Policy Research Working Paper 6648

Specialization, Diversity, and Indian Manufacturing Growth

Ejaz Ghani
William R. Kerr
Ishani Tewari

The World Bank
Poverty Reduction and Economic Management Network
Economic Policy and Debt Department
October 2013

Policy Research Working Paper 6648

Abstract

This paper examines the specialization and diversity of manufacturing industries within Indian districts. Prior to India's recent economic growth and liberalization, specialization levels in 1989 were substantially higher than similar metrics calculated for the United States. From 1989 to 2010, average specialization levels for Indian districts declined to a level that is now quite comparable to the United States. Diversity levels similarly increased. Specialization and diversity levels in India are becoming more persistent with time. Manufacturing plants display higher productivity in districts that display both properties. From 1989 to 2010, manufacturing employment growth was higher in districts that were more specialized at the start of the period.

This paper is a product of the Economic Policy and Debt Department, 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.

Specialization, Diversity, and Indian Manufacturing Growth

Ejaz Ghani, William R. Kerr, and Ishani Tewari

Keywords: Specialization, diversity, growth, development, manufacturing, urban, India.
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 Rahul Gupta and Stephen O'Connell for helpful suggestions/comments and data calculations. Funding for this project was provided by the World Bank. The views expressed here are those of the authors and not of any institution they may be associated with.

1. Introduction

The continued economic development of India has the capacity to better the lives of over one billion people. It is widely recognized that a key aspect of this development process is improvements and enhancement in India's industrial organization, perhaps in close connection with increased urbanization and/or greater formal organization of business. India is emerging from an industrial past with extensive distortions due to government controls and regulations, and many scholars today argue that India still contains extensive misallocation of activity across plants and regions (e.g., Hsieh and Klenow 2009, Ghani et al. 2013a, Desmet et al. 2011). Despite the recent progress of researchers and policy makers towards understanding these issues, many big questions remain unexplored.

This paper takes on one such topic—the specialization and diversity of Indian districts. It has been over two decades since Glaeser et al. (1992) brought the connections between city specialization/diversity and urban growth to the forefront of urban economic analysis. Yet, at this point, we are not aware of any single study that attempts to even measure these specialization and diversity traits for India. We aim to fill that void and to document the connection between these district traits and two key economic outcomes: the productivity of establishments and the employment growth of districts over time. We build on a solid body of research for advanced economies (e.g., Duranton and Puga, 2000), but our analysis is also focused on a country where it is unclear how many of these lessons will continue to hold. India's emergence from its economic liberalization in the early 1990s leaves it at quite a different place from the United States' declining manufacturing sector post-1980 that is most often studied.

The theoretical foundation for the link between industrial composition and growth is forged via technological externalities—that is, the spillovers of innovation and improvements from one firm to another without full compensation. Although there is general agreement on this particular channel, it is not clear whether diversity or specialization is more likely to lead to these externalities. Three theories of dynamic externalities link specialization/diversity and economic development, as depicted by Glaeser et al. (1992). Marshall-Arrow-Romer (MAR) type externalities hold that spillovers come from within industries, but only when concentration is high. Porter (1990) also holds that spillovers emerge from within industries, but only in the presence of competition. Thus, these first two theories favor specialization over diversity for innovation and growth. Jacobs (1969), on the other hand, argues that knowledge spills across industries and that competition is crucial to innovation. Thus, according to Jacobs, industrial variety and diversity are conducive to growth.

Thus, the theory as to whether specialization or diversity most enhances growth is ambiguous, and the empirical literature for the United States is nuanced and sometimes conflicting. This makes empirical measurement for India all the more important. Previous studies emphasize that urban diversity fosters employment growth (Glaeser et al., 1992), new and innovative industries, and the production of less-standardized or non-traditional items (Henderson, 1997a). Specialized cities are often thought better suited for mature industries and the production of standardized and export-oriented products. Some of the channels linking diversity/specialization to economic development are the roles that diversified cities play in fostering innovation (Duranton and Puga, 2001, Feldman and Audretsch, 1999), increasing employment (Glaeser et al., 1992), promoting entrepreneurship (Jacobs, 1969), or reducing costs (Lall et al., 2003). Duranton (2013a) provides a recent review in the developing country context.

This paper documents a detailed set of trends and outcomes for the Indian manufacturing sector. We measure patterns of industrial specialization and diversity in Indian districts from 1989 to 2010. We document that average specialization levels for Indian districts are decreasing over time, from very high levels before liberalization, and are now on par with levels observed in the United States. Specialization in urban areas is particularly strong and follows a similar path. While the levels are comparable to the United States, the persistence of district-level values in India remains very weak. That is, India's spatial distribution of manufacturing is still quite in flux, though stability is increasing with time. Modern industries are quite spatially concentrated, and traditional industries form the highest specialization for most districts. In terms of economic consequences, local specialization and diversity are jointly associated with stronger production functions, and the marginal impacts of these benefits are mostly located outside of the industry leading the local specialization. Perhaps most important, we document a very strong link within India for the initial specialization of districts and their employment growth in manufacturing. The link between diversity and growth appears more non-linear.

The remainder of this paper is as follows: Section 2 describes the data used for this paper and the calculation of our specialization and diversity indices. Section 3 provides an extensive description of the specialization and diversity in Indian manufacturing districts and the trends over time. Section 4 estimates the role of these district traits in manufacturing production functions and the employment growth of districts over the 1989-2010 period. The final section concludes and provides some thoughts about future work on these important topics.

2. Indian Manufacturing Data and Index Calculations

This section begins with a description of the Indian and U.S. manufacturing 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 their important empirical properties.

Indian Manufacturing Data

We employ repeated cross-sectional surveys of manufacturing establishments carried out by the government of India for the fiscal years of 1989, 1994, 2000, 2005, and 2010. In all cases, the survey was undertaken over two fiscal years (e.g., the 1994 survey was conducted during 1994-1995), 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 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 manufacturing employment and establishments, respectively (Ghani et al., 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 total establishments, employment, and output by district. Districts are administrative subdivisions of Indian states or territories that provide meaningful local economic conditions. The average district size is around 5,500 square kilometers—roughly twice the size of a U.S. county—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 like employment, output, and raw materials. When we estimate plant-level production functions, we use the reported traits directly. Most of our analysis considers aggregated measures of manufacturing activity 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. 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.[1]

Our core sample contains 429 districts. This sample is smaller than the total number of districts in India of 630, but it accounts for almost all of the plants, employment, and output in the manufacturing sector throughout the period of study. The reductions from the 630 baseline occur due to requirements that manufacturing employment be observed in every period for the district (i.e., we have balanced panels of districts from 1989 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.

[1] We also considered three-digit level metrics. Using finer industry categories has the challenge of larger numbers of extreme values, especially for the specialization index. The two-digit level balances meaningful industry groups with better measurement.

United States Manufacturing Data

It is useful when discussing the levels of Indian specialization and diversity to benchmark them against other countries. Surprisingly, there is very little tabulation of these metrics internationally. We will compare our Indian metrics with those that we calculate for the United States using 2011 manufacturing data from County Business Patterns (CBP). Our index calculations take the same approach, and we have two CBP data issues to highlight. First, we utilize the three-digit level of the North American Industry Classification System (NAICS). At this level of industry aggregation, NAICS contains 21 manufacturing industries, which is very comparable to the 22 industries at the NIC two-digit level. The NAICS and NIC industries are defined somewhat differently, and to a limited degree this could influence the measures. For example, a more evenly balanced distribution of employment across one set of industries would reduce the likelihood of a very high specialization value occurring. Nevertheless, given the broad comparability of manufacturing industry definitions, we believe this issue to be very minor.

Second and more important, CBP data do not disclose values if tabulations risk violating the confidentiality of individual plants. That is, CBP will not provide the employment for the chemicals industry in a location if there are only one or two chemical plants, as this would be very informative about the underlying plants themselves. This issue is very prominent at the county level, and it becomes increasingly less important with higher levels of spatial aggregation. We thus consider states and metropolitan areas (which we refer to as cities for simplicity). CBP reports the total manufacturing employments for cities, and thus we can calculate the share of the total employment that has been allocated over industries. We report below tabulations that use four levels of aggregation (observation count): states and District of Columbia (51), the top 100 cities in terms of manufacturing employments (100), cities with at least 90% of manufacturing employment allocated (193), and cities with at least 60% of manufacturing employment allocated (746).

Specialization and Diversity Indices

We follow Duranton and Puga (2000) 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). 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 manufacturing 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. 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 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.[2]

Empirical Application of the Specialization and Diversity Indices

The next section provides descriptive statistics of these metrics for Indian manufacturing, 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

[2] 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.

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, 18% of observations have above-median values on both indices, while 18% of observations have below-median values on both indices.[3] The metrics have on average a correlation of about -0.2 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. We observe empirical evidence below that both of these traits can simultaneously promote manufacturing plant productivity.

Second, the metric designs (especially specialization) can yield extreme values. 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. When we move to productivity and 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 3 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 a very undiversified industrial base, while it is easier for large districts to maintain specialization in one industry.

---

[3] In the 2011 U.S. data, 11% of observations have above-median values on both indices, while 12% of observations have below-median values on both indices.

Thus the connections of our indices to these properties are intriguing and may differ from those in advanced economies, 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 manufacturing. As we are the first to depict these patterns for India (and generally among the first studies conducted outside of an advanced economy[4]), we devote extra attention to this section.

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 429 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 five periods. We also calculate the change in specialization and diversity for a district in relative terms using the formula of ($\text{AverageValue}_{2000\text{-}2010}$ - $\text{AverageValue}_{1989\text{-}1994}$)/ $\text{AverageValue}_{1989\text{-}2010.}$ The last two columns provide summary statistics for these metrics.[5]

We focus our descriptive analysis on 2005 due to changes in industry codes between the 2005 and 2010 surveys that make 2010 less comparable to earlier periods (discussed further in Table 6). Looking at 2005, the average specialization value is 6.65, and the median value is 4.49. This suggests that, for the average Indian district, the maximum degree to which one industry's employment exceeds its national share is around 600% (e.g., the industry constitutes 6% of the district's employment relative to 1% nationally). There is a wide variance in this metric, with the standard deviation equal to the mean. The lowest winsorized level in 2005 is 2.1 (e.g., a 2% local

[4] Duranton (2013b) considers specialization and diversity in Colombia.

[5] The change formula follows Davis et al. (1996) by measuring changes relative to the average across periods. This approach limits issues related to mean reversion and reduces the impact of outliers that could result from abnormally low initial values. The value of the change metric is bounded between [-2, 2] and treats positive and negative adjustments symmetrically.

share combined with 1% nationally), while the maximum winsorized level is 40.0 (e.g., a 40% local share combined with 1% nationally).

