WP-2014-033

Local Skill Concentrations and District Employment Growth: A Spatial Simultaneous Equation Approach for India

Ishwarya Balasubramanian

Indira Gandhi Institute of Development Research, Mumbai
August 2014
http://www.igidr.ac.in/pdf/publication/WP-2014-033.pdf

Local Skill Concentrations and District Employment Growth: A Spatial Simultaneous Equation Approach for India

Ishwarya Balasubramanian

Indira Gandhi Institute of Development Research (IGIDR)

General Arun Kumar Vaidya Marg

Goregaon (E), Mumbai- 400065, INDIA

Email(corresponding author): Ishwarya@igidr.ac.in

Abstract

The focus of this paper is to explore the role of spatial distribution of skills in explaining differential growth rates of employment across Indian districts between the years 2001 and 2011 by using data from Census of India. To measure skills across districts, we use the skill-content of occupations and the occupational distribution of workers across districts. We then model employment and population growth simultaneously taking into account spatial correlation of the endogenous variables. We find that a one standard deviation increase in (cognitive) skills is associated with 0.52 standard deviation increase in the growth rate of male main workers and a 0.42 increase in the growth rate of male non-farm main workers. However, female employment has significantly decreased in initially skilled regions.

Keywords: Skills,Occupation,Employment Growth,Education

JEL Code: J24,E24,J21,R12

Acknowledgements:

The author is grateful to Dr. Sripad Motiram and Dr.S. Chandrasekhar for their valuable time, comments and suggestions.

LOCAL SKILL CONCENTRATIONS AND DISTRICT EMPLOYMENT GROWTH

A SPATIAL SIMULTANEOUS EQUATION APPROACH FOR INDIA

Ishwarya Balasubramanian*

ABSTRACT: *The focus of this paper is to explore the role of spatial distribution of skills in explaining differential growth rates of employment across Indian districts between the years 2001 and 2011 by using data from Census of India. To measure skills across districts, we use the skill-content of occupations and the occupational distribution of workers across districts. We then model employment and population growth simultaneously taking into account spatial correlation of the endogenous variables. We find that a one standard deviation increase in (cognitive) skills is associated with 0.52 standard deviation increase in the growth rate of male main workers and a 0.42 increase in the growth rate of male non-farm main workers. However, female employment has significantly decreased in initially skilled regions.*

Keywords: Skills, Occupation, Employment Growth, Education

JEL Codes: J24, E24, J21

1. INTRODUCTION

Employment growth varies substantially across regions within a country. The spatial concentration of skills has been identified as a principal factor that causes these differences. Several studies for the United States (Glaeser et al., 1995; Simon, 1998; Simon and Nardinelli, 2002) and Germany (Poelhekke et al., 2009; Suedekum, 2006) find a robust positive relation between regional skill concentration and employment growth. The main reasons identified for higher employment growth in skilled regions are: in-migration of both skilled and unskilled people, attraction of new firms and investments, and new employment generated through positive externalities (Glaeser and Saiz., 2004; Shapiro, 2006). In this paper, we focus on regional employment growth differences in India and in particular, the role of regional skill concentrations in causing these differences.

India is one of the fastest growing economies in the past two decades (OECD, 2012) and is among the eleven countries in the trillion dollar club with a GDP of US$ 1.8 trillion in 2013 (IMF,2014). However, high growth has not been inclusive and also masks wide regional variations in income levels and poverty (Chaudari and Gupta, 2009). Employment is one of

*Indira Gandhi Institute of Development Research(IGIDR), Mumbai, India; Email: Ishwarya@igidr.ac.in; The author is grateful to Sripad Motiram and S. Chandrasekhar for their valuable time, comments and suggestions.

the major routes through which output translates into well-being and poverty reduction. This calls for analyses of employment growth not only at national level but also at lower levels. However, Indian literature has largely focused on the national level, although a few studies (Ramaswamy, 2007; Thomas, 2014) have analysed state-level patterns. There is virtually no literature that has analysed regions below the state. Besides, the literature has focused on patterns of employment without establishing the reasons behind these patterns. Our paper fills this gap by analysing employment growth across districts using Census of India data for the years 2001 and 2011.

Employment data available for India specify only total number of workers (including self-employed and those with regular and casual jobs) in a given year. Employment growth in a given period is hence only the net addition to total workers which is the difference between total workers in the final year and initial year. It is this that we refer to as employment growth. Given that additional employment includes self-employed workers, the growth of regional population in itself is a crucial determinant of regional employment growth. Population and employment evolve simultaneously and failing to account for this leads to incorrect estimates. A large literature following Carlino and Mills (1987) does model employment and population growth simultaneously. However, this literature mainly focuses on the causality between employment and population more than their determinants. In particular, none of these studies address the role of skill concentrations while modelling employment and population. In this paper, we attempt to formally explore this relation between skills and district-level employment growth in a simultaneous framework for India. There has been no attempt so far to explore this link in India despite the skill-biased demand shifts at the national level which has implications for regional employment, given large differences in the regional distribution of skills. Our paper addresses this lacuna.

To measure skills in a district, rather than educational attainment (which is more commonly used), we use data on employment by *occupation* (from Census of India 2001) and combine it with the skill-content of occupations, which we derive from O-NET (Occupational Information Network, developed by the US Department of Labour).We then examine how the initial distribution of skills (in the year 2001) affects employment growth across districts over the period 2001-2011 and find a positive relation between skill concentration and male employment growth.

The rest of the paper is organised as follows: Section 2 reviews the relevant literature. Section 3 describes the data and methodology used to compute the skill index. Section 4 describes the distribution of skills across demographic and economic groups as well as the spatial pattern of skills. Section 5 describes the methodology to estimate employment growth and the results of our skill-employment growth relation. Section 6 concludes.

