 period. Furthermore, diversity, but not specialization, is closely tied to generating employment in high-growth industries.

Second, our extensions uncover some interesting nuances. We find that the strongest expressions of this diversity-growth link are in settings not often linked to the theories of diversity's benefits. The effects are sharper in rural areas of districts, in districts with low population density, and among the unorganized/informal sector. This contrasts with studies emphasizing the urban impact of diversity and its roles for fostering innovation (e.g., Henderson, 1997a; Duranton and Puga, 2001; Feldman and Audretsch, 1999), promoting urban entrepreneurship (Jacobs, 1969), or reducing costs (Lall et al., 2003). Duranton (2013a) provides a recent review in the developing country context. Looking across industries, higher diversity is associated with lower concentration of employment growth in a few top industries. In other words, more diversified districts tend generate growth that is spread more broadly across industries. While we are not able yet to quantify all of the channels through which initial diversity is linked to local growth, we show in many ways the depth and breadth of local growth that diversity connects to. The final section concludes with some thoughts about how these mechanisms can be further explored.

These patterns uncovered by our analysis suggest that the employment growth correlated with district diversity was "inclusive" in nature, that is, with disproportionately larger effects in less developed areas, sectors and industries. Recent empirical evidence (Anand, Tulin and Kuman, 2014) shows that high growth in the past decade has fueled poverty reduction, especially so in rural areas. Much of the existing work documenting the distributional aspects of economic growth in India links it to social spending especially on education, bank credit, and other macroeconomic policies like trade liberalization (Topalova, 2010). Although in absolute terms poverty levels in India have fallen, there still remain large disparities in relative levels of consumption across states as well as urban and rural areas. Thus, policy makers continue to maintain a close focus on the distributional aspects of economic growth. The Eleventh and Twelfth five years plans strongly reflect this strong agenda of inclusive development. In this paper, we add a novel correlate of employment growth—industrial diversity of districts—to policy agendas centered on inclusive growth. Diversity in industrial bases can aid India's efforts to double its growth rate from 5% to 10% and to do so in a broad and inclusive manner.

2. Indian Establishment Data and Index Calculations

This section begins with a description of the Indian manufacturing and services data that we use in our study. We then outline how we calculate our indices of district-level specialization and diversity and describe some of the important empirical properties of these indices. These depictions draw on the work in Ghani, Kerr and Tewari (2013), and we note in particular the deviations that occur in our present study compared to the earlier work. The earlier paper also showed the broad convergence and comparability of India's specialization and diversity values to those in the United States for the manufacturing sector in recent periods, which we do not carry forward to this study.[2]

Indian Manufacturing Data

We employ repeated cross-sectional surveys of manufacturing establishments carried out by the government of India for the fiscal years of 2000, 2005, and 2010. In all cases, the survey was undertaken over two fiscal years (e.g., the 2000 survey was conducted during 2000-2001), but we will only refer to the initial year for simplicity. The organized and unorganized sectors of Indian manufacturing are surveyed separately, as described next. In every period except the last one, our surveys for the two organized and unorganized sectors were undertaken contemporaneously. In the last period, we combine the 2009-2010 survey for the organized sector with the 2010-2011 survey for the unorganized sector. We will again refer to this period as 2010 for simplicity.

The organized sector comprises establishments with more than 10 workers if the establishment uses electricity. If the establishment does not use electricity, the threshold is 20 workers or more. These establishments are required to register under the India Factories Act of 1948. The unorganized manufacturing sector is, by default, comprised of establishments which fall outside the scope of the Factories Act. The organized sector accounts for over 80% of India's manufacturing output, while the unorganized sector accounts for over 80% and 99% of Indian

[2] Throughout this study, we use the term "sector" in two ways: to signify manufacturing and services sectors broadly defined and/or organized and unorganized establishment groups. We use the term "industry" to denote individual industries (e.g., textiles; hotels and restaurants) within these sectors.

manufacturing employment and establishments, respectively (Ghani, Kerr and O'Connell, 2013a).

The organized manufacturing sector is surveyed by the Central Statistical Organization through the Annual Survey of Industries (ASI). Our data for the unorganized sector come from the National Sample Statistics (NSS). These surveys are used for many published reports on the state of Indian businesses and government agency monitoring of the Indian economy. The typical survey collects data from over 150,000 Indian establishments. In this respect, the surveys are comparable to the Annual Survey of Manufacturing conducted in the United States, with the Indian sampling frame being about three times larger.

Establishments are surveyed with state and four-digit National Industry Classification (NIC) stratification. The surveys provide sample weights that we use to construct population-level estimates of employment by district. Districts are administrative subdivisions of Indian states or territories that provide meaningful local economic conditions and policy choices. The average district size is roughly twice the size of a U.S. county (average size of ~5,500 square kilometers) and there is substantial variability in district size (standard deviation of ~5,500 square kilometers). Indian districts can be effectively considered as self-contained labor markets and, to some degree, economic units.

Our surveys record economic characteristics of plants, such as employment, output, and raw materials. Our analysis considers aggregated measures of manufacturing employment in locations. For this purpose, we sum the activity of plants up to the district or district-industry level, combining the organized and unorganized sectors and using sample weights to prepare population estimates. We use the two-digit level of the NIC system for calculating industrial specialization and diversity for districts. This level of aggregation contains 22 manufacturing industries. This level of aggregation meaningfully balances the identification of distinct industries and economic activities with not being so granular as to create exception outliers in specialization measures.

Our core sample contains 554 districts. This district count is larger than that analyzed in Ghani, Kerr and Tewari (2013), which included 429 districts, for two reasons. First, by considering the 2000 to 2010 period, we do not need to restrict the sample to districts that were surveyed in the 1980s and 1990s as in our longer panel. Second, we retain districts that record

services employment (described next) even without a registered manufacturing base (and vice versa). In some analyses below, we focus in greater detail on the growth of specific sectors, with additional restrictions in sample size.

Even with these expansions, the current sample is smaller than the total number of districts in India of 630. Nevertheless, the sample accounts for almost all of the employment in the manufacturing sector throughout the period of study. The reductions from the 630 baseline occur due to requirements that manufacturing or services employment be observed in every period for the district (i.e., we have balanced panels of districts from 2000 to 2010). Even with these requirements, some districts have a small number of observations, and this could be worrisome given that our data do not constitute a complete census of Indian businesses and have state-level survey stratification. We consider several checks below (e.g., excluding smaller districts) to verify that the results discussed are robust to these considerations.

Indian Services Data

We also consider the services sector in this study. In the services sector, there is no legal distinction between organized and unorganized establishments as in manufacturing. Service establishments, regardless of size or other characteristics, are not required to register and thus are all officially unorganized. The NSS surveys for the services sector are thus comprehensive. Establishments are again surveyed with state and four-digit NIC stratification. The surveys provide sample weights that we use to construct population-level estimates of services employment by district in 2000, 2005, and 2010. We consider 16 two-digit industries within the services sector for most of our empirical work, similar to manufacturing. Additional data development steps are comparable to those described for the manufacturing sector. It is worth noting that the Indian services sector has a relatively broad definition, such that the NIC classification structure contains industries that in other settings would be considered part of retail trade (e.g., restaurants) or transportation.

Specialization and Diversity Indices

We follow Duranton and Puga (2000) and Ghani, Kerr and Tewari (2013) in the indices that we use to measure specialization and diversity for Indian districts. Before providing the

formulas, we outline the key building blocks required. We index districts by *d* and industries by *j*. We calculate $s_{d,j}$ as the share of employment in district *d* contained in industry *j* (thus, $s_{d,j}$ shares sum to 100% across all industries for each district, inclusive of both manufacturing and services). Industries come in many different sizes, and we often normalize the size of industries by their overall presence nationally. To do so, we calculate s_j as the share of employment nationally in industry *j* (thus, s_j shares sum to 100% across all industries for India as a whole).

We divide $s_{d,j}$ by s_j as normalization for our first index on specialization. By doing so, we identify the extent to which an industry *j* is more or less represented in district *d* than what would have been expected had the industry been distributed across districts strictly in accordance to the industry's relative proportions for India as a whole. Said differently, we compare the observed industry distribution in district *d* to what we would have anticipated had the district contained an industry distribution that mimicked India's overall manufacturing and services base. By definition, this ratio is bounded at zero for each district-industry, which occurs if the district does not contain any employment for a given industry. The metric does not have a strict upper bound and is a function of both district sizes and industry shares. A very large value of this ratio can occur in small industries that dominate employment in a very specialized district. We discuss this feature further below.

We measure the relative specialization of a district through the formula:

$$Specialization_d = \max_j \left(\frac{s_{d,j}}{s_j} \right).$$

In this formula, we look across industries within each district to find the highest value of the employment share ratio. This highest value can occur in either a manufacturing or services industry. By definition, the specialization index of a district must be greater than or equal to one. To see this, note that if a district exactly mirrors India as a whole, the ratio $s_{d,j} / s_j$ is equal to one for every industry; thus the maximum ratio observed in the district is also one. If any employment share is then reallocated from one industry to another, then one of those two industries will have a ratio that exceeds one, yielding a maximum value that is greater than one.

The maximum value of the district specialization index has the properties of the individual ratios described above.

We also measure the relative diversity of a district through the formula:

$$Diversity_d = \frac{1}{\sum_j \left| s_{d,j} - s_j \right|}.$$

In this formula, we first calculate the absolute difference between $s_{d,j}$ and s_j to measure the degree to which a given industry is over- or under-represented in the district on a share basis. We then sum across industries. This sum represents the share of the district's employment that would need to be reallocated across industries in order for the district to have the same industrial employment proportions as India does nationally (double counting the deviations).

We then take the inverse of this sum such that a larger value of the diversity index indicates that less employment needs to be reallocated in order for the district to resemble India as a whole. Considering the extreme values of the index can again illustrate its properties. If a district has all of its manufacturing or services employment in one industry that is very small in size nationally, the denominator of the index becomes large, starting to approach 200%. In such cases, the diversity index as a whole takes a very small value that approaches zero. On the other hand, if the district exactly mirrors India as a whole, then the denominator of the index becomes very small, staring to approach 0%. In these cases, the diversity index as a whole takes a very large value, indicative of substantial spread in the employment of a district across industries.[3]

[3] It is worth noting two other properties of the diversity index. First, some measures of concentration like an HHI index or the Ellison and Glaeser (1997) geographic concentration metric consider squared deviations. This approach penalizes distances from the overall baseline in a non-linear manner compared to our linear approach. Thus, our framework treats as equal one industry being 10% away from its national share or ten industries being 1% away from their national shares. Second, we measure deviations from the national employment shares of industries for India, with the maximum diversity being achieved when a district looks like India as a whole. An alternative approach would be to compare industry shares to what one would achieve if every industry held the same proportion of employment in the district, regardless of India's overall sizes for industries.