Is a value of 6.65 high or low? Table 1c compares the 2005 Indian values to what we observe in the United States. Looking at Panel A, the mean and median values for specialization increase as one includes more spatial areas with smaller sizes. This is not surprising given the properties of the indices that we highlighted in the previous section. Recalling that the typical Indian district is about the size of two U.S. counties, we believe that Indian data most closely compare to the last two columns, which include 193 and 746 U.S. cities, respectively. It is very interesting that the Indian level of specialization falls right around what one would expect based upon the U.S. levels, despite the substantial differences between the two countries.[6]

Returning to Table 1a, this comparability of Indian specialization to U.S. values is not present at the start of our sample period. Even with the annual winsorization, the 1989 mean value is significantly higher than in 2005 and is declining consistently in the interim surveys. The same holds true for the other metrics regarding specialization like standard deviation and maximum values. Thus, around the time of liberalization, we estimate India's manufacturing distribution to have been more specialized in most districts and to have displayed a much larger spatial variation than the U.S. data. India has converged towards the U.S. baseline in the two decades since then.

Panel B of Table 1a considers the relative diversity index. The average diversity in 2005 is 1.41, which again is in general agreement to the United States' average level in 2011. The longitudinal pattern for the Indian diversity index is similar to those in Panel A—decreasing specialization corresponding to increasing diversity—but the patterns are more muted, likely in large part due to the diversity index's metric design being more stable than the design of the specialization index. For both specialization and diversity, we see some re-widening of the distributions in 2010 (greater standard deviation). 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.

[6] We have not winsorized the indices for the United States since we want to report multiple levels of aggregation. Nonetheless, the same comparability is also evident for the median values that are unaffected by winsorization.

Figures 1 and 2 graph the index values against each other using 2005 Indian data and 2011 U.S. city data (60% allocation), respectively. The bubble size in each graph represents the size of the local manufacturing basis, and the horizontal and vertical axes provide the median value for each index. In Figure 2, we winsorize U.S. values at the levels of the Indian winsorized values in 2005 so that the scales are comparable. Both graphs show a negative correlation, such that higher diversity is associated with reduced specialization. The Indian correlation is weaker at -0.2 than the U.S. correlation of -0.4. Visually, the U.S. data exhibit a much more regular pattern in terms of this trade-off than the Indian data do.

In advanced economies, specialization patterns of cities are quite stable over time (Henderson, 1997). The substantial trend discussed in Table 1a suggests that this finding may not hold in India. Since 1989, India has experienced rapid economic growth, major infrastructure investments, a greater openness to foreign trade, and the adjustment of many industrial policies regarding location choice (e.g., reduced effort to have industry locate in lagging regions, development of special economic zones). These and other factors may weaken the persistence of district specialization and diversity levels.

Tables 2a-3b show the limited persistence of Indian specialization and diversity levels. For the specialization index in Panel A of Table 2a, there is a 0.2-0.3 correlation of district values across adjacent surveys that are about five years apart. By itself, this correlation over a short time period is pretty small. Across two surveys, or about ten years in duration, these correlations are neither economically or statistically meaningful. Thus, there is rapid change in specialization levels for the period that we consider. Panel B shows a similarly rapid decline for the diversity index, although there is less additional attenuation over the ten-year span. Table 2b looks pretty similar when we isolate urban areas of districts. In both tables, there is some evidence that the patterns may be showing greater persistence in the later periods than earlier, which we quantify next.

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 grouping districts into quintiles of specialization in 1989. For example, the top row is for the 20% of districts that have the lowest specialization in 1989, while the fifth row documents the 20% of districts with the highest 1989 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 2.57 by 1994, and then 2.86 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 3.53 by 1994, and then to 3.19 by 2010. Had there been no movement in the rankings of districts, the values would have continued to look like the 1989 column. Had there been perfect mobility, the values for all districts would be 3.0 as the initial ordering of districts in 1989 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 1989 (i.e., 5.0-1.0). This spread decreases to 0.96 by 1994 (i.e., 3.53-2.57), 0.64 by 2000, 0.21 by 2005, and 0.33 for 2010. For diversity, the initial 4.0 spread similarly decreases to 0.73 by 1994, 0.29 by 2000, 0.37 by 2005, and 0.50 for 2010. This fairly rapid adjustment is indicative of the substantial changes occurring to the Indian manufacturing sector that we noted above. Duranton (2013b) also observes a very low degree of persistence in the production structure of Colombian cities.

Table 3b takes a second approach by considering how the initial 1989→1994 transition compares to that observed across adjacent surveys in later periods. This helps shed further light on whether or not the Indian distribution of specialization and diversity is stabilizing with time, which the data have hinted at on several occasions. We build Table 3b by reordering districts in each survey and calculating new quintiles. We then examine in the next survey the ranks of these cities similar to the initial 1989→1994 transition.

The patterns in Table 3b generally speak to increased stability. This is observed by noting the widening distribution span moving left to right across the table. From 1989 to 1994, the average district in the lowest initial quintile climbed to an average value of 2.57 in 1994, while the average district in the top initial quintile fell to 3.53. From 2005 to 2010, the average district in the lowest initial quintile in 2005 climbed only to 2.41 by 2010, while the average district in the top initial quintile only fell to 3.82. Thus, the substantial decrease in spread to 0.96 during 1989→1994 noted in Table 3a was weaker at 1.42 during 2005→2010. Likewise, the decrease in the diversity spread to 0.73 during 1989→1994 noted in Panel B of Table 3a was weaker at 1.48 during 2005→2010. On the whole, the initial rapid changes in Indian distributions appear to be stabilizing, with perhaps particular stability emerging at the top of the distribution.[7]

Finally, Figures 3-6 graph the district-level patterns in the Indian data across districts. 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 2000-2010 values against 1989-1994 (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 3 of the 12 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 1 of the 24 districts that form an extreme average value on the specialization index in Panels A and B of Table 4a is also an extreme value on the corresponding diversity lists in Table 4b.

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 calculate for an industry the weighted-average specialization and 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 employments for an industry are in districts that are highly specialized, we will measure a high average specialization value for the industry.

In Table 5, we report the levels of these values for industries and their changes over time. The clearest pattern is that more advanced industries like office, accounting and computing machinery (NIC 30) and radio, television, and communication equipment and apparatus (NIC

[7] A comparison point from U.S. states for 2005→2011 is a transition set of specialization of 1.2, 2.3, 3.0, 3.9, and 4.6 (lowest initial quintile to highest initial quintile). For diversity, the transition set is extremely stable at 1.1, 2.1, 3.2, 3.7, and 4.9. Unfortunately, the MSA definitions in the CBP data are not consistent enough to calculate at a lower spatial level.

32) are located in more specialized districts. At the bottom of the table, we group industries into "traditional" or "modern" and find that the latter is generally located in more specialized areas than the former.[8] The traditional versus modern differences are stable over time. By contrast, and showing more of the distinctions between the specialization and diversity indices, there are no differences evident with respect to the average diversity values of the districts in which these two groups of industries locate.

Table 6 provides a second industry perspective. 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 1989-1994 to 2000-2010. While modern industries tend to be in very specialized locations, this concentration means that on average for our 429 districts, it is a traditional industry that is responsible for the local specialization value. Roughly two-thirds of districts have their specialization in a traditional industry, and this share is climbing slightly over time. The computer and communication industries, by contrast, form the specialized industry for typically 10-15 districts over the sample period, which is a substantially smaller count than many of the larger traditional industries (average count across industries is 20).

The bottom of Table 6 documents the correlation of these counts between adjacent surveys. While the patterns across the 1989 to 2005 surveys are quite stable, more substantial differences emerge at the industry level for the 2010 survey compared to earlier years. 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. This uncertainty is why we take the 2005 levels in Table 1a for comparison to the United States. 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.

[8] These groupings follow Ghani et al. (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. Specific values are documented in the notes to Table 5.

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.[9] 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 manufacturing specific covariates 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 weakness is observed in the limited number of regressors that pick up a statistically significant coefficient (often only two out of the 14 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).

Second, there are two traits of districts that consistently link to these industrial traits. The literacy rate of places is closely linked to their initial levels and subsequent changes. Districts with high literacy rates in 2000 start with both a more specialized and diversified industrial base. It is important to recognize that these partial correlations are conditional on other district traits like education shares, population levels, and consumption per capita. Places with high literacy rates also shift towards less specialization over time.[10] Second, the manufacturing share of the

[9] Appendix Tables 1 and 2 report univariate correlations between the index values and district and industry traits, respectively. Caution should be exercised about over-emphasizing any single correlation, given the lack of other controls.

[10] This connection to literacy rates is also intriguing given that Ghani et al. (2013b) find in contemporaneous work 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). This connection to literacy rates is similarly robust to controlling for many other traits of districts.

district's employment in 2000 matters for the changes observed. Areas with a high concentration of local employment in this sector are observed to have 1989-2010 shifts toward specialization and away from diversity. In the upcoming growth analyses, we will control for the traits modeled in Table 9, 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). Duranton (2013b) also observes this for Colombia. For India, at least with respect to manufacturing, we do not see this pattern. These null patterns are also evident with population density.[11] 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.

4. Economic Productivity and Growth with Specialization and Diversity

This section describes our empirical exercises regarding economic outcomes. We first consider production functions for manufacturing establishments that estimate how the specialization and diversity levels of the district in which a plant is located link to its productivity. We then consider whether the growth of a district's manufacturing employment links to the district's specialization or diversity.

Manufacturing Production Function Estimations

Table 8 analyzes whether local specialization or diversity relates to the productivity of Indian manufacturing establishments. We use the 2000 survey for this exercise. We estimate a

[11] Our framework does not yet allow us to compare fully to the United States and Colombia in this regard. For example, we have yet to include services in our framework, and the non-traded nature of services is one reason given for the size-specialization relationship. (It is worth noting, however, that most of the informal sector employment that defines our manufacturing metrics comes in effectively non-traded forms.) Big cities, where firm headquarters and service firms are often based, tend to specialize in business services. Henderson (1997) finds that medium cities have more mature manufacturing industries (but not newer technology-centric manufacturing).

simple production function with log output Y of each establishment i in a district d and industry j as the dependent variable. The specification takes the form:

$$Y_{i,j,d} = \beta \cdot Specialization_d + \delta \cdot Diversity_d + \mu \cdot X_i + \phi \cdot Z_d + \gamma_j + \varepsilon_{i,j,d}.$$

Our core regressors are the specialization and diversity measures for each plant's district. For these estimations, we use a form of indices that measures specialization and diversity on a ten-point scale ranging from zero to nine based upon the decile of a district in the overall distribution for India. This transformation makes the β and δ parameters easier to interpret as described below. This transformation also implicitly suppresses some of the variation in specialization index value vis-à-vis the diversity index.