2. SKILL-GROWTH RELATION: LITERATURE REVIEW

It has been long established that skilled human capital leads to productivity gains and hence is an important source of long run growth (Romer,1986; Lucas, 1988). Since Glaeser et al (1995) showed that initial human capital had a significant positive effect on subsequent population (and income) growth, other studies have empirically established that concentration of skills (proxied by education) led not only to higher regional output but also to employment. Many studies examined the impact of human capital on employment growth across U.S metropolitan areas and found that the effect was large to the tune of 0.6 percent increase in employment growth for a 10 percent increase in the metropolitan area's human capital concentration (Shapiro, 2006). The impact increased over time (Simon, 1998) and is observed even for a very long period of time (Simon and Nardinelli, 2002). The impact of initial share of human capital on employment growth of different skill groups is however contested. In Germany, Suedekum (2006) found evidence for convergence of skill levels (growth of low-skilled jobs increased and that of high-skilled jobs decreased) in initially skilled cities suggesting that externalities (knowledge spillovers associated with skilled labour) are not strong enough. However, Poelhekke (2009) and other studies in the U.S [Berry and Glaeser, 2005; Wheeler, 2006] find evidence for divergence .i.e., initially skilled cities tend to become more skilled over time.

These studies and others that focus on the determinants of employment growth fail to recognise the simultaneity with population growth. Steinnes and Fischer(1974) were among the earliest to empirically model population and employment growth simultaneously. However, Carlino and Mills (1987) developed a theoretical framework[1] to formalise this relationship. Since then, it has been commonly accepted that both population and employment are endogenous in regional economic models. However, this literature focuses largely on the causality between jobs and people. None of the studies adopting this framework examine the relation between skills and employment growth.

There is also much debate on the empirical measurement of skills. Most of the literature has employed education (human capital) as a proxy for skills. These measures capture the availability of knowledge (or skills) in the labour market but fail to capture if the supply of skills have a corresponding demand (e.g., there might be educated people who may not be usefully employed as there is no demand for their skills). Also, the skills acquired through education do not often match those required on the job market, and to a great extent, work-related skills are acquired on the job. Hence, the actual skills that are used in the labour market are better captured by occupations rather than education. However, this required information on the skill-requirements of occupations which is made available only recently by survey based sources such as O-NET. With the availability of such measures, a new approach known as 'job-requirements approach' emerged which combines data on the proportion of people in each occupation in a region with the skill-content of that particular occupation. This measure also has the advantage of being able to differentiate along the horizontal dimensions (cognitive, motor skills etc.) of skills which the education based measures cannot (Bacolod et al., 2009)[2].We use the job-requirements approach in this paper and we use data from O-NET[3], following others who adopt this approach.

It is to be noted that these two approaches differ in terms of how the possession of a degree or a job translates into skill of the individual. While it is straightforward in the case of education since it is a fixed/durable investment, job tasks are not fixed worker attributes and hence it does not logically follow that the person with a certain job that requires certain skills *actually* possesses these skills. Autor and Handel (2013) provide a formal theoretical framework (based on Roy's (1951) self-selection framework) that links the tasks/skills required in the *job* to the human capital/skills of the *worker* and argue that workers self-select themselves into occupations that give the highest reward to their set of skills. Hence, the job-requirements approach to a large extent reflects the actual skills possessed by the job-holder. However, the major limitation with this approach is that it requires information on job characteristics for each worker. But, such information is hardly collected[4] and information on job characteristics is available only at the level of occupations (from sources like O-NET). This amounts to ignoring within-occupation heterogeneity. Despite this limitation, this approach is a useful alternative to measure regional skills. We develop skill indices for districts of India using their occupational profiles. The methodology for the same is discussed in the following section.

3. DATA AND METHODOLOGY

A district is considered to be more skilled if it has a higher proportion of people engaged in higher skilled occupations. Hence, calculating the district skill index requires information on the district's occupational distribution of employment and the skill-content of occupations. Data on employment by occupation is available in Census of India 2001 where occupations are coded according to the National Classification of Occupations (NCO) 2004. NCO 2004 classifies 2945 occupations at six digit (occupation), 4 digit (family), 3 digit (group), 2 digit (sub-division) and one digit (division) levels. In the 2001 census, occupation data has been presented for the first time at a 4-digit level. Data is available on the number of main workers according to a detailed list of 434 (4-digit) occupations for every district, separately for rural and urban areas, and for males and females. These occupations constitute 47, 29 and 94 percent of total, rural and urban employment, respectively. The remaining workers are constituted solely by cultivators and agricultural labourers.

Each of these occupations requires several skills in varying degrees. An occupation is considered to be more skilled (with respect to a given skill) if it requires that skill to a higher degree. Instead of considering every skill separately, it is useful to combine related skills (e.g., reading, writing, speaking) into skill indices (say cognitive index) that represent different dimensions. O-NET allows us to do this by providing information on various skills required for an occupation, grouped under broad categories (which help in combining similar skill variables into indices) and also quantifying the degree to which these skills are required in each occupation. O-NET is a survey based database which uses inputs from employees and occupational analysts to describe each occupation in terms of several characteristics/variables. The detailed list of variables and their description can be found in the content model of ONET[5]. There are 52, 35 and 33 variables under the categories 'abilities', 'skills' and 'knowledge' (which capture aspects of human capital), respectively. Since most of these variables are correlated and in some cases, capture similar concepts, we focus only on the 'abilities' category. The 52 abilities are organised into four broad groups: cognitive, psychomotor, physical and sensory abilities. This allows us to combine all the variables under the group 'cognitive' into a cognitive index and those under the group 'physical' into a physical index. We restrict attention to these two indices in this paper.