Empirical Application of the Specialization and Diversity Indices

The next section provides descriptive statistics of these metrics for Indian manufacturing and services, and we highlight first three important issues for our empirical application. First, it is important to note that the indices are related to each other but also not redundant. The two indices measure different things. The specialization index identifies the extreme value for a district across all industries, while the diversity index considers the district's composition as a whole. It is quite feasible for a district to have specialization and diversity values that are both above average. In the 2000 Indian data, about 20% of observations have above-median values on both indices, while 20% of observations have below-median values on both indices. The metrics have on average a correlation of about -0.14 across the surveys. This correlation is statistically significant at a 10% level, such that higher values of the specialization index are associated with lower district-level diversity values. The correlation, however, is sufficiently low that we can model the two indices together.

Second, the metric designs (especially specialization) can yield extreme values. This is even more prominent in this study compared to our prior work because we have incorporated more industries, and the specialization metric takes the maximum of abnormal concentration. Likewise, the sampled nature of our Indian data may contribute to outliers among small districts. Thus, we winsorize our specialization and diversity metrics at their 2% and 98% values. This winsorization is done separately for each year. We use these winsorized values for descriptive statistics in the next section, except where noted otherwise, to focus on the more meaningful variations and trends in the data. As in our prior work, when we move to growth estimations, we primarily report results that use a ten-point scale for the decile in which a district's specialization or diversity falls compared to the whole set of Indian districts. For example, a district receives a value of three if its specialization level is between the 30th and 39th percentile for India. This approach makes the scales and variances of our indices more comparable and aids in interpretation, although we derive quite similar results with other approaches.

Finally, it is important to note that the indices do not directly relate to or build upon other properties of districts (e.g., size, income per capita, etc.). Both measures are calculated over industry distributions within districts and thus do not build upon these features specifically. This

is not to say, however, that the indices are orthogonal to these properties either. For example, it is increasingly difficult for large districts with lots of employment to have an exceptionally undiversified industrial base, while it is easier for large districts to maintain specialization in one industry. Thus the connections of our indices to these properties are intriguing, and we quantify these relationships below.

3. The Evolution of Indian Specialization and Diversity

This section describes the levels and trends in the specialization and diversity of Indian districts when considering both manufacturing and services. Duranton (2013b) considers specialization and diversity in Colombia across multiple sectors, but these types of analyses have otherwise been very rarely undertaken. We devote attention to showing the patterns as they compare to our longer series in the prior work that considered manufacturing only.

Levels, Trends and Persistence of Indian Specialization and Diversity

Table 1a begins with descriptive statistics for our indices by survey. Panel A provides the specialization index, and Panel B provides the diversity index. Within each panel and survey year, we provide the mean, standard deviation, min, max, and median values across our 554 districts. The statistics are calculated after the winsorization noted earlier. In addition to these annual values, we calculate the average specialization and diversity for a district across the three periods. We also calculate the change in specialization and diversity for a district in relative terms to its initial value. The last two columns provide summary statistics for these metrics. Similar depictions are found when measuring changes relative to average values over the period.

Looking at 2005, the average specialization value is 10.66, and the median value is 6.55. This suggests that, for the average Indian district, the maximum degree to which one industry's employment exceeds its national share is around 10-fold (e.g., the industry constitutes 10% of the district's employment relative to 1% nationally). There is a wide variance in this metric, with the standard deviation greater than the mean. The lowest winsorized level in 2005 is 2.9 (e.g., a 3% local share combined with 1% nationally), while the maximum winsorized level is 80 (e.g., a 80% local share combined with 1% nationally).

The values are substantially higher than those observed by Ghani, Kerr and Tewari (2013) for the manufacturing sector only. For 2005, the earlier study found a mean value of 6.65 and a median value of 4.49. This study also discussed the comparability of the 6.65 value to what is observed with U.S. manufacturing data. The higher values in this study for specialization indices are not surprising. As discussed in the methodology section above, the specialization index measures the extreme value observed over a set of industries. By almost doubling the underlying set of industries, we greatly increase the potential for the metric to capture larger values.

Figure 1a compares the mean and median series for manufacturing only compared to manufacturing and services. While the levels differences exist, the trends in the post 2000 period are broadly comparable, especially for the median. An important subject of Ghani, Kerr and Tewari (2013) was the dramatic decline in specialization levels from manufacturing in 1989. Around the time of liberalization, India's manufacturing distribution displayed more specialization and greater spatial variation in specialization rates than the U.S. data. India has converged towards the U.S. baseline in the two decades since then. As the services data only begin in 2000, we cannot observe this longer trend in our current work.

Panel B of Table 1a considers the relative diversity index. The average diversity in 2005 is 1.38, which is very close to the 1.41 derived with just manufacturing data in the earlier study. Unlike the construction of the specialization index, it is quite possible for the diversity index to decline in average value following the addition of the services industries. Similar to the specialization index, we observe a re-widening of the distributions in 2010 (greater standard deviation) after a decline from 2000 to 2005.

Figure 1b shows the remarkably tight connection between trends in overall district diversity levels and those present in just manufacturing alone. This stability is both striking and encouraging for our empirical work and its broad robustness. Table 1b shows a very similar set of patterns when considering only the urban areas of districts, with the additional observation that the specialization of urban areas tends to exceed that for the district as a whole, while diversity levels are very similar.

Figure 2 graphs the specialization and diversity index values against each other using 2000 Indian data. The bubble size in each graph represents the size of the local manufacturing

and services employment, and the horizontal and vertical axes provide the median value for each index. The graph demonstrates a negative correlation, such that higher diversity is associated with reduced specialization, but also confirms that the correlation is modest.

We next evaluate the degree to which specialization and diversity patterns are stable over time (Henderson, 1997). When looking at the manufacturing sector by itself, our earlier study found very little persistence—even at the frequency of a decade, there was almost no correlation in the longitudinal values of districts. Tables 2a and 2b show a stronger degree of persistence of Indian specialization and diversity levels when also modeling services in the local industrial base. This persistence is still modest in overall scope, but nevertheless noticeably stronger. For the specialization index in Panel A of Table 2a, there is a 0.2-0.4 correlation of district values across adjacent surveys that are about five years apart. This correlation over a short time period is somewhat stronger than for manufacturing alone. [4] Across 2000→2010, or about ten years in duration, the correlation remains economically and statistically meaningful at 0.2. This is significantly larger than the rapid change in specialization levels observed for manufacturing alone for 2000→2010. Panel B shows even greater persistence for the diversity index, and Table 2b looks pretty similar when we isolate urban areas of districts.

Pausing for a moment on the persistence evaluation, Table 2c compares the specialization and diversity index values derived using both sectors with that evident in manufacturing and services data only. We observe two broad patterns. First, our indices using the combined data closely correlate with each sector individually. The connection is a bit tighter on the specialization index to the services sector, while the link is a bit tighter on the diversity index to the manufacturing sector. These patterns were foreshadowed by Figures 1a and 1b, and the most important point is that neither sector is dominating the index design. Second, the negative correlation noted earlier between specialization and diversity index values holds within each sector, while being uncorrelated across sectors. We further discuss these correlation and sector-level values below when modeling them as regression covariates.

Tables 3a-b show an alternative approach for measuring persistence. We develop a transition matrix to follow cohorts of districts over time. In Panel A of Table 3a, we start by

[4] In the longer series for manufacturing alone, the stability of index values across five-year intervals increases with time. This connects to the broader stability we observe for 2000-2010 with the combined data.

grouping districts into quintiles of specialization in 2000. For example, the top row is for the 20% of districts that have the lowest specialization in 2000, while the fifth row documents the 20% of districts with the highest 2000 specialization index values. Moving across the columns, we keep these district groups the same and calculate the average value for the group in terms of their future quintiles by survey. Thus, the group of districts that constitutes the lowest 1989 quintile of specialization has an average quintile value of 2.32 by 2005, and then 2.82 by 2010. In words, this group is moving substantially up the distribution of districts in terms of relative specialization. By contrast, the bottom row of Panel A shows that the group with the highest initial specialization values declines to an average quintile value of 3.68 by 2005, and then to 3.11 by 2010. Had there been no movement in the rankings of districts, the values would have continued to look like the 2000 column. Had there been perfect mobility, the values for all districts would be 3.0 as the initial ordering of districts in 2000 would not be systematically related to their ranks in future periods.

This transition matrix thus finds quite rapid compression of the initial distribution of specialization. By definition, the quintile spread is 4.0 in 2000 (i.e., 5.0-1.0). This spread decreases to 1.36 by 2005 (i.e., 3.68-2.32) and 0.29 for 2010. For diversity, the initial 4.0 spread similarly decreases to 1.15 by 2005 and 0.88 for 2010. Compared to Tables 2a and 2b, this analysis of transition matrices retains more of the fluidity in index ordering noted for the manufacturing sector alone in Ghani, Kerr and Tewari (2013). Table 3b also shows this by looking at the 2005→2010 transition if we reorder districts in 2005. This again points to rapid adjustments, especially for specialization, across adjacent surveys. Duranton (2013b) also observes a very low degree of persistence in the production structure of Colombian cities. Across long horizons, it appears that India's trends for specialization and diversity are stabilizing, but within the more localized variation after 2000, the trend seems less certain.

. Tables 4a and 4b also document the districts showing extreme values for the specialization and diversity indices, respectively. In both tables, we list the highest and lowest average values across the full period (Panels A and B), and the major increases or declines when comparing 2010 to 2000 (Panels C and D). For the purposes of this table, we report values before winsorization, so the extreme values on this table differ from those listed in Table 1a. As described above, our index values do not depend directly on the size or economic advancement

of a region. To this end, no more than four of the twelve districts for any of the eight lists provided are in the same Indian state. Likewise, specialization and diversity are related, but not one-for-one. Only two of the 24 districts that form an extreme average value on the specialization index in Panels A and B of Table 4a are also an extreme value on the corresponding diversity lists in Table 4b. Interestingly, some of the major urban areas like Mumbai and Bangalore have undergone the largest declines in specialization.

Industry-Oriented Perspective

The specialization and diversity metrics are defined for districts, but it is also useful to take an industry-oriented perspective. To do so, we first calculate for an industry the weighted-average diversity index values for the districts in which the industry resides. We weight by the employment levels of the industry across districts. Thus, if most of the employment for an industry is in districts that are highly diversified, we will measure a high average diversity value for the industry. As the specialization measure has a less direct interpretation in this regard, due to its extreme value calculation, we only consider the diversity index with this approach.

In Table 5, we report the levels of these values for industries and their changes over time. There is some variation across industries, but the differences are not extreme. In tabulations available upon request, we group industries into "traditional" or "modern" and find that the overall diversity levels appear higher for modern industries.[5] Thus modern industries are distributed across locations that show higher levels of diversity in their industrial bases.