We include a vector X_i of plant inputs into the production function: log employees, log book values of capital, and log costs of materials. We exclude plants with missing values for these metrics, which increases the relative weight of organized-sector facilities in the estimations. We also control for other district traits described below with a vector of district-level controls Z_d. Regressions include three-digit industry fixed effects γ_j to capture regular differences in production techniques and spatial locations across industries. We use establishment weights from the surveys and cluster standard errors by district.

Column 1 provides a baseline estimation of the plant-level production function before district-level conditions are incorporated. These underlying parameters for the production function, emphasizing employees and materials, are very stable across estimations. Columns 2 and 3 show that simple connections do not exist for the production function to specialization or diversity in isolation. Interestingly, Column 4 finds, however, that both specialization and diversity correlate with greater productivity when jointly modeled. The 0.015 coefficients can be interpreted as quantifying that each increase in decile of specialization or diversity (e.g., moving from the 8th to the 9th decile) is associated with a 1.5% increase in the conditional output of the plant.

While we have yet to come to terms with why both features of the district's industrial base need to be modeled, we do know that this pairing is not simply reflecting other aspects of the district landscape. Column 5, for example, shows very similar results when controlling for log manufacturing employment in district per square kilometer, log manufacturing employment in plant's district-industry per square kilometer, log share of local manufacturing employment in the unorganized sector, and log share of district-industry manufacturing employment in the unorganized sector. The first two of these controls are often associated with the urbanization and agglomeration premiums of dense locations, and we use a per square kilometer normalization as Indian districts vary in spatial size.[12] Column 6 likewise finds very similar results when further modeling the manufacturing share of local employment. Controlling for these extra covariates does not impact the reduced significance for specialization or diversity when they are modeled independently of each other.

Columns 7-9 consider heterogeneity in the sample. Columns 7 and 8 again split the Indian sample by traditional versus modern sectors, finding that the specialization effect is present in both sectors, but the diversity impact is concentrated in variation over traditional sectors. Column 9 determines for each district its specialized industry and then enters a dummy variable for that industry. We further interact this dummy variable with the two specialization and diversity indices. Interestingly, the indicator variable itself suggests an upwards productivity shift for the plants in the specialized industry, although this effect is not statistically significant. On the other hand, the interactions suggest that the marginal productivity gains to further specialization and diversity are associated with plants outside of the specialized industry itself. A similar unreported exercise introduces an indicator variable for plants in the organized sector and their interactions with district specialization and diversity. Organized plants demonstrate substantially higher productivity than unorganized plants, but most of the marginal productivity gains to further specialization and diversity are associated with plants in the unorganized sector.

We stress that these estimations only document partial correlations, and we have not identified an exogenous shifter in local specialization or diversity. The rapid changes noted in the earlier section may allow for such a metric to be identified in the future. Nevertheless, even in their current form these estimations show a potentially important link of the local industrial organization to the productivity of Indian plants not understood before.

[12] For example, Ciccone and Hall (1996), Duranton and Puga (2004), and Rosenthal and Strange (2004, 2011). Indian studies include Lall et al. (2004), Lall and Mengistae (2005), Deichmann et al. (2008), and Fernandes and Sharma (2011).

Growth of Manufacturing Bases in Districts

Moving from the productivity regressions, Table 9 considers the potential link between specialization and growth. The topic of specialization/diversity and city growth has been often studied over the past two decades (Duranton, 2013a), and we seek to provide early evidence on this link within Indian manufacturing. 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 1989-1994 data. The outcome variable is the log employment growth in manufacturing for the district from 1989 to 2010. Regressions include a vector of state fixed effects χ_s to account for differences in regional growth in manufacturing for India across the period. The vector Z_d contains the additional district-level controls that we modeled in Table 7. We weight districts by their log population in 2001 and report robust standard errors.

The first column shows a large partial correlation for the specialization index. An increase of one decile in the specialization index is associated with 9% higher employment growth (e.g., moving from a growth rate of 2.00% to 2.18%). This is a quite substantial boost that Columns 2 and 3 show is not present with the diversity levels of districts. 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, with the main exception being a robust relationship between education shares and higher growth. Regardless, the inclusion of these controls does not impact the specialization 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 manufacturing data start to become 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 1989 and the national rate of growth by industry across the period.

Column 8 replaces the ten-point measure with 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 quartile of initial specialization. We also find quantitatively similar results when considering raw index values. Interestingly, the diversity indicator variables suggest that the general null result observed on this margin may be due to non-linear features of the data, as the third quartile's coefficient is quite strong. Finally, Column 9 shows that these results are associated with the initial specialization value and not the change in specialization over the period.

In addition to these variants, we have performed several other robustness checks. First, if using average index values across the full period, the coefficients for specialization and diversity are 0.102 (0.032) and -0.009 (0.038), respectively. Using just the 1989 value for initial specialization lowers the coefficient to 0.047 (0.022), which is to be expected given the dramatic changes observed earlier during the initial 1989→1994 adjustment. Adding an additional control for the average employment level in manufacturing across the 1989-2010 period in a district yields a specialization coefficient of 0.082 (0.031). Finally, we observe employment growth both within the specialized industry for a district and also outside of it. The coefficient when isolating the specialized industry of districts is 0.115 (0.049), while it is 0.074 (0.032) when focusing on all industries in a district other than the specialization district.

On the other hand, there are some limits to these findings. While employment growth displays a clear pattern, the relationship of initial specialization to output growth is more tenuous and depends significantly on data preparation steps like deflator use. Second, when looking at changes in employment by period, we do not observe a positive relationship in all cases, with again some modeling choices that do not matter when examining the whole period (e.g., the inclusion of state fixed effects) starting to become important when considering each period separately. These break-outs suggest the most important growth period was the 2000→2005 period.

Table 10 further extends these regressions to consider the specialization and diversity differences across urban and rural areas of districts. Interestingly, we see employment growth for the district as a whole is most closely connected to both traits in urban areas and to rural specialization. The latter may be especially related to the movement of organized plants to rural areas in many Indian districts (e.g., Ghani et al. 2012a), while the former is more closely associated with traditional city dynamics that can favor both industrial forms.

To summarize, there is a very clear and robust relationship between specialization and growth in the Indian manufacturing data over the 1989-2010 period. Similar to the productivity estimations, these relationships remain non-causal in nature. They do, however, provide new guidance as to how Indian specialization and diversity link to local growth, which we can begin to compare with other countries.

5. Conclusions

The industrial specialization and diversity trends documented in this paper for India are very exciting. They show evidence of a rapid change from specialized, undiversified manufacturing districts prior to India's liberalization to a pattern today that has a distribution more closely resembling that of the United States. While India's economic geography remains in flux, the very rapid shifts during the 1990s have segued to more modest adjustments during the last decade. Looking forward, modern industries are quite spatially concentrated in India, which has the corollary that most districts derive their highest specialization from a traditional industry. In terms of economic consequences, local specialization and diversity are jointly associated with stronger production functions. The marginal impacts of these benefits are mostly located outside of the industry leading the local specialization. In terms of manufacturing employment growth, the greatest gains over the 1989 to 2010 period are observed in districts that had the highest initial specialization levels, with diversity perhaps showing a non-linear role.

We see several very promising avenues for further study. First, there remain some intriguing issues and question marks within the patterns that we have outlined. We have noted the general movements of Indian districts towards less specialization and greater diversity; we have also observed, however, that districts with higher initial specialization exhibited greater manufacturing employment growth. These facts are not necessarily at odds with each other, as much of the growth could have come outside of the initial specialization industry, just as we observed broader marginal productivity spillovers. But it is not clear why this would be the case, and it hints that very interesting patterns may exist when jointly studying districts at the regional level. For example, it is known that Indian government policy prior to the deregulations promoted industrial placement in less developed locations in the name of distributional equality. These patterns are perhaps showing how this initial, artificial placement unwound itself, with a general movement of regional industry towards a focal point (somehow determined). Our current framework treats each district as a separate entity, and answering these types of questions will necessitate modeling the spatial distances between districts of various types. Such an extension would also allow us to use plant age data contained in our surveys to measure the degree to which maturing industries move from diversified districts to specialized locations (e.g., Duranton and Puga, 2001).

Second, 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., 2010a,b), but we do not know their role in terms of the specialization and diversity of India's districts. For example, highly tradable goods may have also been the goods that could have formed the basis for the most specialized districts before India's reforms. 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. Likewise, Ellison et al. (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.

Finally, manufacturing is a natural starting point, especially given its study in many countries. India, however, has relied extensively on services growth for much of its economic development, and it is critically important in this context for us to learn about how this sector's patterns are similar or different to those in manufacturing (e.g., does services growth link to specialized districts for services). This extension would also allow us to compare India more closely with the work on Colombia by Duranton (2013b), who finds some interesting differences between manufacturing and services performance around specialization dynamics. This added perspective will provide a richer foundation for understanding growth in Indian cities.

References

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

Ciccone, Antonio, and Robert Hall, "Productivity and the Density of Economic Activity," *American Economic Review* 86:1 (1996), 54–70.

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, William Kerr and Stephen O'Connell, "Spatial Determinants of Entrepreneurship in India," National Bureau of Economic Research Working Paper No. 17514 (2011).

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, 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).

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 (2010a), 1727-1767.

Goldberg, Pinelopi, Amit Khandelwal, Nina Pavcnik, and Petia Topalova, "Multi-product Firms and Product Turnover in the Developing World: Evidence from India," *Review of Economics and Statistics* 92:4 (2010b), 1042-1049.

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.

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

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," McKinsey & Company, April 2010.