The skill index of each *district* can then calculated as the weighted sum of the percentage of people employed in each occupation in that district, the weights being skill indices for the occupations[6].

$$\boldsymbol{I_j} = \sum_{i} \boldsymbol{E_{ij}} \quad \boldsymbol{S_i}$$

[j x 1]   [j x i]   [i x 1]

Where $\boldsymbol{j}$ – index for districts ($\boldsymbol{j}$ = 1 to 593)

$\boldsymbol{i}$– index for occupations ($\boldsymbol{i}$=1 to 409 /411 when the skill index is calculated for the non-farm workforce /entire workforce including cultivators and agricultural labourers)

$\boldsymbol{I_j}$ is the skill index for each district $\boldsymbol{j}$

$\boldsymbol{S_i}$ is the skill index of each occupation $\boldsymbol{i}$

$\boldsymbol{E_{ij}}$ is the employment in occupation $\boldsymbol{i}$ and district $\boldsymbol{j}$ as a percentage of total employment in the district. Total employment refers to the total main workers in the district[7].

Computing a skill index $\boldsymbol{S_i}$ for occupation $\boldsymbol{i}$, entails two steps: One, quantifing *each* skill required for the occupation and two, combining all the skills into a few skill indices using an appropriate weighting scheme. To quantify each of these skills/abilities for a given occupation (say journalist), the respondents[8] are asked to rate *each* of these variables (e.g., 'Writing' skill) on a scale of 1 to 5 on the *importance* of the skill (Not important to Extremely Important) to the occupation. If the skill gets a rank of atleast 2 (Somewhat important), it is ranked on a scale of 1 to 7 on the *level* of skill needed[9]. For each occupation, the ratings of the respondents are then combined to give one value of importance and level of *every* skill/variable. Following Feser (2003), we take a product of the importance and level variables so that there is additional variation that is introduced which has the effect of the occupations becoming more distinct from each other in terms of skills and abilities[10]. To combine variables, we square the individual variables/skills and then take a mean (assuming equal weights for all skills within a group). The mean skill index (by taking the mean of all the variables) captures the '*breadth*' of skills required and the mean squared skill index (by taking the mean of the *squares* of all the variables) captures the '*depth*' of skills required (Feser, 2003).That is, if an occupation requires a broad base of skills even if the overall level and importance of the skills are lower, the former tends to be higher. An occupation that requires a greater depth of certain skills (even if it needs only a few skills) will tend to have a higher mean squared skill index.

Since O-NET is survey based, it essentially reflects the skill requirements of occupations in the US. Ideally, we require information on skill requirements for Indian occupations. Since this is not available, we make use of O-NET data as the best available alternative. A limitation in doing so is the assumption that the requirements of US occupations (as described by O-NET) are similar to the requirements of corresponding occupations in India. This assumption would be unrealistic for occupations that are highly technology-intensive in the US, but not in India (e.g. farming). However, this limitation may not be serious since the ranking of occupations in terms of skill-content is likely to be similar in both countries even if the skill-content of occupations is not estimated correctly. Moreover, some studies have applied O-NET for other countries, e.g., Aedo et al (2013) who analyses thirty countries (including India). Also, the detailed descriptions of occupations were used carefully to match occupations as accurately as possible. After we arrived at skill indices for Indian occupations, we checked for inconsistencies and dropped those occupations which seemed contrary to expectations. We also checked how well the skills fit with Indian economic and demographic characteristics.

Occupations in ONET are coded according to SOC (Standard Occupational Classification) used by the Bureau of Labour Statistics[11] whereas census 2001 follows NCO-2004 occupation code which is based on International Standard for Classification of Occupations (ISCO) 1988[12]. Using available concordance tables, the census occupations are first matched with the ONET occupations[13]. Given the greater detail in the SOC titles, most census occupations correspond to more than one ONET occupation. For these occupations, skill values for each skill variable are calculated as the average of values of all SOC occupations matched for a particular census occupation. Also, some census occupations do not find direct/close ONET matches and hence had to be ignored[14]. As a result, a total of 409 4-digit occupations are used for the analysis.[15]

4. DISTRIBUTION OF SKILLS

Once we calculate the skill index for each occupation, we can characterize the skills across demographic groups (by gender, age, religion, education and so on) and economic groups (by income, industry) if we have information on the occupation of individuals. We can also characterize the spatial (or geographical) distribution of skills if we have information on the proportion of people in each occupation in a given district as explained in the previous section. NSS data gives information on the occupation of individuals whereas Census has district-level data on the occupation profile of people. We first use NSS data to characterize skills across demographic and economic groups which is discussed in section 4.1. We then use census of India 2001 data to characterize the spatial distribution of skills in section 4.2.

