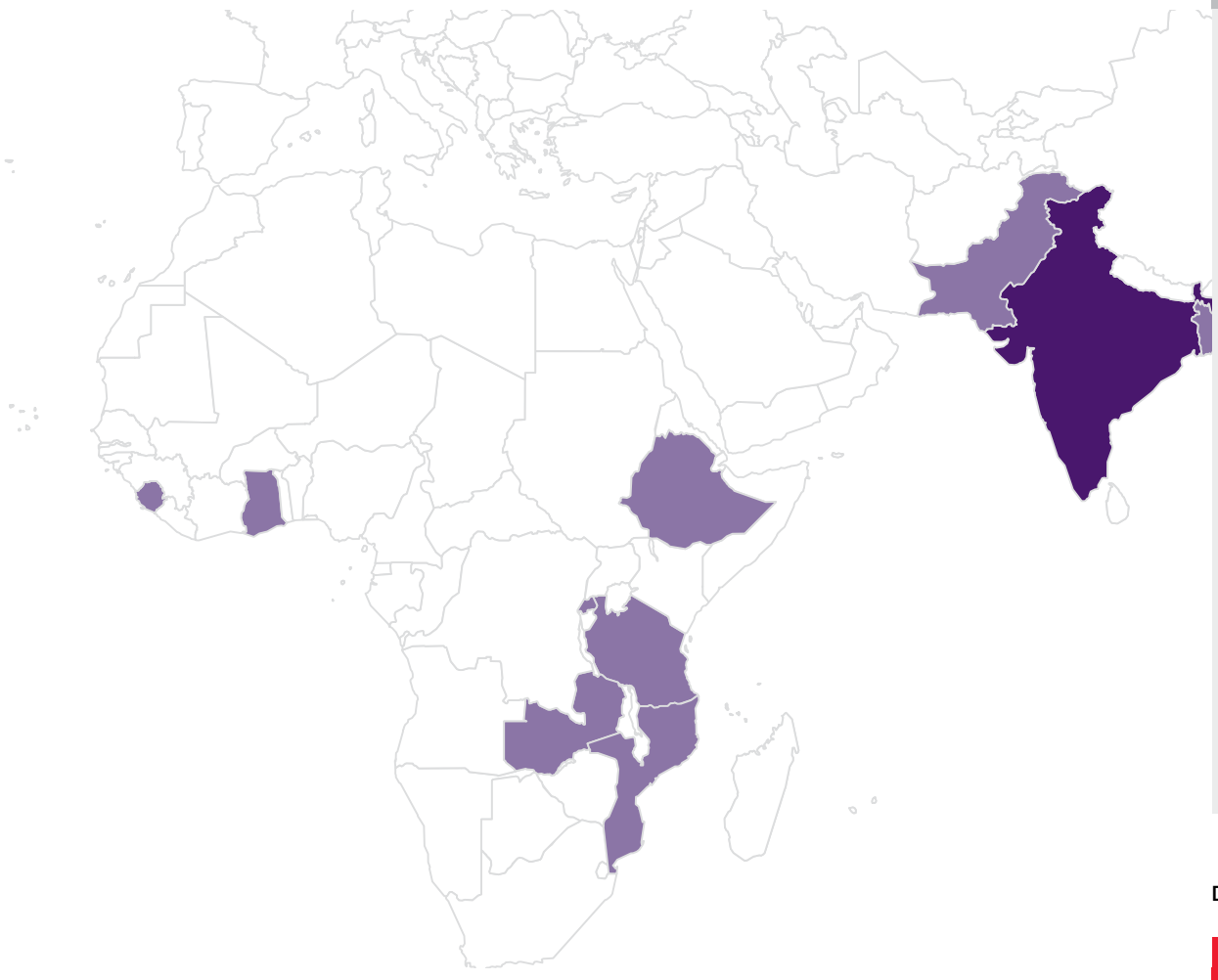
IGC International Growth Centre

Determinants of Clusters in Indian Manufacturing: The Role of Infrastructure, Governance, Education, and Industrial Policy

Ana M. Fernandes[a] and **Gunjan Sharma**[b]

[a] The World Bank (contact: afernandes@worldbank.org), [b] University of Missouri (contact: sharmag@missouri.edu)

Working Paper 12/0606
August 2012



International Growth Centre
London School of Economics and Political Science
4th Floor, Tower Two
Houghton Street
London WC2A 2AE
United Kingdom

For media or communications enquiries, please contact Mazida Khatun
mazida.khatun@theigc.org

Directed and Organised by

LSE UNIVERSITY OF OXFORD

www.theigc.org

Determinants of Clusters in Indian Manufacturing: The Role of Infrastructure, Governance, Education, and Industrial Policy

Ana M. Fernandes (The World Bank)[1]

Gunjan Sharma (University of Missouri)[2]

First Draft: March 2012[3]

Abstract

This paper investigates the determinants of spatial concentration and entry within manufacturing across states in India. Using an unbalanced panel of 180 industries spread across 16 major Indian states over the time period 1985-2007, we estimate the effect of location (state) characteristics interacted with characteristics that make industries naturally more prone to concentrate in locations (states) of certain types on spatial concentration and entry. The results show that governance, infrastructure and the availability of skilled labor are important determinants of increased concentration and entry. Moreover, the estimates indicate that state characteristics associated with lower distance to foreign markets, lower costs of accessing domestic suppliers, or lower costs of doing business matter for the impact of licensing, FDI and trade reforms on concentration and new entry. There is also evidence that less substitutable inputs (e.g., roads) raise spatial concentration while more substitutable inputs (e.g., electricity) do not.

Keywords: Spatial Concentration, Agglomeration Economies, India, Infrastructure, Trade Policy, FDI, Industrial Licensing.

JEL Classification codes: R11, R12, O47.

[1] Corresponding author: Ana Margarida Fernandes. The World Bank. Development Research Group. 1818 H Street NW, Washington DC, 20433. E-mail: afernandes@worldbank.org.

[2] Corresponding author Gunjan Sharma, University of Missouri, 228 Professional Building, Columbia MO. Email: sharmag@missouri.edu.

[3] Support from the International Growth Center in the form of an in-country grant is gratefully acknowledged. Support from the governments of Norway, Sweden and the United Kingdom through the Multi-Donor Trust Fund for Trade and Development is also gratefully acknowledged. We would like to thank seminar participants at the University of Missouri and the IGC-ISI India Development Policy Conference in December 2011 for comments. We thank Ann Harrison and Shanti Nataraj for providing us with tariff data and Anumapa Sen for providing us with infrastructure data. We thank Brock Williams for excellent research assistance. The findings expressed in this paper are those of the authors and do not necessarily represent the views of the International Growth Center nor the World Bank.

1. Introduction

The benefits of industrial concentration for economic growth are well established empirically (Ciccone and Hall, 1996; Cingano and Schivardo, 2004; Brülhart and Sbergami, 2009). The geographic variation in industrialization influences to a large extent the geographical variation in average incomes and poverty levels in developing countries (Lall and Chakravorty, 2005). At the same time entrepreneurship is seen as a key contributor to job creation and economic growth. Several studies provide evidence of the link between entrepreneurship and income levels at the aggregate level (Acs et al., 2008; Audretsch, 2007; Baumol and Straum 2007; Baumol, 2010; Klapper and Love, 2011) and increasingly at the regional level (Audretsch and Keilbach, 2005; Li et al., 2009; Dejardin, 2011; Samila and Sorenson, 2011).[4] Therefore, understanding the process of spatial industrial concentration and new firm entry is relevant for economic development. At the same time regional employment dynamism is highly coveted by local policy-makers. In order to attract manufacturing industries and promote the formation of industrial clusters, policy-makers need to know the factors that correlate with local employment dynamism.

The process of industrialization has been accompanied by spatial concentration of economic activity in Europe, and more recently in East Asia and Latin America. This has also been the case in India for certain industries and certain time periods (Fernandes and Sharma, 2010). The highly regulated nature of the spatial allocation of manufacturing activity which prevailed in India up to the 1980s – with an aim at promoting equality across regions – and the subsequent large-scale deregulations and policy reforms provide a unique environment to explore the determinants of spatial manufacturing distribution. Since industrialization is still at an early stage in India and the country‟s industrial landscape is still adjusting to the policy reforms of the 1990s, location characteristics have the potential to have a much more substantial impact in determining spatial concentration in India than would be the case in mature developed countries with more rigid industrial structures (Ghani et al., 2011).

Manufacturing activity is extremely concentrated in India but few studies have examined the factors that explain this concentration. This paper examines the determinants of industrial location in India using data from the Annual Survey of Industries over a very long and rich time period from 1985-1986 to 2007-2008 considering both the net expansion in employment of incumbent firms as well as employment in new firms. We study the role comparative advantage factors, new economic geography (NEG) mechanisms, and business environment factors following an empirical specification along the lines of those proposed by Midelfart-Knarvik et al. (2000) and Ellison and Glaeser (1999). The specification estimates the effect of location characteristics interacted with characteristics that make industries naturally more prone to locate

[4] The concept of „entrepreneurship‟ used in this paper will capture the entry of new firms defined as firms that initiate their formal activity. See Klapper and Love (2011) for a discussion on alternative definitions of entrepreneurship.

in locations of certain types. The specification allows us to test several hypotheses regarding the determinants of clustering of manufacturing industries in India. For example, industries that are dependent on intermediate inputs chose to locate in states with larger input-output networks. Similarly, industries that are more dependent on infrastructure inputs such as electricity and transport may chose to locate in states with better infrastructure.

Our main findings can be summarized as follows. First, our estimates show strong positive effects of comparative advantage linked to skilled labor as expected, indicating that industries that rely more on skilled labor concentrate and exhibit higher entry rates in states with a more abundant skilled labor force. In contrast, we find a negative and significant counter-intuitive effect of comparative advantage linked to unskilled labor, even after controlling for possible congestion forces. These two findings combined suggest that skilled labor is the least substitutable input in the production process and its relative abundance in a state is a critical determinant of firms‟ location choices. Second, better business environment in terms of infrastructure or governance increases manufacturing employment shares and new entry across Indian states for industries that are more dependent on infrastructure or on contracts. However, different types of infrastructure have differential effects. Transport infrastructure which is less substitutable by firms - state roads or highways need to be relied upon - has a strong positive effect on concentration whereas electricity infrastructure has a negative effect. Third, our evidence suggests little to no effects of NEG mechanisms and Marshallian economies linked to market access and input-output linkages on either concentration or entry rates. Fourth, after the 1991 Indian market-oriented policy reforms the importance of transport infrastructure falls (in some cases becomes negative) in explaining concentration and new entry. However, there is industry heterogeneity in the sense that FDI-liberalized industries continue to exhibit higher concentration and entry in states with better transport infrastructure after 1991. Fifth, our estimates show that the negative effect of unskilled labor abundance is verified only in the post-reform period but is entirely driven by the industries facing stronger import competition that may be more capital-intensive and thus less likely to rely on the availability of cheap unskilled labor for their location decisions. Finally, we show that state characteristics associated with lower distance to foreign markets, lower costs of accessing domestic suppliers, or lower costs of doing business matter for the impact of the policy reforms on concentration and new entry. In particular, we find that tariff-liberalized and FDI-liberalized industries tend to locate in states with lower a manufacturing base suggesting the weakening of domestic input-output linkages and increased importance of access to imported inputs in the wake of the 1991 reforms.

Despite our use of a long and rich panel of industry-state cells, our control for industry and location fixed effects in all specifications, and our consideration of an extensive list of potential determinants of location, we caution that our findings should be interpreted as strong correlations but are not guaranteed to be causal. The reason is that some of our determinants may themselves be influenced by the concentration of manufacturing activity. That is, the coefficients on the comparative advantage factors reflect the equilibrium relationship between geographic

concentration and localization and urbanization economies. Hence we want our results to be viewed as carefully established robust correlations and we leave for future research the establishment of definitive causal links between local conditions and concentration in manufacturing for example by exploiting natural experiment-types of events. We should note that the rate of entry of new firms by industry-location which is one of our dependent variables is presumably less endogenous because for a new firm, location attributes – including the industrial structure – can be safely taken as given at the time of entry (Rosenthal and Strange 2003; Ghani et al., 2011; Jofre-Monseny et al., 2011).

Our paper contributes to the emerging area of agglomeration and entrepreneurship in developing countries by examining how local conditions and industry characteristics interact to determine clustering in India over a period of more than two decades. Relative to the previous literature discussed in Section 2, our contribution is to consider an unbalanced panel of Indian industries and states over a very long and eventful policy reform period thus exploiting within-industry within-state variation over time to identify the effects of various determinants of concentration while accounting for the possibility of omitted variables at the industry or state levels. Also we employ a richer set of covariates than were included in previous studies using a related approach for Argentina and Brazil. In particular we include natural cost advantage-related variables that were purposely excluded from a prior study on India by Ghani et al. (2011). Moreover, the long time span of our data allow us to compare the role of these natural cost-advantage variables in the pre- and post-reform periods in India.

The rest of the paper is organized as follows. In Section 2 we review the literature on the determinants of industrial concentration. In Section 3 we describe the data and present our empirical approach. Section 4 provides a rich set of descriptive statistics on industrial concentration across Indian states. Section 5 presents the baseline estimation results and Section 6 focuses on robustness results. Section 7 presents extension results while Section 8 examines the role of policy reforms. Section 9 concludes.

2. Determinants of Industrial Location: Literature Review

The literature on industrial and firm location and its determinants is vast and rich and a comprehensive review is beyond the scope of this paper. Fujita et al. (1999) Rosenthal and Strange (2001), Henderson (2003), Combes and Overman (2004), Mayer and Mayer (2004) provide extensive reviews while Deichmann et al. (2008) focus on industrial location in developing countries. Most studies focus on the determinants of location for firms and industries in the manufacturing sector, which is also the focus of our analysis.[5] Below we discuss the

[5] Ghani et al. (2011) - whose manufacturing-related results are discussed throughout the section - consider also the location of services firms in the context of India.

various types of spatial factors determining concentration and new firm entry and in doing so provide evidence from key studies in the literature.

The starting point for examining the determinants of firm or industry location are location choice models which is generally assumed to be made by profit-maximizing agents, where agents compare the potential profitability of different locations. The profitability of a firm in a location is a function of cost-related factors (including the availability and cost of inputs) and demand-related factors (market access factors such as the size of the local consumer market or the distance to markets). Both the costs and the demand can be influenced also by the existing industrial structure capturing previous location choices of other firms (agglomeration economies) and by location characteristics related to a large extent to the availability of public goods and services as well as to the general business environment in the location. We now delve into these factors providing available evidence from key studies in the industrial location literature.

The endowments and natural geography of a location - „natural advantages" according to Ellison and Glaeser (1999) or comparative advantage factors in trade theory - can play an important role in determining the spatial concentration of manufacturing activity. Natural cost advantages can relate to natural resource endowments, labor endowments, or transportation costs linked to a particular geographical location (e.g., lower costs for coastal locations). Sub-national locations" natural advantages are akin to the endowment-driven comparative advantage concept in an international trade context. These natural advantages can help to explain both industrial concentration as well as entry rates in certain locations or clusters. It should be note that regarding labor, both availability and quality can play a role in determining industrial concentration in a location and while both have a „natural advantage" dimension to them at the same time both can also be influenced by government policy at the central or local levels. Ellison and Glaeser (1999) show that a large range of natural advantage factors –electricity, gas, coal, agricultural products, lumber endowments as well as variables capturing differences in the relative prices of different types of labor - explain a large portion of the geographic concentration of U.S. manufacturing industries across states in 1987. Sanguinetti and Martincus (2009) show that in Argentina, industries using intensively oil and mineral inputs have a significant tendency to locate in provinces with more abundant natural resources.

Focusing on the case of India, Lall and Chakravorty (2005) show that new private industrial investments are biased towards coastal districts after the 1991 structural policy reforms in India. Ghani et al. (2011) examine the spatial determinants of entrepreneurship in India. They explain the employment in young firms across each industry-district cell in 2005 calculated based on Annual Survey of Industries data – the same data that we will employ in our analysis as detailed in Section 4. Considering several district characteristics as determinants of firm entry, they show that education is a key factor in explaining entry in a district while a location farther away from the major Indian cities discourages entry.

The concept of agglomeration economies introduced by Marshall (1920) suggests that a location‟s industrial structure can play a critical role in explaining subsequent industrial concentration and the entry of new firms. Agglomeration economies relate to the benefits arising from the proximity to local suppliers and customers, the access to an appropriately skilled or specialized local labor pool, and the transmission of knowledge across closely located firms. The benefit from expansion or locating near existing clusters of activity is that agglomeration economies can reduce costs, for example the proximity to suppliers and customers can reduce transportation costs and increase productivity.[6] New Economic Geography (NEG) models take into account some of the Marshallian mechanisms but their emphasis is on the role of trade costs in determining industrial concentration. Firm location decisions are determined by two opposing forces. Centripetal or agglomeration forces result from the interaction between increasing returns to scale, market access (the location of demand) and trade costs or input-output linkages across producers. Centrifugal or dispersion forces relate to the trade costs to serve a dispersed consumer market but also to the competition for scarce local resources, i.e, congestion costs (related to rent or commuting costs) associated with large industrial agglomerations. Population density in a location captures both the market size but also congestion.

Many studies have examined the importance of Marshallian economies for industrial concentration and new firm entry in developed countries, and increasingly in developing countries. A small literature tests the implications of NEG models for spatial concentration of domestic employment or of foreign firms (or investments) within countries. Some key findings in those studies are presented below.[7] Glaeser and Kerr (2009) and Jofre-Monseny et al. (2011) examine the determinants of location of new manufacturing firm entry across U.S. and Spanish cities, respectively, estimating specifications whose dependent variable is the number of new firms in an industry-location. Marshallian agglomeration mechanisms are the factors of focus in Jofre-Monseny et al. (2011) and their findings show that new firm entry is significantly higher in locations with a strong presence of industries using similar workers and in locations with a large presence of input suppliers. Glaeser and Kerr (2009) show some role of the local presence of supplier industries and a critical role of the presence of many small suppliers – the Chinitz (1961) effect – in explaining new firm entry in U.S. industries. They also show that the abundance of industries employing similar types of workers fosters new firm entry in a U.S. state. Amiti and Javorcik (2008) examine the determinants of the spatial distribution of foreign firm entry in China, with a focus on NEG models‟ mechanisms. Market access and supplier

[6] There are different theories on the types of externalities involved in this process, some arguing that they are stronger when local clusters include firms in a single industry –localization or Marshall (1890)-Arrow (1962)-Romer (1986) externalities – others claiming that they are stronger when local clusters include firms in a diverse set of industries –urbanization or Jacobs (1969) externalities.