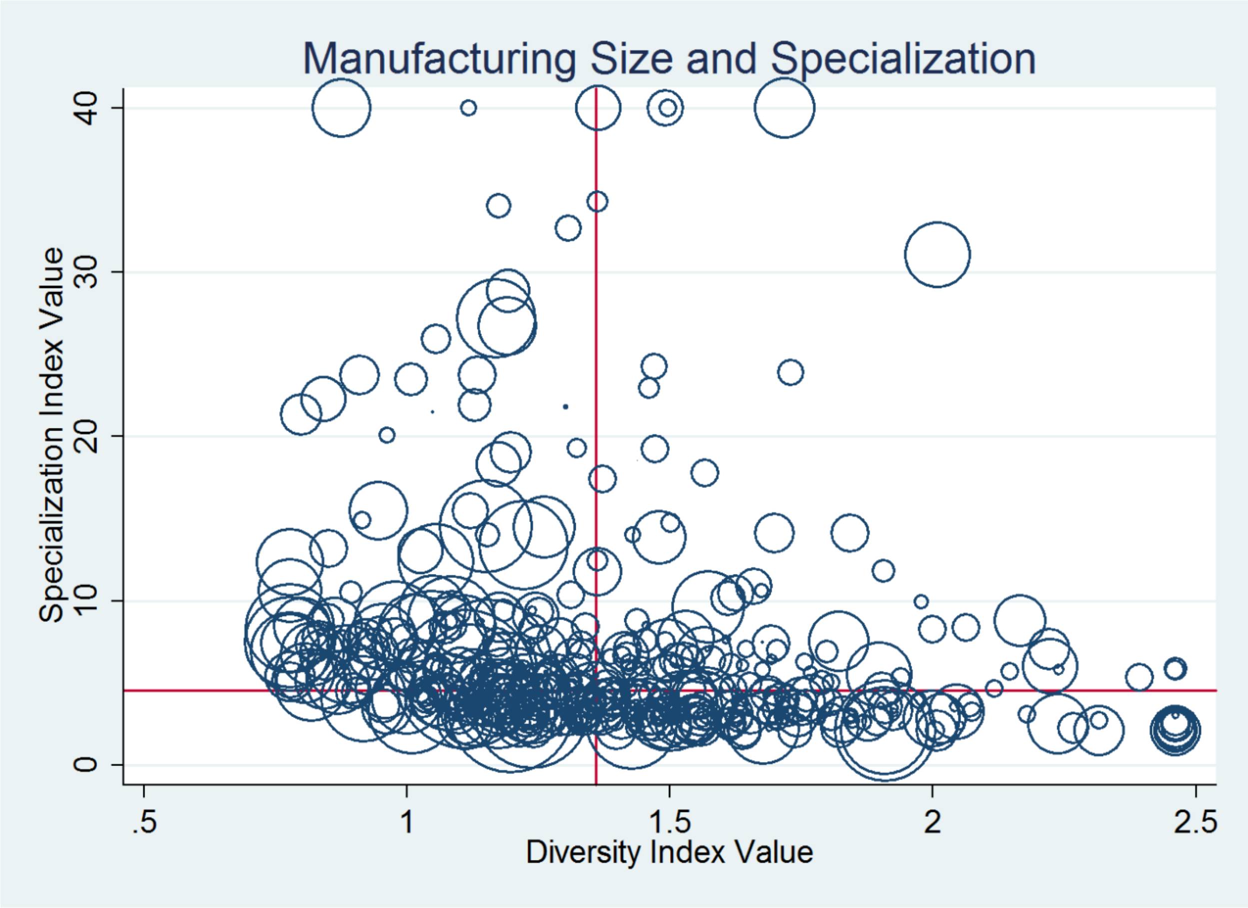
McKinsey Global Institute, "The Shifting Urban Economic Landscape: What Does it Mean for Cities?" McKinsey & Company, July 2012.

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.

Rosenthal, Stuart, and William Strange, “Female Entrepreneurship, Agglomeration, and a New Spatial Mismatch,” *Review of Economics and Statistics*, forthcoming (2011).

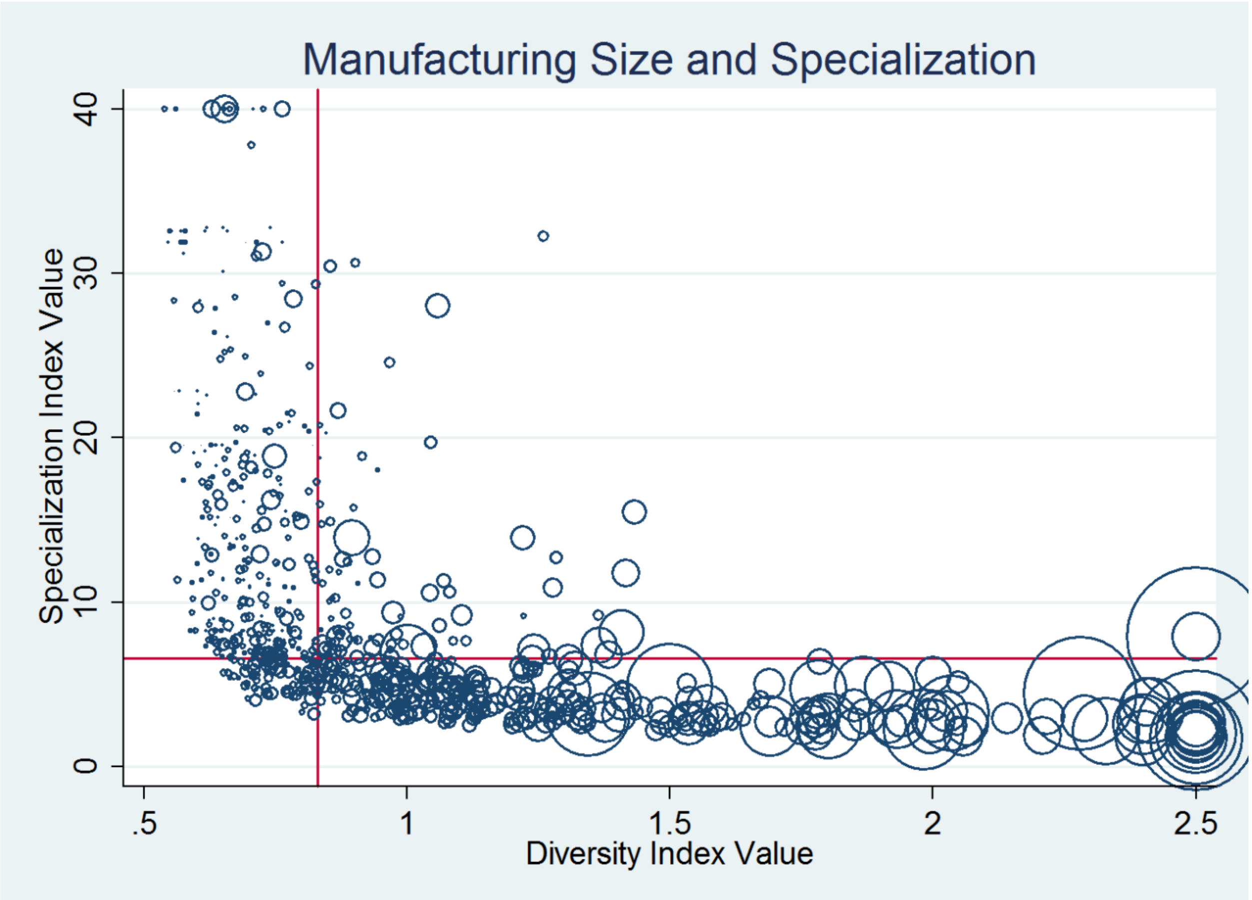
World Bank, “Planning, Connecting, and Financing Cities—Now,” Urbanization Review Flagship Report, The World Bank, 2013.

Figure 1: India Specialization and Diversity, 2005



Notes: Each observation is a district in 2005. Bubble size indicates log manufacturing 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.

Figure 2: United States Specialization and Diversity, 2011



Notes: Each observation is a city in 2011 that contains at least 60% of its manufacturing employment allocated in county business patterns. Bubble size indicates manufacturing employment in the city. Values for specialization and diversity are winsorized for the graph to match the Indian values in Figure 1. The red axis lines document the median values for each index.