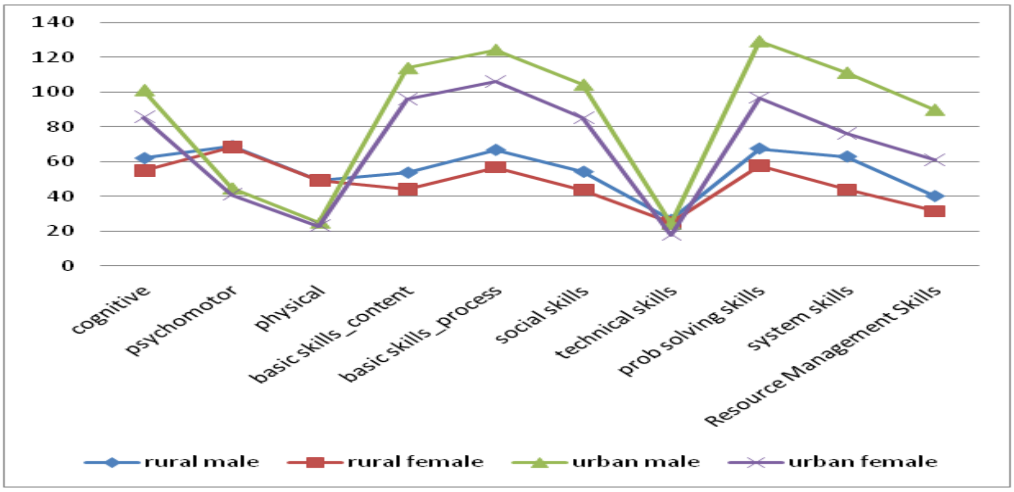
4.1. SKILLS ACROSS DEMOGRAPHIC AND ECONOMIC GROUPS

Using NSS data, we first examine how skills are distributed across different education categories. Educational attainment is the commonly used proxy for skill as the completion of a bachelor’s degree signifies that the holder may have some sort of cognitive human capital. If our skill mapping is correct, we should expect skills to be higher at higher levels of educational attainment. The skill indices for different education categories are plotted in Figure A1.

For both rural and urban areas, it can be clearly seen that the cognitive skills, basic skills and social skills are higher for higher levels of education; technical skills are constant across education categories whereas motor skills and physical skills decline with higher levels of education. The absolute values of the indices are much higher in urban areas than that in rural areas. This validates our skill mapping to a great extent.

In Figure 1, we plot the skill indices by gender and location. Two observations are clear – One, in both rural and urban areas, there is gender bias in the skills possessed and the bias is stronger for cognitive sort of skills. This is consistent with the commonly known fact that women in India are engaged in less skilled occupations than men. Two, the urban values, again for cognitive sort of skills are considerably higher than those of the rural values. The urban female has considerably high skills compared to the rural male in spite of the gender bias. This is also consistent with what we would expect given that majority of the rural population in engaged in agriculture.

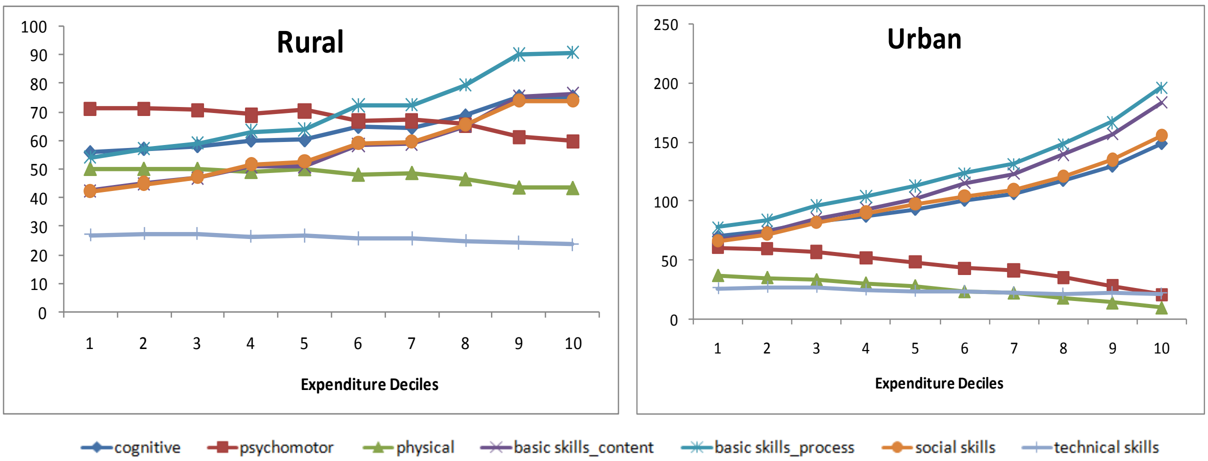
Figure 1: Skill indices by gender and location



Source: Author's own calculations based on data from NSSO Employment-Unemployment Survey (round 66)

Using NSS data, we can also examine how skills are distributed across industries and across income groups. In Figure 2, we show how the skills are distributed across different expenditure deciles[16]. As one would expect, the cognitive sort of skills is higher in higher expenditure groups and physical and motor skills tend to decrease at higher deciles. This pattern is much clearer in the urban scenario. Again, the distribution of skills fits very well with the economic characteristics. These patterns are observations (rather than findings) which are in line with our intuition and therefore serve the purpose of validating our skill mapping using the ONET data.

Figure 2: Skill Indices Across Expenditure Deciles



Source: Author's own calculations based on data from NSSO Employment-Unemployment Survey (round 66)

4.2. SPATIAL DISTRIBUTION OF SKILLS

To see the geographical distribution of skills, we map the value taken by two skills indices – cognitive and physical skills across districts for the entire workforce and the non-farm workforce only (see Figure A2). The first observation is that the two skills have almost similar distribution across districts when we consider only the non-farm workforce (Figure A2c and A2d). Both the cognitive and physical skills are concentrated in states like Jammu and Kashmir, Punjab, Haryana, Uttarakhand, Gujarat, North Eastern and Southern states. However, when the entire workforce is considered, the geographical distributions of the two skills are not similar. Cognitive skills are concentrated in the North East, South and some parts of North India, whereas physical skills are concentrated in Central and Eastern India. Another observation is that the distribution of cognitive skills across districts looks almost identical when we consider the entire workforce and the non-farm workforce. Hence, it is the physical skills that change drastically when the farm workforce is included because of the high intensity of physical skills that are primarily needed in farm occupations.