[7] We focus here on studies that examine the importance of Marshallian externalities using disaggregated industry-location level data on employment or firm entry. A different approach used to test the importance of Marshallian economies is to use industry-level measures of concentration – for example agglomeration indexes as in Ellison and Glaeser (1997) thus subsuming the location dimension into the calculation of those measures. See for example Ellison and Glaeser (1997) for evidence on U.S. industries and Fernandes and Sharma (2010) for evidence on Indian industries.

access - in particular the presence of customers and suppliers in the same province - are shown to be the most important factors determining the location of foreign firms in China, while production costs matter but to a much lesser degree.[8]

Focusing on the determinants of location choices of different types of manufacturing investments in India, Lall and Chakravorty (2005) show that new private (domestic or foreign) investments in India in 1998 are biased towards existing industrial clusters. Mukim and Nunnenkamp (2010) examine the spatial distribution of foreign investment projects in India between 1991 and 2005 and show that foreign investors prefer to locate where other foreign investors are already present and in industrially diverse locations. Ghani et al. (2011) show that agglomeration economies captured by the presence of incumbent firms as well as a measure of labor pooling influence to an important degree the entry of new firms at the industry-district level in 2005. They also show the importance of the presence of small suppliers in the location in fostering new entry, an effect predicted by Chinitz (1961). Once they replace the district characteristics in their specifications by district fixed effects and thus focus only on district-industry variation the strength of the effects of agglomeration economies increases.[9] Interestingly, the effects of agglomeration economies increases in their more stringent specification controlling for industry-district fixed effects using a balanced panel covering 1989 and 2005. Ghani et al. (2011) also show that high population density discourage entry in a district, providing some evidence of the centrifugal NEG forces at work. Overall, Ghani et al. (2011) conclude in favor of a strong link between entrepreneurship and local industrial conditions in India.

The importance of location characteristics including the general business environment for the concentration of industries and investments is self-evident. For developing countries in particular, the availability of infrastructure as well as its quality or reliability, are particularly important factors affecting a firm‟s expected profitability from operating in a given location (Deichmann, et al., 1998). Within infrastructure, power can be a particularly important driver of geographic specialization given that it is the most important factor of production after labor and capital. The standard or quality of governance in a particular location may also affect the spatial concentration of industrial activity: better governance can reduce uncertainty, reduce firm costs (via lower corruption), and in general, reduce the costs of doing business. The stringency of labor regulations and the degree of enforcement of property rights are two dimensions of governance that may play a particularly relevant role in determining location.

Amiti and Javorcik (2008) show the importance of the availability of infrastructure in explaining the spatial distribution of foreign firm entry in China. Michielsen (2011) shows that regional

[8] The importance of customers and suppliers in the same province is consistent with important inter-provincial barriers to trade.

[9] Ghani et al. (2011) use a couple of years of data from the Annual Survey of Industries – which is also our main data source – and explain the employment in young firms in each industry-district cell in 2005 calculated using ASI data – the same dataset we employ in our analysis - based on district characteristics, agglomeration measures by industry-district, while controlling also for incumbent employment in the industry-district and for industry fixed effects.

endowments of coal, natural gas, oil and hydro-power affect significantly the location of energy-intensive U.S. industries. Rothenberg (2011) shows that substantial road improvements in Indonesia led to a dispersion of manufacturing employment on average but with important differences across industries. The spatial concentration of producers of perishable goods that deteriorate rapidly in transit and hence need to be consumed near where they are produced did not change much with road improvements whereas the dispersion of employment of producers of durable goods declined substantially.

For the case of India, a positive and significant effect of the supply of local infrastructure services on the „competitiveness" of Indian cities is shown by Lall et al. (2010), where „competitiveness" is measured by the share of the city in national private manufacturing investment. Mukim and Nunnenkamp (2010) show that foreign investors prefer to locate in Indian districts with better infrastructure. Ghani et al. (2011) show that education is a key factor systematically linked to entrepreneurship in a district while stringent labor regulations discourage entry. The effect of labor regulations that raise monitoring costs, and reduce flexibility with consequences for productivity and inter-regional wage and growth disparities is also documented for India by Besley and Burgess (2004). Chemin (2010) shows that better judicial efficiency in an Indian state – a proxy for a better local enforcement of property rights – increases the probability of firms investing on machinery and equipment.

The literature above considers concentration and new firm entry in the formal manufacturing sector. However, the informal sector tends to account for a very large share of economic activity in developing countries and it is thus important to understand the determinants of the spatial distribution of activity in that sector. Two recent studies examine this issue for India using data from the National Sample Survey (NSS) that we will also use in our analysis as detailed in Section 5. Mukim (2011) examines the drivers of spatial clustering of new firms in the informal sector in India in 2004 and shows that localization (proxied by the share of firms in the industry-district) has a negative effect on entry, reflecting perhaps a fear of competition within the industry, while the strength of linkages to suppliers as well as district population density have a strong positive effect on entry.[10] The level of education and the quality and availability of infrastructure play no significant role for the clustering of firms in the informal sector. Ejaz et al. (2011) estimate the spatial determinants of informal entrepreneurship measured by the employment in young informal firms in each industry-district cell in 2005 and find a very strong effect from district population, which captures the overall size of the labor market as well as from the physical infrastructure in the district and the strength of the local banking environment.[11] They also provide evidence of agglomeration economies in that local industrial

[10] The study estimates a count model whose dependent variable is the number of new firms in the informal sector in a district in 2004 and the determinants are several economic geography-related and business environment-related factors.

[11] The specification considered for the new entry of firms in the informal sector is similar to that described above for the entry of new firms in the formal sector.

conditions by incumbent firms - labor market strength, the input-supplier strength and the output-customer strength - are strongly positively correlated with new entry of informal firms.

To study the spatial distribution of manufacturing activity in India we want to follow an approach that allows for certain state characteristics to complement industry characteristics in explaining location decisions. We draw heavily on the approach proposed by Midelfart-Knarvik et al. (2000) who explain industrial location across countries based on a model that combines comparative advantage cost factors and geographical forces based on mechanisms present in new trade NEG models. To facilitate the transition to our approach where we will replace countries by sub-national geographical units in India (states), we discuss the model in terms of „locations" instead of countries. Some key features of the model are as follows. Locations differ in endowments and trade among locations is costly with transport costs varying with distance. Industries use intermediate factors to produce differentiated final goods. In the presence of transport costs, both supply and demand considerations matter to determine the location of different industries.[12] Also, since prices and demand for intermediate goods differ across locations, industries tend to locate near their supplier and customer industries. The fundamental feature of Midelfart-Knarvik et al."s model is that it allows the various effects to vary across industries by considering characteristics that lead industries to concentrate in locations of different types generalizing the standard trade theorems of Rybczynski and Heckscher-Ohlin.[13] The empirical specification derived from the model shows how location characteristics demand (endowments or geography) interact with industry characteristics (factor intensity or transport costs) to determine the industrial production structure across locations.[14] Both location characteristics and industry characteristics can influence costs and demand patterns (whether it is demand for final or intermediate goods). This approach is closely related to the model proposed by Ellison and Glaeser (1999) to explain the spatial distribution of an industry whereby firms choose locations maximizing profits which depend on local cost advantages and industry characteristics. Local characteristics considered afford natural advantages in terms of natural resources, transportation costs and labor inputs. They combine cost differences with industry"s intensity of factor use to estimate the spatial distribution of manufacturing activities expected due to cost differences and population distributions.

Our empirical specification examines the determinants of regional concentration of manufacturing industries in India over time adapting and extending the specifications of

[12] In particular, the location of demand - the market potential - shapes the industrial structure of a location.

[13] The Rybczynski theorem states that a rise in the endowment of a factor will lead to a more than proportional expansion of the output in the sector which uses that factor intensively, and an absolute decline of the output of the other good, at constant relative goods prices (in a model with 2 factors, 2 goods, and 2 countries). The Heckscher-Ohlin theorem states that a country exports goods that use its abundant factors intensively and imports goods that use its scarce factors intensively (in a model where countries differ only in their factor endowments).

[14] The simple intuition behind the mechanisms at work is that in principle all industries would like to locate in the „most attractive" locations – those with larger endowments, better market access and more proximity to suppliers but it is impossible for a location to absorb all industries. Thus, the industries going to a given location are those that benefit most from the endowments relatively abundant in that region.

Midelfart-Knarvik et al. (2000) and Ellison and Glaeser (1999).The use of this type of empirical specification to explain industrial location in India is appropriate given the spatial dimension of the country and the relatively high transport costs across states in India. This approach has been used to examine the effects of trade policy in shaping the sub-national economic geography of Argentina and Brazil by Sanguinetti and Martincus (2009) and Martincus (2010), respectively. As will be detailed in Section 4.2, we allow concentration to depend on a series of interactions between location characteristics and industry characteristics. One way of viewing this type of specification is that it imposes cross-industry restrictions whereby the sensitivity of location decisions to the cost of a given input is required to be related to the intensity with which the industry uses that input. The conjectures tested by this type of specification can be illustrated by several examples. Firms in industries relying on energy-intensive techniques would be likely to concentrate in locations rich in sources of electricity while firms in export-oriented industries would be likely to concentrate in coastal locations. If an industry‟s production reduces the weight of material inputs in the final good, the preferred location would be near suppliers whereas if transportation of finished products is difficult the preferred location would be near buyers.

3. Data and Empirical Approach

3.1 Data

Our main source of data is the plant-level data from the Annual Survey of Industries (ASI) conducted by the Central Statistical Organization (CSO), a department of the Ministry of Programme Planning and Implementation of the Government of India for all consecutive years in the period 1980-1981 until 2007-2008, with the exception of 1995-96 when the survey was not conducted.[15] To our knowledge, our paper is the first to use this very long time frame for India which covers all the Indian reforms of the 1980s, the major reform episode of 1991, as well as the fast-growing period of the 2000s. The survey covers all factories registered under the Factories Act of 1948 - defined as units employing 20 or more workers - which corresponds to the formal or organized sector in Indian manufacturing. The ASI frame can be classified into 2 sectors: the „census sector" and the „sample sector". Units in the census sector - larger firms - are covered with a sampling probability of one while units in the sample sector - smaller firms - are covered with sampling probabilities lower than one. The ASI dataset includes multipliers (the inverse of the sampling probabilities) for each firm that we use in the calculation of any variable based on information in that dataset so that we ensure representativeness of the entire formal manufacturing sector. It should be noted that due to restrictions on some independent variables in our analysis, the longest sample period that can be used for our empirical analysis is from 1985-1986 to 2007-2008.

[15] Note that although we use the word plant and firm interchangeably throughout the paper, the ASI data covers establishments, i.e., plants, not firms.

Some important caveats about the ASI dataset must be noted. First, the dataset do not contain a plant identifier. That is, the dataset is a repeated yearly cross-section of plants, not a panel of plants.[16] Second, due to multiple changes in industrial classification occurring during the sample period a great effort was put into concording these classifications over time. In order to use as many years of data as possible, we use the 3-digit National Industry Classification (NIC 1987) level of industry disaggregation. Third, the most disaggregated location unit we can use is an Indian state. The ASI does collect information about the district of location of each plant but does not provide this information to users in order to protect the plant"s identity. Fourth, as mentioned in Bollard et al. (2010) and Fernandes and Sharma (2010) the ASI data are extremely noisy in some years.

The ASI data includes, for each plant, information on total employment which we aggregate up to the industry-state level use to construct our measures of the spatial manufacturing patterns in India: employment concentration and of the entry of new firms. The concentration of industry j in state s is measured by the employment share below:

$$share_{jst} = \frac{L_{jst}}{\sum_{s=1}^{S} L_{jst}} \quad (1)$$

where L_{jst} is employment in industry j in state s in year t and S is the total number of states.[17]

The concentration of employment in Eq. (1) is a „net" measure that covers both expansion and contraction of incumbent firms in an industry and location but also firm entry and exit in that industry and location. It also of interest to examine a measure capturing more directly the entry of new firms in an industry and location. While we cannot observe exactly when a plant enters into an industry and location because the ASI dataset is not a panel of plants, it does include information on plant age. Thus, as a proxy for the entry of new firms, we use the employment in young firms in an industry and state, where young is defined as a plant that is less than three years old, following the approach used by Ghani et al. (2011). The rate of entry of new firms in industry j and state s is given by:

$$entry_{jst} = \frac{\sum_{i=1}^{N} I(age_{ijst} \leq 3) * L_{ijst}}{\sum_{i=1}^{N} L_{ijst}} \quad (2)$$

where $I(age_{ijst} \leq 3)$ is an indicator variable for plants under three years of age and L_{ijst} is employment of plant i in industry j in state s in year t and N is the total number of plants in industry j in state s in year t.[18]

The determinants of concentration and new firm entry that we will use in our empirical analysis are described in Section 4.2. The industry characteristics are constructed based on variables

[16] This limitation restricts the modeling approach for our location decision. Rather than modeling firms" location choices, we use more aggregate measures of spatial concentration and new entry at the industry-state level.
[17] Therefore, the denominator of Eq. (1) is total employment of industry j in year t.
[18] Therefore, the denominator of Eq. (2) is total employment of industry j in state s in year t.

included in the ASI dataset whereas the state characteristics are based on multiple data sources, most notably the State Domestic Product dataset produced by the EPW Foundation. Table 1 provides a complete list of all the variables used in our empirical analysis (and their names that will be used consistently throughout the paper), their descriptions, details on the data sources and time coverage for each of the variables used. Appendix Tables 1 and 2 present the means of our main explanatory variables over time.

For our outcome variables of interest as well as for most of our regressors, the industrial classification used is 3-digit NIC 1987 (exceptions will be noted) which covers 180 industries.[19] Also, due to data limitations, we consider only the subset of 16 Indian states listed in Table 2 but these are the largest Indian states which account for 94% of formal manufacturing employment and 92% of formal manufacturing output in the country. Note that although only 16 states are considered, the employment shares are calculated based on the total number of states in order to address an econometric challenge discussed in Section 6. Further, three new states were created during the sample period: Uttarakhand carved out of Uttar Pradesh, Chattisgarh out of Madhya Pradesh and Jharkhand out of Bihar. In the analysis we treat the new states as if they were part of the original state.

3.2 Empirical Approach

Our empirical specification explains either the concentration of industry j - $share_{jst}$ - or the entry of new firms in industry j - $entry_{jst}$ - across Indian states s based on a rich vector of interaction terms between location characteristics $LOCATION_{st}$ and industry characteristics $INDUSTRY_{jt}$:

$$share_{jst} = \alpha_o + INDUSTRY_{jt} * REGION_{st} * \alpha + I^j + I^s + I^t + \varepsilon_{jst} \quad (3)$$

$$entry_{jst} = \beta_o + INDUSTRY_{jt} * REGION_{st} * \beta + I^j + I^s + I^t + u_{jst} \quad (4)$$

where I^j, I^s, and I^t are, respectively, industry, state, and year fixed effects, and ε_{jst} and u_{jst} are independent and identically distributed (i.i.d.) residuals which can represent a random element in firm location decisions or a set of unobservable local attributes. The industry and state fixed effects included account for unobserved heterogeneity across locations and industries that may affect concentration or new firm entry and also account for the levels of the industry and state characteristics whose interaction terms are our main variables of interest.[20] However, we will also consider other types of fixed effects in our empirical specifications either at the industry-

[19] Due to the changes in the 3-digit NIC classification that occurred over our sample period and the difficulties experienced in establishing concordances of the classifications over time the textiles-related industries (NIC 231 through 237) are collapsed into a single industry NIC 230 over the sample period and the same occurs for NIC 241 through 247 collapsed into NIC 240 and NIC 251 through 259 collapsed into NIC 250.

[20] This type of approach was actually initially proposed by Rajan and Zingales (1998) to examine whether industries that are relatively more dependent on external financing grow faster in countries with more developed financial markets.

year and state-year levels or at the industry-state level and we will also consider the alternative of including also the levels of both state and industry characteristics.

In generic terms, a term in the interaction vector $INDUSTRY_{jt} * REGION_{st}$ should be viewed as the comparative advantage that location s has for industry j. If location s has a desirable characteristic A, this would in principle make all industries locate entirely there. However, given physical constraints and the congestion costs associated with an unique-location equilibrium, not all industries will locate entirely in that location. The industries will take the constraints and costs into account in choosing their location and only those that benefit the most from the desirable characteristic A will locate in s.

We describe below various components of the interaction vector used in our baseline specifications and where appropriate we discuss also the interaction terms used in robustness specifications. The variables that build these interaction terms are listed and defined in Table 1.

The first set of interaction terms we consider correspond to the natural advantage type of effects in Ellison and Glaeser (1999) and follows the reasoning in the Rybczynski and Heckscher-Ohlin trade theorems, i.e., the fact that industries are expected to concentrate in locations that are relatively abundant in the factors that they use relatively more intensively in their production processes. We consider an interaction term multiplying unskilled labor intensity at the industry-level by unskilled labor abundance at the location-level. Unskilled labor intensity is proxied by the average ratio of unskilled (non-production) labor costs (wages plus benefits) to the value of production in the industry (UNSKILLED INTENSITY) whereas unskilled labor abundance is proxied by the ratio of the average wage per formal production worker in India to the average wage per formal production worker in the state (UNSKABUN).[21] We consider also an interaction term multiplying skilled labor intensity at the industry-level by skilled labor abundance at the location-level. Skilled labor intensity is proxied by the average employment share of skilled (non-production) workers in the industry (SKILL INTENSITY) whereas skilled labor abundance is proxied by the share of formal manufacturing skilled (non-production) workers in the state in formal manufacturing skilled (non-production) workers in India (SKABUN).[22]

The next set of interactions relates to the mechanisms in NEG models and to Marshallian economies. One of the interaction terms multiplies the degree of returns to scale at the industry-level by the market size at the location-level to capture the possibility that industries with

[21] Our proxy is actually the inverse of the measure used by Sanguinetti and Martincus (2009).