Table 6 provides a second industry perspective to consider specialization rates. We count by survey the number of times that each industry is responsible for the specialization value of its district. We also tabulate the average of these counts for an industry across all surveys and the changes in these counts from 2000 to 2010. Some advanced industries like office, accounting and computing machinery (NIC 30) and radio, television, and communication equipment and apparatus (NIC 32) are located in more specialized districts. Thus, a traditional industry is

[5] The manufacturing groupings follow Ghani, Kerr and O'Connell (2013a) and simply classify an industry as being modern if its unorganized share is the less than the unweighted average of the unorganized share across industries in the manufacturing sector in 2000. We similarly group the services sector through its median value in 2000. Appendix Tables 1a-2b repeat Tables 5 and 6 when considering manufacturing and services by themselves.

responsible for the local specialization value in most districts. Roughly three-quarters of districts have their specialization in a traditional industry, especially in the manufacturing sector. The computer and communication industries form the specialized industry for fewer than 20 districts over the sample period, which is a substantially smaller count than many of the larger traditional industries (average count across industries is about 24). Advanced industries like computer and related activities (NIC 72) and research and development (NIC 73) are rarely the source of a district's specialization due to the extreme degree to which they are specialized in certain locations. We later analyze whether these pockets of concentration for modern services are especially linked to growth.

The bottom of Table 6 documents the correlation of these counts between adjacent surveys. While the patterns across the 2000 and 2005 surveys are quite stable, more substantial differences emerge at the industry level for the 2010 survey compared to earlier years. Ghani, Kerr and Tewari (2013) show this break with manufacturing alone. This is likely due to the change from the NIC-2004 industry codes in the 2005 data to the NIC-2008 industry codes in the 2010 data. The fact that the reduced correlation occurs in both sectors suggests that it is most likely a definitional issue rather than a true economic pattern. Most of the remaining analyses only employ the 2010 data to calculate growth rates at the district level and do not depend upon industry definitions, thus minimizing the importance of this issue.

Correlations to District Traits

Table 7 documents multivariate analyses of the district traits that associate with specialization and diversity levels, as well as changes in these values over time. Outcome variables are indicated by column headers. These dependent variables are expressed in a ten-point scale for the levels metrics, representing deciles of the raw index values, and in unit standard deviation for the change metrics. District traits are taken primarily from the 2001 Population Census, with some specific covariates about the sizes of the organized and unorganized sectors also calculated directly from the ASI and NSS. We transform non-logarithm explanatory variables to have unit standard deviation for interpretation. Estimations include state fixed effects, weight districts by their log 2001 population, and report robust standard errors.

Several important observations can be made from Table 7. First and most important, this substantial battery of district traits has pretty weak predictive power for industrial specialization and diversity. This was true for the manufacturing sector as well in the longer panel. This weakness is observed in the limited number of regressors that pick up a statistically significant coefficient (often only two out of the 13 regressors modeled) and the low R-Squared value overall. To an important degree, this is the outcome of the metric design that makes these traits more or less independent of factors like district size. Given this independence, we can expect to find broad stability in our upcoming regressions with and without district covariates being modeled (which will hold true).[6]

Second, as in our study of the manufacturing sector, worker human capital is important. Districts with high education levels in 2000 display more specialized and diversified industrial bases. Literacy also displays some connection. Literacy was the stronger predictor for manufacturing specialization and diversity in Ghani, Kerr and Tewari (2013), while education plays the sharper role in our new sample that includes services. Either way, these two measures of human capital continue to display the strongest partial correlations, even conditional on other district traits like population levels and consumption per capita. Ghani, Kerr and O'Connell (2013b) also find a connection between local literacy rates and greater use of a wider variety of distinct material inputs in Indian plants (overall and relative to plant size).

In the upcoming growth analyses, we will control for the traits modeled in Table 7, so that we are measuring the impact of specialization and diversity in a district's industrial base over-and-beyond these correlates. While this approach better isolates the role of specialization and diversity, we should not forget these two basic connections when using stricter econometric frameworks.

By contrast, the absence of some correlations is striking. In the United States, larger cities tend to be less specialized and more diversified (e.g., Black and Henderson, 1998, Duranton and Puga, 2000, Henderson, 1997). Big cities, where firm headquarters and service firms are often based, tend to specialize in business services. Duranton (2013b) also observes this size

[6] Our earlier study controlled as well for the manufacturing share of the local district work force. We exclude this variable in our current work given the high degree to which our two sectors account for local activity. Our estimations are sensitive to how such a control is constructed.

relationship for Colombia. For India, we do not see this pattern, and this held true when just considering manufacturing in Ghani, Kerr and Tewari (2013) too. These null patterns are also evident with population density, although the point estimate starts to become larger in size. It is also striking that infrastructure metrics and banking conditions tend to have limited correlation with the specialization and diversity metrics. Finally, overall urbanization levels of districts are not clearly linked. Further inquiry about the lack of predictive power for these metrics is warranted given their observation in other settings.

4. Employment Growth with Specialization and Diversity

This section describes our empirical exercises regarding how employment growth for a district links to its initial specialization or diversity. The topic of specialization/diversity and city growth has been often been studied over the past two decades (Duranton, 2013a), and we seek to provide evidence on this link across Indian manufacturing and services. We particularly study the inclusive nature of local growth given its pressing importance for India.

We estimate a cross-sectional growth equation at the district level of the form:

$$Growth_d = \beta \cdot Specialization_{d,t0} + \delta \cdot Diversity_{d,t0} + \chi_s + \phi \cdot Z_d + \varepsilon_d \, .$$

Our core regressors are again the specialization and diversity measures that utilize a ten-point scale. The *t0* subscript signifies that we measure these attributes at the start of the sample using the 2000 data. The outcome variable is the log employment growth in manufacturing and services for the district from 2000 to 2010. Regressions include a vector of state fixed effects χ_s to account for systematic differences in regional specialization and diversity evident in Figures 3 and 4, differences in regional growth rates in these sectors for India across the period, and similar. The vector Z_d contains the additional district-level controls that we modeled in Table 7. These controls, among other things, include measures like initial population size and income levels to capture convergence processes. We weight districts by their log population in 2001 and report robust standard errors.

The first column of Table 8 shows a lack of correlation for the specialization index, while Column 2 finds a strong and significant role for the diversity index. An increase of one decile in the diversity index is associated with 7% higher employment growth (e.g., moving from a

growth rate of 2.00% to 2.14%). This substantial boost holds in the joint analysis in Column 3, and similar elasticities are evident without the state fixed effects, too.

Column 4 further shows that the results are robust to including district-level covariates taken from the 2001 Census discussed earlier. These covariates tend to have weak multivariate explanatory power, and the inclusion of these controls does not impact the diversity index's connection to growth. Columns 5 and 6 show similar results when we do not weight observations or when we exclude districts with less than one million people in population. The latter check is important given that the exclusion of small districts guards against cases where our data start to become too thin for measuring the index values, and it is noteworthy that the standard errors remain consistent despite the reduction in sample size. Column 7 shows similar results when we include a control for the expected growth based upon the industrial distribution for the district in 2000 and the national rate of growth by industry across the period.

Column 8 replaces the ten-point measure with two indicator variables for districts being in the 50^{th}-75^{th} percentile range or in the 76^{th}-99^{th} percentile range. These estimations measure coefficients relative to the bottom half of the distribution. The results indicate that the most substantial differences for growth come from districts being in the top half of initial diversity. We also find quantitatively similar results when considering raw index values.

We will next move to a variety of growth decompositions, but before doing so it is valuable to compare these patterns against our earlier results. Ghani, Kerr and Tewari (2013) identify a different relationship—that manufacturing specialization in the 1989-1994 period was important for explaining the employment growth in manufacturing from 1989 to 2010. Our current work emphasizes instead local diversity. We have investigated how to reconcile these features, and we are able to mostly close the gap. The biggest contributor to the different outcomes is simply moving to a 2000-2010 analysis. There were exceptionally strong adjustments in India's manufacturing base and its spatial allocation during the 1990s, reflective of the null correlation in the index values over decadal frequencies discussed earlier. By measuring our present study from 2000 onwards, we pick India up at very different point in its economic emergence. Unfortunately, we will never be able to construct similar services indices for 1989 to compare. A second and smaller factor is removal of manufacturing growth during the 1990s for India that was responsible for a big portion of the employment growth relationship. A

priori, the current estimates likely provide the better foundation for predicting India's development going forward due to their more recent and comprehensive focus.

Nevertheless, in Column 9, we show both effects in a joint estimation. In addition to our specialization and diversity metrics estimated over the manufacturing and services sectors combined, we include the specialization and diversity index values for manufacturing and services specifically (Table 2c). The individual metrics for services are small and statistically insignificant and not reported. The overall diversity metric grows somewhat in value, while that for manufacturing is negative. Interesting, we observe some evidence for manufacturing specialization still playing a role in local growth in this augmented specification. We report but do not overly push these estimations given the high correlations across variants of the indices noted earlier. The estimation nonetheless helps narrow the differences between the two studies and provides some evidence for the different paths to city growth. Manufacturing specialization appears to have been one path, and broader local diversity is an even stronger one in the 2000-2010 period.[7]

Table 9 provides several decompositions of the employment growth relationship. Column 1 starts by repeating the base specification (Column 4 of Table 8). In the remaining columns, we keep the core specialization and diversity index values constant. We instead consider growth among different types of industries and establishments. We have the added complication that an industry may not be recorded in 2010, which is an undefined log growth rate. In these cases, we recode these values with the lowest observed growth for the variable in question. This allows us to maintain a consistent sample across the specifications.

Columns 2 and 3 separately consider employment growth that is occurring in manufacturing and services industries. The diversity index connects with both sectors on an equal footing. The employment growth in services is more precisely measured, while the manufacturing coefficient falls short of being statistically significant. However, the coefficient estimates are otherwise quite comparable and in unreported work we have found a broad basis to this work. In this analysis, and some that follow, the coefficient estimates do not necessarily

[7] By contrast, we also considered augmenting the manufacturing productivity regressions from Ghani, Kerr and Tewari (2013) with district-wide metrics. In this case, the manufacturing traits of the local area solidly outperformed the metrics that combined manufacturing and services.

bracket the overall effect as the fixed effects and covariates are allowed to adjust over estimations.

Columns 4 and 5 separately consider employment growth within and outside the industry that was the most important to the district's employment base in 2000. We define this most important industry for each district in 2000 through shares of local employment. Similar the manufacturing/services split, the coefficients across the two estimations are similar, but they are much more precisely estimated for growth external to the most important industry in 2000. This broader industry group is where most of the population works and accounts for most of the total job growth observed. These estimations have an intuitive feel to them—diversity promoting growth in many local industries, rather than just the one that initially led the city.