Figure 3: Average Specialization Values, 1989-2010

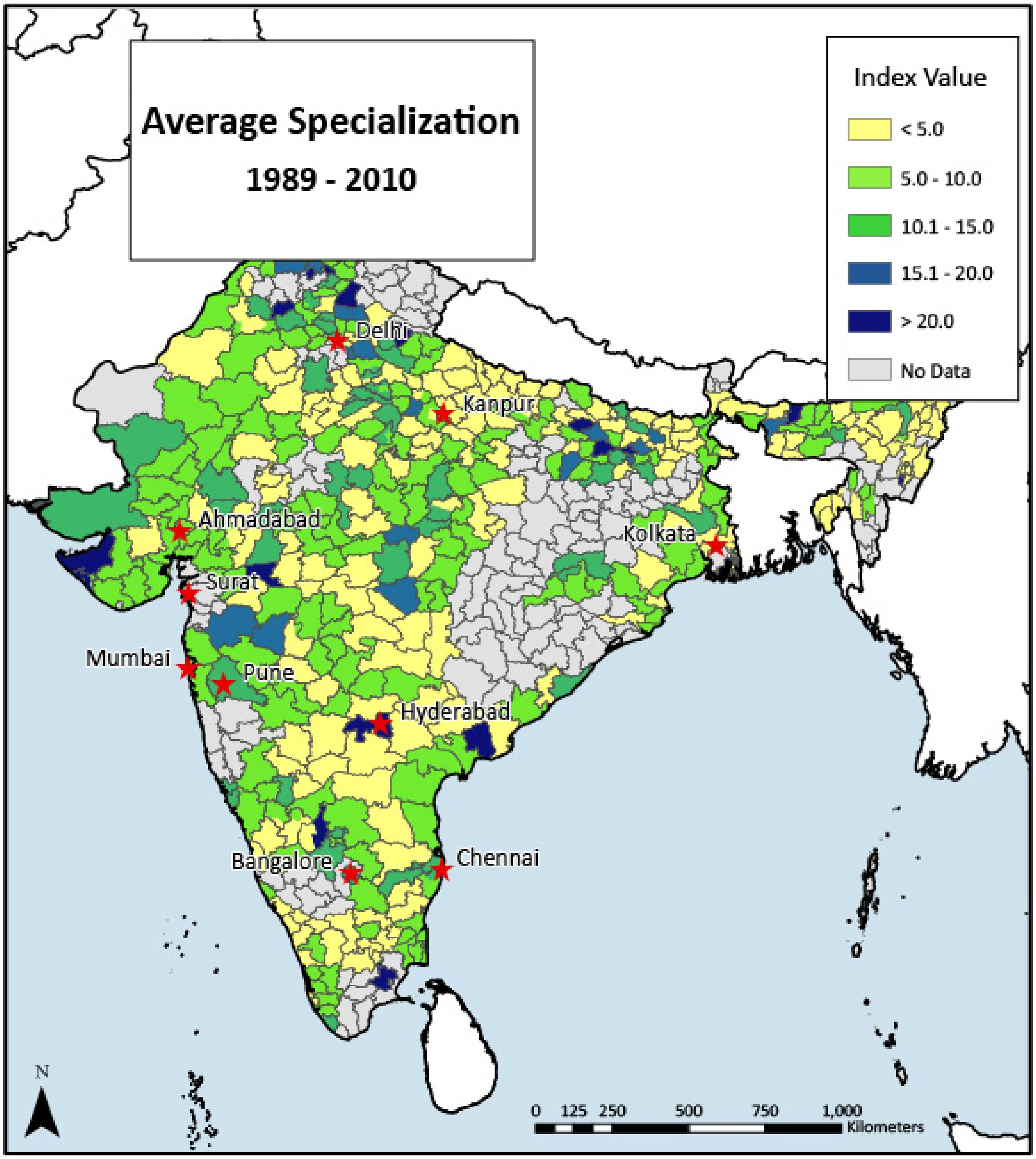


Figure 4: Change in Specialization Values, 1989/1994 to 2000/2010

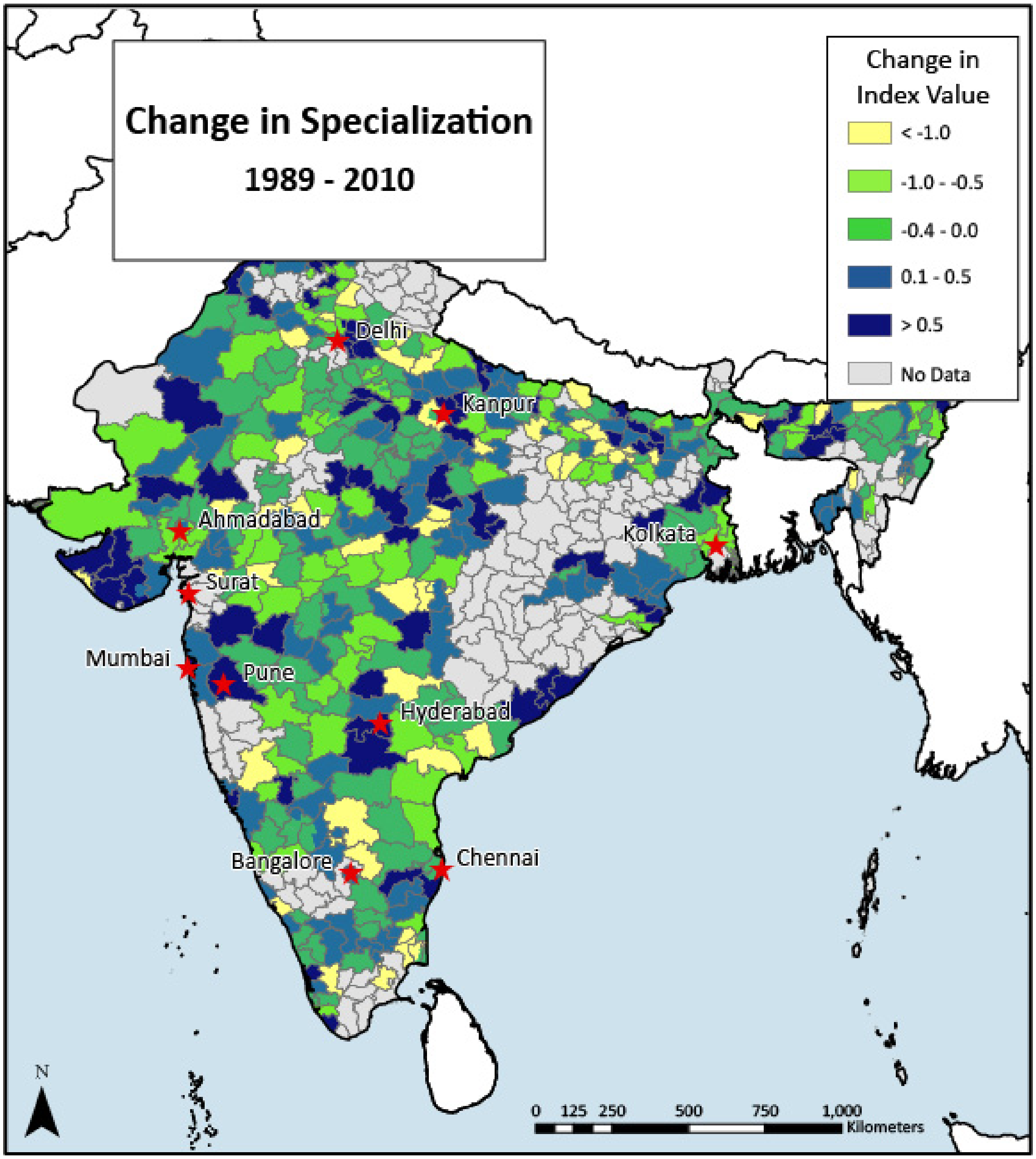


Figure 5: Average Diversity Values, 1989-2010

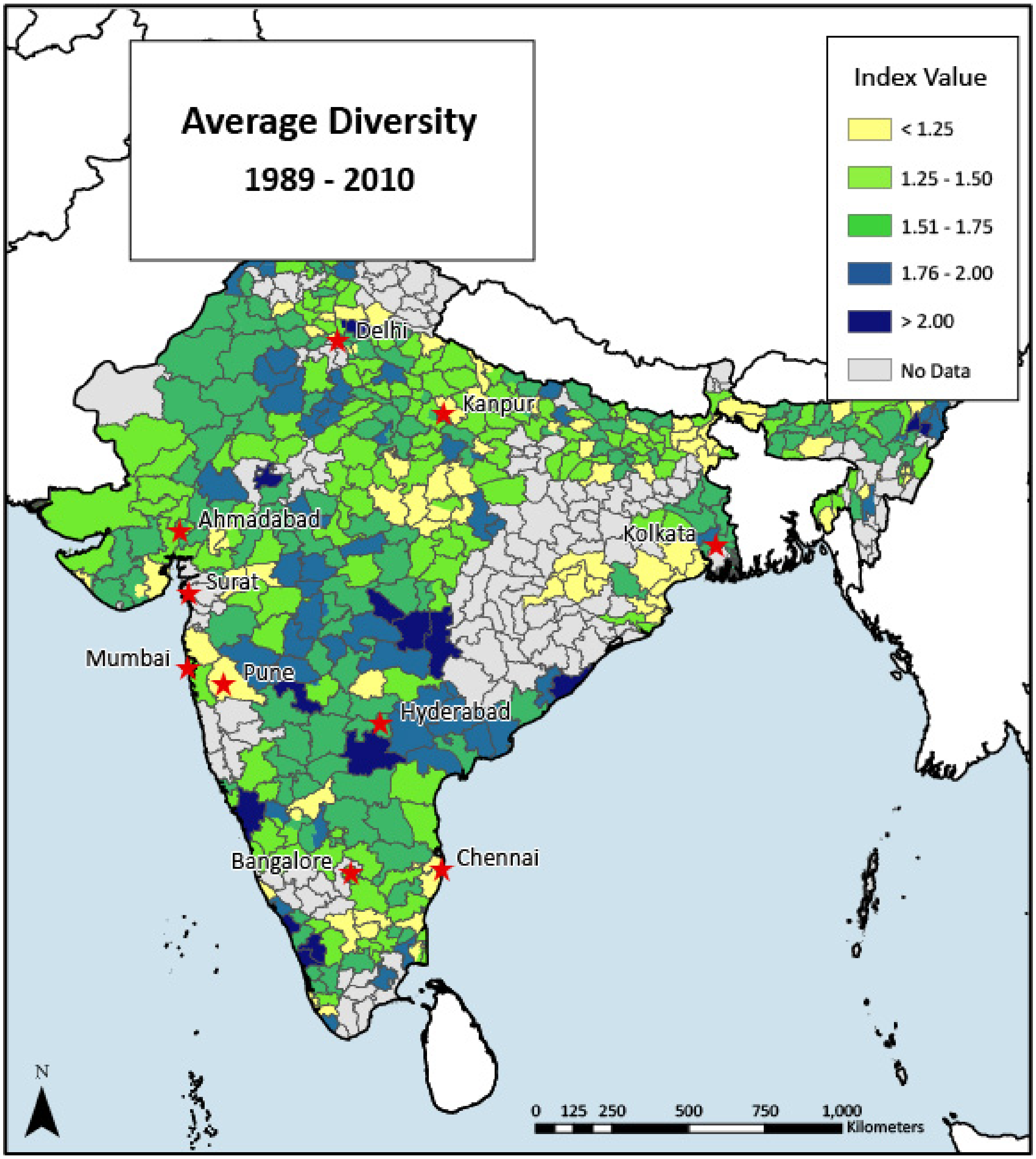


Figure 6: Change in Diversity Values, 1989/1994 to 2000/2010

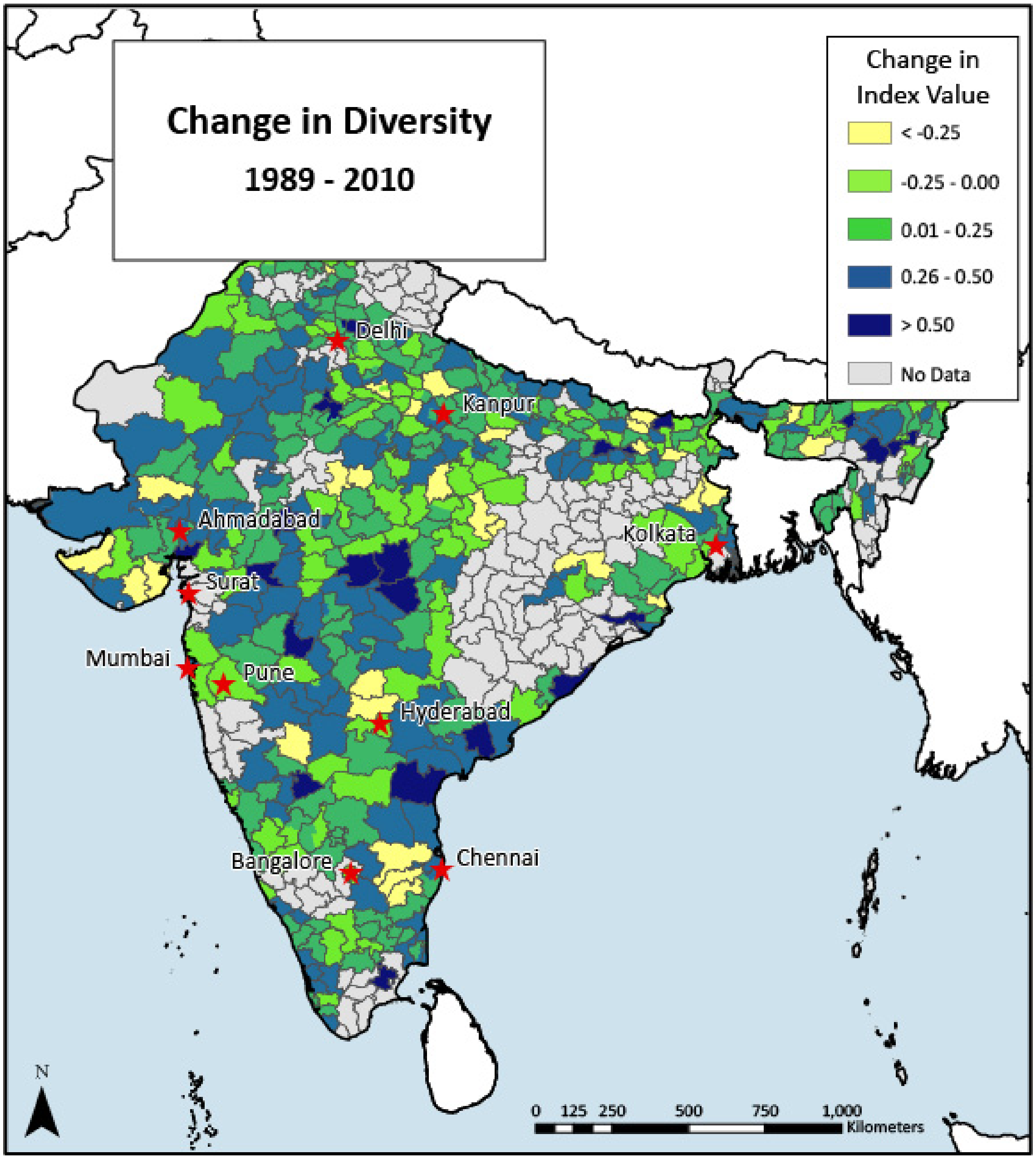


Table 1a: Descriptive statistics on specialization/diversity in manufacturing

	1989	1994	2000	2005	2010	Average	Change
A. Specialization index for districts							
Mean value	10.17	7.61	6.44	6.65	6.82	7.54	-0.15
Standard deviation	17.19	9.49	6.09	6.54	10.36	5.40	0.70
Minimum value	2.35	1.95	2.02	2.10	1.75	2.51	-1.86
Maximum value	128.27	74.34	33.71	39.98	65.44	39.39	1.75
Median value	5.43	4.42	4.34	4.49	4.02	5.75	-0.15
B. Diversity index for districts							
Mean value	1.33	1.43	1.55	1.41	1.78	1.50	0.13
Standard deviation	0.33	0.41	0.43	0.37	0.64	0.26	0.26
Minimum value	0.80	0.82	0.83	0.78	0.82	0.91	-0.71
Maximum value	2.12	2.47	2.56	2.46	3.46	2.36	0.83
Median value	1.27	1.35	1.51	1.36	1.69	1.49	0.15

Notes: Descriptive statistics taken from Annual Survey of Industries and National Sample Statistics. The index of relative specialization measures the maximum for a district of its share of employment in a sector compared to the national share of employment in that sector. The index of relative diversity is the inverse sum across all sectors of the absolute value of the difference between a sector'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 429 districts across the 1989-2010 period. The Change column describes the properties of the differences for districts between their mean index values in 2000-2010 compared to 1989-1994, relative to the overall period average.

Table 1b: Table 1a considering urban areas only

	1989	1994	2000	2005	2010	Average	Change
A. Specialization index for districts							
Mean value	12.71	12.54	10.33	9.40	7.84	10.56	-0.22
Standard deviation	17.50	14.96	13.35	9.60	10.23	7.31	0.74
Minimum value	2.54	2.53	2.20	2.34	1.88	2.65	-1.86
Maximum value	128.27	89.45	79.72	54.41	60.72	52.85	1.78
Median value	7.31	7.61	5.58	5.88	4.73	8.43	-0.26
B. Diversity index for districts							
Mean value	1.37	1.33	1.45	1.43	1.84	1.48	0.15
Standard deviation	0.35	0.36	0.40	0.35	0.62	0.23	0.25
Minimum value	0.81	0.79	0.80	0.83	0.86	0.91	-0.68
Maximum value	2.33	2.23	2.59	2.39	3.44	2.16	0.90
Median value	1.30	1.26	1.39	1.38	1.75	1.48	0.16

Notes: See Table 1a.

Table 1c: Comparison to specialization/diversity in U.S. manufacturing

	Indian districts in 2010 sample	2011 States and District of Columbia	2011 cities with the 100 largest manufacturing employments	2011 cities with 90% threshold on allocated CBP employment	2011 cities with 60% threshold on allocated CBP employment
Count	429	51	100	193	746
Average mfg. establishments	31,162	11,593	1,875	1,081	359
Average mfg. employment	82,845	417,412	61,176	33,489	10,780
A. Specialization index					
Mean value	6.65	3.73	4.58	5.95	9.96
Standard deviation	6.54	2.27	6.80	5.48	10.06
Minimum value	2.10	1.51	1.70	1.79	1.70
Maximum value	39.98	11.67	68.68	43.22	86.85
Median value	4.49	2.95	3.22	4.34	6.58
B. Diversity index					
Mean value	1.41	2.31	1.79	1.44	0.98
Standard deviation	0.37	0.98	0.63	0.61	0.45
Minimum value	0.78	0.67	0.65	0.61	0.54
Maximum value	2.46	4.80	3.69	3.61	3.69
Median value	1.36	2.20	1.77	1.28	0.83

Notes: Table provides comparison points of specialization and diversity from 2011 County Business Patterns (CBP) file. For smaller cities, CBP does not disclose employment values for some industries where the presence in the city is very small, due to confidentiality for individual plants. Columns 3-5 take alternative strategies for inclusion of cities based upon what share of the city's overall manufacturing employment is allocated.

Table 2a: Persistence of specialization/diversity levels

	1989	1994	2000	2005	2010
A. Specialization index for districts - Correlation over time of district values					
1989 value	1.00				
1994 value	0.18*	1.00			
2000 value	0.05	0.16*	1.00		
2005 value	-0.03	-0.02	0.31*	1.00	
2010 value	0.04	0.01	0.04	0.25*	1.00