[22] In future work we would like to estimate also specifications where we proxy for skilled labor abundance using a measure of the proportion of the population in a state with a high-school degree for example. While Ellison and Glaser (1999) consider also a rich set of natural resources as potential determinants of location, we are unable to do so in our analysis of location in India due to data restrictions. This could potentially lead to an omitted variables problem that we will try to address in our most stringent specification where we include industry-state interaction fixed effects, which is a reasonable approach for natural resources, safe for any heavy depletion or a discovery of unknown sources.

increasing returns to scale will concentrate in locations with larger markets or a larger market potential, as predicted by NEG models with inter-location trade costs. Returns to scale are proxied by the average real capital per plant in the industry (IRS) while the size of the market is proxied by wealth in the state - and thus potential demand - measured by the state‟s gross domestic product (GDP) per capita (MKT). Another interaction term addresses the strength of the input-output linkages in explaining location. The idea is that manufacturing industries that utilize more manufacturing intermediate inputs in their production processes will benefit from concentrating in locations with a large presence of manufacturing suppliers that allows them to obtain cheaper inputs and a greater variety of inputs. The utilization of intermediate inputs from other manufacturing industries is proxied by the ratio of manufacturing intermediate inputs to the value of gross sales in the industry (MATS) while the size of the local manufacturing base is proxied by the share of the total GDP in the state that originates in manufacturing (BASE).

The final set of interactions relates to business environment - more directly policy-related - aspects.[23] Transport infrastructure is expected to reduce costs and affect concentration for all industries but particularly so for industries that rely more heavily on roads (or other means of transport). Our interaction term multiplies a measure of the availability or quality of transport infrastructure at the state-level by a measure of transport intensity at the industry-level. However, we should note that since better transport infrastructure reduces transport costs within and across locations, its consequences for the spatial distribution of manufacturing activity are ambiguous. Better transport infrastructure can lead to dispersion by inducing industries to locate farther from areas where other industries are concentrated because they can still purchase inputs and supply their outputs cost-effectively. This is in fact the principle behind most regional development policies that propose transport infrastructure investments as a means to better integrate lagging regions and in doing so promote their development. The availability or quality of transport infrastructure is proxied alternatively by state expenditures on transport, communications and storage (TRANSPORT) in Section 6, and by the length of highways in the state (HIGHWAY) in Section 7. The transport intensity is proxied by the industry‟s average ratio of inventories to sales (TRANSPORT INTENSITY). A low value of this ratio implies that the industry produces a more perishable good and hence faces greater transport cost per unit distance. This implies that the industry would be less spatially concentrated than others. Electricity infrastructure is another potentially important driver of manufacturing location that should matter particularly for industries that rely on highly electricity-intensive technologies. Our interaction term multiplies a measure of the availability or quality of electricity infrastructure at the state-level by a measure of energy intensity at the industry-level. The availability or quality of electricity infrastructure is proxied by the length of the electricity transmission lines in the state (TRANSMISSION) while

[23] Due to data limitations, our empirical specification does not control explicitly for state‟s regional policies such as subsidies or tax breaks that could be used to influence the location of manufacturing activity.

the electricity intensity is measured by the average ratio of electricity consumption to gross sales in the industry (ELECTRIC).[24]

The quality of governance can cover a wide range of aspects from the ability to enforce contracts (property rights) to the ease of doing business in the state and it is likely to play a major role in determining spatial concentration but is more likely to do so when industries rely more extensively on contracts and on business interactions with other economic agents. Our interaction term multiplies a measure of the quality of the governance at the state-level by a measure of contract intensity at the industry-level. The perceived quality of governance provided by the state government is proxied by a measure of reported crime, the negative of the number of murders per capita (GOVERN). The reliance on contracts and business interactions is proxied by the ratio of materials to gross sales in the industry (MATS) with the rationale that industries that utilize a larger amount of materials need to interact with a large number of suppliers and engage in more contracts.

One final general remark about our specification is that given the still relatively early stage of industrialization of India and the fact that the country‟s industrial landscape is still adjusting to the policy reforms following the 1991 liberalization – as shown by Fernandes and Sharma (2010) - location characteristics have the potential to have a much more substantial impact in determining spatial concentration in India than would be the case in mature developed countries (Ghani et al., 2011).

4. Descriptive Evidence

A starting point for our analysis is the establishment of the presence of concentration or clustering of manufacturing activity across industries in India. Fernandes and Sharma (2010) provide such evidence for the period ending in 1999-2000. They show that the distribution of industries across Indian states (as measured by the Ellison and Glaeser (1997) index) is highly skewed and that certain industries are particularly concentrated but the ranking of concentration levels swifts quite importantly between 1980 and 1990.[25]

To provide some insights onto the concentration of manufacturing employment across Indian states and over time, we show in Table 2 the share of manufacturing output accounted for by each of the top 10 states - in terms of their manufacturing employment in 2007 - in selected sample years from 1985 to 2005. The top 5 states account for close to 60% of total

[24] Describing the electricity interaction term as part of the business environment aspects is driven by the actual measures we have that capture produced/transmitted electricity. If we had information on natural resource endowments in terms of coal, natural gas, oil, or hydro-power across Indian states we could consider the interaction between such endowments and industry-level energy intensity as part of the natural advantage terms. This is the approach taken by Michielsen (2011) for the U.S..

[25] For examples of the most and least concentrated industries in India see Fernandes and Sharma (2010).

manufacturing employment while the top 10 states account for roughly 85% of manufacturing employment in any of the years, and the shares of individual states are also quite stable over time.[26] Note that in most cases, each states‟ share of manufacturing employment is higher than its share of the total Indian population, providing some preliminary evidence of excess concentration of manufacturing. This tendency is also noticeable in Table 3 where we show for selected sample years, the average employment share (Eq. (1)) and the average entry rates (Eq. (2)) in each state, with the average being taken across industries *j*.

Table 3 provides averages calculated excluding as well as including zeroes pointing out to an important feature of our data – not every industry is present in every state (in every year). Table 4 shows the proportion of manufacturing industries (out of the total of 180) in each state with no employment in the formal sector. The proportion ranges from a minimum of 12.57% (Maharastra in 1996) to a maximum of 73.77 (Assam in 2001). Further, for almost all states this proportion is increasing over time, that is, the number of different manufacturing industries located in each state is falling over time. The de-industrialization process is much faster for some states - Orissa, Assam, Bihar, Punjab, Kerala and Rajasthan - but the proportion is increasing for all states. The problem of zeros is even higher when we consider entry (i.e., employment by young plants) in each industry-state cell in Table 5. The proportion of empty cells for entry ranges from 29% (Uttar Pradesh in 1989 and 1990) to 93% (Himachal Pradesh in 1987). Entry becomes less diversified between the 1980s and the 1990s - that is, the proportion of empty cells rises - but slightly more diversified between the 1990s and 2000s. The distributions of our main dependent variables reflect this characteristic of the data as well. Figures 1 and 2 show the distribution of employment shares and entry rates across 3-digit industries within each state in 1996 including the zeroes. The figures reflect the large number of zeros as well as some degree of heterogeneity in both employment shares and entry rates. That is, our dependent variables are clustered around zero and contain outliers. In order to deal with this problem we will winsorize our dependent as was described in Section 3.

5. Baseline Estimation Results

Several data and econometric challenges need to be discussed related to the estimation of Eqs. (3) and (4). A first key challenge concerns the appropriate estimation sample to use. Our dependent variables - the share and new firm entry measures in Eqs. (1) and (2) - are calculated for an unbalanced panel of 41429 observations based on 180 industries, 16 states, and 22 years. However, as discussed in Section 4 not all industries locate in all states in all years, the potential total number of industry-state-year cells in our dataset would in fact be 63360 (180 industries*16 states*22 years). Hence, there is an important fraction of zeros for both the employment share –

[26] Appendix Table 3 shows for all 16 Indian states the manufacturing employment levels in all firms and in new firms (i.e., entry) in selected sample years.

40% of the potential total - as well as for the entry rate – 60% of the potential total. This censoring problem has a key meaning in the context of our paper as it indicates that firms in certain industries chose to not locate in certain states. Ignoring the zeros in the estimation could therefore lead to biased estimates of the effects of certain location determinants. We should note that zero employment for an industry in a state may represent the true absence of the industry in that state but since ASI data covers only formal manufacturing we cannot discard the possibility that informal plants in that industry could be located in that state but are not captured by our measure. To address the potential biases introduced by censoring our preferred specifications will include the cells with zeros (although we also show for comparison results ignoring the zeros).

A second challenge is that both of our dependent variables are fractional measures restricted to vary between 0 and 1. If OLS is used for the estimation of Eqs. (3) and (4) and the variance of the error term is large, then the estimated coefficients could be biased. To address this problem, we use Tobit estimation for all our specifications.

A third challenge is that the shares of an industry"s employment add up to 1 when summed across states.[27] This is a problem that has been ignored in previous studies but implies that errors may have a specific structure within each industry that would affect inference. To address this problem we would need to drop one state for each industry, but we are already implicitly doing this because due to data limitations we focus only on 16 Indian states for which we have location characteristics although the shares are calculated relative to manufacturing employment in all Indian states. Moreover, we will cluster standard errors by industry in most of our specifications to account for any potential cross-industry correlations in unobservables. Additionally, we will also discuss the results using employment levels instead of employment shares as dependent variables in Section 7. Finally, since the data are noisy and contain outliers we use a winsoring technique to minimize the effect of potential outlier observations in all our specifications. This technique replaces the top and bottom 1% observations from both tails of the distribution of any of our dependent variables by, respectively, the value of the variable at the 99th percentile and at the 1st percentile.

Table 6 presents the baseline results from estimating Eq. (3) (columns (1)-(2)) and Eq. (4) (columns ((3)-(4)) including as determinants of location the interaction terms described in Section 4 with a one-year lag. In calculating the interaction variables $INDUSTRY_{jt} * REGION_{st}$ note that the location characteristics are defined such that a larger value indicates a more „attractive" location, and the industry characteristics are defined such that a larger value indicates a higher dependence of the industry on the location factor. In order to make comparisons across the effects of the various determinants easier, the interaction variables enter the regressions in a standardized format, i.e., for each variable we take its original value and subtract its sample mean and then divide that by its sample standard deviation. Our tables report White

[27] We thank Jeff Milyo and Peter Mueser for pointing this problem to us.

heteroskedasticity-robust standard errors clustered by 3-digit industry for all coefficients. Our specifications include state and industry fixed effects - to account for unobserved heterogeneity across states and industries - which imply that our coefficients on the interaction terms are identified within locations and industries. The year fixed effects account for cyclical patterns in manufacturing concentration or new entry in India overall.

Columns (1) and (3) of Table 6 present the estimates based on samples that do not include zeroes in the dependent variables while columns (2) and (4) present the estimates based on samples including zeroes in the dependent variables, as can be seen by the very different number of observations across the two groups of columns. In all cases, since qualitatively the effects of the interaction terms do not differ when zeroes are included, we base our comments in what follows on those results that are our preferred specifications.

The evidence on the importance of the comparative advantage factors for location is mixed. The estimated coefficients on the interaction term linked to skilled labor are positive and significant as expected, indicating that industries that rely more on skilled labor concentrate relatively more and exhibit more new firm entry in states with a more abundant skilled workforce. However, the estimated coefficient on the interaction term linked to unskilled labor is negative and significant which suggests the counter-intuitive result that unskilled labor-intensive industries chose to concentrate and exhibit higher new entry in states with relatively higher unskilled wages than the Indian average. A problem with our choice of the average unskilled wage measure to measure unskilled labor abundance is that lower average unskilled wages may in fact be proxying for a more pro-business environment in the state (if they are driven by lower minimum wages) or may be proxying for a lower quality of the unskilled workers in the state.[28]

We also test the robustness of this counter-intuitive result estimating a specification that allows the unskilled labor comparative advantage factor to have a separate effect for states whose relative unskilled wage ratio is lower than the median unskilled wage ratio across all Indian states, i.e., states with higher than median unskilled labor abundance. The unreported results show that this separate effect is positive and significant, i.e., states with higher than median unskilled labor abundance have a greater concentration of manufacturing and higher new firm entry. However the net effect of the interaction between unskilled labor endowments and unskilled labor intensity on the average state remains negative.

[28] To account for these possibilities, we plan to estimate new specifications where in addition to the labor interaction term we include also (i) an interaction between an index of the flexibility of labor regulations in the state – where flexibility is associated with a more pro-business environment following Ahsan and Pages (2009) – with the unskilled labor intensity of the industry and (ii) an interaction between a measure of the quality of the manual workers in the state – the average number of years of education of manual workers following Amoranto and Chun (2011) – with the unskilled labor intensity of the industry. By including these two interaction terms, the average unskilled wage in the state should be closer to its original role of capturing the abundance of unskilled labor in the state.

The results on the interaction terms linked to NEG mechanisms and Marshallian economies in Table 6 are weak. Industries with increasing returns to scale do not have a tendency to locate nor to have higher entry rates in states with larger market potential. If anything, the estimates in column (2) suggest significantly lower concentration (at the 10% confidence level) in states with larger market potential. The absence of evidence of this NEG mechanism is similar to what was found in the case of Argentina by Sanguinetti and Martincus (2009) and in prior work on India by Fernandes and Sharma (2010).[29]

The estimates in columns (2) and (4) of Table 6 suggest that industries that rely more intensively on manufacturing intermediate inputs exhibit no tendency to locate in states with a larger manufacturing base.[30] If anything, the estimates in column (2) suggest significantly lower concentration (at the 10% confidence level) in states with a larger manufacturing base. One way to interpret this counter-intuitive result is that it suggests that input-output linkages are not important in determining location in India, i.e., firms do not try to locate near suppliers. Alternatively, the states with a large industrial base may be suffering from congestion and congestion prevents the benefits from locating near suppliers to materialize. To address this possibility, we estimate a specification similar to that used for the labor comparative advantage factor where we allow for the interaction to have a separate effect for states whose industrial base is higher than the median across states. The unreported results from this new specification do not provide support for the congestion hypothesis (the separate effect is insignificant). In yet another specification, we include in addition to the interaction term, another term which is the product of the industry manufacturing inputs‟ intensity with the square of the size of the manufacturing base. However, we find no significant second order effects of the size of the manufacturing base on spatial concentration nor new firm entry in India.

The business environment interaction terms have a significant role in explaining spatial concentration of and new firm entry into manufacturing industries in India. According to the estimates in Table 6, industries that are more transport dependent are significantly more likely to locate in states with larger expenditures on transport infrastructure and new firms in those industries also chose to enter more frequently in those states. One potential problem with this result is that in addition to proxying for the availability and quality of the transport infrastructure network, expenditures on transport infrastructure can also proxy for other attractive features of the states linked for example to a pro-business regulatory regime. Hence we will present in Section 7 results using alternative proxies for the availability and quality of the infrastructure, extending the analysis to consider electricity in addition to transport infrastructure. Finally, industries that are more reliant on contracts locate in states with better governance and new entrants in those industries also prefer to locate in those states. This makes sense given recent

[29] This finding could be due to the proxy used for market potential in the state. In future work we will construct an alternative measure of market potential that accounts for the access to the markets in other states weighted by distance, as in Amiti and Javorcik (2008).

[30] New firm entry does not relate significantly to the size of the manufacturing base in the state.

research that points to the importance of governance quality in enhancing the gains from economic liberalization (Ahsan, 2010).

6. Robustness

We conduct a series of robustness checks to our baseline results in Table 6. Table 7 presents the results from re-estimating Eqs. (3) and (4) but replacing the state and industry fixed effects by the levels of the state characteristics and of the industry characteristics in columns (1) and (2).[31] The estimates suggest our baseline findings are mostly maintained unchanged: we find a positive effect of the comparative advantage factor related to skilled labor but a counter-intuitive negative effect of the comparative advantage factor related to unskilled labor. Good infrastructure and governance still play a role - though weaker than in Table 6 - in explaining spatial concentration of, and new firm entry into industries more reliant on business environment factors. A different but more intuitive result is now obtained in column (2) in that new firm entry is higher for industries more reliant on manufacturing inputs in states with a larger manufacturing base. However the corresponding coefficient is significant only at the 10% confidence level.

In columns (3) and (4) of Table 7, in addition to controlling for time-invariant state unobservables by including state fixed effects, we include regional trends in the specifications to account for movements that may be common across geographic areas larger than states. For example, South Indian states have been rapidly industrializing over the past 20 years, while North Indian states have been de-industrializing.[32] The estimates show that the coefficients on the interaction terms are essentially unchanged in terms of signs and significance once regional trends are included.

Columns (5) and (6) of Table 7 present the results from a more stringent specification where Eqs. (3) and (4) are re-estimated including state-year and industry-year fixed effects.[33] The estimates are qualitatively similar to our baseline findings, the only exception being the weaker role of transport infrastructure in Table 7. Columns (7) and (8) of Table 7 show that some of the interaction terms have weaker significance when robust standard errors are clustered by state rather than by industry. This finding is not surprising given the much small number of states relative to the number of industries in the sample.

Table 8 presents the results from re-estimating Eqs. (3) and (4) using alternative proxies to construct the interaction terms linked to NEG mechanisms and Marshallian economies for which

[31] The coefficient estimates on the levels of the state and industry characteristics are available from the authors upon request.

[32] We define four regions according to the following groups of states: region 1 (AP, KAR, KER, TN), region 2 (MAH, GUJ, RAJ, MP), region 3 (BIH, ASS, WB, ORS, UP), and region 4 (HAR, PUN, HP).

[33] Year fixed effects are dropped since they are collinear with the interaction fixed effects. The coefficient estimates on the levels of the state and industry characteristics in columns (1) and (2) are available from the authors upon request.

we obtained counter-intuitive results in Table 2. To proxy for the wealth of a state, we use the state‟s electricity consumption per capita (ELEC MKT) instead of the state‟s GDP per capita, but in columns (1) and (2) the interaction with the proxy for increasing returns to scale exhibits negative and significant effects in the employment shares and the entry rates regressions.[34] To proxy for the manufacturing base in a state, we use the volume of electricity sales to the manufacturing sector in the state (ELEC BASE) and the interaction with the proxy for manufacturing input usage still shows a negative and significant coefficient in the employment share regression in column (1) as was the case in Table 6.