Columns 6 and 7 split employment growth in establishments by their size. This decomposition allows us to observe the types of firms that are supporting the employment growth. The split at ten employees mimics the demarcation between the organized and unorganized sectors for manufacturing. Interestingly, the bulk of the growth is among small establishments that are reflective of the unorganized sector. Establishments of this size account for over 80% of India's employment, and the outcomes for these small establishments closely mimic the main effect. Districts with more diverse industrial bases are mostly expanding employment throughout the unorganized and informal sectors.

Columns 8 and 9 split the employment growth in the urban and rural areas of districts. The surprising outcome here is that the employment growth associated with district diversity is strongest in rural areas. Given the other findings in Table 9—especially the role of small-scale establishments in Column 6—this rural connection is not implausible. It does indicate however that the connection of diversity to growth is likely coming through channels beyond the celebrated accounts about diversity launching new combinations of ideas. Instead, diversity may be offering better matches of workers and industries, insurance against declines in a single industry, and similar roles that link it to inclusive growth.

Column 10 examines how diversified employment growth relates to diversity and specialization. To measure this, we construct a "concentration ratio" which is the ratio of employment growth of a district's top 3 industries to overall district employment growth (we only keep industries in a district that have experienced positive employment growth for this

exercise). Both specialization and diversity decrease the share of employment in a district's largest growth industries, with diversity having a slightly larger effect. The last column in Table 9 examines whether diversity or specialization increases the probability that a district will experience growth in a high growth-industry, which we define as the top 10 industries in terms of employment growth between 2000 and 2010. Indeed, diversity but not specialization affects this probability in a positive and significant manner.

Table 10 next considers a different set of decompositions. These specifications are more aggressive in empirical design in that we often partition the sample into subsets of districts for analysis. We build specifically on this observed importance of employment growth in rural areas. We consider how the effects are moderated by distance from big cities and district population density. Columns 2 and 3 partition by whether a district is within 250 km of the seven largest cities in India. We find very uniform effects upon this dimension. Thus, the connection of diversity to growth is not confined to places near India's big cities.

Columns 4-6 alternatively partition the sample by district population density in 2001. Our earlier results controlled for population density, and this test instead looks for whether effects are different in sub-groups. This is of active policy interest given the perceived importance by some observers of intermediate cities, or even more specifically specialized intermediate cities, in growth. We find the link of diversity to growth is present in all levels, with its strongest expression coming in the bottom third of districts in terms of population density. We find similarly stable results if instead estimating a stacked regression with interactions for our indices by density levels. The most important take-away is again that diversity's impact is not confined to just big cities or dense locales.

The final three columns test a particular phenomenon of specialization for India. Many have observed the importance of modern services in the special development and growth of places like Bangalore. To test the importance of this effect, we introduce indicator variables in Columns 7-9 for whether in 2000 a district ranked among the top five districts in terms of the share of modern services in the local industrial base or among the top five districts in terms of the absolute size of its modern services base. There does appear to be a particular connection in the data for initial clusters around modern services to experience strong subsequent employment growth. The introduction of these controls does not reduce the strength of the estimated

relationship for services diversity. These correlations are also evident when using an alternative definition of modern services that classifies firms as modern if they had high ICT use (using data from the 2010-11 NSS survey).

5. Conclusions

India's economic geography is a very lively research topic, full of many nuances and twists befitting the rapidly developing country and its present and past policy frameworks. This study simultaneously resolves questions on some dimensions and opens up questions on other dimensions. In terms of what it resolves, the prior study by Ghani, Kerr and Tewari (2013) found some interesting patterns with respect to manufacturing specialization and diversification. Yet, strong conclusions were not forthcoming given the exclusive focus on one sector. By shortening the time frame and incorporating additional surveys, the present study provides a more secure base. It confirms the comparable specialization and diversity levels to what is observed in the United States, the trend for convergence and then modest re-expansion of the distributions, the relative roles of modern and traditional industries, and so on.

On the other hand, we have several findings that deserve further attention. First, we have yet to find a consistent link in India between the size or density of districts and their diversity patterns. This was true in our earlier work for manufacturing, and it continues to hold with services and our joint analyses. This pattern is different than what is observed in other places like Colombia and the United States. It is worth further investigation as to the source of this deviation. Such an extension would also allow us to compare India more closely with the work on Colombia by Duranton (2013b), who finds differences between manufacturing and services performance around specialization dynamics. This added perspective will provide a richer foundation for understanding growth in Indian cities. Detailed case studies around the growth of modern services in the extreme cases identified of initial employment are also warranted.

Second, and perhaps even more intriguing, is our connection of diversity to employment growth in rural areas of districts and districts with low population density. While we find a broad and robust connection of diversity to local employment growth in many settings, the fact that the sharpest expressions occur in these unexpected places is surprising to us. Beyond the standard

arguments about diversity in nursery cities and innovation, we need to understand the role of diversity outside the most prominent clusters. This is key to refining our conclusion about diversity's role for inclusive growth. The ability to observe inputs into plants with the Indian data may provide an empirical foothold. The sub-contracting of work between organized and unorganized sectors might also be at play (e.g., Mukim, 2013). It is also worth investigating whether onerous land-use regulations and building codes in India cities (Sridhar, 2010) is pushing some of the growth linked to diversity outwards, especially as local infrastructure improves connectivity between urban and rural areas within districts.

Finally, we would like to evaluate the longitudinal changes highlighted in this study in terms of the impact of specific infrastructure projects (e.g., the Golden Quadrangle project) or trade liberalizations. These factors have been shown to play an important role for Indian manufacturing (e.g., Goldberg et al., 2010), but we do not know their role in terms of the specialization and diversity of India's districts. Tracing out the economic geography of these economic shocks is important, and the rich longitudinal data for India provide a unique laboratory for doing so in a developing economy. This would be useful for understanding the allocation of activity in the Indian economy and how it can be improved (e.g., Hsieh and Klenow, 2009; Ghani, Goswami and Kerr, 2012a,b; Desmet et al., 2011). Likewise, Ellison, Glaeser and Kerr (2010) consider coagglomeration and the inter-linkages of industries within a local area. It would be interesting to quantify what happens to related industries located in specialized or diversified districts when these dramatic changes occur. It would also be important to depict how the traded versus non-traded nature of manufacturing and services industries modulates or amplifies these external shocks.

References

Acemoglu, Daron, and Fabrizio Zilibotti, "Was Prometheus Unbound by Chance? Risk, Diversification, and Growth," *Journal of Political Economy* 105:4 (1997), 709-751.

Anand, Rahul, Volodmyr Tulin, and Naresh Kumar, "India: Defining and Explaining Inclusive Growth and Poverty Reduction," IMF Working Paper 14/63 (2014).

Black, Duncan, and J. Vernon Henderson, "Urban Evolution in the USA," *Journal of Economic Geography* 3:4 (2003), 343–372.

Deichmann, Uwe, Somik Lall, Stephen Redding, and Anthony Venables, "Industrial Location in Developing Countries," *World Bank Research Observer* 23:2 (2008), 219–246.

Desmet, Klaus, Ejaz Ghani, Stephen O'Connell, and Esteban Rossi-Hansberg, "The Spatial Development of India," World Bank Policy Research Working Paper No. 6060 (2012).

Duranton, Gilles, "Growth through Cities in Developing Countries," Working Paper (2013a).

Duranton, Gilles, "Specialisation, Diversity, and Sectoral Growth in Colombian Cities," Working Paper (2013b).

Duranton, Gilles, and Diego Puga, "Diversity and Specialization in Cities: Why, Where and When Does It Matter?" *Urban Studies* 37:3 (2000), 533–555.

Duranton, Gilles, and Diego Puga, "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economic Review* 91:5 (2001), 1454–1477.

Duranton, Gilles, and Diego Puga, "Micro-Foundations of Urban Agglomeration Economies," in *Handbook of Regional and Urban Economics*, Volume 4, eds. J. Vernon Henderson and Jacques François Thisse (Amsterdam: North-Holland, 2004), 2063–2117.

Ellison, Glenn, and Edward Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* 105 (1997), 889–927.

Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review* 100:3 (2010), 1195–1213.

Feldman, Maryann and David Audretsch, "Innovation in Cities: Science-Based Diversity, Specialization and Localized Competition," *European Economic Review* 43:2 (1999), 409–429.

Fernandes, Ana and Gunjam Sharma, "Together We Stand? Agglomeration in Indian Manufacturing," World Bank Policy Research Working Paper No. 6062 (2012).

Ghani, Ejaz, Arti Goswami, and William Kerr, "Is India's Manufacturing Sector Moving Away From Cities?" National Bureau of Economic Research Working Paper No. 17992 (2012a).

Ghani, Ejaz, Arti Goswami, and William Kerr, "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," National Bureau of Economic Research Working Paper No. 18524 (2012b).

Ghani, Ejaz, William Kerr, and Stephen O'Connell, "The Exceptional Persistence of India's Unorganized Sector," World Bank Policy Research Working Paper No. 6454 (2013a).

Ghani, Ejaz, William Kerr, and Stephen O'Connell, "Input Usage and Productivity in Indian Manufacturing Plants," World Bank Policy Research Working Paper (2013b).

Ghani, Ejaz, William Kerr, and Ishani Tewari, "Specialization, Diversity, and Indian Manufacturing Growth," World Bank Policy Research Working Paper (2013).

Glaeser, Edward, Heidi Kallal, José Scheinkman, and Andrei Shleifer, "Growth in Cities," *Journal of Political Economy* 100:6 (1992), 1126–1152.

Goldberg, Pinelopi, Amit Khandelwal, Nina Pavcnik, and Petia Topalova, "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India," *Quarterly Journal of Economics* 125:4 (2010), 1727-1767.

Haussmann, Ricardo, "The Specialization Myth," http://www.social-europe.eu/2014/01/specialization-myth/ (2014).

Henderson, J. Vernon, "Externalities and Industrial Development," *Journal of Urban Economics* 42:3 (1997a), 449–470.

Henderson, J. Vernon, "Medium Size Cities," *Regional Science and Urban Economics* 27:6 (1997b), 583–612.

Henderson, J. Vernon, Ari Kuncoro, and Matt Turner, "Industrial Development in Cities," *Journal of Political Economy* 103:5 (1995), 1067–1090.

Hsieh, Chang-Tai, and Peter Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics* 124:4 (2009), 1403–1448.

Imbs, Jean, and Romain Wacziarg, "Stages of Diversification," *American Economic Review* 93:1 (2003), 63-86.

Jacobs, Jane, *The Economy of Cities* (New York: Random House, 1969).

Kaulich, Florian, "Diversification vs. Specialization as Alternative Strategies For Economic Development: Can We Settle a Debate by Looking at the Empirical Evidence?" United Nations Industrial Development Organization Working Paper 03/2012 (2012).

Lall, Somik, Jun Koo, and Sanjoy Chakravorty, "Diversity Matters: The Economic Geography of Industry Location in India," World Bank Policy Research Working Paper No. 3072 (2003).