B. Diversity index for districts - Correlation over time of district values					
1989 value	1.00				
1994 value	0.16*	1.00			
2000 value	0.11*	0.24*	1.00		
2005 value	0.10*	0.07	0.22*	1.00	
2010 value	0.12*	0.11*	0.22*	0.34*	1.00

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

Table 2b: Table 2a considering urban areas only

	1989	1994	2000	2005	2010
A. Specialization index for districts - Correlation over time of district values					
1989 value	1.00				
1994 value	0.21*	1.00			
2000 value	0.03	0.17*	1.00		
2005 value	-0.03	0.06	0.27*	1.00	
2010 value	0.03	0.06	0.25	0.30*	1.00
B. Diversity index for districts - Correlation over time of district values					
1989 value	1.00				
1994 value	0.18*	1.00			
2000 value	0.10*	0.18*	1.00		
2005 value	-0.03	0.05	0.13*	1.00	
2010 value	0.10*	0.13*	0.07	0.22*	1.00

Notes: See Table 2a.

Table 3a: Transition matrix of specialization/diversity levels

	1989	1994	2000	2005	2010
A. Specialization index for districts - Average quintile value by time period					
Lowest initial quintile	1.00	2.57	2.70	2.93	2.86
2nd initial quintile	2.00	2.60	2.79	3.01	2.80
3rd initial quintile	3.00	3.14	3.12	2.88	2.99
4th initial quintile	4.00	3.14	3.03	3.01	3.14
Highest initial quintile	5.00	3.53	3.34	3.14	3.19
Quintile gap	4.00	0.96	0.64	0.21	0.33
B. Diversity index for districts - Average quintile value by time period					
Lowest initial quintile	1.00	2.64	2.83	2.85	2.74
2nd initial quintile	2.00	3.03	2.87	2.92	2.81
3rd initial quintile	3.00	2.80	3.05	2.94	3.08
4th initial quintile	4.00	3.14	3.12	3.05	3.09
Highest initial quintile	5.00	3.36	3.12	3.22	3.25
Quintile gap	4.00	0.73	0.29	0.37	0.50

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 1989. 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

	1989→1994	1994→2000	2000→2005	2005→2010
A. Specialization index for districts - Average quintile value in end period				
Lowest quintile in start period	2.57	2.67	2.57	2.41
2nd quintile in start period	2.60	2.77	2.49	2.67
3rd quintile in start period	3.14	3.06	3.00	2.71
4th quintile in start period	3.14	3.15	3.23	3.37
Highest quintile in start period	3.53	3.33	3.69	3.82
Quintile gap	0.96	0.65	1.12	1.42
B. Diversity index for districts - Average quintile value in end period				
Lowest quintile in start period	2.64	2.62	2.47	2.20
2nd quintile in start period	3.03	2.65	2.92	2.71
3rd quintile in start period	2.80	2.97	2.95	3.14
4th quintile in start period	3.14	3.23	3.14	3.26
Highest quintile in start period	3.36	3.52	3.51	3.68
Quintile gap	0.73	0.90	1.04	1.48

Notes: See Table 3a. The table provides the equivalent of the 1989→1994 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.	Average value Diversity	District	State	Change Special.
A. Districts with highest average values across 1989-2010				C. Districts with largest increases from 1989		
Dibrugarh	ASSAM	178.74	1.05	Dibrugarh	ASSAM	1.91
Jhabua	MADHYA PRADESH	98.50	1.65	Jhabua	MADHYA PRADESH	1.85
Bargarh	ORISSA	64.65	1.22	Bongaigaon	ASSAM	1.82
Patna	BIHAR	63.86	1.66	Ri Bhoi	MEGHALAYA	1.77
Saharanpur	UTTAR PRADESH	51.93	1.37	Bargarh	ORISSA	1.74
Lakhimpur	ASSAM	50.92	1.71	Jamnagar	GUJARAT	1.70
Belgaum	KARNATAKA	41.45	1.35	Saharsa	BIHAR	1.65
Jamnagar	GUJARAT	40.09	1.42	Begusarai	BIHAR	1.64
Rampur	UTTAR PRADESH	39.39	1.58	Tinsukia	ASSAM	1.55
Ri Bhoi	MEGHALAYA	37.71	1.55	Solan	HIMACHAL PRADESH	1.54
Siwan	BIHAR	37.35	1.17	Mumbai Suburban	MAHARASHTRA	1.53
Fatehgarh Sahib	PUNJAB	37.08	1.26	Nizamabad	ANDHRA PRADESH	1.43
B. Districts with lowest average values across 1989-2010				D. Districts with largest declines from 1989		
Mandi	HIMACHAL PRADESH	2.51	1.60	West Godavari	ANDHRA PRADESH	-1.86
Kokrajhar	ASSAM	2.55	1.74	Saharanpur	UTTAR PRADESH	-1.83
Mokokchung	NAGALAND	2.64	1.87	Sivaganga	TAMIL NADU	-1.78
Raichur	KARNATAKA	2.64	1.65	Bishnupur	MANIPUR	-1.77
Adilabad	ANDHRA PRADESH	2.82	1.82	Lakhimpur	ASSAM	-1.75
Zunheboto	NAGALAND	2.84	2.02	Siwan	BIHAR	-1.74
Nawada	BIHAR	2.87	1.61	Porbandar	GUJARAT	-1.69
Kurnool	ANDHRA PRADESH	2.88	1.64	Rampur	UTTAR PRADESH	-1.61
Muzaffarpur	BIHAR	2.88	1.66	Sheikhpura	BIHAR	-1.61
Ambedkar Nagar	UTTAR PRADESH	2.93	1.61	Barwani	MADHYA PRADESH	-1.60
Wokha	NAGALAND	2.99	2.23	Saran	BIHAR	-1.55
Wayanad	KERALA	3.02	1.70	Srinagar	JAMMU & KASHMIR	-1.54