One of the econometric challenges with the estimation of Eq. (3) identified in Section 5 is the potential problem from using industry shares of employment that sum up to one across Indian states. While the results presented so far are unlikely to suffer from this problem since they are based on a sample of 16 states while the employment shares are calculated based on all states, we still pursued this issue further and estimated a variant of Eq. (3) where log total employment in all firms in an industry and state (log of the numerator in Eq. (1)) is the dependent variable and we include as a control the size of the industry in India as a whole (as measured by log total employment).[35] The corresponding baseline and robustness results are shown in Appendix Tables 4 and 5. The qualitative patterns in these results are similar to those shown in Tables 6 through 8. In particular, we find a significant positive effect of the comparative advantage factor linked to skilled labor and positive albeit weaker effects of the transport infrastructure and governance factors. Although our entry rates in Eq. (2) do not suffer from the same adding up problem, we also present in Appendix Tables 4 and 5 the results from estimating a variant of Eq. (4) where log employment in new firms in an industry and state (log of the numerator in Eq. (2)) is the dependent variable and we also include as a control the size of the industry in India as a whole (as measured by log total employment). The results are also qualitatively similar to those shown in Tables 6 through 8.

7. Extensions

7.1 Infrastructure

In addition to considering alternative proxies for the interaction terms linked to NEG mechanisms and Marshallian economies - described above - in Table 8, we consider the role of electricity (instead of transport) in determining spatial concentration of manufacturing in India,

[34] Due to restrictions on the availability of data on the electricity state-level variables, the sample period of the regressions in columns (1)-(4) is 1985-2005 instead of 1985-2007 as in Tables 2 and 3.

[35] As noted there are several industry-state cells with no employment. In order to include these observations in the specifications that use log employment as dependent variable, we add a small number of employees to each empty cell. In particular, we add 50 employees to each empty cell since that is the threshold employment level that defines a formal plant in India that would be sampled under the ASI framework. A similar procedure is followed for the entry measure.

our infrastructure interaction term in columns (3) and (4) is constructed multiplying the length of the electricity transmission lines in the state (TRANSMISSION) by the average ratio of electricity consumption to gross sales in the industry (ELECTRIC) (as was described in Section 5). In contrast to the positive effects of transport infrastructure on location in Table 6, the coefficient on the electricity infrastructure interaction term is now negative in both the employment share and the entry rate regressions - significant in column (3) - indicating that industries that use electricity more intensively exhibit higher concentration and entry in states with a smaller electricity transmission line grid. This counter-intuitive finding could be linked to the unobserved quality of electricity infrastructure (for example, as reflected in prices of electricity or hours of uninterrupted provision of electricity) or could indicate a problem of congestion. States with more electricity transmission lines may be over-crowded and despite being wider, the electricity grid is not able to respond to the heavy demand from electricity users in the state. To address the possibility of congestion, we estimate a separate specification where we allow for the electricity interaction term to have a separate effect for states whose length of the electricity transmission lines is above the median across states. The unreported results from this specification do not provide any evidence of congestion (the separate effect is insignificant).

A different interpretation of our results – positive effects of infrastructure when measured by state expenditures on transport infrastructure and negative effects of infrastructure when measured by physical infrastructure – could be that the state transport expenditure levels provide firms with information signals about the quality of the business and regulatory environment in the state. That is, firms may choose their location based not on actual stocks of infrastructure but instead on signals contained in the state expenditure levels.[36] To investigate further the role of physical infrastructure for location, we consider a different proxy for the availability of transport infrastructure in the state – the length of the highway network (HIGHWAY) – interacted with our previous proxy for the industry dependence on transport (TRANSPORT INTENSITY). Since this proxy is available only for the 1991-2000 period, we present in columns (5)-(8) of Table 8 the results from estimating the specifications in columns (1)-(4) on that reduced sample to verify whether the effects of the infrastructure proxies are sensitive to the sample period used. The estimates show that the positive effect of transport and the negative effect of electricity are maintained – albeit weaker in significance – in the 1991-2000 sample. Columns (9) and (10) of Table 8 show that the length of the highway network in a state is positively and significantly correlated with the spatial concentration of industries that have greater reliance on transport networks.[37] So state level expenditures do not seem to be used as signals about the quality of the regulatory regime. What the evidence suggests is that Indian firms respond differently to different types of infrastructure. In particular, firms may chose location based on the infrastructure for which they have less outside options and must rely on the state. Large plants in

[36] We thank Adnan Khan for suggesting this possibility.

[37] The same qualitative pattern is obtained when using a measure of the combined road and highway network in a state as the proxy for the availability of transport infrastructure though the effects are weaker.

India can and do often have private in-house electricity generation capabilities but they cannot circumvent the need to use public roads and highways.

7.2 Initial Conditions

Table 9 presents results from estimating variants of Eqs. (3) and (4) that allow the initial conditions in a location or in an industry to affect the employment share and new entry rates of manufacturing over time. Specifically, we fix either the location characteristics - in columns (1) and (2) - or the industry characteristics - in columns (3) and (4) - at their values in 1980 prior to the sample period. These specifications also help us address potential endogeneity problems as will be discussed in Section 7.3. A few interesting patterns emerge from these results. In the specification where state characteristics are held constant at their 1980 levels, the positive and significant effect of the comparative advantage factor related to skilled labor continues to hold: industries that use skilled labor intensively tend to concentrate in states that had larger abundance of skilled labor in 1980. Industries that rely more on contracts tend to concentrate in states that had better governance in 1980. One of the interaction terms linked to NEG mechanisms and Marshallian economies that showed counter-intuitive effects when using current state characteristics now has its expected positive sign in columns (1) and (2). The significant coefficient in column (2) indicates that new entry in industries that rely more intensively on manufacturing intermediate inputs is higher in states that had a larger manufacturing base in 1980.

Some of the results based on fixed industry characteristics in columns (3) and (4) are weaker. However, industries that used skilled labor intensively in 1980 tend to concentrate in states that have larger abundance of skilled labor over time in column (4) and moreover unskilled labor-intensive industries in 1980 also tend to concentrate in states that have larger unskilled labor abundance.

7.3 Discussion of Endogeneity

Although our paper does not claim to provide estimates of causal links between location determinants and location patterns but rather to establish robust correlations, it is important to discuss some potential simultaneity concerns with our specifications and the ways in which we address those. Studies whose focus is on examining the importance of the interaction terms linked to NEG mechanisms and to Marshallian economies for the clustering of manufacturing firms are particularly hampered by potential biases given the endogeneity of the linkages that form between clustered firms. Ellison and Glaeser (1997) argue that if a study estimates a positive coefficient for the number of firms or the total employment in an industry when explaining contemporaneous and even subsequent clustering, it is impossible to identify whether

firms are attracted by a location unobservable characteristics or whether they truly derive benefits from proximity to one another. However, in the case of our study we are interested in identifying the important correlates of spatial concentration and the entry of new firms in India, whether they originate in comparative advantage forces, NEG mechanisms or Marshallian economies, or aspects of the business environment and thus we are not so concerned with simultaneity problems.

Nevertheless, we pursue several strategies to mitigate the potential simultaneity between location characteristics and spatial concentration of and new firm entry into manufacturing in India and and location characteristics mediated by industry characteristics, and hence estimate robust correlations across those. First, all our specifications include one-year lagged location characteristics-industry characteristics interaction terms to mitigate the potential simultaneity and allow for the spatial effects to operate with a lag. Moreover, the results just discussed in Section 7.2 from specifications that fix either the location characteristics or the industry characteristics at their values at the beginning of the sample period also address endogeneity problems. The skilled labor coefficient estimate shown in columns (1) and (3) of Table 6 is intuitive and expected but could be upward biased if skilled workers - which tend to be more spatially mobile - were able to move to states with better economic opportunities linked to a higher concentration of manufacturing employment. The specifications where location characteristics are fixed at their values at beginning of the sample period values allow us to mitigate this problem.

Second, the inclusion in our specifications in Table 6 of a rich set of industry, state, and year fixed effects accounts for unobserved heterogeneity at all levels that could bias our coefficients. In particular, state fixed effects account for unobserved characteristics that improve a location‟s attractiveness related for instance to natural endowments, aspects of the regulatory framework, or even cultural issues. However, it is important to account for omitted factors at the industry-location level that may influence spatial concentration and are not captured by separate location fixed effects and industry fixed effects. Given the location-related content of some of India‟s industrial policies this is a particularly relevant concern. This is the most stringent specification that we consider and we present the results in Appendix Table 6. Note that due to the incidental parameters problem in the Tobit model with individual fixed effects – in our case an „individual‟ is an industry-location – we use instead OLS estimation for these specifications.[38] The estimates indicate that in this panel estimation framework the strong determinants of manufacturing location and new firm entry continue to be the business environment-related factors, both transport infrastructure and governance. The importance of skilled labor reduces to almost zero once we control for industry-state fixed effects and this confirms our intuition that location characteristics may evolve to respond to greater concentration leading to endogeneity problems: that is, as manufacturing concentration grows in a given state, skilled workers are attracted to that state.

[38] See Greene (2004) on the biases in the maximum likelihood estimates of fixed effects Tobit models.

Third, using the count of new firms in an industry-location addresses some of the endogeneity problems in the study of location determinants because for a new firm, location attributes – including industrial structure - are exogenous at the time of entry (Rosenthal and Strange 2003; Ghani et al., 2011; Jofre-Monseny et al., 2011). For example, the state of Bangalore in India has a very large concentration of information technology (IT) firms in part because of the availability of a local pool of skilled labor. As the concentration of IT firms in Bangalore rises, more skilled labor will be attracted to the state and thus the estimated effect of skilled labor on concentration could be biased. However, the entry of new IT firms in Bangalore is less likely to create skilled labor movements to Bangalore. Factors are more likely to move across locations based on the stock (existing clusters) of manufacturing concentration rather than the flow (new entry).

8. The role of policy reforms in spatial concentration and entry

8.1 Indian policy reforms and their expected effects

Our sample period covers India‟s major policy reform episode of 1991. Following a major balance of payments crisis in 1990 India was forced to take an adjustment loan from the International Monetary Fund which required the government to engage in substantial market-oriented reforms in the areas of industrial licensing, trade and foreign direct investment (FDI) policies.[39] The most significant de-licensing episode occurred in 1991 whereby licensing was removed on industries accounting for 84% of total manufacturing output. Average tariffs fell by 43 percentage points between 1990 and 1996, and non-tariff trade barriers were rationalized and scaled down. The FDI regime was liberalized after 1991, with foreign ownership rates of up to 51% being allowed for a group of industries and a relaxation of restrictions on brand names, remittances of dividends and the proportion of local content of output.

Conceptually, industrial de-licensing, trade and FDI liberalization may lead to more or less spatial concentration of manufacturing depending on several forces, as we discuss below. Regarding the impact of de-licensing, the industrial policy regime was willfully used to settle industry in backward, non-industrialized areas of India. Upon de-licensing, firms are expected to choose locations based on cost criteria: e.g., areas with good infrastructure, input networks, and an appropriately skilled labor force. Therefore, industrial de-licensing could be linked to an increase in spatial concentration. But at the same time, since the licensing regime created artificial industrial clusters in backward areas, as de-licensing occurred, these clusters would likely break up, leading to a decline in the spatial concentration of Indian manufacturing.

[39] The foreign exchange regime was also liberalized with the adoption of a flexible exchange rate regime in 1991. Chamarbagwalla and Sharma (2011), Fernandes and Sharma (2010), and Nataraj (2011) provide further details on the policy reforms.

Regarding the impact of trade liberalization, most NEG models predict that it leads to (i) increased importance of foreign demand (exports) with a potential lowering of the incentives for domestic firms to locate near domestic consumers that now represent a smaller share of sales, (ii) increased importance of foreign supply (imports) and a decline in the weight of backward and forward linkages with the potential to lower the incentives for domestic firms to locate near other firms for input-output linkages since imports now represent a larger share of supply to domestic consumers.[40] Whether these mechanisms result in higher or lower spatial concentration of manufacturing within the liberalizing country depends on the NEG model. Paluzie (2001) and Monfort (2000) show that trade liberalization favors spatial concentration of manufacturing as firms‟ incentives to locate in the periphery away from domestic competitors to be sheltered when serving the local market are reversed given that now competition in the periphery arises also from imports. Brülhart et al. (2004) and Crozet and Koenig (2004) show that trade liberalization fosters spatial concentration in the border region with better access to international markets: domestic firms are attracted to that region to reap the full benefits from improved access to foreign demand while domestic consumers agglomerate in that region to access imported goods. However, Krugman and Elizondo (1996) predict that trade liberalization favors spatial dispersion of manufacturing. The rationale for that prediction is that due to the increased importance of foreign demand and supply and the decline in that of backward and forward linkages, the actual location of domestic firms and consumers matters less and thus in the presence of congestion costs (assumed to be independent of trade openness) manufacturing disperses. The same prediction is obtained by Behrens (2007) in a model with endogenous competition effects.

Regarding the impact of FDI liberalization, if new foreign firms aim to supply the domestic market, they will tend to locate close to large markets according to NEG models. But if transport costs are low, then they are indifferent about their location since they can serve domestic customers from anywhere, possibly contributing to spatial dispersion. If new foreign firms aim to exploit production advantages (e.g., cheap labor) to target export markets, the optimal location choice is also unclear as they may either have an incentive to locate away from main industrial centers where wages are higher – leading to spatial dispersion – or they may prefer to locate near major centers to use a pool of skilled labor or be near ports or major borders – increasing spatial concentration. The degree of usage of domestic versus imported inputs can also affect foreign firms‟ location choices: if they rely heavily on domestic input-output linkages they will tend to locate close to suppliers, which may or may not be located in existing industrial clusters. FDI liberalization can also affect domestic firms‟ location choices: dispersion could result if they locate away from foreign firms to avoid competition in the domestic market but further concentration could result if instead they locate near foreign firms due to input-output linkages.

[40] See Brülhart (2011) for a review of the NEG literature on the spatial effects of international trade. NEG models assume that the location choices of firms and consumers are determined by opposing agglomeration and dispersion forces. The agglomeration (centripetal) forces result from the interaction between IRS, market size, trading costs, and/or backward and forward linkages among producers. The dispersion (centrifugal) forces can arise from the costs that firms face to reach an exogenously dispersed demand of immobile consumers (working in the agricultural sector) or from congestion costs (rent and commuting costs) associated with large industrial agglomerations.

Often new foreign firms are explicitly attracted to be part of clusters such as special economic zones. But at the same time many policy-makers believe that FDI can support the development of disadvantaged regions, which would result in a dispersion of manufacturing activity.

Ge (2009) shows that access to foreign trade and FDI is a driving force of unbalanced spatial distribution in China. In particular, industries dependent on foreign trade and FDI are more likely to locate in regions with better access to foreign markets. Sanguinetti and Martincus (2009) show that lower tariffs in Argentina were associated with the de-concentration of manufacturing activities away from the Buenos Aires region, which they interpret as indicative that with the opening to trade, demand and cost linkages weakened and agglomeration diseconomies such as high commuting costs or high land rents prevailed. Martincus (2010) shows more open industries in Brazil (in terms of trade volumes) tended to locate in states near the country's largest trading partner (Argentina) and this tendency increased with the trade 1990s‟ liberalization. Amiti and Javorcik (2008) show that both market access and supplier access were critical in determining the location of new FDI inflows across Chinese provinces.

8.2 Estimated effects of Indian policy reforms

While the Indian market-oriented policy reforms were implemented differentially across industries and over time, they were not implemented differentially across locations. It is important, however, to examine whether the relationship between spatial concentration or new firm entry and state and industry characteristics was influenced by the policy reforms. We follow three different approaches to address this possibility and present them in sequence below.

As a first approach, we estimate a variant of Eqs. (3) and (4) where the natural advantage factors, the NEG mechanisms and Marshallian economies, and the business environment factors enter by themselves and interacted with a dummy identifying the years post-reform i.e., after 1991. The results are presented in columns (1) and (2) of Table 10. By allowing the location determinants to have different effects in the post-reform period, this approach can potentially provide evidence on the channels through which policy reforms affected location patterns in India and inform on whether locational advantages became more or less important after the reforms.

The labor comparative advantage factor per se now has the expected positive and significant effect on concentration and entry but the coefficients for the post-reform period are either insignificant – for employment shares – or negative and significant – for entry rates. In Table 9 we showed that when industry characteristics are held constant at their 1980 values, the unskilled labor comparative advantage factor also has an expected positive effect on employment shares. These two findings suggest that the counter-intuitive negative effect of the unskilled labor abundance in Table 6 may be explained by a fall in unskilled labor intensity in Indian industries over time particularly post-reform, which would be the prediction from traditional trade models for an unskilled labor-abundant country such as India opening to trade. Indeed, the average

across industries of our unskilled labor intensity ratio exhibits a clear decline over time during our sample period, as shown in Appendix Table 1. Moreover, Indian firms seem to be substituting away from labor and towards capital following the policy reforms, as is shown by Chamarbagwalla and Sharma (2011).

Our estimates in columns (1) and (2) of Table 10 show that the NEG mechanisms and the Marshallian economies have an insignificant effect on employment shares and entry rates after the Indian policy reforms. This could be understood in the context of the NEG models discussed above that suggest that trade reform leads to a weakening of the market access and input-output linkages effects in determining firm location. However, note that the effects are also absent or negative prior to the reforms, as was the case in Tables 6 through 8.

The estimates in columns (1) and (2) also show that transport infrastructure per se has a positive and significant effect in the employment shares and the entry rates regressions, but the coefficient is negative and significant for the post-reform period. This would suggest that after reforms firms in more transport-dependent industries are significantly more likely to locate in states with smaller expenditures on transport infrastructure. If we believe that Indian firms‟ location decisions can be more responsive to economic factors post-reforms, this result would appear very counter-intuitive. However, this result can be rationalized considering the evidence discussed below that post-reform manufacturing tends to concentrate relative more in backward states, as these are the states with poorer infrastructure.