Lall, Somik, and Taye Mengistae, "The Impact of Business Environment and Economic Geography on Plant Level Productivity: An Analysis of Indian Industry," World Bank Policy Research Working Paper No. 3664 (2005).

Lall, Somik, Zmarak Shalizi, and Uwe Deichmann, "Agglomeration Economies and Productivity in Indian Industry," *Journal of Development Economics* 73:2 (2004), 643–673.

McKinsey Global Institute, India's Urban Awakening: Building Inclusive Cities, Sustaining Economic Growth, Report (2010).

Mukim, Megha, "Coagglomeration of Formal and Informal Industry: Evidence from India," World Bank Working Paper 6622 (2013).

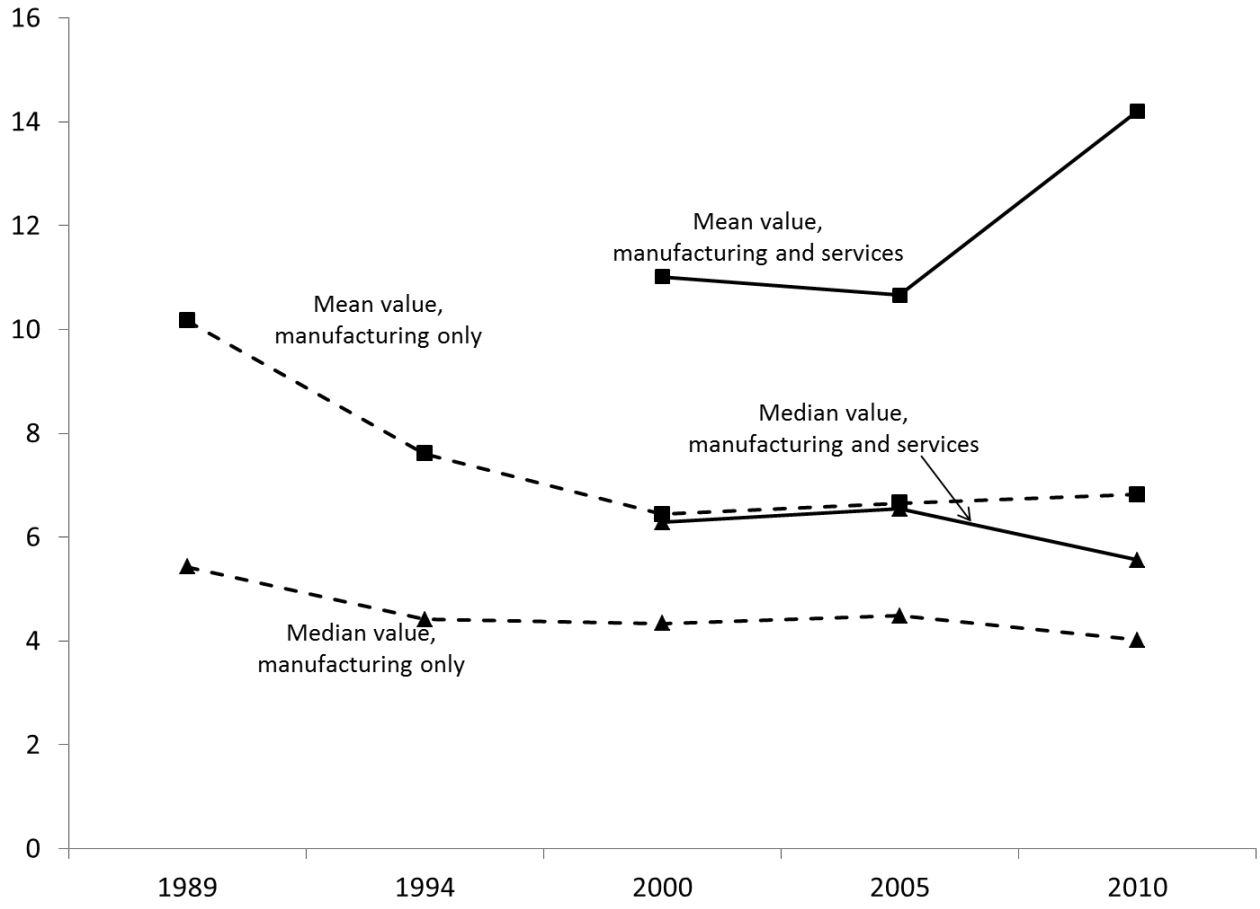
Rosenthal, Stuart, and William Strange, "Evidence on the Nature and Sources of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, Volume 4, eds. J. Vernon Henderson and Jacques François Thisse (Amsterdam: North-Holland, 2004), 2119–2171.

Sridhar, Kala Seetharam, "Impact of Land Use Regulations: Evidence from India's Cities," *Urban Studies* 47:7 (2010), 1541-1569.

Topalova, Petia, "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics* 2:4 (2010), 1-41.

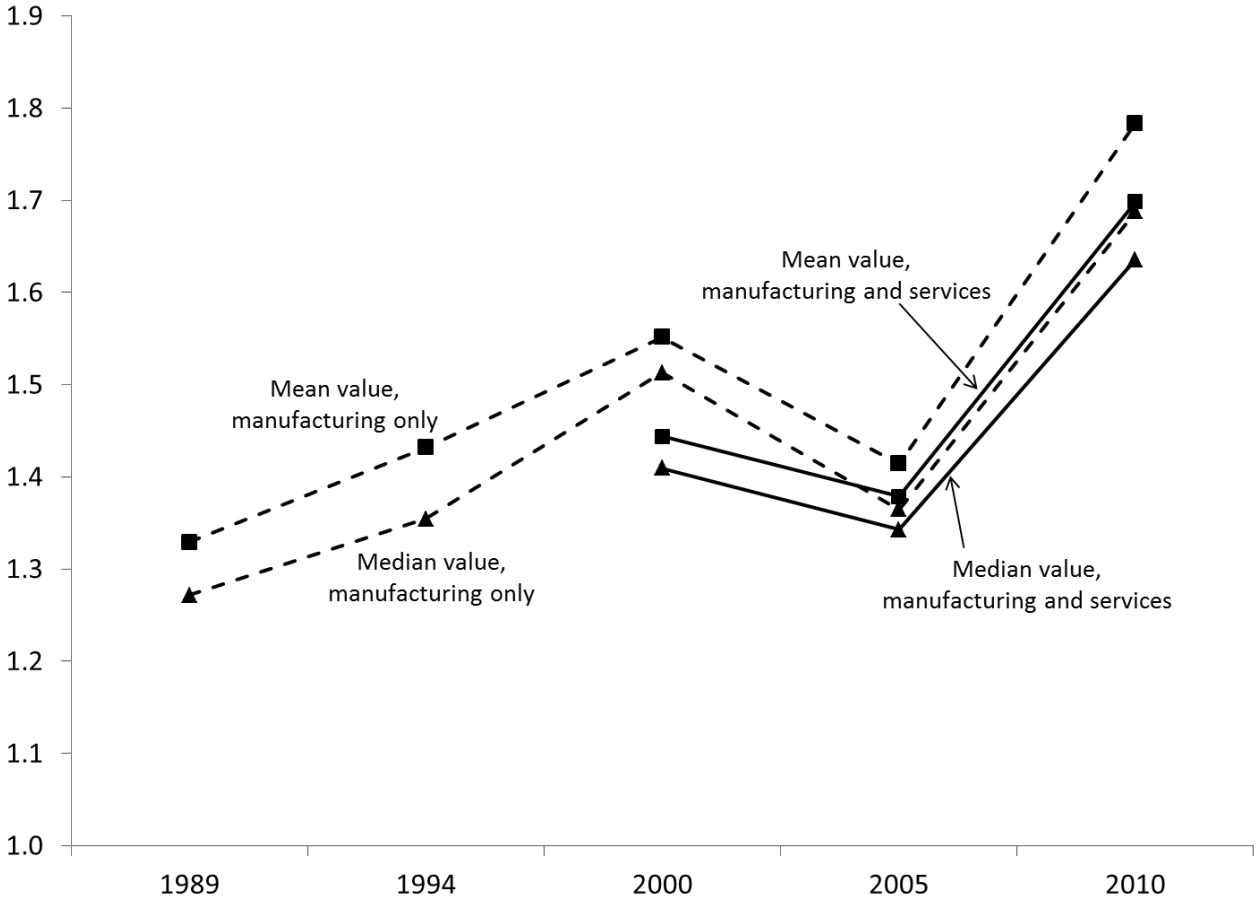
World Bank, "Planning, Connecting, and Financing Cities—Now," Urbanization Review Flagship Report, The World Bank, 2013.

Figure 1a: India specialization index value comparisons



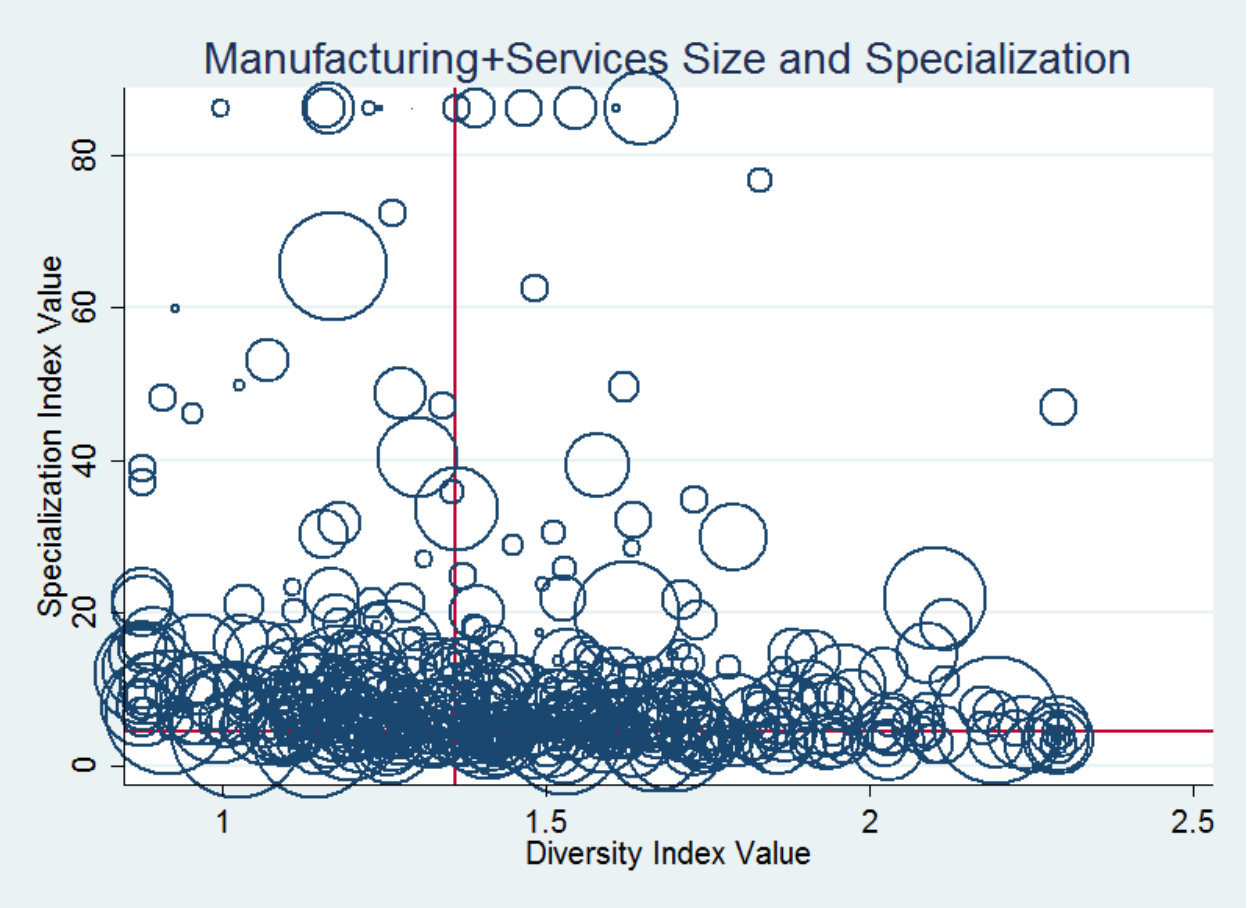
Notes: Figure compares the values and trends derived for the specialization index. The longer series are calculated for the manufacturing sector only from 1989 to 2010. The shorter series are jointly calculated over industries in both the manufacturing and services sectors.