This is also clear when we compare our skill index with the conventional measure of educational attainment (see Figure A3). We calculate this (education) index as the total workers who have a technical degree, technical diploma or a graduate degree as a proportion of total main workers. Again, this index is calculated for both the non-farm workforce as well as the entire workforce. The education index calculated for the non-farm workforce only is positively correlated with both the cognitive and physical index (Figure A3a). However, the education index calculated for the entire workforce is positively correlated with the education index but negatively correlated with the physical index (Figure A3b). Educated districts are thus high on cognitive skills and low on physical when one considers the entire workforce. However, educated districts are high on both the skills when one considers the non-farm workforce only.

The interpretation of this could be that non-farm occupations are those which need some of all skills. So, when there is higher proportion of non-farm occupations in a district, there are higher cognitive skills as well as higher motor skills, physical skills and so on. On the other hand, farm occupations primarily needs physical skills and hence when farm occupations are higher (probably in districts with lesser education), physical skills are higher but cognitive skills are lower. Using an ‘education index’, one cannot distinguish between the different skills present in a region. This horizontal dimension is what the common measure using educational attainment does not capture.

The maps are intended to give a broad overview of the distribution of skills. The spatial allocation is quite different between rural and urban areas and between male and female workers. Also, we have considered only the employed workers in calculating the skill index. If instead, we calculate the proportion with respect to total labour force, the skill index would be lower in high-unemployment districts.

V. REGRESSION RESULTS

In this section, we examine how the initial (2001) distribution of skills affects employment growth across districts over the period 2001-2011.

1. Methodology

As already noted, we model employment and population growth simultaneously a 'la Carlino and Mills (1987). Further,since our observations are spatial units, we need to take into account the possibilityof spatial/auto correlation in each equation. Hence, we use a Spatial Autoregressive model (SAR) where the dependent variable Y depends on the values of Y for neighbouring units (spatially lagged values of Y). SAR takes into account these spatial spillovers in Y by introducing a spatial lag term on the RHS which is a weighted average of the values of Y observed for neighbouring (as defined by the weights) spatial units. A generalised version of the SAR model allows for spatial autoregressive errors (errors are also generated by a spatial autoregressive process) and is known as SARAR model. In the Carlino-Mills type simultaneous equations model where there are two dependent variables (Y1 and Y2), there is another dimension of spatial interaction as introduced by Boarnet (1994) where Y1 is affected by spatially lagged values of Y2 and vice versa.

To illustrate all the above forms of simultaneity, consider the following general model. W is the weight matrix which defines which spatial units are neighbours based on pre-determined criteria.

$$Y_1 = \beta_1 X_1 + \gamma_{21} Y_2 + \phi_{21} W Y_2 + \rho_{11} W Y_1 + \varepsilon_1$$

$$Y_2 = \beta_2 X_2 + \gamma_{12} Y_1 + \phi_{12} W Y_1 + \rho_{22} W Y_2 + \varepsilon_2$$

In the above system of simultaneous equations, there are three forms of simultaneity: feedback simultaneity (γ), spatial autoregressive lag simultaneity (ρ) and spatial cross regressive lag simultaneity (ϕ) (Rey and Boarnet, 2004).

The general model for population and employment growth can be written as

$$dP = \alpha_p + \lambda_p P_{t-1} + \beta_p C + \gamma_p dE + \rho_p WdP + \phi_{1p}(I+W)E_{t-1} + \phi_{2p}(I+W)dE + \varepsilon_1 \text{--(1)}$$

$$dE = \alpha_e + \lambda_e E_{t-1} + \beta_e D + \gamma_e dP + \rho_e WdE + \phi_{1e}(I+W)P_{t-1} + \phi_{2e}(I+W)dP + \varepsilon_2 \text{--(2)}$$

where:

dP - $\ln(P_t) - \ln(P_{t-1})$ where P is a n x 1 vector of population in year t and t-1

dE - $\ln(E_t) - \ln(E_{t-1})$ where E is a n x 1 vector of main workers in year t and t-1

C – n x j matrix of j exogenous variables affecting population growth in period t-1

D – n x k matrix of k exogenous variables affecting employment growth in period t-1

W – n x n positive matrix, which takes $w_{ij} \neq 0$ if i and j are neighbours and 0 otherwise. We use a contiguity matrix here where $w_{ij} = 1$ if i and j share a border or a vertex. The rows of the matrix are then standardised such that they add to one.

I - n x n identity matrix

n – number of observations/districts

Carlino-Mills (1987) introduced the feedback structure (γ) by developing a model in which equilibrium population and employment are determined simultaneously. Boarnet (1994) augmented the C-M model by introducing the spatialcross regressive lagged terms of initial population (or employment) and population (or employment) change in the employment (or population) change equation (captured by the parameters ϕ_1 and ϕ_2). The term $(I+W)E$ is called labour market variable which measures the sum of the particular location's employment value plus the weighted averages of the employment values of the neighbouring locations. Hence, population is not only affected by employment change in the census area but also that of neighbouring areas. Similarly, firms are attracted not only to labour pool of the area that they are located in but also to that around the census area. Later studies incorporated the SAR interactions (dependence of the dependent variable on the neighbouring values of itself captured by ρ) to the Boarnet and C-M model (e.g., Henry at al, 1999). The following table summarizes the various models.