As a second approach, we estimate a variant of Eqs. (3) and (4) where the natural advantage factors, the NEG mechanisms and Marshallian economies, and the business environment factors enter by themselves and interacted with each of three industry-level policy variables: a measure of the proportion of employment that is de-licensed in each industry and year, the negative of a tariff measure (whose increase indicates trade liberalization), and a measure of the proportion of employment exposed to FDI in each industry and year.[41] Further details on the measures are provided in Table 1. Since the specifications include an extremely large number of interactions, Table 11 presents only the policy-related interaction terms that have a significant effect on either employment shares or entry rates, in addition to the baseline interaction factors.[42]

The estimates in column (1) of Table 11 show that the unskilled labor comparative advantage factor has a positive and significant effect of the employment shares of industries where more of the employment was de-licensed or exposed to FDI. However, the employment shares of industries experiencing stronger tariff reductions are substantially lower in more unskilled labor-abundant states. One possible interpretation is that the industries most affected by the trade reforms in India were the relatively more capital-intensive industries and would not respond to unskilled labor abundance of a location. The skilled labor comparative advantage factor has a stronger positive effect in fostering new firm entry in industries experiencing stronger FDI

[41] The specifications include also the industry-level policy measures per se.
[42] The full specification is available from the authors upon request.

liberalization. This finding is actually expected given the well established evidence of higher skill intensity of foreign firms.

Interestingly, column (1) of Table 11 shows that, as FDI is liberalized, industries with higher economies of scale tend to concentrate in states with larger market potential and industries that rely more intensively on manufacturing intermediate inputs tend to concentrate and exhibit higher entry rates in states with a larger manufacturing base. This implies that access to local demand and to local suppliers are more important determinants of firm location in FDI-liberalized industries, suggesting that foreign firms are serving local markets and interacting with domestic firms through input-output linkages. However, new entry rates are lower in states with a larger manufacturing base for industries that rely more intensively on manufacturing inputs as trade is liberalized. An interpretation for this finding is that since reductions in output tariffs were accompanied by substantial input tariff reductions in India – as shown by Harrison et al. (2011) – this implies that the importance of domestic input-output linkages declines.

Column (2) of Table 11 shows that the transport infrastructure interaction term has a negative and significant effect on entry rates in industries experiencing more de-licensing. This finding is in line with the evidence from Table 10 that after the reforms, firms in more transport-dependent industries tend to locate in states with smaller expenditures on transport infrastructure. However, the positive and significant effect on new entry of the infrastructure interaction term in FDI-liberalized industries suggests that such pattern was not verified across the board for all industries.

As a third approach, we estimate a variant of Eqs. (3) and (4) where in addition to the natural advantage factors, the NEG mechanisms and Marshallian economies, and the business environment factors we add interaction terms that multiply an industry-level policy measure by a state characteristic that captures the state-specific impact of a policy reform.[43] Policy reforms may favor concentration or new entry in states with particular characteristics, namely those associated with lower distance to foreign markets, lower costs of accessing domestic suppliers, or lower costs of doing business. For example, trade liberalization may have had differential effects across states depending on their trade costs to connect to foreign markets. The corresponding results will be presented in columns (1) and (2) of Table 12. We include interactions between the de-licensing variable and (i) the number of seaports in each state proxying for the access to foreign markets, (ii) the share of total GDP in the state that originates in manufacturing proxying for the size of the manufacturing base, and (iii) a dummy for backward states that accounts for the fact that an important goal of the licensing regime was to create industrial bases in these states.[44] For both the tariff variable and the FDI liberalization variable, we include interactions between each of those and (i) the number of seaports in each state and (ii) the share of the total GDP in the state that originates in manufacturing. Columns (3)

[43] The specifications include also the industry-level policy measures per se.

[44] We use the list of "Industrially Backward States" as notified by the Ministry of Industry, Government of India during the 1980s: Assam, Himachal Pradesh, Madhya Pradesh, Orissa, Rajasthan, Uttar Pradesh, and West Bengal.

and (4) of Table 12 add to the specifications in columns (1) and (2) an additional set of interactions between a time trend and those three state characteristics: (i) the number of seaports in each state, (ii) the share of total GDP in the state that originates in manufacturing, and (iii) the dummy for backward states.

The estimates in column (1) of Table 12 show a significantly lower concentration of more de-licensed industries in states with better foreign market access through ports but a higher concentration in states with a larger manufacturing base. This finding suggests a potential fear of stronger import competition and an increased importance of domestic forward and backward linkages for firms in de-licensed industries. The estimates also show significantly lower employment shares of de-licensed industries in backward states, which is exactly what is expected of firms that are no longer constrained by law to locate in disadvantaged states as was the case while the industrial licensing regime was in place.[45] Note that the interaction terms with the de-licensing have no effect on the entry of new firms. This finding mirrors the evidence in prior studies on India that the de-licensing regime mostly affected incumbent firms and not the entry of new firms (Bollard et al. 2011; Sharma, 2006).

Also, columns (1) and (2) point to a significantly lower concentration and lower entry rates in industries experiencing more substantive trade liberalization in states with a larger manufacturing base. This finding can be rationalized by some of the NEG predictions discussed in Section 8.1 whereby the role of domestic forward and backward linkages for concentration is weakened as foreign markets gain importance upon reductions in trade costs.

Finally, the estimates in columns (1) and (2) suggest that FDI-liberalized industries tend to concentrate more and exhibit higher entry rates in states with better foreign market access through ports but tend to locate away from states with a larger manufacturing base. These findings point to the importance of access to imported inputs for firms in FDI-liberalized industries.[46]

Columns (3) and (4) of Table 12 show some interesting patterns for the interaction of time trends with state characteristics. Entry rates into manufacturing increase over the sample period in backward states such as Himachal Pradesh, Rajasthan, Uttar Pradesh, and in states with better foreign market access. In contrast, both employment shares and entry rates decline over time in states with larger manufacturing bases. The estimated coefficients on the interactions with the de-licensing and the FDI liberalization variables in columns (3) and (4) are qualitatively similar to those in columns (1) and (2) while the coefficients on the interactions with the trade liberalization variable are now insignificant.

[45] See Sharma (2006) for additional details on the industrial licensing policy.

[46] The location in states with more ports for industries with more FDI liberalization could also be linked to the increased importance of exports for those industries. However, in contrast to China, serving as an export platform is not the major motivation for FDI into India.

9. Conclusion

In this paper we use a long panel of 180 industries across 16 states in India over a 27-year time span to investigate the determinants of spatial concentration and entry into manufacturing. To our knowledge we are the first to use such a long time span of ASI data on formal manufacturing, and to have assembled a detailed dataset of state-level characteristics that allows us to control for important determinants of manufacturing firms‟ location choices.

Our estimates show strong positive effects of comparative advantage linked to skilled labor as expected, indicating that industries that rely more on skilled labor concentrate and exhibit higher entry rates in states with a more abundant skilled labor force. In contrast, we find a negative and significant counter-intuitive effect of unskilled labor abundance, even after controlling for possible congestion forces. These two findings combined suggest that skilled labor is the least substitutable input in the production process and thus firms make their location depend on the availability of skilled labor in a state. This result is consistent with the decline in labor intensity in Indian manufacturing and the evidence of skill-biased technological change as a result of market-oriented reforms (Chamarbagwala and Sharma, 2011). An important policy implication of this result is that employment generation in the formal manufacturing sector can be enhanced by increasing the skills of the labor force (via increased investment in education).

A consistent finding of our analysis is that better business environment in terms of better infrastructure or better governance increases manufacturing employment shares and new entry across Indian states for industries that are more dependent on infrastructure or on contracts. However, different types of infrastructure have differential effects. Transport infrastructure which is less substitutable by firms - state roads or highways need to be relied upon - has a strong positive effect on concentration whereas electricity infrastructure has a negative effect. Our results with respect to electricity provision (measured by the length of electric transmission lines in a state) are particularly interesting in this regard because large manufacturing plants in India do have in-house power generation facilities. This raises the policy issue of whether these plants could become more productive if their states became a reliable, high quality supplier of electricity, this process would presumably free up some of the plants‟ productive assets which were being used for electricity generation.

Our analysis finds significant, though heterogeneous effects of the Indian market-oriented reforms of the 1980s and 1990s. After the 1991 policy reforms the importance of transport infrastructure falls (in some cases becomes negative) in explaining concentration and new entry. However, there is industry heterogeneity in the sense that FDI-liberalized industries continue to exhibit higher concentration and entry in states with better transport infrastructure after 1991. Further, our estimates show that the negative effect of unskilled labor abundance is verified only in the post-reform period but is entirely driven by the industries facing stronger import competition which may be more capital-intensive and thus less likely to rely on unskilled labor abundance for their location decisions. Finally, we show that state characteristics associated with

lower distance to foreign markets, lower costs of accessing domestic suppliers, or lower costs of doing business matter for the impact of policy reforms on concentration and new entry. In particular, we find that tariff-liberalized and FDI-liberalized industries tend to locate in states with lower a manufacturing base suggesting the weakening of domestic input-output linkages and increased importance of access to imported inputs in the wake of the 1991 reforms. These results reinforce the importance of controlling for various market-oriented reforms separately, rather than lumping them together. The mechanisms through which the reforms affect concentration and entry are varied and nuanced.

References

Acs, Z. J., S. Desai, and J. Hessels. (2008). "Entrepreneurship, Economic Development, and Institutions," *Small Business Economics* 31, 219–34.

Ahsan, A. and C. Pages (2009). "Are all Labor Regulations Equal? Evidence from Indian Manufacturing," *Journal of Comparative Economics* 37, 62-75.

Ahsan, R. (2010). Input Tariffs, Speed of Contract Enforcement, and the Productivity of Firms in India. Syracuse University, Mimeo.

Amiti, M. and B. Javorcik (2008). "Trade Costs and Location of Foreign Firms in China," *Journal of Development Economics* 85, 129-149.

Arrow, K. (1962). "The Economic Implications of Learning by Doing," *Review of Economic Studies* 29, 155-173.

Audrestsch, D. (2007). "Entrepreneurship Capital and Economic Growth," *Oxford Review of Economic Policy* 23, 63-78.

Audretsch, D. and M. Keilbach (2005). "Entrepreneurship Capital and Regional Growth," *The Annals of Regional Science* 39, 457-469.

Baumol, W. (2010). *The Microtheory of Innovative Entrepreneurship*. Princeton, NJ: Princeton University Press.

Baumol, W. and R. Strom (2007). "Entrepreneurship and Economic Growth," *Strategic Entrepreneurship Journal* 1, 233-37.

Bollard, A., Klenow, P., and G. Sharma (2010). "India's Mysterious Manufacturing Miracle," Mimeo.

Brülhart, M. (2011). "The Spatial Effects of Trade Openness: A Survey," *Review of World Economics* 147, 59-83.

Brülhart, M., Crozet, M., and P. Koenig (2004). "Enlargement and the EU Periphery: The Impact of Changing Market Potential," *The World Economy* 27, 853-875.

Brülhart, M. and F. Sbergami (2009). "Agglomeration and Growth: Cross-Country Evidence," *Journal of Urban Economics* 65, 48-63.

Chamarbagwalla, R. and G. Sharma (2011). "Industrial Deregulation, Trade Liberalization, and Skill Upgrading in India," *Journal of Development Economics* 96 314-336.

Chemin, M. (2010). "Does Court Speed Shape Economic Activity? Evidence from a Judicial Reform in India," *Journal of Law, Economics, and Organization* forthcoming.

Ciccone, A. and R. Hall (1996). "Productivity and Density of Economic Activity," *American Economic Review* 86, 54-70.

Cingano, F. and F. Schivardi (2004). "Identifying the Sources of Local Productivity Growth," *Journal of the European Economic Association* 2, 720-742.

Combes, P. and H. Overman (2004). "Chapter 64 The Spatial Distribution of Economic activities in the European Union," in *Handbook of Regional and Urban Economics: Cities and Geography*, Volume 4, ed. by J. Henderson. and J. Thisse, Elsevier, 2845-2910.

Deichmann, U., Lall, S., Redding, S. and A. Venables (2008). "Industrial Location in Developing Countries," *World Bank Research Observer* 23, 219-246.

Dejardin, M. (2011) "Linking Net Entry to Regional Economic Growth," *Small Business Economics* 36, 443-460.

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

Ellison, G. and E. Glaeser (1999). "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review* 89, 311-316.

Fernandes, A. and G. Sharma (2010). "Together we stand? Agglomeration in Indian Manufacturing," World Bank mimeo.

Fujita, M., P. Krugman, and A. Venables (1999). *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge MA, MIT Press.

Harrison, A., Martin, L., and S. Nataraj (2011). "Learning versus Stealing: How Important are Market-Share Reallocations to India"s Productivity Growth?," NBER Working Paper No. 16733.

Ge, Y. (2009). "Globalization and Industry Agglomeration in China," *World Development* 37, 550-559.

Ghani, E, Kerr, W., and S. O"Connell (2011). "Spatial Determinants of Entrepreneurship in India," NBER Working Paper No. 17514.

Glaeser, E. and W. Kerr (2009). "Local Industrial Conditions and Entrepreneurship: How much of the Spatial Distribution Can We Explain?" *Journal of Economics and Management Strategy* 18, 623-663.

Greene, W. (2004). "Fixed Effects and Bias Due to the Incidental Parameters Problem in the Tobit Model," *Econometric Reviews* 23, 125-147.

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

Jofre-Monseny, J., Marín-López, R., and E. Viladecans-Marsal (2011). "The Mechanisms of Agglomeration: Evidence from the Effect of Inter-Industry Relations on the Location of New Firms," *Journal of Urban Economics* 70, 61-74.

Klapper, L. and I. Love (2011). "Entrepreneurship and Development: the Role of Information Asymmetries," *World Bank Economic Review* 25, 448-455.

Krugman, P. and R. Elizondo (1996). "Trade Policy and the Third World Metropolis," *Journal of Development Economics* 49, 137-150.

Lall, S. and S. Chakravorty (2005). "Industrial Location and Spatial Inequality: Theory and Evidence from India," *Review of Development Economics* 9, 47-68.

Li, H., Yang, Z., Yao, X., and J. Zhang (2009). "Entrepreneurship and Growth: Evidence from China," Discussion Paper 00022, Chinese University of Hong Kong, Department of Economics.

Marshall, A. (1920). *Principles of Economics*. London, Macmillan.

Martincus, C. (2010). "Spatial Effects of Trade Policy: Evidence from Brazil," *Journal of Regional Science* 50, 541-569.

Mayer, K. and T. Mayer (2004). "The Empirics and Agglomeration and Trade," Chapter 59 in *Handbook of Regional and Urban Economics: Cities and Geography*, volume 4, ed. by J. Henderson. and J. Thisse, Elsevier, 2609-2669.

Midelfart-Knarvik, K., Overman, H. and A. Venables (2000). "Comparative Advantage and Economic Geography: Estimating the Location of Production in the EU." CEPR Research Discussion Paper 2618.

Monfort, P. and R. Nicolini (2000). "Regional Convergence and International Integration," *Journal of Urban Economics* 48, 286-306.

Mukim, M. (2011). "Industry and the Urge to Cluster: a Study of the Informal Sector in India. Spatial Economics Research Centre, Discussion Paper 72, London School of Economics.

Mukim, M. and P. Nunnenkamp. (2010). "The Location Choices of Foreign Investors: A District-level Analysis," Kiel Working Paper, no. 1628, Kiel Institute for the World Economy.

Michielsen, T. (2011). "The Distribution of Energy-Intensive Sectors in the U.S.," CentERWorking paper 2011-075, Tilburg University.

Paluzie, E. (2011). "Trade Policy and Regional Inequalities," *Papers in Regional Science* 80:67-85.

Rajan, R., and L. Zingales (1998). "Financial Dependence and Growth," *American Economic Review* 88, 559-86.

Romer, P. (1986). "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94, 1002-1037.

Rosenthal, S. and W. Strange (2001). "The Determinants of Agglomeration," *Journal of Urban Economics* 50, 191-229.

Rothenberg, A. (2011). "Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia"s Highways," University of California at Berkeley mimeo.

Samila, S. and O. Sorenson (2011). "Venture Capital, Entrepreneurship, and Economic Growth," *Review of Economics and Statistics* 93, 338–349.

Sanguinetti, P. and C. Martincus (2009). "Tariffs and Manufacturing Location in Argentina," *Regional Science and Urban Economics* 39, 155-167.

Sen, A. and T. Jamasb (2010). "The Economic Effects of Electricity Deregulation: An Empirical Analysis of Indian States," Electricity Policy Research Group Working Paper 1001, Cambridge Working Paper in Economics 1005, University of Cambridge.

Sivadasan, J. (2009). "Barriers to Competition and Productivity: Evidence from India," The *B.E. Jounal of Economic Analysis and Policy* 9(1 (Advances)).

Sharma, G. (2006). "Competing or Collaborating Siblings: Industrial and Trade policies in India," University of Missouri Working Paper 06-10.

Sharma, G. (2009). "Crime and Inequality in India," University of Missouri mimeo.