Notes: See Table 1a. Panels A and B report average values across the 1989-2010 period. Panels C and D report changes from 1989-1994 to 2000-2010 relative to the average value across the full period. 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 1989-2005				C. Districts with largest increases since 1989		
Mahbubnagar	ANDHRA PRADESH	3.85	2.40	Prakasam	ANDHRA PRADESH	0.94
Datia	MADHYA PRADESH	10.95	2.33	Meerut	UTTAR PRADESH	0.90
Meerut	UTTAR PRADESH	9.97	2.30	Chhindwara	MADHYA PRADESH	0.85
Kozhikode	KERALA	6.78	2.24	Ujjain	MADHYA PRADESH	0.83
Wokha	NAGALAND	2.99	2.23	Sivaganga	TAMIL NADU	0.79
Palghat	KERALA	6.81	2.22	Srikakulam	ANDHRA PRADESH	0.79
Srikakulam	ANDHRA PRADESH	3.34	2.19	Patna	BIHAR	0.72
Hoshiarpur	PUNJAB	5.52	2.16	Sapaul	BIHAR	0.67
Wardha	MAHARASHTRA	4.69	2.13	Karauli	RAJASTHAN	0.66
Neemuch	MADHYA PRADESH	4.30	2.12	Pondicherry	PONDICHERRY	0.66
Osmanabad	MAHARASHTRA	5.52	2.10	Nagpur	MAHARASHTRA	0.66
Firozpur	PUNJAB	8.07	2.09	Cuttack	ORISSA	0.64
B. Districts with lowest average values across 1989-2005				D. Districts with largest declines since 1989		
Damoh	MADHYA PRADESH	7.71	0.89	Birbhum	WEST BENGAL	-0.83
Sagar	MADHYA PRADESH	6.60	0.91	Nizamabad	ANDHRA PRADESH	-0.66
Kasaragod	KERALA	8.02	0.93	Pratapgarh	UTTAR PRADESH	-0.58
Maldah	WEST BENGAL	7.20	0.95	Murshidabad	WEST BENGAL	-0.56
Sambalpur	ORISSA	18.65	0.96	Shahdol	MADHYA PRADESH	-0.52
Debagarh	ORISSA	6.99	0.99	Amreli	GUJARAT	-0.51
Jabalpur	MADHYA PRADESH	12.16	0.99	West Khasi Hills	MEGHALAYA	-0.45
Jalpaiguri	WEST BENGAL	4.59	1.00	Panchkula	HARYANA	-0.43
Quilon	KERALA	5.03	1.00	Sundargarh	ORISSA	-0.43
Bhavnagar	GUJARAT	6.14	1.01	Deoria	UTTAR PRADESH	-0.40
North West Delhi	DELHI	12.06	1.02	Bhadrak	ORISSA	-0.40
Nizamabad	ANDHRA PRADESH	11.21	1.03	Vellore	TAMIL NADU	-0.40

Notes: See Table 4a.

Table 5: Specialization/diversity levels for industries based upon district locations

NIC	Industry Description	Specialization 1989-1994	Specialization 2000-2010	Specialization Change	Diversity 1989-1994	Diversity 2000-2010	Diversity Change
15	Food products and beverages	1.37	1.34	-0.02	1.39	1.59	0.13
16	Tobacco products	0.84	1.10	0.27	1.37	1.51	0.10
17	Textiles	0.94	0.86	-0.09	1.39	1.59	0.13
18	Wearing apparel; dressing and dyeing of fur	1.39	13.97	1.64	1.40	1.76	0.23
19	Leather; luggage, handbags, saddlery, harness and footwear	1.71	1.48	-0.14	1.41	1.61	0.13
20	Wood and wood products, except furniture; straw and plating	1.26	59.25	1.92	1.39	1.65	0.17
21	Paper and paper products	2.07	24.05	1.68	1.40	1.56	0.11
22	Publishing, printing and reproduction of recorded media	1.30	1.07	-0.20	1.39	1.58	0.13
23	Coke, refined petroleum and nuclear fuel	4.99	5.38	0.08	1.37	1.53	0.11
24	Chemicals and chemical products	1.32	1.04	-0.23	1.39	1.58	0.13
25	Rubber and plastic products	1.49	1.22	-0.20	1.40	1.57	0.11
26	Other non-metallic mineral products	1.41	1.40	-0.01	1.39	1.58	0.13
27	Basic metals	0.95	1.67	0.55	1.39	1.56	0.12
28	Fabricated metal products, except machinery and equipments	1.32	1.08	-0.20	1.39	1.58	0.13
29	Machinery and equipment, n.e.c.	1.40	1.15	-0.20	1.39	1.58	0.13
30	Office, accounting and computing machinery	6.70	97.93	1.74	1.38	1.62	0.16
31	Electrical machinery and apparatus, n.e.c.	1.60	1.34	-0.18	1.41	1.58	0.11
32	Radio, television, and communication equipment and apparatus	5.10	31.57	1.44	1.43	1.46	0.02
33	Medical, precision and optical instruments, watches and clocks	3.03	3.06	0.01	1.42	1.60	0.12
34	Motor vehicles, trailers and semi-trailers	2.18	1.82	-0.18	1.43	1.61	0.12
35	Other transport equipment	1.86	1.96	0.05	1.43	1.59	0.10
36	Furniture, manufacturing n.e.c.	1.02	1.11	0.08	1.39	1.59	0.13
	Traditional	1.31	1.34	0.02	1.39	1.48	0.06
	Modern	2.08	2.06	-0.01	1.40	1.48	0.05

Notes: See Table 1a. "n.e.c." stands for Not Elsewhere Classified. Change is measured in relative terms to average values across 1989-2010. "Modern" industries are comprised of the following (NIC 98 2-digit): 23, 24, 25, 27, 29, 30, 31, 32, 33, 34, 35. Average values for Traditional/Modern are calculated by aggregating all plants in these industries before indices are measured.

Table 6: Count of districts where the industry produces specialization value

NIC	Industry Description	1989	1994	2000	2005	2010	Average	Change
15	Food products and beverages	37	41	38	42	57	43.0	-0.04
16	Tobacco products	20	17	24	29	29	23.8	0.43
17	Textiles	31	29	29	41	28	31.6	0.13
18	Wearing apparel; dressing and dyeing of fur	16	5	20	18	2	12.2	1.05
19	Leather; luggage, handbags, saddlery, harness and footwear	33	29	36	20	23	28.2	-0.08
20	Wood and wood products, except furniture; straw and plating	34	44	34	37	6	31.0	-0.08
21	Paper and paper products	23	20	18	18	2	16.2	-0.14
22	Publishing, printing and reproduction of recorded media	12	14	19	16	21	16.4	0.38
23	Coke, refined petroleum and nuclear fuel	18	17	16	25	19	19.0	0.17
24	Chemicals and chemical products	14	13	13	12	4	11.2	-0.14
25	Rubber and plastic products	11	12	7	10	12	10.4	-0.23
26	Other non-metallic mineral products	34	36	37	39	58	40.8	0.06
27	Basic metals	4	16	16	20	19	15.0	0.85
28	Fabricated metal products, except machinery and equipments	25	17	18	12	21	18.6	-0.27
29	Machinery and equipment, n.e.c.	26	22	17	9	9	16.6	-0.47
30	Office, accounting and computing machinery	8	9	9	13	2	8.2	0.35
31	Electrical machinery and apparatus, n.e.c.	12	12	7	12	14	11.4	-0.21
32	Radio, television, and communication equipment and apparatus	15	13	13	12	7	12.0	-0.14
33	Medical, precision and optical instruments, watches and clocks	9	11	11	10	44	17.0	0.15
34	Motor vehicles, trailers and semi-trailers	12	14	16	11	7	12.0	0.04
35	Other transport equipment	15	16	14	7	7	11.8	-0.35
36	Furniture, manufacturing n.e.c.	20	22	17	16	38	22.6	-0.13
	Correlation to previous survey		0.88	0.86	0.83	0.55		
	Total count for Traditional	297	286	297	300	299	295.8	
	Total count for Modern	132	143	132	129	130	133.2	

Notes: See Table 5. Change compares the 1989-1994 period to the 2000-2010 period.