Table 1: Taxonomy of Models

Cross Section (No Simultaneity)	$\rho = 0$; $\phi_1 = 0$; $\phi_2 = 0$; $\gamma = 0$	OLS
SARAR	$\gamma = 0$; $\phi_1 = 0$; $\phi_2 = 0$	MLE/GMM
Carlino-Mills	$\rho = 0$; $\phi_1 = 0$; $\phi_2 = 0$	2SLS
Boarnet	$\rho = 0$; $\gamma = 0$	2SLS
CM-SARAR	$\phi_1 = 0$; $\phi_2 = 0$	2SLS/GMM
B-SARAR	$\gamma = 0$	2SLS/GMM

Ignoring endogeniety and estimating either of the equations using OLS will yield inconsistent estimates.The simultaneous equation system of the *C-M model* hence has to be estimated by 2 stage least squares (2SLS) where the endogenous explanatory variable dE (or dP) is regressed on all the exogenous variables in the system. The predicted value of the first stage regression is then used in each of the regressions. In the *Boarnet model*, there are two ways to get the predicted values of (I+W) dE and (I+W) dP in the first stage of 2SLS. The traditional method is to get the predicted value of dE and dP and then multiply it with (I+W) to get the predicted values of (I+W) dE and (I+W) dP. Henry et al (2001) and others use a different method where they get the predicted values of (I+W) dE and (I+W) dP directly by regressing dE (or dP) on all the exogenous variables in the system *plus* the spatial lags. Rey and Boarnet (2004) show, that the latter is a better technique than the traditional method as it ensures that the variables are orthogonal to the residuals. The last two models (CM-SARAR and Boarnet-SARAR) are complex by introducing spatial autoregressive terms in a simultaneous equation framework. Kelijian and Prucha (2004) suggested an estimation procedure for such cases of simultaneity with spatial dependence which combines GMM and IV estimation that yields consistent and asymptotically normal estimates.

2. Variables

We estimate four models described above: CM, Boarnet, CM-SARAR and Boarnet-SARAR. The dependent variable for the employment equation (dE) is the difference in natural logarithms of total main workers in the final year and initial year for each district. The dependent variable for the population equation (dP) is the difference in natural logarithms of the population above 6 years. The matrix D (the exogenous variables for the employment equation) includes the skill index calculated for each district as described earlier. We consider here only the skills of the non-farm workforce. It also includes controls for district characteristics in 2001 such as: urbanization rate, workforce participation rate, proportion of Scheduled Castes(SC) and Tribes (ST), proportion of households with access to electricity, and employment share in nine major secondary and tertiary sectors. It also includes the proportion of people with secondary education, those with technical diploma, technical degree and graduate degree to represent the level of educational attainment in the district(which is used to measure skills in the literature). According to the literature, these variables are expected to have a positive and significant impact on employment growth.

The matrix C consists of instruments for the endogenous population variable i.e., the exogenous variables not included in the employment equation but included in the population equation. It includes district amenities such as proportion of households with access to electricity, toilets and water from any source within premises, number of hospitals, schools, colleges, cinema theatres, banks and credit societies in the district.[17]

Data for all the variables are from the 2001 census except for the employment data for the year 2011 which is from the 2011 census[18]. In 2001, there were 593 districts which rose to 640 districts in 2011. Accordingly, the 2011 figures on employment for those districts which underwent boundary changes were adjusted[19].

Results of the employment equation for the four models are given in Table 1. The labour market for male and female workers is very different in India. Female employment in India depends largely on the social and cultural environment and has been highly fluctuating in response to distress conditions (Abraham, 2009). So, we estimate the models for male and female workers separately.The total main workers constitute a major share of farm workers and an increase in total employment may not be indicative of an increase in productive employment. Hence, we also perform similar regressions with the dependent variable as the growth rate of male and female non-farm workers .i.e., main workers excluding cultivators and agricultural labourers.The results of the Boarnet-SARAR model for male (and female) main and non-farm employment growth are presented in Table 2 and that of male (and female) population growth are presented in Table 3.

3. Discussion of Results

First, the initial value of employment is expected to have a negative sign as employment growth is by definition lower when the initial year employment is high. Across all the models and six dependent variables (except for some models for total main workers), the sign of this variable is negative. But, it is statistically significant only in some cases.

There are two variables in our model that describe the linkage between population and employment - population of the residential zone for the Boarnet model and population growth (non-spatial endogenous population growth in C-M model/spatial endogenous population growth in Boarnet model). For all our dependent variables, the population growth variable has a positive and significant coefficient supporting the 'jobs follow people' hypothesis. The initial value of the 'zone' population included in the Boarnet model also has a positive coefficient throughout, but is not statistically significant in most cases.

The CM-SARAR and Boarnet-SARAR models which include the spatial interactions of the dependent variable and the error term generate two values λ and ρ. A positive coefficient on λ indicates spatial autoregressive dependence in the dependent variable i.e., employment growth of one district is affected by employment growth of neighbouring districts. A positive coefficient on ρ indicates spatial autoregressive dependence in the error term i.e., an exogenous shock in one district will affect the employment growth rate of neighbouring districts. This may be due to spatial interactions in the other exogenous variables. The results show that there are significant spatial dependence for all the dependent variables as shown by statistically significant values of λ and ρ. This justifies the use of spatial models for a district level analysis.

Table 1 Cognitive Skills and Employment Growth Rate (2001-2011)

	Carlino-Mills	*Boarnet*	*Carlino-Mills SARAR*	*Boarnet-SARAR*
Main Workers(log)	-0.000	-0.005	0.005	0.007
	(0.007)	(0.010)	(0.005)	(0.008)
Endogenous Population Growth	0.501		0.506	
(Predicted)	(0.155)**		(0.075)**	
Endogenous Population Growth		0.392		0.741
of *residential zone*(predicted)		(0.110)**		(0.055)**
Population of the *residential*		0.003		-0.002
zone(log)		(0.005)		(0.004)
Urbanization rate	0.009	0.025	-0.027	-0.048
	(0.053)	(0.051)	(0.039)	(0.044)
Scheduled Caste(SC) Population	-0.043	-0.054	-0.032	-0.031
	(0.064)	(0.066)	(0.047)	(0.060)
Scheduled Tribe(ST) Population	0.023	0.022	0.001	0.004
	(0.025)	(0.026)	(0.019)	(0.024)
Workforce Participation rate	0.153	0.122	0.041	0.111