Table 1. Variables

Variable Name	Description	Definition	Yearly Coverage	Source
State Characteristics *(s designates a state)*				
UNSKABUN	Unskilled labor Abundance	$\frac{National\ unskilled\ wage\ rate_t}{Unskilled\ wage\ rate_{st}}$	1985-2007	ASI, authors‟ calculations
SKABUN	Skilled labor Abundance	$\frac{Number\ of\ skilled\ workers_{st}}{Total\ number\ of\ workers_{st}}$	1985-2007	ASI, authors‟ calculations
MKT	Market potential	Per capita state domestic product at constant prices	1985-2007	SDP Dataset, EPW Foundation
BASE	Manufacturing base	State domestic product from manufacturing sector	1985-2007	SDP Dataset, EPW Foundation
TRANSPORT	Infrastructure	State expenditure on transport, storage and communication	1985-2007	SDP Dataset, EPW Foundation
GOVERN	Governance quality	(-1)*Per capita murder rate in a state	1985-2007	National Crime Records Bureau Annual Publications
HIGHWAY	Highway infrastructure	Number of kilometers of highways in a state	1985-2005	Sen and Jamasb (2010)
TRANSMISSION	Electricity infrastructure	Length of electricity transmission and distribution lines in a state	1985-2005	Sen and Jamasb (2010)
ELECBASE	Manufacturing base in terms of electricity	Number of kilowatt hours of electricity consumed by the manufacturing sector divided by total population in a state	1985-2005	Sen and Jamasb (2010)
ELECMKT	Market potential in terms of electricity	Number of kilowatt hours of electricity consumed divided by total population in a state	1985-2005	Sen and Jamasb (2010)
Industry Characteristics *(j designates an industry)*				
SKILLINTENSITY	Intensity of use of skilled labor	$\frac{Number\ of\ skilled\ workers_{jt}}{Number\ of\ workers_{jt}}$	1985-2007	ASI, authors‟ calculations

UNSKILLED INTENSITY	Intensity of use of unskilled labor	$\frac{Wage\ bill\ of\ unskilled\ workers_{jt}}{Sales_{jt}}$	1985-2007	ASI, authors" calculations
IRS	Returns to scale	Average capital stock per plant in an industry	1985-2007	ASI, authors" calculations
MATS	Intensity of use of intermediate inputs	$Materials\ purchases_{jt} / Sales_{jt}$	1985-2007	ASI, authors" calculations
TRANSPORT INTENSITY	Intensity of use of transport	$Value\ of\ inventories_{jt} / Sales_{jt}$	1985-2007	ASI, authors" calculations
ELECTRIC	Intensity of use of electricity	$Electricity\ purchases_{jt} / Sales_{jt}$	1985-2007	ASI, authors" calculations
FDI	FDI Deregulation	Proportion of output in the 3-digit industry that is de-regulated (weighted sum across 4-digit industries)	1985-2007	Sivadasan (2009), author calculations
DEL	De-licensing	Proportion of output in the 3-digit industry that is de-licensed (weighted sum across 4-digit industries)	1985-2007	Sharma (2006), author calculations
TAR	Trade liberalization	-1*Tariff rate in the 3-digit industry	1985-2007	Harrison, Martin & Natraj (2011)

Table 2. Shares of States in Manufacturing Employment in Selected Sample Years

	State Share of Manufacturing Employment (%)					Average Population Share (%)	Employment in 2007 (in hundred thousands)
State	**1985**	**1989**	**1994**	**2000**	**2005**		
Maharastra (MAH)	16.70	15.44	14.87	14.68	14.64	10.80	251
Tamil Nadu (TN)	12.14	12.13	13.83	13.26	14.34	7.32	224
Andra Pradesh (AP)	9.86	11.28	11.87	11.92	11.30	8.75	177
Gujarat (GUJ)	10.16	9.89	9.51	9.23	9.94	5.67	154
Uttar Pradesh (UP)	8.92	9.78	8.62	8.01	9.05	14.89	141
Top 5 sum	57.78	58.53	58.70	57.10	59.27	47.44	
West Bengal (WB)	10.53	8.83	8.09	7.55	5.77	9.12	126
Karnataka (KAR)	5.07	5.42	5.48	5.55	6.45	6.05	91
Punjab (PUN)	4.17	4.71	4.46	5.20	5.56	2.74	77
Madhya Pradesh (MP)	4.63	4.87	4.87	4.71	3.98	6.76	74
Haryana (HAR)	3.31	3.21	3.47	4.42	5.02	2.30	64
Top 10 sum	85.49	85.57	85.07	84.52	86.06	74.42	

Table 3. Average Employment Shares and Entry Rates per State in Selected Sample Years

Panel A. Average Share of State Employment in Existing Industries - Ignoring Zeros (%)

Year	Andra Pradesh (AP)	Assam (ASS)	Bihar (BIH)	Gujarat (GUJ)	Haryana (HAR)	Himachal Pradesh (HP)	Karnataka (KAR)	Kerala (KER)	Maharastra (MAH)	Madhya Pradesh (MP)	Orissa (ORS)	Punjab (PUN)	Rajasthan (RAJ)	Tamil Nadu (TN)	Uttar Pradesh (UP)	West Bengal (WB)
1985	0.0696	0.0100	0.0347	0.0708	0.0276	0.0022	0.0540	0.0253	0.1659	0.0257	0.0102	0.0328	0.0211	0.1128	0.0637	0.0822
1989	0.0696	0.0104	0.0366	0.0754	0.0288	0.0041	0.0568	0.0250	0.1528	0.0296	0.0121	0.0368	0.0254	0.1122	0.0884	0.0885
1994	0.0751	0.0108	0.0294	0.0745	0.0350	0.0047	0.0480	0.0264	0.1530	0.0297	0.0134	0.0346	0.0265	0.1379	0.0895	0.0750
2000	0.0651	0.0101	0.0339	0.0748	0.0413	0.0083	0.0621	0.0357	0.1409	0.0291	0.0152	0.0334	0.0330	0.0925	0.0744	0.0534
2005	0.0703	0.0095	0.0212	0.0913	0.0402	0.0079	0.0739	0.0317	0.1247	0.0283	0.0128	0.0373	0.0408	0.1045	0.0932	0.0486

Panel B. Average Share of State Employment in All Industries - Including Zeros (%)

Year	Andra Pradesh (AP)	Assam (ASS)	Bihar (BIH)	Gujarat (GUJ)	Haryana (HAR)	Himachal Pradesh (HP)	Karnataka (KAR)	Kerala (KER)	Maharastra (MAH)	Madhya Pradesh (MP)	Orissa (ORS)	Punjab (PUN)	Rajasthan (RAJ)	Tamil Nadu (TN)	Uttar Pradesh (UP)	West Bengal (WB)
1985	0.0943	0.0269	0.0562	0.0982	0.0471	0.0068	0.0760	0.0421	0.2123	0.0413	0.0208	0.0595	0.0379	0.1485	0.0890	0.1140
1989	0.0910	0.0254	0.0577	0.0985	0.0471	0.0109	0.0732	0.0394	0.1915	0.0440	0.0206	0.0607	0.0387	0.1446	0.1115	0.1141
1994	0.0942	0.0242	0.0476	0.0988	0.0520	0.0118	0.0636	0.0383	0.1892	0.0435	0.0227	0.0551	0.0376	0.1728	0.1137	0.0980
2000	0.1104	0.0430	0.0795	0.1291	0.0796	0.0261	0.1024	0.0752	0.2097	0.0620	0.0434	0.0744	0.0617	0.1511	0.1204	0.1040
2005	0.1005	0.0223	0.0370	0.1247	0.0624	0.0150	0.0972	0.0517	0.1630	0.0414	0.0253	0.0615	0.0607	0.1396	0.1263	0.0717

Panel C. Entry Rate = Share of New Plants in States' Employment - Ignoring Zeros (%)

Year	Andra Pradesh (AP)	Assam (ASS)	Bihar (BIH)	Gujarat (GUJ)	Haryana (HAR)	Himachal Pradesh (HP)	Karnataka (KAR)	Kerala (KER)	Maharastra (MAH)	Madhya Pradesh (MP)	Orissa (ORS)	Punjab (PUN)	Rajasthan (RAJ)	Tamil Nadu (TN)	Uttar Pradesh (UP)	West Bengal (WB)
1985	0.1497	0.0648	0.0676	0.1332	0.0961	0.0998	0.1209	0.0673	0.0872	0.1009	0.1109	0.1108	0.1074	0.1607	0.2148	0.0463
1989	0.1593	0.0603	0.0686	0.1282	0.0570	0.0732	0.1005	0.0773	0.0945	0.1469	0.1059	0.0860	0.1067	0.1347	0.2996	0.0643
1994	0.1237	0.0617	0.0498	0.1131	0.1192	0.0489	0.1065	0.0923	0.0962	0.0838	0.0821	0.0518	0.1517	0.1516	0.1508	0.0496
2000	0.1100	0.0418	0.0580	0.0602	0.0606	0.1005	0.1040	0.0623	0.0791	0.0639	0.0571	0.0531	0.0891	0.0779	0.1020	0.0438
2005	0.1436	0.1623	0.0952	0.1159	0.0717	0.2074	0.1661	0.0770	0.1015	0.1000	0.0471	0.0722	0.1101	0.1385	0.1718	0.0637

Panel D. Entry Rate = Share of New Plants in States' Employment - Including Zeros (%)

Year	Andra Pradesh (AP)	Assam (ASS)	Bihar (BIH)	Gujarat (GUJ)	Haryana (HAR)	Himachal Pradesh (HP)	Karnataka (KAR)	Kerala (KER)	Maharastra (MAH)	Madhya Pradesh (MP)	Orissa (ORS)	Punjab (PUN)	Rajasthan (RAJ)	Tamil Nadu (TN)	Uttar Pradesh (UP)	West Bengal (WB)
1985	0.2029	0.1745	0.1095	0.1846	0.1644	0.3095	0.1702	0.1120	0.1115	0.1619	0.2256	0.2007	0.1927	0.2116	0.3001	0.0641
1989	0.2082	0.1472	0.1083	0.1676	0.0931	0.1942	0.1295	0.1219	0.1185	0.2185	0.1812	0.1417	0.1627	0.1736	0.3781	0.0829
1994	0.1551	0.1377	0.0807	0.1500	0.1774	0.1242	0.1413	0.1341	0.1190	0.1227	0.1391	0.0824	0.2152	0.1900	0.1917	0.0648
2000	0.1863	0.1779	0.1360	0.1039	0.1167	0.3172	0.1714	0.1311	0.1177	0.1360	0.1633	0.1185	0.1664	0.1273	0.1652	0.0853
2005	0.2053	0.3807	0.1659	0.1583	0.1111	0.3953	0.2187	0.1258	0.1327	0.1465	0.0926	0.1190	0.1638	0.1850	0.2329	0.0940

Table 4. Proportion of Empty Industry Cells per State Over Time - Employment

	% of Empty Cells in Each State (Maximum Possible = 180)															
Year	**AP**	**ASS**	**BIH**	**GUJ**	**HAR**	**HP**	**KAR**	**KER**	**MAH**	**MP**	**ORS**	**PUN**	**RAJ**	**TN**	**UP**	**WB**
1984	26.78	68.31	41.53	30.05	45.90	71.04	27.32	47.54	22.95	38.80	57.38	45.90	44.26	28.42	30.05	28.42
1985	26.23	63.93	40.44	27.87	41.53	68.31	28.96	40.44	21.86	38.80	52.46	44.81	44.26	24.04	28.96	28.42
1986	28.96	68.85	42.62	28.96	41.53	69.40	28.96	46.99	24.59	40.44	54.10	48.63	45.36	26.78	29.51	32.79
1987	26.23	64.48	40.98	26.78	40.44	67.76	27.87	38.25	21.31	33.33	46.99	43.72	35.52	22.40	25.68	26.23
1988	25.68	62.30	40.44	28.96	39.89	66.12	27.32	37.16	24.04	33.88	42.62	42.08	39.34	25.68	27.32	26.23
1989	23.50	59.02	37.16	23.50	38.80	62.30	22.40	36.61	20.22	32.79	42.62	39.89	34.43	22.40	21.31	22.40
1990	22.95	63.39	40.44	23.50	37.70	60.66	25.68	33.88	20.77	32.79	44.26	40.44	33.33	22.95	21.31	25.14
1991	24.59	60.11	39.34	21.86	36.61	58.47	24.04	34.97	20.22	32.79	42.62	37.70	32.79	24.04	19.67	24.59
Avg. 1984-1991	25.61	63.80	40.37	26.43	40.30	65.51	26.57	39.48	21.99	35.45	47.88	42.90	38.66	24.59	25.48	26.78
1992	21.86	62.30	39.89	23.50	35.52	63.39	25.14	34.43	20.22	29.51	43.72	38.80	34.43	20.77	19.67	24.59
1993	22.40	59.56	36.07	24.59	36.07	61.20	22.40	34.97	18.58	30.60	42.08	38.25	36.07	21.86	20.77	24.04
1994	20.22	57.92	38.80	24.59	33.33	61.20	25.14	31.15	19.13	32.24	40.98	37.16	30.60	20.77	21.31	23.50
1996	15.30	56.28	39.89	16.94	29.51	57.92	19.67	29.51	12.57	26.78	40.98	32.79	26.78	15.30	18.58	22.95
1997	16.94	62.30	43.17	21.31	37.16	56.28	22.95	29.51	16.39	31.69	48.09	38.80	31.15	18.58	20.22	27.32
1998	32.79	71.58	55.74	38.80	43.72	69.95	36.61	45.90	31.69	40.44	55.74	46.45	39.34	30.60	35.52	41.53
1999	40.44	77.05	57.92	43.72	52.46	71.58	43.17	55.19	36.07	50.27	60.66	54.10	47.54	40.98	38.80	43.72
2000	40.98	76.50	57.38	42.08	48.09	68.31	39.34	52.46	32.79	53.01	65.03	55.19	46.45	38.80	38.25	48.63
Avg. 1992-2000	26.37	65.44	46.11	29.44	39.48	63.73	29.30	39.14	23.43	36.82	49.66	42.69	36.54	25.96	26.64	32.04
2001	38.25	73.77	56.83	43.17	47.54	68.31	39.34	53.01	31.15	45.90	61.75	52.46	52.46	34.97	36.61	42.62
2002	26.78	63.39	42.08	30.60	37.70	53.55	28.42	40.98	25.14	33.33	49.18	43.72	34.43	27.32	26.23	32.24
2003	38.80	72.13	56.28	39.34	44.26	66.67	37.70	53.01	32.79	43.72	60.66	51.91	45.36	34.43	34.97	42.08
2004	32.79	71.04	55.19	31.15	45.36	64.48	34.97	49.73	29.51	42.08	60.11	46.45	42.62	30.05	32.79	38.80
2005	30.05	57.38	42.62	26.78	35.52	47.54	24.04	38.80	23.50	31.69	49.18	39.34	32.79	25.14	26.23	32.24
2006	27.32	55.74	44.26	30.60	33.88	51.37	28.42	39.34	22.95	33.33	49.73	36.61	31.69	26.78	22.95	29.51
2007	26.23	56.83	45.90	30.05	28.96	44.26	26.78	41.53	26.78	31.15	49.73	38.80	33.88	30.05	22.40	29.51
Avg. 2001-2007	31.46	64.32	49.02	33.10	39.03	56.60	31.38	45.20	27.40	37.31	54.33	44.18	39.03	29.82	28.88	35.28

Table 5. Proportion of Empty Industry Cells per State Over Time – Entry

	% of Empty Cells in Each State (Maximum Possible = 180)															
Year	**AP**	**ASS**	**BIH**	**GUJ**	**HAR**	**HP**	**KAR**	**KER**	**MAH**	**MP**	**ORS**	**PUN**	**RAJ**	**TN**	**UP**	**WB**
1984	46.67	90.56	75.00	45.00	71.67	90.56	50.56	76.67	38.33	73.89	77.22	65.56	63.89	38.89	45.56	66.67
1985	47.78	87.22	71.67	46.11	68.33	85.00	48.33	74.44	41.11	65.00	71.67	67.22	66.11	38.89	41.11	62.22
1986	48.33	85.56	73.33	45.00	71.67	91.67	48.89	76.11	40.56	68.33	76.11	67.22	71.67	40.56	40.00	66.67
1987	43.33	85.56	69.44	43.89	70.00	93.33	45.56	71.11	39.44	61.11	69.44	65.00	65.00	38.33	37.78	60.56
1988	42.22	83.33	76.11	46.67	71.11	90.56	48.33	68.89	43.33	54.44	70.00	65.00	66.67	42.22	33.33	58.33
1989	40.56	85.56	74.44	41.11	72.22	88.33	49.44	67.22	37.78	53.89	71.11	62.78	64.44	38.89	29.44	62.78
1990	43.33	85.00	75.56	42.22	76.67	89.44	50.00	65.56	35.56	55.00	67.78	64.44	57.22	39.44	29.44	61.11
1991	43.33	83.89	73.89	41.67	73.89	86.11	50.56	66.11	40.00	56.67	70.56	65.56	57.78	38.89	30.00	63.33
Avg. 1984-1991	44.44	85.83	73.68	43.96	71.94	89.38	48.96	70.76	39.51	61.04	71.74	65.35	64.10	39.51	35.83	62.71
1992	38.33	87.78	72.22	45.00	70.00	90.00	47.78	63.33	38.89	57.78	73.33	65.56	61.67	35.00	34.44	63.33
1993	39.44	88.33	74.44	45.56	60.56	88.89	53.33	66.67	35.00	63.33	71.11	66.11	60.56	40.00	33.33	63.89
1994	41.11	87.78	77.22	42.78	57.22	87.22	48.89	65.00	37.22	63.89	71.11	64.44	54.44	32.22	37.22	62.78
1996	45.56	81.67	80.00	33.89	50.56	80.56	45.00	53.89	37.22	56.11	74.44	57.78	55.00	36.67	37.78	71.11
1997	65.56	91.67	85.00	51.67	72.78	81.11	58.89	71.11	52.78	76.11	84.44	77.78	68.33	53.33	52.78	79.44
1998	65.00	93.89	87.22	69.44	73.89	84.44	66.67	71.11	63.89	71.67	87.22	76.67	67.78	53.33	63.33	86.67
1999	60.56	90.00	85.56	67.22	68.33	81.11	65.00	71.11	61.67	74.44	82.78	73.33	65.00	65.56	53.89	80.00
2000	58.89	88.89	79.44	63.89	70.00	78.33	59.44	72.22	55.56	71.67	80.56	71.11	67.22	58.33	53.89	78.33
Avg. 1992-2000	51.81	88.75	80.14	52.43	65.42	83.96	55.63	66.81	47.78	66.88	78.13	69.10	62.50	46.81	45.83	73.19
2001	60.00	88.89	77.78	66.11	75.00	81.67	53.89	73.33	51.11	68.33	80.56	74.44	70.56	56.67	53.89	72.78
2002	53.33	81.67	75.00	58.89	70.00	78.33	58.33	75.00	55.00	68.89	77.22	78.33	63.33	55.56	47.78	73.33
2003	52.22	82.22	78.33	58.89	76.11	81.67	50.56	73.89	54.44	70.00	81.11	75.00	59.44	58.89	48.33	67.22
2004	63.33	79.44	80.56	55.56	78.33	76.67	55.00	77.78	56.11	73.89	82.22	72.78	66.11	56.11	46.11	77.22
2005	47.22	73.33	73.33	50.00	72.78	65.56	49.44	72.78	46.11	64.44	85.56	67.78	60.56	43.33	41.67	70.00
2006	50.00	72.22	75.00	45.56	63.89	68.89	49.44	72.22	46.11	59.44	83.89	66.67	57.22	51.11	33.89	66.67
2007	42.78	74.44	75.56	43.89	62.22	55.56	48.33	70.00	43.89	51.67	82.22	66.11	53.89	50.56	38.33	66.11
Avg. 2001-2007	52.70	78.89	76.51	54.13	71.19	72.62	52.14	73.57	50.40	65.24	81.83	71.59	61.59	53.17	44.29	70.48