Figure 1b: India diversity index value comparisons



Notes: See Figure 1a.

Figure 2: India specialization and diversity index values, 2000



Notes: Each observation is a district in 2000. Bubble size indicates log manufacturing and services employment in the district. Values for specialization and diversity are winsorized at their 2% and 98% values. The red axis lines document the median values for each index.

Table 1a: Descriptive statistics on district specialization/diversity

	2000	2005	2010	Average	Change
A. Specialization index for districts					
Mean value	11.01	10.66	14.21	11.96	1.22
Standard deviation	14.73	12.27	26.75	12.81	5.16
Minimum value	2.61	2.86	2.25	3.10	-0.97
Maximum value	86.03	79.77	149.39	105.07	46.98
Median value	6.28	6.55	5.56	7.51	-0.10
B. Diversity index for districts					
Mean value	1.44	1.38	1.70	1.51	0.22
Standard deviation	0.32	0.27	0.49	0.26	0.39
Minimum value	0.88	0.83	0.88	0.86	-0.52
Maximum value	2.29	2.11	2.89	2.34	1.52
Median value	1.41	1.34	1.64	1.48	0.16

Notes: Descriptive statistics taken from Annual Survey of Industries and National Sample Statistics combining the manufacturing and services sectors. The index of relative specialization measures the maximum for a district of its share of employment in an industry compared to the national share of employment in that industry. The index of relative diversity is the inverse sum across all industries in both sectors of the absolute value of the difference between an industry's district share and its national share. Indices are winsorized at the 2%/98% values by year. The Average column describes the properties of the mean estimates of specialization and diversity for districts across the 2000-2010 period. The Change column describes the properties of the differences for districts between their mean index values in 2010 compared to 2000 relative to the 2000 base value. The sample includes 554 districts.

Table 1b: Table 1a considering urban areas only

	2000	2005	2010	Average	Change
A. Specialization index for districts					
Mean value	10.86	10.92	14.05	11.94	1.11
Standard deviation	13.09	10.41	29.80	12.97	5.36
Minimum value	2.56	2.70	2.38	2.82	-0.97
Maximum value	75.65	53.72	187.82	98.96	62.06
Median value	6.26	7.02	5.82	7.88	-0.08
B. Diversity index for districts					
Mean value	1.43	1.45	1.68	1.52	0.22
Standard deviation	0.29	0.28	0.45	0.23	0.40
Minimum value	0.92	0.93	0.93	0.93	-0.53
Maximum value	2.20	2.12	2.95	2.31	1.75
Median value	1.40	1.40	1.62	1.50	0.15

Notes: See Table 1a. The sample includes the urban portion of 540 districts that have urban areas.

Table 2a: Persistence of specialization/diversity

	2000	2005	2010
A. Specialization index for districts - Correlation over time			
2000 value	1		
2005 value	0.3752*	1	
2010 value	0.1530*	0.2086*	1
B. Diversity index for districts - Correlation over time			
2000 value	1		
2005 value	0.2813*	1	
2010 value	0.2061*	0.3831*	1

Notes: See Table 1a. An asterisk denotes a correlation is statistically significant at the 10% level.

Table 2b: Table 2a considering urban areas only

	2000	2005	2010
A. Specialization index for districts - Correlation over time			
2000 value	1		
2005 value	0.3458*	1	
2010 value	0.1950*	0.1630*	1
B. Diversity index for districts - Correlation over time			
2000 value	1		
2005 value	0.2417*	1	
2010 value	0.0628	0.1779*	1

Notes: See Table 2a.

Table 2c: Correlation of specialization/diversity levels in 2000 by sector

	Total spec	Mfg spec	Services spec	Total div	Mfg div	Services div
Total specialization	1					
Mfg specialization	0.6115*	1				
Services	0.8229*	0.0246	1			
Total diversity	-0.1427*	-0.1447*	-0.0500	1		
Mfg diversity	-0.0989*	-0.2464*	0.0174	0.6617*	1	
Services diversity	-0.1026*	0.0706	-0.124*	0.3945*	0.0917*	1

Notes: See Table 2a.

Table 3a: Transition matrix of specialization/diversity levels

	2000	2005	2010
A. Specialization index for districts - Average quintile value by time period			
Lowest initial quintile	1.00	2.32	2.82
2nd initial quintile	2.00	2.75	2.67
3rd initial quintile	3.00	3.08	3.16
4th initial quintile	4.00	3.15	3.23
Highest initial quintile	5.00	3.68	3.11
Quintile gap	4.00	1.36	0.29
B. Diversity index for districts - Average quintile value by time period			
Lowest initial quintile	1.00	2.38	2.63
2nd initial quintile	2.00	3.01	2.91
3rd initial quintile	3.00	2.94	2.93
4th initial quintile	4.00	3.14	3.01
Highest initial quintile	5.00	3.53	3.51
Quintile gap	4.00	1.15	0.88

Notes: See Table 1a. Districts are divided into quintiles for each period based upon index values. Each row documents a cohort's evolution based upon initial value in 2000. Moving from left to right along a row, an increase in the value indicates that the average index value for that cohort is increasing from one period to the next.

Table 3b: Initial transition matrix on a period-by-period basis

	2000→2005	2005→2010
A. Specialization index for districts - Average quintile value in end period		
Lowest quintile in start	2.32	2.69
2nd quintile in start period	2.75	2.77
3rd quintile in start period	3.08	2.92
4th quintile in start period	3.15	3.14
Highest quintile in start	3.68	3.47
Quintile gap	1.36	0.78
B. Diversity index for districts - Average quintile value in end period		
Lowest quintile in start	2.38	2.38
2nd quintile in start period	3.01	2.74
3rd quintile in start period	2.94	2.84
4th quintile in start period	3.14	3.22
Highest quintile in start	3.53	3.82
Quintile gap	1.15	1.44

Notes: See Table 3a. The table provides the one-step transition for each period by reordering districts after each survey into their quintiles at that time.

Table 4a: Detailed specialization levels for districts

District	State	Average value Special.	Diversity	District	State	Change Special.
A. Districts with highest average values across 2000-2010				C. Districts with largest increases from 2000		
Kavaratti	LAKSHADWEEP	253.88	1.32	Darjiling	WEST BENGAL	156.11
Darjiling	WEST BENGAL	195.44	1.57	Wokha	NAGALAND	155.17
Panchkula	HARYANA	185.85	1.43	Gorakhpur	UTTAR PRADESH	66.12
Wokha	NAGALAND	164.01	2.04	Jaunpur	UTTAR PRADESH	58.02
Dhuburi	ASSAM	152.27	1.12	Cannanore	KERALA	42.25
Bhopal	MADHYA PRADESH	118.66	1.51	Lakhimpur	ASSAM	35.98
Ri Bhoi	MEGHALAYA	102.71	1.37	Solan	HIMACHAL PRADESH	28.85
Gorakhpur	UTTAR PRADESH	93.95	1.54	Giridih	JHARKHAND	24.38
Saharsa	BIHAR	86.19	1.41	Bhopal	MADHYA PRADESH	23.41
Dewas	MADHYA PRADESH	80.01	1.56	Sonitpur	ASSAM	23.00
Jaunpur	UTTAR PRADESH	75.53	1.62	Goalpara	ASSAM	22.18
Jamnagar	GUJARAT	70.00	1.31	Nanded	MAHARASHTRA	21.54
B. Districts with lowest average values across 2000-2010				D. Districts with largest declines from 2000		
Bhiwani	HARYANA	3.10	1.80	Etah	UTTAR PRADESH	-0.98
Nellore	ANDHRA PRADESH	3.10	1.72	Mumbai Suburban	MAHARASHTRA	-0.96
Patna	BIHAR	3.17	1.50	Nuapada	ORISSA	-0.96
Amritsar	PUNJAB	3.19	2.06	West Nimar	MADHYA PRADESH	-0.95
Srikakulam	ANDHRA PRADESH	3.28	2.17	Sitamarhi	BIHAR	-0.95
Trichur	KERALA	3.28	2.32	Jhalawar	RAJASTHAN	-0.94
Bikaner	RAJASTHAN	3.30	1.60	Kota	RAJASTHAN	-0.94
Khammam	ANDHRA PRADESH	3.33	1.83	Kamrup	ASSAM	-0.93
Yavatmal	MAHARASHTRA	3.34	1.66	Sawai Madhopur	RAJASTHAN	-0.93
Mau	UTTAR PRADESH	3.34	1.49	Rayagada	ORISSA	-0.92
Sangli	MAHARASHTRA	3.35	1.89	Bangalore Urban	KARNATAKA	-0.92
Adilabad	ANDHRA PRADESH	3.43	2.06	Bundi	RAJASTHAN	-0.91

Notes: See Table 1a. Panels A and B report average values across the 2000-2010 period. Panels C and D report changes from 2000 to 2010 relative to the 2000 value. Values in this table have not been winsorized annually as in Table 1a.