(WPR)	(0.115)	(0.111)	(0.080)	(0.098)
Household Manufacturing	-0.465	-0.481	-0.146	-0.214
	(0.150)**	(0.152)**	(0.118)	(0.134)
Non-Household Manufacturing	-0.064	0.007	0.009	-0.069
	(0.160)	(0.146)	(0.111)	(0.120)
Electricity, Gas and Water Supply	-0.324	-0.153	0.242	-0.021
	(0.822)	(0.844)	(0.676)	(0.723)
Construction	-0.198	-0.121	-0.320	-0.486
	(0.328)	(0.331)	(0.251)	(0.290)
Wholesale and Retail Trade	-0.094	-0.145	0.141	0.195
	(0.300)	(0.302)	(0.231)	(0.267)
Hotels and Restaurants	0.006	-0.381	0.207	0.433
	(1.108)	(1.126)	(0.829)	(0.983)
Transport, Storage and Communications	0.203	0.226	-0.172	0.021
	(0.524)	(0.545)	(0.414)	(0.482)
Finance, Real Estate and Business	0.642	0.621	-0.072	0.123
	(0.417)	(0.427)	(0.325)	(0.390)
Public administration, Defence,	-0.452	-0.442	-0.042	-0.131
Education, Health,other services	(0.137)**	(0.143)**	(0.111)	(0.132)
Cognitive skills	**0.144**	**0.139**	**0.075**	**0.100**
	(0.029)**	**(0.029)****	**(0.023)****	**(0.027)****
Secondary Education/	-0.512	-0.544	-0.179	-0.239
Below Graduate	(0.206)*	(0.210)**	(0.157)	(0.191)
Technical Diploma	-1.117	-1.350	0.356	0.256
	(0.975)	(0.968)	(0.707)	(0.880)
Graduate Degree and above	0.094	0.062	0.126	0.156
	(0.290)	(0.296)	(0.213)	(0.273)
Technical Degree	1.992	2.221	1.270	1.354
	(1.491)	(1.516)	(1.138)	(1.352)
Proportion of households with	0.073	0.069	0.030	0.061
access to electricity	(0.027)**	(0.027)*	(0.019)	(0.025)*
Constant	-0.437	-0.396	-0.383	-0.478
	(0.144)**	(0.135)**	(0.100)**	(0.113)**
lambda(λ)			0.744	0.449
			(0.054)**	(0.066)**
rho(ρ)			-0.383	0.175
			(0.106)**	(0.082)*
N	592	592	592	592

Notes: * Significant at 5 percent level; ** Significant at 1percent level;Values in brackets are standard errors; Urbanization rate,SC population and ST population and WPR denote the urban population, SC population, ST population and total main workers respectively as a proportion of total population in the year 2001. The industry variables are calculated as the proportion of total main workers employed in the particular industry in the year 2001. The education variables are calculated as the proportion of total main workers with a given level of education.

Table 2 Cognitive Skills and Male and Female Employment Growth Rate (2001-2011)

Boarnet-SARAR model	*Male Employment Gr. Rate*		*Female Employment Gr.Rate*	
	Main Workers	*Non-Farm*	*Main Workers*	*Non-Farm*
Main Workers(log)	-0.000	-0.014	-0.021	-0.114
	(0.007)	(0.009)	(0.017)	(0.016)**
Endogenous Population Growth	0.802	0.964	0.362	0.473
of *residential zone*(predicted)	(0.047)**	(0.060)**	(0.108)**	(0.113)**
Population of the *residential*	-0.001	0.004	0.008	0.035
zone(log)	(0.004)	(0.005)	(0.009)	(0.009)**
Urbanization rate	-0.040	-0.112	0.036	-0.069
	(0.038)	(0.049)*	(0.088)	(0.089)
Scheduled Caste(SC) Population	-0.029	0.061	-0.083	-0.153
	(0.054)	(0.069)	(0.109)	(0.111)
Scheduled Tribe(ST) Population	0.003	0.013	0.020	-0.098
	(0.021)	(0.027)	(0.044)	(0.046)*
Workforce Participation rate	0.150	-0.209	-0.376	-0.023
(WPR)	(0.085)	(0.107)	(0.160)*	(0.141)
Household Manufacturing	-0.339	-0.520	0.215	0.130
	(0.117)**	(0.148)**	(0.261)	(0.277)
Non-Household Manufacturing	-0.115	0.119	0.157	0.581
	(0.105)	(0.135)	(0.228)	(0.235)*
Electricity, Gas and Water Supply	-0.293	-2.234	0.748	0.212
	(0.603)	(0.779)**	(1.559)	(1.637)
Construction	-0.665	-0.452	0.396	0.828
	(0.251)**	(0.324)	(0.578)	(0.598)
Wholesale and Retail Trade	0.123	0.046	-0.112	-0.780
	(0.228)	(0.293)	(0.563)	(0.595)
Hotels and Restaurants	0.127	0.921	-0.338	-0.753
	(0.858)	(1.097)	(1.897)	(1.969)
Transport, Storage and Communications	-0.085	-0.040	0.832	2.950
	(0.414)	(0.533)	(0.966)	(0.963)**
Finance, Real Estate and Business	0.085	1.731	0.707	2.275
	(0.342)	(0.444)**	(0.773)	(0.795)**
Public administration, Defence,	-0.230	0.009	0.117	0.322
Education, Health,other services	(0.116)*	(0.147)	(0.256)	(0.254)
Cognitive skills	**0.131**	**0.115**	**-0.046**	**-0.154**
	(0.023)**	**(0.030)****	**(0.051)**	**(0.057)****
Secondary Education/	-0.207	-0.344	-0.495	-0.384
Below Graduate	(0.169)	(0.215)	(0.358)	(0.356)
Technical Diploma	0.429	0.205	-1.611	-1.533
	(0.803)	(1.025)	(1.589)	(1.638)
Graduate Degree and above	0.158	-0.403	0.105	0.026
	(0.251)	(0.322)	(0.483)	(0.510)
Technical Degree	0.091	-0.564	5.900	4.964
	(1.187)	(1.519)	(2.586)*	(2.634)
Proportion of households with	0.106	0.119	0.027	-0.023
access to electricity	(0.022)**	(0.027)**	(0.044)	(0.046)
Constant	-0.522	-0.242	0.120	0.668
	(0.101)**	(0.126)	(0.202)	(0.229)**
lambda(λ)	0.349	-0.079	0.754	0.635
	(0.064)**	(0.080)	(0.059)**	(0.053)**
rho(ρ)	0.370	0.331	-0.322	-0.358
	(0.065)**	(0.074)**	(0.109)**	(0.104)**
N	592	592	592	592