Table 6. Baseline Results

Industry Characteristic	State Characteristic	Industry-State Employment Share Ignoring Zeros	Industry-State Employment Share Including Zeros	Industry-State Entry Rate Ignoring Zeros	Industry-State Entry Rate Including Zeros
		Tobit	Tobit	Tobit	Tobit
		(1)	(2)	(3)	(4)
IRS	MKT	-0.0005	-0.0067*	0.0060	-0.0027
		(0.0043)	(0.0040)	(0.0099)	(0.0086)
MATS	BASE	-0.0009	-0.0008*	0.0035	0.0009
		(0.0009)	(0.0004)	(0.0022)	(0.0013)
SKILL INTENSITY	SKABUN	0.0132***	0.0113***	0.0023	0.0066*
		(0.0029)	(0.0024)	(0.0033)	(0.0039)
UNSKILLED INTENSITY	UNSKABUN	-0.0221***	-0.0257***	-0.0093	-0.0291**
		(0.0036)	(0.0037)	(0.0118)	(0.0117)
TRANSPORT INTENSITY	TRANSPORT	0.0010	0.0017**	0.0052**	0.0077**
		(0.0010)	(0.0008)	(0.0026)	(0.0030)
MATS	GOVERN	0.0051***	0.0022**	0.0268***	0.0191***
		(0.0015)	(0.0010)	(0.0033)	(0.0019)
Industry Fixed Effects		Yes	Yes	Yes	Yes
State Fixed Effects		Yes	Yes	Yes	Yes
Year Fixed Effects		Yes	Yes	Yes	Yes
N. Observations		37631	61305	37629	61305

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

Table 7. Robustness Results

Industry Characteristic	State Characteristic	Industry-State Employment Share Including Zeros Tobit (1)	Industry-State Entry Rate Including Zeros Tobit (2)	Industry-State Employment Share Including Zeros Tobit (3)	Industry-State Entry Rate Including Zeros Tobit (4)	Industry-State Employment Share Including Zeros OLS (5)	Industry-State Entry Rate Including Zeros OLS (6)	Industry-State Employment Share Including Zeros Tobit (7)	Industry-State Entry Rate Including Zeros Tobit (8)
IRS	MKT	-0.0065*	-0.0007	-0.0067*	-0.0029	-0.0040*	0.0025*	-0.0067**	-0.0027
		(0.0039)	(0.0087)	(0.0040)	(0.0087)	(0.0024)	(0.0014)	(0.0028)	(0.0065)
MATS	BASE	-0.0008**	0.0019*	-0.0008*	0.0009	-0.0010***	-0.0000	-0.0008	0.0009
		(0.0004)	(0.0011)	(0.0004)	(0.0013)	(0.0002)	(0.0004)	(0.0008)	(0.0019)
SKILL INTENSITY	SKABUN	0.0116***	0.0094**	0.0112***	0.0056	0.0091***	0.0019	0.0113***	0.0066
		(0.0024)	(0.0039)	(0.0024)	(0.0039)	(0.0016)	(0.0012)	(0.0020)	(0.0060)
UNSKILLED INTENSITY	UNSKABUN	-0.0181***	-0.0371***	-0.0268***	-0.0383***	-0.0053	-0.0341***	-0.0257**	-0.0291
		(0.0036)	(0.0127)	(0.0037)	(0.0120)	(0.0047)	(0.0126)	(0.0119)	(0.0235)
TRANSPORT	TRANSPORT	0.0011	0.0050*	0.0016*	0.0069**	0.0006	0.0007	0.0017***	0.0077*
		(0.0009)	(0.0028)	(0.0008)	(0.0029)	(0.0011)	(0.0010)	(0.0004)	(0.0046)
MATS	GOVERN	0.0021*	0.0292***	0.0022**	0.0191***	0.0016***	0.0034***	0.0022	0.0191***
		(0.0011)	(0.0070)	(0.0010)	(0.0020)	(0.0006)	(0.0008)	(0.0032)	(0.0057)
Industry Characteristics in Levels		Yes	Yes						
State Characteristics in Levels		Yes	Yes						
Industry Fixed Effects				Yes	Yes			Yes	Yes
State Fixed Effects				Yes	Yes			Yes	Yes
Year Fixed Effects		Yes	Yes	Yes	Yes			Yes	Yes
Regional Trend				Yes	Yes				
Industry-Year Fixed Effects						Yes	Yes		
State-Year Fixed Effects						Yes	Yes		
N. Observations		61305	61305	61305	61305	61305	61305	61305	61305
R-Squared						0.2091	0.2574		

Notes: Robust standard errors in parentheses clustered by industry in columns (1)-(6) and clustered by state in columns (7)-(8). ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

Table 8. Infrastructure Proxies Results

Industry Characteristic	State Characteristic	Industry-State Employment Share Including Zeros Tobit 1985-2005 (1)	Industry-State Entry Rate Including Zeros Tobit 1985-2005 (2)	Industry-State Employment Share Including Zeros Tobit 1985-2005 (3)	Industry-State Entry Rate Including Zeros Tobit 1985-2005 (4)	Industry-State Employment Share Including Zeros Tobit 1991-2000 (5)	Industry-State Entry Rate Including Zeros Tobit 1991-2000 (6)	Industry-State Employment Share Including Zeros Tobit 1991-2000 (7)	Industry-State Entry Rate Including Zeros Tobit 1991-2000 (8)	Industry-State Employment Share Including Zeros Tobit 1991-2000 (9)	Industry-State Entry Rate Including Zeros Tobit 1991-2000 (10)
IRS	ELEC MKT	-0.0050**	-0.0085*	-0.0046**	-0.0074	-0.0080**	-0.0114	-0.0077*	-0.0103	-0.0079**	-0.0106
		(0.0022)	(0.0048)	(0.0022)	(0.0050)	(0.0039)	(0.0090)	(0.0039)	(0.0090)	(0.0038)	(0.0086)
MATS	ELEC BASE	-0.0005**	0.0004	-0.0005**	0.0003	-0.0006***	-0.0008	-0.0006***	-0.0008	-0.0006***	-0.0007
		(0.0002)	(0.0012)	(0.0002)	(0.0011)	(0.0002)	(0.0007)	(0.0002)	(0.0007)	(0.0001)	(0.0007)
SKILL INTENSITY	SKABUN	0.0118***	0.0071*	0.0117***	0.0067*	0.0108***	0.0098**	0.0105***	0.0092**	0.0111***	0.0106**
		(0.0026)	(0.0038)	(0.0026)	(0.0039)	(0.0024)	(0.0045)	(0.0025)	(0.0045)	(0.0024)	(0.0049)
UNSKILLED INTENSITY	UNSKABUN	-0.0219***	-0.0193	-0.0203***	-0.0196	-0.0107***	-0.0227	-0.0086**	-0.0224	-0.0105***	-0.0225
		(0.0038)	(0.0122)	(0.0034)	(0.0121)	(0.0039)	(0.0140)	(0.0034)	(0.0140)	(0.0040)	(0.0143)
TRANSPORT INTENSITY	TRANSPORT	0.0016*	0.0075**			0.0010	0.0020				
		(0.0009)	(0.0029)			(0.0020)	(0.0063)				
MATS	GOVERN	0.0018*	0.0177***	0.0019*	0.0175***	0.0021***	0.0159***	0.0022***	0.0159***	0.0022***	0.0157***
		(0.0011)	(0.0018)	(0.0011)	(0.0018)	(0.0008)	(0.0020)	(0.0008)	(0.0020)	(0.0007)	(0.0018)
ELECTRIC	TRANSMISSION			-0.0119**	-0.0104			-0.0070	-0.0192**		
				(0.0052)	(0.0073)			(0.0044)	(0.0077)		
TRANSPORT INTENSITY	HIGHWAY									0.0040**	0.0068
										(0.0017)	(0.0042)
Industry Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. Observations		51972	51972	46162	46162	21960	21960	20370	20370	19215	19215

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row.

Table 9. Results Fixing State or Industry Characteristics at 1980 Levels

Industry Characteristic	State Characteristic	Industry-State Employment Share Including Zeros Tobit (1)	Industry-State Entry Rate Including Zeros Tobit (2)	Industry-State Employment Share Including Zeros Tobit (3)	Industry-State Entry Rate Including Zeros Tobit (4)
IRS	MKT in 1980	-0.0040	0.0196		
		(0.0045)	(0.0132)		
MATS	BASE in 1980	0.0013	0.0075**		
		(0.0011)	(0.0030)		
SKILL INTENSITY	SKABUN in 1980	0.0103***	0.0041		
		(0.0022)	(0.0029)		
UNSKILLED INTENSITY	UNSKABUN in 1980	-0.0520***	-0.0500**		
		(0.0065)	(0.0229)		
TRANSPORT INTENSITY	TRANSPORT in 1980	0.0004	-0.0044		
		(0.0023)	(0.0078)		
MATS	GOVERN in 1980	0.0004	0.0132***		
		(0.0007)	(0.0018)		
IRS in 1980	MKT			-0.0037*	0.0000
				(0.0020)	(0.0063)
MATS in 1980	BASE			-0.0026	0.0141
				(0.0073)	(0.0145)
SKILL INTENSITY in 1980	SKABUN			0.0107***	0.0070
				(0.0037)	(0.0059)
UNSKILLED INTENSITY in 1980	UNSKABUN			0.0067*	-0.0057
				(0.0039)	(0.0190)
TRANSPORT INTENSITY in 1980	TRANSPORT			-0.0017	-0.0063***
				(0.0012)	(0.0022)
MATS in 1980	GOVERN			-0.0129	0.0131
				(0.0093)	(0.0185)
Industry Fixed Effects				Yes	Yes
State Fixed Effects				Yes	Yes
Year Fixed Effects				Yes	Yes
N. Observations		67344	67344	67161	67161

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

Table 10. Results Including Interactions with Post-Reform Period

Industry Characteristic	State Characteristic	Policy	Industry-State Employment Share Including Zeros Tobit (1)	Industry-State Entry Rate Including Zeros Tobit (2)
IRS	MKT		-0.0133**	0.0053
			(0.0053)	(0.0204)
MATS	BASE		-0.0044	0.0235
			(0.0366)	(0.0607)
SKILL INTENSITY	SKABUN		0.0088**	0.0089
			(0.0041)	(0.0069)
LABOR INTENSITY	UNSKABUN		0.0237*	0.3195***
			(0.0126)	(0.0274)
TRANSPORT INTENSITY	TRANSPORT		0.0090**	0.0413***
			(0.0037)	(0.0108)
MATS	GOVERN		-0.0078	0.0075
			(0.0068)	(0.0201)
IRS	MKT	POST-1991	0.0085	-0.0090
			(0.0059)	(0.0207)
MATS	BASE	POST-1991	0.0040	-0.0225
			(0.0364)	(0.0605)
SKILL INTENSITY	SKABUN	POST-1991	0.0052	0.0033
			(0.0034)	(0.0080)
UNSKILLED INTENSITY	UNSKABUN	POST-1991	0.0126	-0.2835***
			(0.0107)	(0.0291)
TRANSPORT INTENSITY	TRANSPORT	POST-1991	-0.0126***	-0.0427***
			(0.0039)	(0.0119)
MATS	GOVERN	POST-1991	0.0101	0.0107
			(0.0067)	(0.0199)
Industry Fixed Effects			Yes	Yes
State Fixed Effects			Yes	Yes
Year Fixed Effects			Yes	Yes
N. Observations			61305	61305

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

Table 11. Results Including Interactions with Policy Variables

Industry Characteristic	State Characteristic	Policy	Employment Share Including Zeros Tobit (1)	Entry Rate Including Zeros Tobit (2)
IRS	MKT		-0.0095**	-0.0106
			(0.0043)	(0.0194)
MATS	BASE		-0.0054	-0.0037
			(0.0163)	(0.0389)
SKILL INTENSITY	SKABUN		0.0123***	0.0090
			(0.0037)	(0.0074)
LABOR INTENSITY	LABUN		-0.0775***	-0.0594**
			(0.0109)	(0.0266)
TRANSPORT INTENSITY	TRANSPORT		0.0009	0.0328***
			(0.0040)	(0.0107)
MATS	GOVERN		0.0020	0.0385*
			(0.0063)	(0.0223)
LABOR INTENSITY	LABUN	DEL	0.0625***	0.0457
			(0.0121)	(0.0295)
TRANSPORT INTENSITY	TRANSPORT	DEL	0.0009	-0.0299***
			(0.0040)	(0.0105)
IRS	MKT	FDI	0.0032***	-0.0050
			(0.0012)	(0.0047)
MATS	BASE	FDI	0.0080**	0.0164***
			(0.0039)	(0.0061)
SKILL INTENSITY	SKABUN	FDI	0.0014	0.0045*
			(0.0012)	(0.0025)
LABOR INTENSITY	LABUN	FDI	0.0081***	0.0137
			(0.0028)	(0.0083)
TRANSPORT INTENSITY	TRANSPORT	FDI	0.0005	0.0110***
			(0.0010)	(0.0033)
MATS	BASE	TAR	-0.0058	-0.0327*
			(0.0062)	(0.0177)
LABOR INTENSITY	LABUN	TAR	-0.0182***	-0.0711***
			(0.0056)	(0.0178)
		DEL	0.0153***	0.0246*
			(0.0039)	(0.0144)
		FDI	0.0002	-0.0053
			(0.0011)	(0.0049)
		TAR	-0.0010	-0.0179*
			(0.0019)	(0.0092)
Industry Fixed Effects			Yes	Yes
State Fixed Effects			Yes	Yes
Year Fixed Effects			Yes	Yes
N. Observations			58560	58560

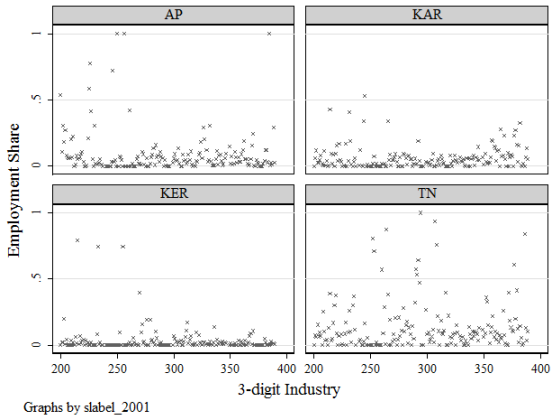
Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The specifications include also the other interactions between policy variables and the location determinants. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

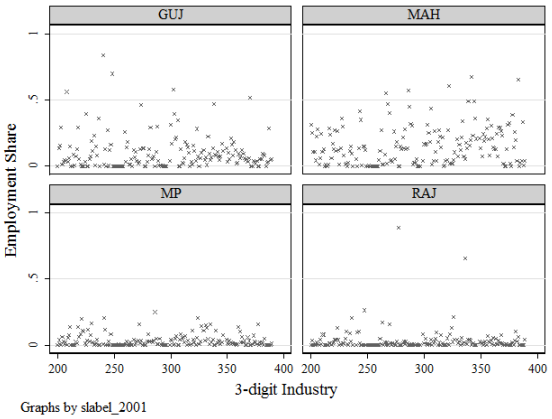
Table 12. Results Including Interactions Between Policy Variables and State Characteristics

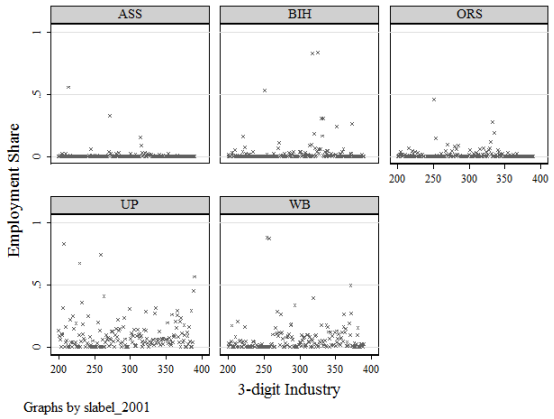
Industry Characteristic	State Characteristic	Policy	Industry-State Employment Share Including Zeros Tobit (1)	Industry-State Entry Rate Including Zeros Tobit (2)
IRS	MKT		-0.0095**	-0.0106
			(0.0043)	(0.0194)
MATS	BASE		-0.0054	-0.0037
			(0.0163)	(0.0389)
SKILL INTENSITY	SKABUN		0.0123***	0.0090
			(0.0037)	(0.0074)
LABOR INTENSITY	UNSKABUN		-0.0775***	-0.0594**
			(0.0109)	(0.0266)
TRANSPORT INTENSITY	TRANSPORT		0.0009	0.0328***
			(0.0040)	(0.0107)
MATS	GOVERN		0.0020	0.0385*
			(0.0063)	(0.0223)
UNSKILLED INTENSITY	LABUN	DEL	0.0625***	0.0457
			(0.0121)	(0.0295)
TRANSPORT INTENSITY	TRANSPORT	DEL	0.0009	-0.0299***
			(0.0040)	(0.0105)
IRS	MKT	FDI	0.0032***	-0.0050
			(0.0012)	(0.0047)
MATS	BASE	FDI	0.0080**	0.0164***
			(0.0039)	(0.0061)
SKILL INTENSITY	SKABUN	FDI	0.0014	0.0045*
			(0.0012)	(0.0025)
UNSKILLED INTENSITY	UNSKABUN	FDI	0.0081***	0.0137
			(0.0028)	(0.0083)
TRANSPORT INTENSITY	TRANSPORT	FDI	0.0005	0.0110***
			(0.0010)	(0.0033)
MATS	BASE	TAR	-0.0058	-0.0327*
			(0.0062)	(0.0177)
UNSKILLED INTENSITY	UNSKABUN	TAR	-0.0182***	-0.0711***
			(0.0056)	(0.0178)
		DEL	0.0153***	0.0246*
			(0.0039)	(0.0144)
		FDI	0.0002	-0.0053
			(0.0011)	(0.0049)
		TAR	-0.0010	-0.0179*
			(0.0019)	(0.0092)
Industry Fixed Effects			Yes	Yes
State Fixed Effects			Yes	Yes
Year Fixed Effects			Yes	Yes
N. Observations			58560	58560