Table 4b: Detailed diversity levels for districts

District	State	Average value Special.	Average value Diversity	District	State	Change Diversity
A. Districts with highest average values across 2000-2010				C. Districts with largest increases from 2000		
Mahbubnagar	ANDHRA PRADESH	3.44	2.52	Neemuch	MADHYA PRADESH	1.79
Palghat	KERALA	5.36	2.38	Wardha	MAHARASHTRA	1.55
Trichur	KERALA	3.28	2.32	Kinnaur	HIMACHAL PRADESH	1.52
Jalna	MAHARASHTRA	4.43	2.29	Udaipur	RAJASTHAN	1.50
Sivaganga	TAMIL NADU	4.21	2.27	Pulwama	JAMMU & KASHMIR	1.46
North 24 Parganas	WEST BENGAL	8.62	2.26	Mahbubnagar	ANDHRA PRADESH	1.42
Thanjavur	TAMIL NADU	5.75	2.21	Tuensang	NAGALAND	1.37
Buldana	MAHARASHTRA	5.38	2.21	Bilaspur H	HIMACHAL PRADESH	1.29
Neemuch	MADHYA PRADESH	3.88	2.19	Guntur	ANDHRA PRADESH	1.28
Madurai	TAMIL NADU	3.87	2.18	Jhalawar	RAJASTHAN	1.28
Hugli	WEST BENGAL	12.52	2.17	Sikar	RAJASTHAN	1.26
Srikakulam	ANDHRA PRADESH	3.28	2.17	Jaisalmer	RAJASTHAN	1.25
B. Districts with lowest average values across 2000-2010				D. Districts with largest declines from 2000		
Nizamabad	ANDHRA PRADESH	13.82	0.79	Murshidabad	WEST BENGAL	-0.60
Daman	DAMAN & DIU	32.59	0.81	Dibrugarh	ASSAM	-0.53
Damoh	MADHYA PRADESH	12.93	0.84	Lucknow	UTTAR PRADESH	-0.52
Bhavnagar	GUJARAT	9.63	0.90	Singhbhum	JHARKHAND	-0.49
Sagar	MADHYA PRADESH	11.46	0.91	Dakshin Kannad	KARNATAKA	-0.48
Bagdam	JAMMU & KASHMIR	7.92	0.96	Gondiya	MAHARASHTRA	-0.48
Kasaragod	KERALA	9.72	0.97	Nayagarh	ORISSA	-0.46
Gurgaon	HARYANA	18.04	1.04	New Delhi	DELHI	-0.46
Sambalpur	ORISSA	9.23	1.04	Shahjahanpur	UTTAR PRADESH	-0.45
Debagarh	ORISSA	12.79	1.05	Jalaun	UTTAR PRADESH	-0.44
Balangir	ORISSA	7.06	1.06	Jalor	RAJASTHAN	-0.44
Tikamgarh	MADHYA PRADESH	8.05	1.07	Banswara	RAJASTHAN	-0.43

Notes: See Table 4a.

Table 5: Index levels for industries based upon district locations

NIC	Industry Description	Diversity		
		2000	2010	Change
15	Food products and beverages	1.45	1.71	0.18
16	Tobacco products	1.43	1.69	0.18
17	Textiles	1.45	1.71	0.18
18	Wearing apparel; dressing and dyeing of fur	1.45	1.95	0.35
19	Leather; luggage, handbags, saddlery, harness and footwear	1.49	1.77	0.19
20	Wood and wood products, except furniture; straw and plating	1.45	2.05	0.42
21	Paper and paper products	1.48	1.83	0.23
22	Publishing, printing and reproduction of recorded media	1.46	1.73	0.19
23	Coke, refined petroleum and nuclear fuel	1.43	1.71	0.20
24	Chemicals and chemical products	1.49	1.73	0.16
25	Rubber and plastic products	1.48	1.73	0.18
26	Other non-metallic mineral products	1.46	1.72	0.18
27	Basic metals	1.44	1.74	0.20
28	Fabricated metal products, except machinery and equipments	1.45	1.71	0.17
29	Machinery and equipment, n.e.c.	1.49	1.76	0.18
30	Office, accounting and computing machinery	1.35	1.63	0.21
31	Electrical machinery and apparatus, n.e.c.	1.49	1.75	0.17
32	Radio, television, and communication equipment and apparatus	1.49	1.67	0.12
33	Medical, precision and optical instruments, watches and clocks	1.50	1.77	0.18
34	Motor vehicles, trailers and semi-trailers	1.52	1.83	0.20
35	Other transport equipment	1.51	1.75	0.16
36	Furniture, manufacturing n.e.c.	1.45	1.71	0.18

Table 5, continued

NIC	Industry Description	Diversity 2000	Diversity 2010	Diversity Change
55	Hotels and restaurants	1.46	1.70	0.17
60	Land transport (via pipelines)	1.46	1.71	0.17
61	Water transport	1.47	1.67	0.13
63	Supporting transport activities, travel agencies	1.48	1.74	0.18
64	Post and telecommunications	1.46	1.72	0.18
70	Real estate activities	1.52	1.76	0.16
71	Renting of machinery, personal goods	1.47	1.89	0.29
72	Computer and related activities	1.47	1.69	0.15
73	Research and development	1.50	1.97	0.32
74	Other business activities	1.46	1.71	0.17
80	Education	1.46	1.90	0.30
85	Health and social work	1.46	1.72	0.18
90	Sewage and refuse disposal, sanitation, etc.	1.48	1.84	0.24
91	Activities of membership organizations n.e.c.	1.49	1.73	0.16
92	Recreational, cultural and sporting activities	1.47	1.73	0.18
93	Other service activities	1.46	1.71	0.17

Notes: See Table 1a. "n.e.c." stands for Not Elsewhere Classified.

Table 6: Count of districts by specialization

NIC	Industry Description	2000	2005	2010	Average	Change
15	Food products and beverages	21	24	19	21.3	-0.10
16	Tobacco products	23	24	28	25.0	0.22
17	Textiles	24	27	14	21.7	-0.42
18	Wearing apparel; dressing and dyeing of fur	10	8	2	6.7	-0.80
19	Leather; luggage, handbags, saddlery, harness and footwear	28	13	17	19.3	-0.39
20	Wood and wood products, except furniture; straw and plating	26	31	4	20.3	-0.85
21	Paper and paper products	14	12	2	9.3	-0.86
22	Publishing, printing and reproduction of recorded media	13	6	9	9.3	-0.31
23	Coke, refined petroleum and nuclear fuel	12	22	14	16.0	0.17
24	Chemicals and chemical products	14	11	7	10.7	-0.50
25	Rubber and plastic products	6	12	9	9.0	0.50
26	Other non-metallic mineral products	26	35	40	33.7	0.54
27	Basic metals	17	16	18	17.0	0.06
28	Fabricated metal products, except machinery and equipments	14	11	11	12.0	-0.21
29	Machinery and equipment, n.e.c.	18	9	7	11.3	-0.61
30	Office, accounting and computing machinery	10	10	2	7.3	-0.80
31	Electrical machinery and apparatus, n.e.c.	6	9	9	8.0	0.50
32	Radio, television, and communication equipment and apparatus	12	9	5	8.7	-0.58
33	Medical, precision and optical instruments, watches and clocks	9	6	22	12.3	1.44
34	Motor vehicles, trailers and semi-trailers	13	13	10	12.0	-0.23
35	Other transport equipment	16	7	9	10.7	-0.44
36	Furniture, manufacturing n.e.c.	15	12	22	16.3	0.47

Table 6, continued

NIC	Industry Description	2000	2005	2010	Average	Change
55	Hotels and restaurants	17	16	17	16.7	0.00
60	Land transport (via pipelines)	22	11	20	17.7	-0.09
61	Water transport	15	19	21	18.3	0.40
63	Supporting transport activities, travel agencies	8	11	15	11.3	0.88
64	Post and telecommunications	7	11	27	15.0	2.86
70	Real estate activities	10	15	21	15.3	1.10
71	Renting of machinery, personal goods	12	15	6	11.0	-0.50
72	Computer and related activities	4	1	17	7.3	3.25
73	Research and development	3	1	3	2.3	0.00
74	Other business activities	3	5	16	8.0	4.33
80	Education	21	15	4	13.3	-0.81
85	Health and social work	20	6	13	13.0	-0.35
90	Sewage and refuse disposal, sanitation, etc.	8	20	17	15.0	1.13
91	Activities of membership organizations n.e.c.	27	28	34	29.7	0.26
92	Recreational, cultural and sporting activities	16	16	27	19.7	0.69
93	Other service activities	14	15	16	15.0	0.14
	Correlation to previous survey		0.67	0.47		

Notes: See Table 5. Change compares the 2000 and 2010 periods.

Table 7: Multivariate estimations of district traits and specialization/diversity

	Specialization index			Diversity index		
	Average	Initial	Change	Average	Initial	Change
	(1)	(2)	(3)	(4)	(5)	(6)
Log of district population	-0.286	-0.196	-0.079	0.022	-0.024	-0.157
	(0.392)	(0.383)	(0.162)	(0.355)	(0.433)	(0.153)
Log of district population density	0.709	0.213	0.358*	0.396	0.163	0.288
	(0.491)	(0.480)	(0.181)	(0.445)	(0.580)	(0.261)
Educated worker share (% pop graduate)	0.643*	0.131	0.246*	0.523*	0.145	0.122
	(0.284)	(0.311)	(0.105)	(0.243)	(0.317)	(0.117)
Age profile (demographic dividend)	-0.235	0.238	-0.276	-0.350	-0.164	-0.043
	(0.395)	(0.374)	(0.173)	(0.345)	(0.427)	(0.157)
Literacy rate	0.416	-0.235	0.275+	0.614*	0.010	0.186
	(0.403)	(0.375)	(0.160)	(0.281)	(0.374)	(0.133)
Infrastructure: paved roads	-0.974*	-0.155	-0.348+	0.318	0.750+	-0.226
	(0.453)	(0.445)	(0.193)	(0.370)	(0.440)	(0.144)
Infrastructure: electricity	0.123	0.415	-0.159	-0.566*	-0.013	-0.235*
	(0.342)	(0.322)	(0.098)	(0.280)	(0.308)	(0.096)
Strength of household banking	-0.096	-0.081	0.046	0.114	0.665+	-0.256*
	(0.357)	(0.342)	(0.130)	(0.309)	(0.346)	(0.124)
Log travel time to closest large city	0.006	-0.238	0.159+	-0.092	0.184	-0.008
	(0.318)	(0.356)	(0.090)	(0.168)	(0.249)	(0.085)
Urbanization rate (% urban)	-0.713	-0.359	-0.328	-0.630	0.037	-0.319
	(0.541)	(0.502)	(0.203)	(0.507)	(0.595)	(0.242)
Log per capita consumption	-0.168	0.521	-0.183	-0.396	-1.604	0.660
	(1.183)	(1.093)	(0.442)	(0.900)	(1.197)	(0.416)
Log organized employment	0.133	0.435	-0.044	-0.068	-0.142	0.012
	(0.312)	(0.288)	(0.112)	(0.245)	(0.323)	(0.119)
Log unorganized employment	0.037	-0.298	-0.114	-0.567	-0.961*	0.226
	(0.514)	(0.500)	(0.158)	(0.410)	(0.476)	(0.179)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	339	339	339	339	339	339
Adjusted R-squared	0.159	0.180	0.140	0.408	0.299	0.234

Notes: Table documents multivariate correlations between district traits and specialization and diversity index values. Average and initial index values are expressed on a ten-point scale representing deciles. Change values are expressed in unit standard deviations. District traits are from the 2001 Population Census and ASI/NSS. District traits are expressed in log values or unit standard deviations for interpretation. Estimations include state fixed effects, weight districts by log district population, and report robust standard errors. + significant at 10% level; * significant at 5% level; ** significant at 1% level.