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	-0.159	-0.053	0.064	-0.546	-0.101	-0.221
	(0.457)	(0.455)	(0.155)	(0.376)	(0.444)	(0.175)
Log of district population density	-0.058	0.709	0.106	-0.032	-0.216	-0.174
	(0.506)	(0.508)	(0.195)	(0.492)	(0.550)	(0.205)
Educated worker share (% pop graduate)	0.163	-0.255	0.208+	0.275	-0.374	0.146
	(0.347)	(0.310)	(0.121)	(0.259)	(0.363)	(0.111)
Age profile (demographic dividend)	0.183	-0.521	0.170	-0.670	-0.224	-0.012
	(0.490)	(0.530)	(0.183)	(0.491)	(0.497)	(0.178)
Literacy rate	-0.181	0.717+	-0.428*	1.079**	0.888*	0.148
	(0.417)	(0.389)	(0.165)	(0.372)	(0.441)	(0.138)
Infrastructure: electricity access	-0.159	-0.325	0.002	-0.211	-0.470	0.102
	(0.429)	(0.448)	(0.173)	(0.388)	(0.466)	(0.148)
Infrastructure: paved roads	-0.377	0.245	0.158	-0.211	-0.425	-0.160
	(0.361)	(0.363)	(0.138)	(0.317)	(0.377)	(0.134)
Strength of household banking	0.119	-0.359	-0.038	-0.285	0.088	0.116
	(0.384)	(0.423)	(0.134)	(0.346)	(0.399)	(0.142)
Log travel time to closest large city	-0.237	-0.311	-0.055	-0.531**	-0.122	-0.036
	(0.234)	(0.332)	(0.140)	(0.181)	(0.294)	(0.118)
Manufacturing share of local employment	0.124	0.044	0.237*	-0.971**	-0.370	-0.399**
	(0.347)	(0.320)	(0.106)	(0.302)	(0.290)	(0.127)
Urbanization rate (% urban)	0.461	-1.429*	-0.179	-0.801	0.577	0.140
	(0.563)	(0.576)	(0.224)	(0.575)	(0.621)	(0.218)
Log per capita consumption	1.239	1.938+	0.066	-0.089	0.964	-0.577
	(1.127)	(1.142)	(0.433)	(1.047)	(1.199)	(0.407)
Log organized employment	0.559	-0.192	0.032	0.050	0.277	0.119
	(0.353)	(0.344)	(0.134)	(0.288)	(0.352)	(0.103)
Log unorganized employment	-1.006+	0.343	-0.019	-0.205	-0.580	-0.119
	(0.522)	(0.502)	(0.189)	(0.487)	(0.536)	(0.185)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	281	281	281	281	281	281
Adjusted R-squared	0.233	0.174	0.161	0.367	0.203	0.181

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 manufacturing production functions and district specialization/diversity

	DV: Log output in manufacturing establishment, 2000 sample								
	Full	Full	Full	Full	Full	Full	Traditional	Modern	Full
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log employment in establishment	0.510	0.512	0.509	0.513	0.508	0.508	0.514	0.393	0.512
	(0.028)**	(0.028)**	(0.028)**	(0.028)**	(0.025)**	(0.025)**	(0.028)**	(0.058)**	(0.028)**
Log capital in establishment	0.059	0.059	0.058	0.058	0.059	0.059	0.057	0.080	0.057
	(0.008)**	(0.008)**	(0.008)**	(0.008)**	(0.008)**	(0.008)**	(0.008)**	(0.015)**	(0.008)**
Log materials in establishment	0.508	0.507	0.509	0.507	0.507	0.507	0.503	0.568	0.507
	(0.016)**	(0.016)**	(0.016)**	(0.016)**	(0.015)**	(0.015)**	(0.016)**	(0.018)**	(0.015)**
District specialization index		0.007		0.015	0.013	0.013	0.015	0.015	0.018
		(0.006)		(0.005)**	(0.005)*	(0.005)**	(0.005)**	(0.010)	(0.004)**
District diversity index			0.008	0.015	0.014	0.015	0.016	0.003	0.016
			(0.006)	(0.006)**	(0.005)*	(0.005)**	(0.006)**	(0.008)	(0.006)**
Manufacturing share of local						0.012			
employment						(0.016)			
(0,1) Plant in specialized industry x									-0.024
District specialization index									(0.024)
(0,1) Plant in specialized industry x									-0.012
District diversity index									(0.018)
(0,1) Plant in specialized industry									0.201
									(0.155)
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional urban traits					Yes	Yes			
Observations	137,090	137,090	137,090	137,090	137,090	137,090	124,022	13,068	137,090

Notes: Estimations consider simple production functions for manufacturing establishments in India. Indian data are taken from Annual Survey of Industries and National Sample Statistics. Specialization and diversity indices are on a ten-point scale representing deciles for districts. Additional urban traits include Log manufacturing employment in district; Log manufacturing employment in a plant's district-industry; Log share of local manufacturing employment in the unorganized sector; and Log share of district-industry manufacturing employment in the unorganized sector. Estimations report standard errors clustered by district, include industry fixed effects, and weight observations by sample weights. + significant at 10% level; * significant at 5% level; ** significant at 1% level.

Table 9: Estimations of log manufacturing employment growth and district specialization/diversity

	DV: Log growth in district manufacturing employment from 1989-2010								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
District specialization index (1989/94)	0.093		0.097	0.076	0.074	0.094	0.076		0.096
	(0.032)**		(0.034)**	(0.034)*	(0.034)*	(0.037)*	(0.034)*		(0.040)*
District specialization growth									0.018
									(0.153)
District specialization index in								-0.054	
in 50th-75th percentile								(0.208)	
District specialization index in								0.513	
in 76th-99th percentile								(0.236)*	
District diversity index (1989/94)		-0.016	0.015	0.037	0.033	0.039	0.037		
		(0.029)	(0.031)	(0.033)	(0.033)	(0.032)	(0.033)		
District diversity index in								0.500	
in 50th-75th percentile								(0.236)*	
District diversity index in								-0.020	
in 76th-99th percentile								(0.207)	
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	
Drop districts with <1 mil population						Yes			
Industry projection control							Yes		
Observations	276	276	276	273	273	229	273	273	276

Notes: Estimations quantify the relationship between district employment growth in manufacturing and district specialization and diversity indices. 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 10: Table 9 with urban and rural distinctions

	DV: Log growth in district manufacturing employment from 1989-2010						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Urban district specialization index (1989/94)	0.052		0.071				0.077
	(0.029)+		(0.033)*				(0.031)*
Urban district diversity index (1989/94)		0.021	0.050				0.083
		(0.029)	(0.033)				(0.030)**
Rural district specialization index (1989/94)				0.107		0.108	0.066
				(0.035)**		(0.037)**	(0.034)+
Rural district diversity index (1989/94)					-0.028	0.004	-0.014
					(0.032)	(0.033)	(0.029)
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	272	272	272	277	277	277	265

Notes: See Table 9. + significant at 10% level; * significant at 5% level; ** significant at 1% level.

App. Table 1: Correlation between district traits and specialization/diversity values

	Specialization index			Diversity index		
	Average	Initial	Change	Average	Initial	Change
	(1)	(2)	(3)	(4)	(5)	(6)
A. District traits from 2001 Population Census						
Log population	0.023	-0.016	0.043	-0.118*	-0.111*	0.019
Log population density	0.148*	0.041	0.076	-0.207*	-0.131*	-0.041
Age profile (demographic dividend)	0.040	0.032	0.034	0.135*	0.035	0.075
Share of population in scheduled caste	0.028	0.071	-0.008	-0.011	0.009	-0.006
Share of population in scheduled tribe	-0.148*	-0.095	-0.041	0.106*	0.061	0.006
Female labor force participation rate	-0.207*	-0.100*	-0.051	0.197*	0.073	0.103*
Educated worker share (% pop graduate)	0.218*	0.079	0.093	-0.041	-0.126*	0.110*
Literacy rate	0.057	0.107*	-0.026	0.222*	0.070	0.135*
Infrastructure: electricity access	-0.110*	-0.115*	0.043	-0.119*	-0.009	-0.137*
Infrastructure: paved roads	0.085	0.089	-0.009	0.138*	-0.021	0.182*
Travel time to nearest of ten largest cities	-0.122*	-0.022	-0.069	-0.007	0.019	-0.018
Strength of household banking sector	0.091	0.050	0.077	0.068	0.024	0.067
Urbanization rate (% urban)	0.159*	0.075	0.042	-0.177*	-0.153*	0.000
Log average income per capita	0.124*	0.106*	0.044	0.217*	0.098*	0.100*
Manufacturing share of local employment	0.010	-0.013	0.059	-0.034	0.054	-0.099*
B. District traits for manufacturing sector from ASI/NSS across 1989-2010						
Log growth in establishments	-0.157*	-0.231*	0.232*	-0.025	0.031	-0.043
Log growth in employment	0.073	-0.061	0.132*	-0.006	0.017	-0.036
Log growth in output	0.173*	0.156*	-0.084*	0.028	0.044	-0.042
Organized sector share of establishments	0.120*	0.211*	-0.224*	0.009	-0.082*	0.047*
Organized sector share of employment	0.161*	0.070	0.110*	0.146*	0.075	0.064
Organized sector share of output	0.392*	0.348*	-0.070	0.040	-0.067	0.120*

Notes: Table documents univariate 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 in Panel A are from the 2001 Population Census. District traits are expressed in log values or percentage point values as indicated. District traits in Panel B are calculated from the ASI/NSS data over the 1989-2010 period. An asterisk denotes a correlation is statistically significant at the 10% level.

App. Table 2: Correlation between industry traits and specialization/diversity values

	Specialization index			Diversity index		
	Average	Initial	Change	Average	Initial	Change
	(1)	(2)	(3)	(4)	(5)	(6)
A. Industry traits from ASI/NSS in 2000						
Log labor intensity	-0.169	-0.075	-0.179	0.227	0.236	-0.063
Log capital intensity	0.137	0.315	-0.139	-0.041	-0.166	0.062
Log materials intensity	0.104	0.222	-0.163	0.021	-0.126	0.212
Log average wage	0.022	0.124	-0.171	0.008	-0.068	0.055
Log financial dependency	-0.172	-0.032	-0.277	0.260	0.183	0.146
Log import dependency	0.430*	0.664*	0.072	0.235	0.338	-0.173
Share of establ. that are female-owned	-0.258	-0.468*	0.271	-0.247	-0.255	-0.163
Share of establ. that are household-based	-0.134	-0.604*	0.417*	-0.373*	-0.136	-0.408*
B. Industry traits from ASI/NSS across 1989-2010						
Log growth in establishments	0.617*	0.132	0.814*	-0.281	-0.158	-0.295
Log growth in employment	-0.601*	-0.235	-0.766*	0.075	-0.027	0.241
Log growth in output	-0.666*	-0.166	-0.830*	0.205	0.033	0.347
Organized sector share of establishments	0.426*	0.554*	0.179	-0.109	-0.081	-0.210
Organized sector share of employment	0.092	0.510*	-0.343	0.313	0.038	0.334
Organized sector share of output	0.415*	0.582*	-0.081	0.177	0.217	-0.004

Notes: Table documents univariate correlations between industry traits and specialization and diversity index values. Specialization and diversity index values use the employment distributions of industries across districts as shown in Table 5. Average and initial index values are expressed on a ten-point scale representing deciles. Change values are expressed in unit standard deviations. Industry traits in Panel A are from the 2000 ASI/NSS, measured at the national level. Intensity measures are measured relative to industry sales. Industry traits in Panel B are calculated from the ASI/NSS data over the 1989-2010 period. An asterisk denotes a correlation is statistically significant at the 10% level.