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

Figure 1. Employment Shares of All Industries in Each State in 1996







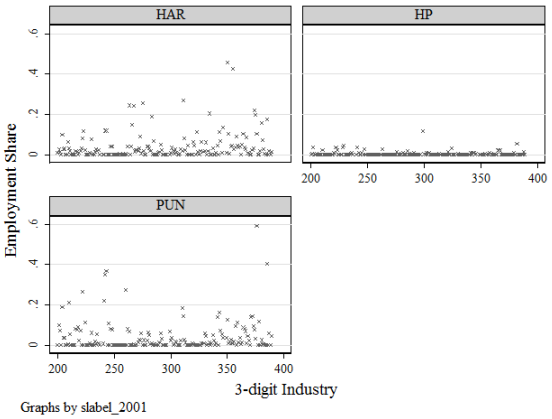
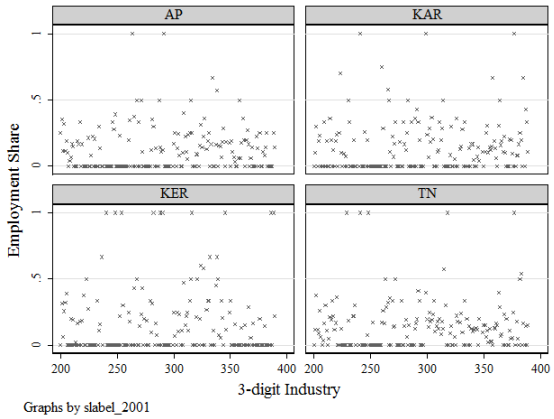
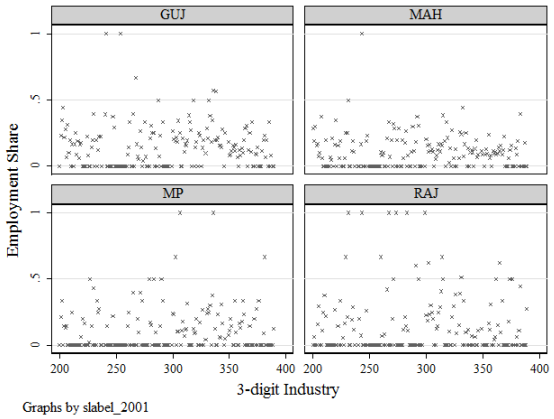
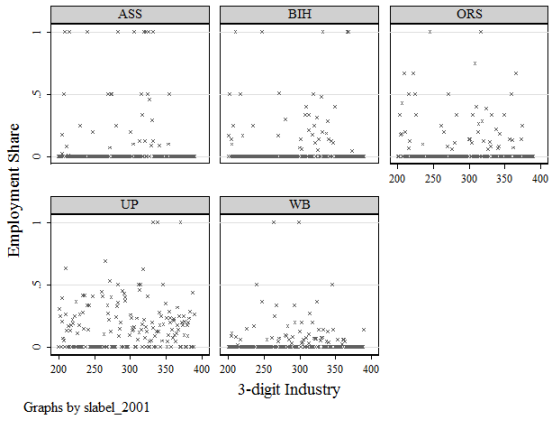
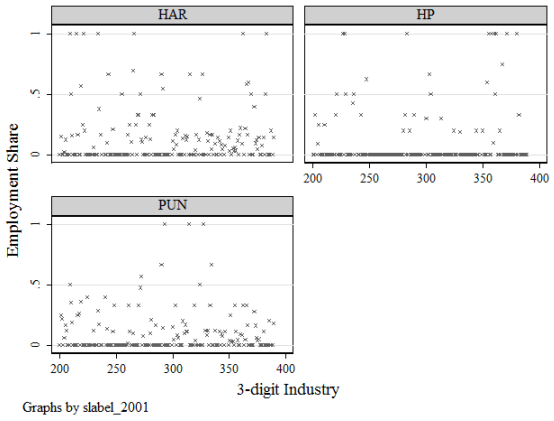


Figure 2. Entry Rates of All Industries in Each State in 1996









Appendix

Appendix Table 1. Annual Averages of Industry Characteristics

Year	SKILL INTENSITY	UNSKILLED INTENSITY	IRS	MATS	TRANSPORT INTENSITY	ELECTRIC	FDI	DEL	TARIFFS
1985	0.329	0.073	137.891	0.615	0.013	0.037	0.000	0.179	91.996
1986	0.334	0.072	133.567	0.616	0.008	0.033	0.000	0.179	99.511
1987	0.333	0.071	133.919	0.618	0.013	0.032	0.000	0.239	96.312
1988	0.336	0.068	139.395	0.626	0.018	0.030	0.000	0.227	97.930
1989	0.342	0.063	124.002	0.643	0.009	0.026	0.000	0.235	95.727
1990	0.344	0.064	112.395	0.633	0.010	0.022	0.000	0.239	96.462
1991	0.347	0.059	118.078	0.633	0.009	0.016	0.000	0.896	96.383
1992	0.357	0.063	117.521	0.627	0.016	0.048	0.139	0.923	63.989
1993	0.362	0.054	108.705	0.613	0.006	0.017	0.139	0.917	63.837
1994	0.361	0.053	110.274	0.622	0.006	0.015	0.140	0.928	64.631
1995							0.000	0.000	53.650
1996	0.372	0.103	127.997	0.920	0.012	0.000	0.140	0.950	42.895
1997	0.375	0.095	143.861	0.749	0.008	0.000	0.140	0.939	34.492
1998	0.394	0.051	180.451	0.591	0.004	0.013	0.140	0.814	35.054
1999	0.404	0.049	160.826	0.600	0.005	0.013	0.139	0.786	36.033
2000	0.412	0.047	135.111	0.595	0.007	0.012	0.139	0.781	35.269
2001	0.415	0.045	147.087	0.599	0.000	0.012	0.138	0.808	34.155
2002	0.408	0.052	145.748	0.619	0.002	0.012	0.140	0.830	30.459
2003	0.425	0.042	142.647	0.615	0.005	0.010	0.135	0.792	30.643
2004	0.425	0.046	149.643	0.625	0.005	0.008	0.140	0.819	30.646
2005	0.434	0.038	159.013	0.614	0.008	0.008	0.140	0.835	20.449
2006	0.437	0.036	160.792	0.621	0.004	0.008	0.140	0.830	17.038
2007	0.447	0.035	172.626	0.625	0.009	0.007	0.140	0.830	16.667

Appendix Table 2. Annual Averages of State Characteristics

Year	MKT	BASE	SKABUN	UNSKABUN	TRANSPORT	GOVERN	HIGHWAY	TRANSMISSION	ELEC MKT	ELEC BASE	POPULATION
1985	7180	31200000000	0.23	1.16	11400000000	0.00	7472	131875	153	2609	30190113
1986	7290	32400000000	0.22	1.19	12200000000	0.00	7505	138464	164	2772	30851444
1987	7520	34200000000	0.23	1.18	13000000000	0.00	7536	146549	178	2906	31517282
1988	8130	39300000000	0.23	1.18	13600000000	0.00	8405	155567	190	2855	32152834
1989	8408	41600000000	0.22	1.18	14600000000	0.00	8862	165675	209	3107	32836340
1990	8606	44700000000	0.23	1.18	15200000000	0.00	8956	180792	223	3319	33525868
1991	8755	44300000000	0.24	1.11	16300000000	0.00	9145	185248	237	3460	34185691
1992	9066	48200000000	0.24	1.12	17400000000	0.00	9203	186246	241	3584	36281693
1993	9733	47500000000	0.25	1.17	17200000000	0.00	9289	192878	265	3697	32261500
1994	10159	53200000000	0.24	1.12	18600000000	0.00	9456	198872	272	3871	32984225
1995	10467	58700000000	0.24	1.12	20600000000	0.00	9523	204539	288	4101	33663395
1996	11060	62300000000	0.23	1.13	22500000000	0.00	9597	207140	299	4283	34297296
1997	11476	66000000000	0.22	1.24	24600000000	0.00	9726	209088	302	4251	34915608
1998	11815	66400000000	0.24	1.13	26300000000	0.00	9908	225135	324	4270	35541284
1999	12302	72100000000	0.23	1.15	28700000000	0.00	10569	232365	337	4266	36175685
2000	12485	74600000000	0.23	1.18	30900000000	0.00	10573	233506	333	4325	41491548
2001	13876	70300000000	0.23	1.16	36600000000	0.00	10934	238190	347	4350	32018651
2002	14344	75900000000	0.23	1.18	40900000000	0.00	11104	238183	387	4536	30983767
2003	13650	81300000000	0.22	1.19	47200000000	0.00	11104	250681	627	4310	35046758
2004	13957	87300000000	0.22	1.15	53700000000	0.00	11104	247991	706	5001	36899077
2005	12869	86800000000	0.21	1.15	50700000000	0.00	11104	258887	740	5405	36744385
2006	15702	97800000000	0.21	1.17	57800000000	0.00	11104				32689787
2007	16507	107000000000	0.21	1.18	67700000000	0.00	11104				34362998

Appendix Table 3. Total Employment and Total Entry Across States in Selected Sample Years

Panel A. Total Employment in All Firms in State (in `0000s)

Year	AP	ASS	BIH	GUJ	HAR	HP	KAR	KER	MAH	MP	ORS	PUN	RAJ	TN	UP	WB
1985	60.13	8.65	29.01	62.00	19.88	1.41	30.81	19.90	102.08	27.99	11.01	25.11	16.12	74.17	54.37	64.26
1989	74.49	9.12	29.74	65.29	20.63	1.73	35.62	21.84	102.25	31.79	11.55	30.56	16.93	80.22	64.63	58.26
1994	87.10	8.99	28.38	69.65	24.99	2.78	39.97	31.07	109.20	35.29	13.89	32.16	20.20	101.53	63.17	59.21
2000	80.89	8.81	22.41	62.25	28.77	3.87	36.92	24.98	100.30	30.60	12.25	33.95	21.28	90.24	53.98	50.42
2005	82.76	9.48	19.58	72.78	36.35	5.15	47.11	24.18	107.45	28.78	13.50	40.28	25.96	105.26	66.25	41.95

Panel B. Total Employment in New Firms in State (in `0000s)

Year	AP	ASS	BIH	GUJ	HAR	HP	KAR	KER	MAH	MP	ORS	PUN	RAJ	TN	UP	WB
1985	11.01	0.75	1.99	7.70	2.18	0.51	3.25	1.00	6.72	2.31	1.60	3.50	2.41	10.31	6.09	1.85
1989	8.51	0.78	1.39	8.31	1.09	0.28	3.97	1.13	6.72	3.51	1.41	2.31	2.09	9.44	8.59	1.57
1994	8.48	0.24	1.26	10.37	2.49	0.39	5.99	2.90	8.76	4.49	1.27	1.66	3.29	18.37	6.32	1.52
2000	5.44	0.92	1.52	5.74	2.53	0.71	4.47	1.44	10.24	1.89	1.09	2.68	2.33	11.19	6.60	2.78
2005	6.81	1.50	1.91	8.34	3.56	1.74	8.20	1.11	10.32	2.04	2.29	3.08	2.75	10.31	11.34	2.87

Appendix Table 4. Baseline Results using Employment Levels

Industry Characteristic	State Characteristic		Industry-State Log Employment in All Firms Ignoring Zeros OLS (1)	Industry-State Log Employment in All Firms Including Zeros OLS (2)	Industry-State Log Employment in New Firms Ignoring Zeros OLS (3)	Industry-State Log Employment in New Firms Including Zeros OLS (4)
IRS	MKT		-0.0789*	-0.0702**	-0.0224	-0.0060
			(0.0451)	(0.0313)	(0.0432)	(0.0399)
MATS	BASE		0.0025	-0.0076***	0.0397	-0.0038
			(0.0106)	(0.0026)	(0.0370)	(0.0050)
SKILL INTENSITY	SKABUN		0.1235***	0.0979***	0.1220***	0.1236***
			(0.0318)	(0.0242)	(0.0307)	(0.0316)
LABOR INTENSITY	LABUN		-0.3242***	-0.2311***	-0.0621	-0.2102***
			(0.0453)	(0.0337)	(0.0618)	(0.0580)
TRANSPORT INTENSITY	TRANSPORT		0.0042	0.0021	-0.0060	0.0349**
			(0.0071)	(0.0082)	(0.0093)	(0.0139)
MATS	GOVERN		0.0272	0.0163*	0.0353	0.0307***
			(0.0208)	(0.0086)	(0.0370)	(0.0069)
		Industry Log Employment in All Firms	0.6951***	0.3854***	0.4937***	0.4469***
			(0.0333)	(0.0414)	(0.0451)	(0.0497)
Industry Fixed Effects						
State Fixed Effects						
Year Fixed Effects						
N. Observations			37631	51052	23033	51052

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

Appendix Table 5. Robustness Results using Employment Levels

Industry Characteristic	State Characteristic		Industry-State Log Employment in All Firms Including OLS (1)	Industry-State Log Employment in New Firms Including OLS (2)	Industry-State Log Employment in All Firms Including OLS (3)	Industry-State Log Employment in New Firms Including OLS (4)	Industry-State Log Employment in All Firms Including OLS (5)	Industry-State Log Employment in New Firms Including OLS (6)	Industry-State Log Employment in All Firms Including OLS (7)	Industry-State Log Employment in New Firms Including OLS (8)	Industry-State Log Employment in All Firms Including OLS (9)	Industry-State Log Employment in New Firms Including OLS (10)
IRS	MKT		-0.0688**	0.0027	-0.0707**	-0.0073	-0.0900**	-0.0586				
			(0.0317)	(0.0331)	(0.0313)	(0.0401)	(0.0415)	(0.0425)				
MATS	BASE		-0.0130***	-0.0145***	-0.0077***	-0.0039	-0.0131***	-0.0132***				
			(0.0027)	(0.0051)	(0.0026)	(0.0048)	(0.0029)	(0.0043)				
SKILL INTENSITY	SKABUN		0.1027***	0.1408***	0.0957***	0.1172***	0.1032***	0.1420***	0.0776***	0.1093***	0.0769***	0.1085***
			(0.0245)	(0.0321)	(0.0242)	(0.0318)	(0.0276)	(0.0365)	(0.0236)	(0.0340)	(0.0250)	(0.0341)
LABOR INTENSITY	LABUN		-0.1611***	-0.0579	-0.2494***	-0.2640***	-0.1526**	-0.5933***	-0.0303	-0.1521**	-0.0285	-0.1564**
			(0.0372)	(0.0677)	(0.0331)	(0.0578)	(0.0646)	(0.1258)	(0.0318)	(0.0693)	(0.0322)	(0.0691)
TRANSPORT INTENSITY	TRANSPORT		-0.0049	0.0196	0.0006	0.0305**	-0.0106	0.0136				
			(0.0082)	(0.0135)	(0.0081)	(0.0137)	(0.0139)	(0.0192)				
MATS	GOVERN		0.0090	0.0173**	0.0162*	0.0306***	0.0094	0.0137**	0.0172***	0.0292***	0.0166**	0.0313***
			(0.0088)	(0.0068)	(0.0086)	(0.0069)	(0.0087)	(0.0066)	(0.0055)	(0.0037)	(0.0071)	(0.0047)
IRS	ELEC MKT								-0.1027***	-0.1128***	-0.1040***	-0.1192***
									(0.0316)	(0.0407)	(0.0330)	(0.0416)
MATS	ELEC BASE								-0.0057***	-0.0022	-0.0054***	-0.0011
									(0.0012)	(0.0018)	(0.0013)	(0.0020)
TRANSPORT INTENSITY	HIGHWAY								0.0111	0.0338		
									(0.0132)	(0.0248)		
ELECTRIC	TRANSMISSION										0.0943**	0.0523
											(0.0394)	(0.0484)
		Industry Log Employment in All Firms	0.3880***	0.4637***	0.3854***	0.4469***			0.3405***	0.3798***	0.3545***	0.3948***
			(0.0415)	(0.0488)	(0.0414)	(0.0498)			(0.0480)	(0.0559)	(0.0513)	(0.0600)
Industry Characteristics in Levels			Yes	Yes								
State Characteristics in Levels			Yes	Yes								
Industry Fixed Effects					Yes	Yes			Yes	Yes	Yes	Yes
State Fixed Effects					Yes	Yes			Yes	Yes	Yes	Yes
Year Fixed Effects			Yes	Yes	Yes	Yes			Yes	Yes	Yes	Yes
Regional Trend					Yes	Yes						
Industry-Year Fixed Effects							Yes	Yes				
State-Year Fixed Effects							Yes	Yes				
N. Observations			51052	51052	51052	51052	51052	51052	16380	16380	18435	18435

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.

Appendix Table 6. Results Including Industry-State Fixed Effects

Industry Characteristic	State Characteristic	Industry-State Employment Share Including Zeros	Industry-State Entry Rate Including Zeros
		OLS	OLS
		(5)	(6)
IRS	MKT	-0.0011	-0.0007
		(0.0009)	(0.0017)
MATS	BASE	-0.0012***	0.0002
		(0.0001)	(0.0002)
SKILL INTENSITY	SKABUN	0.0005	-0.0032***
		(0.0007)	(0.0012)
UNSKILLED INTENSITY	UNSKABUN	-0.0138***	-0.0057
		(0.0025)	(0.0046)
TRANSPORT INTENSITY	TRANSPORT	0.0010**	0.0018*
		(0.0004)	(0.0010)
MATS	GOVERN	0.0015***	0.0031***
		(0.0004)	(0.0003)
Year Fixed Effects		Yes	Yes
Industry-State Fixed Effects		Yes	Yes
N. Observations		61305	61305
R-Squared		0.6734	0.2767

Notes: Robust standard errors clustered by industry in parentheses. ***, **, and * indicate significance at 1%, 5%, and 10% confidence levels, respectively. The industry and state characteristics are described in Table 1. The coefficient shown in a row is the effect of the interaction between the industry characteristic and the state characteristic shown in that row. The sample period used in the estimation is 1985-2007.