Table 8: Estimations of log employment growth and overall district specialization/diversity

	DV: Log growth in district manufacturing and services employment from 2000-2010								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
District specialization index (2000)	-0.0073		0.0108	-0.0024	-0.0018	0.0016	-0.0024		-0.0220
	(0.0141)		(0.0142)	(0.0135)	(0.0136)	(0.0131)	(0.0135)		(0.0219)
District specialization index in 50th-75th percentile								0.0793	
								(0.1010)	
District specialization index in 76th-99th percentile								-0.0088	
								(0.1020)	
District specialization index in manufacturing									0.0346+
									(0.0189)
District diversity index (2000)		0.0691**	0.0722**	0.0662**	0.0678**	0.0550**	0.0662**		0.1080**
		(0.0136)	(0.0146)	(0.0146)	(0.0148)	(0.0148)	(0.0146)		(0.0170)
District diversity index in 50th-75th percentile								0.4120**	
								(0.0975)	
District diversity index in 76th-99th percentile								0.3440**	
								(0.0976)	
District diversity index in manufacturing									-0.0610**
									(0.0157)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
District trait covariates				Yes	Yes	Yes	Yes	Yes	Yes
Drop districts with <1 mil population						Yes			
Industry projection control							Yes		
Services indices controls									Yes
Observations	353	353	353	334	334	271	334	334	327

Notes: Estimations quantify the relationship between district employment growth and district specialization and diversity indices. Estimations combine manufacturing and services industries. District-level traits used for covariates are those included in Table 7's estimation. Estimations weight observations by the log of district size and report robust standard errors. + significant at 10% level; * significant at 5% level; ** significant at 1% level.

Table 9: Estimations of log employment growth by subgroup

	Total growth in mfg and services	Growth in mfg sector	Growth in services sector	Growth in most important industry in 2000 for district	Growth external to the most important industry in 2000	Growth in establish. with <=10 employees	Growth in establish. with >10 employees	Growth in urban areas of district	Growth in rural areas of district	Concen-tration ratio of local growth	Growth in high-growth industries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
District specialization index (2000)	-0.0024	0.0014	0.0096	0.0177	-0.0058	0.0192	-0.101**	-0.0159	0.0050	-0.0063*	0.0142
	(0.0135)	(0.0198)	(0.0130)	(0.0799)	(0.0141)	(0.0140)	(0.0256)	(0.0171)	(0.0157)	(0.0027)	(0.0090)
District diversity index (2000)	0.0662**	0.0305	0.0311*	0.0339	0.0347*	0.0644**	0.0344	0.0225	0.0786**	-0.0065*	0.0337**
	(0.0146)	(0.0209)	(0.0128)	(0.0763)	(0.0146)	(0.0151)	(0.0252)	(0.0173)	(0.0168)	(0.0027)	(0.0087)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District trait covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	334	327	334	334	334	334	323	334	331	328	335

Notes: See Table 8. Cases where an industry disappears by 2010 are coded with the lowest observed growth value.

Table 10: Estimations of log employment growth by distance bin, density levels, and high-end services

	Total growth in mfg and services	Districts far from India's largest cities	Districts near to India's largest cities	Districts with low population density	Districts with medium population density	Districts with high population density	Adding metric for modern services share	Adding metric for modern services absolute size	Adding both metrics for modern services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
District specialization index (2000)	-0.0024	-0.0097	0.0474	0.0049	0.0111	-0.0085	-0.0010	-0.0010	-0.0010
	(0.0135)	(0.0157)	(0.0362)	(0.0311)	(0.0241)	(0.0216)	(0.0133)	(0.0134)	(0.0134)
District diversity index (2000)	0.0662**	0.0669**	0.0618+	0.0722*	0.0464*	0.0526*	0.0647**	0.0648**	0.0648**
	(0.0146)	(0.0165)	(0.0365)	(0.0297)	(0.0232)	(0.0222)	(0.0141)	(0.0142)	(0.0142)
Indicator variable for top five districts in terms of modern services share							0.639**		0.639**
							(0.1830)		(0.1840)
Indicator variable for top five districts in terms of modern services size								0.0289	0.0142
								(0.2680)	(0.2680)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District trait covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	334	246	87	107	115	113	335	335	335

Notes: See Table 8.

App. Table 1a: Table 5 with manufacturing only

NIC	Industry Description	Diversity		
		2000	2010	Change
15	Food products and beverages	1.74	2.01	0.16
16	Tobacco products	1.56	1.78	0.14
17	Textiles	1.81	2.02	0.12
18	Wearing apparel; dressing and dyeing of fur	1.70	1.95	0.15
19	Leather; luggage, handbags, saddlery, harness and footwear	1.72	1.90	0.11
20	Wood and wood products, except furniture; straw and plating	1.71	2.15	0.25
21	Paper and paper products	1.64	1.68	0.02
22	Publishing, printing and reproduction of recorded media	1.66	1.91	0.15
23	Coke, refined petroleum and nuclear fuel	1.59	1.77	0.12
24	Chemicals and chemical products	1.87	1.84	-0.01
25	Rubber and plastic products	1.86	1.82	-0.02
26	Other non-metallic mineral products	1.69	1.94	0.15
27	Basic metals	1.67	1.82	0.09
28	Fabricated metal products, except machinery and equipments	1.72	1.97	0.15
29	Machinery and equipment, n.e.c.	1.67	1.85	0.11
30	Office, accounting and computing machinery	1.43	1.48	0.03
31	Electrical machinery and apparatus, n.e.c.	1.66	1.82	0.10
32	Radio, television, and communication equipment and apparatus	1.57	1.51	-0.03
33	Medical, precision and optical instruments, watches and clocks	1.61	1.94	0.20
34	Motor vehicles, trailers and semi-trailers	1.65	1.84	0.12
35	Other transport equipment	1.64	1.78	0.08
36	Furniture, manufacturing n.e.c.	1.69	2.02	0.19

Notes: See Table 5.

App. Table 1b: Table 5 with services only

NIC	Industry Description	Diversity		
		2000	2010	Change
55	Hotels and restaurants	1.97	1.78	-0.10
60	Land transport (via pipelines)	1.78	1.77	0.00
61	Water transport	1.85	1.69	-0.09
63	Supporting transport activities, travel agencies	1.82	1.71	-0.06
64	Post and telecommunications	1.82	1.76	-0.03
70	Real estate activities	1.85	1.72	-0.07
71	Renting of machinery, personal goods	1.79	1.81	0.01
72	Computer and related activities	1.92	1.74	-0.09
73	Research and development	1.93	1.55	-0.20
74	Other business activities	1.78	1.78	0.00
80	Education	1.77	1.84	0.04
85	Health and social work	1.78	1.74	-0.02
90	Sewage and refuse disposal, sanitation, etc.	1.97	1.70	-0.14
91	Activities of membership organizations n.e.c.	1.90	1.70	-0.10
92	Recreational, cultural and sporting activities	1.79	1.73	-0.03
93	Other service activities	1.78	1.77	-0.01

Notes: See Table 5.

App. Table 2a: Table 6 with manufacturing only

NIC	Industry Description	2000	2005	2010	Average	Change
15	Food products and beverages	46	45	69	53.3	0.50
16	Tobacco products	31	31	32	31.3	0.03
17	Textiles	37	46	39	40.7	0.05
18	Wearing apparel; dressing and dyeing of fur	23	21	2	15.3	-0.91
19	Leather; luggage, handbags, saddlery, harness and footwear	39	21	25	28.3	-0.36
20	Wood and wood products, except furniture; straw and plating	41	57	4	34.0	-0.90
21	Paper and paper products	23	22	2	15.7	-0.91
22	Publishing, printing and reproduction of recorded media	20	17	25	20.7	0.25
23	Coke, refined petroleum and nuclear fuel	20	27	20	22.3	0.00
24	Chemicals and chemical products	17	16	12	15.0	-0.29
25	Rubber and plastic products	9	15	14	12.7	0.56
26	Other non-metallic mineral products	43	46	70	53.0	0.63
27	Basic metals	22	25	21	22.7	-0.05
28	Fabricated metal products, except machinery and equipments	22	19	22	21.0	0.00
29	Machinery and equipment, n.e.c.	22	11	12	15.0	-0.45
30	Office, accounting and computing machinery	10	13	2	8.3	-0.80
31	Electrical machinery and apparatus, n.e.c.	8	15	15	12.7	0.88
32	Radio, television, and communication equipment and apparatus	14	13	5	10.7	-0.64
33	Medical, precision and optical instruments, watches and clocks	13	11	51	25.0	2.92
34	Motor vehicles, trailers and semi-trailers	17	14	12	14.3	-0.29
35	Other transport equipment	16	12	12	13.3	-0.25
36	Furniture, manufacturing n.e.c.	25	21	52	32.7	1.08
	Correlation to previous survey		0.85	0.43		

Notes: See Table 6.

App. Table 2b: Table 6 with services only

NIC	Industry Description	2000	2005	2010	Average	Change
55	Hotels and restaurants	42	30	36	36.0	-0.14
60	Land transport (via pipelines)	50	30	45	41.7	-0.10
61	Water transport	21	20	22	21.0	0.05
63	Supporting transport activities, travel agencies	23	19	20	20.7	-0.13
64	Post and telecommunications	26	28	33	29.0	0.27
70	Real estate activities	23	30	33	28.7	0.43
71	Renting of machinery, personal goods	31	28	7	22.0	-0.77
72	Computer and related activities	12	1	26	13.0	1.17
73	Research and development	5	1	4	3.3	-0.20
74	Other business activities	15	18	33	22.0	1.20
80	Education	42	38	5	28.3	-0.88
85	Health and social work	26	13	29	22.7	0.12
90	Sewage and refuse disposal, sanitation, etc.	11	27	23	20.3	1.09
91	Activities of membership organizations n.e.c.	36	41	43	40.0	0.19
92	Recreational, cultural and sporting activities	35	32	39	35.3	0.11
93	Other service activities	28	32	28	29.3	0.00
	Correlation to previous survey		0.73	0.34		

Notes: See Table 6.