Notes: same as above; In the models estimated for male (and female) workers, the variables for population growth, initial employment and workforce participation rate consists only of male (female) population growth, male (female) initial employment and male (female) work participation rate.

Table 4 Male and Female Population Growth Rate (2001-2011)

Boarnet-SARAR model	*Male Population Growth Rate*	*Female Population Growth Rate*
Population(log)	-0.003	-0.011
	(0.006)	(0.005)*
Emdogenous employment (main workers)	0.512	0.047
growth of *labour market zone* (predicted)	(0.026)**	(0.010)**
Employment of the labour market zone(log)	-0.007	-0.000
	(0.002)**	(0.002)
Urbanization rate	0.046	0.050
	(0.022)*	(0.020)*
Cognitive Skills	**-0.017**	**-0.005**
	(0.010)	**(0.008)**
Access to electricity	-0.098	-0.014
	(0.016)**	(0.012)
Access to toilets	-0.104	-0.031
	(0.020)**	(0.013)*
Access to water within premises	0.050	0.015
	(0.020)*	(0.014)
Constant	0.372	0.176
	(0.063)**	(0.061)**
District amenities	Yes	Yes
lambda(λ)	0.452	0.871
	(0.066)**	(0.057)**
rho(ρ)	0.420	-0.585
	(0.057)**	(0.108)**
N	592	592

Notes: same as above; 'Access'variables are in terms of proportion of households with access to electricity, toilets and water. District amenities are as specified in text.

Our main variable of interest is the skill index. This is probably the only variable apart from the above variables whose coefficient has a consistent and statistically significant sign across all the models. However, while the coefficient on this index is positive and significant for total main workers, it takes opposite signs when we consider male and female employment separately. It is positive and significant for male workers but insignificant for female workers. When we consider the growth rate of male and female non-farm employment, the coefficient is positive and significant for male workers but negative and significant for female workers. Overall, it has an insignificant coefficient when non-farm workers are taken together. A one standard deviation increase in (cognitive) skills is associated with 0.52 (0.42) standard deviation increase in the growth rate of male main (non-farm) workers. The results therefore suggest a gendered labour market in India with regard to the skill-growth relation. Initially skilled regions do tend to grow faster in terms of male employment. However, female employment has significantly decreased in initially skilled regions. This is a puzzle to explore further.

The impact of education variables is puzzling. Barring a few coefficients which are positive (but still not significant), the coefficient on every level of education is negative. Since the skill index and the education variables capture a similar concept, we tried specifications with the skill index and the education variables included separately. The education variables were still not significant and the skill variables had similar signs for their coefficients. This is surprising given that the education variables have shown a robust positive relationship with employment growth in earlier studies. However, it is not so surprising given that the earlier studies were for developed countries. The labour markets are very different for developed and developing countries. In developed countries, the actual level of educational attainment of workers may have externality effects and may lure jobs for two reasons. One, the absolute volume of educated workers is much higher in developed countries to have externality effects. Two, the jobs in developed countries are on the average more skilled and hence move to places where there are more educated people who are assumed to have those skills. This is not applicable for developing countries. Only sectors like Information Technology, Finance etc. look for very high skills which the educational attainment is assumed to proxy. But, such sectors constitute a very low share of employment in India.

Hence, our results are contrary to the literature regarding the impact of education variables on employment growth. However, we argue that it is probably not the correct measure to reflect skills in developing countries. The alternative measure that we consider shows a robust positive and significant relation with employment growth which is consistent with the literature in terms of the channels through which skills affect employment growth. Nevertheless, the skills-growth relation is found only for male employment growth. The labour market for female workers in India is too complicated to make any comments on this relation. Especially the period of study spanning 10 years included a widely fluctuating phase of female employment. A more detailed exploration is needed to study female labour market in India.

Regarding the other exogenous variables, for male employment growth rate (main workers and non-farm workers), the coefficient of household manufacturing takes a negative and significant value. This implies that districts with a high share of household manufacturing in 2001 have seen a decline in male employment growth rate. We do not have sufficient evidence to say that the fall in this growth rate of total employment is due to a fall in growth rate of manufacturing employment (for which a sector wise regression is required). However, this does seem likely in line with overwhelming evidence for a fall in manufacturing employment growth in India in the last decade (Mehrotra et al, 2012). The districts with high household manufacturing are those which lost clearly in employment growth than those with non-household manufacturing. The negative coefficient for the construction sector is puzzling given that much of the employment gain in the last decade has been in this sector (Thomas, 2014). However, this probably reflects the initial-level effect and is also consistent with evidence that construction employment has largely increased in rural areas in less-developed regions where initial year employment in construction was lower (ibid). This is also evident from the coefficient of urbanisation which is negative implying that much of the employment gain has been in rural areas.

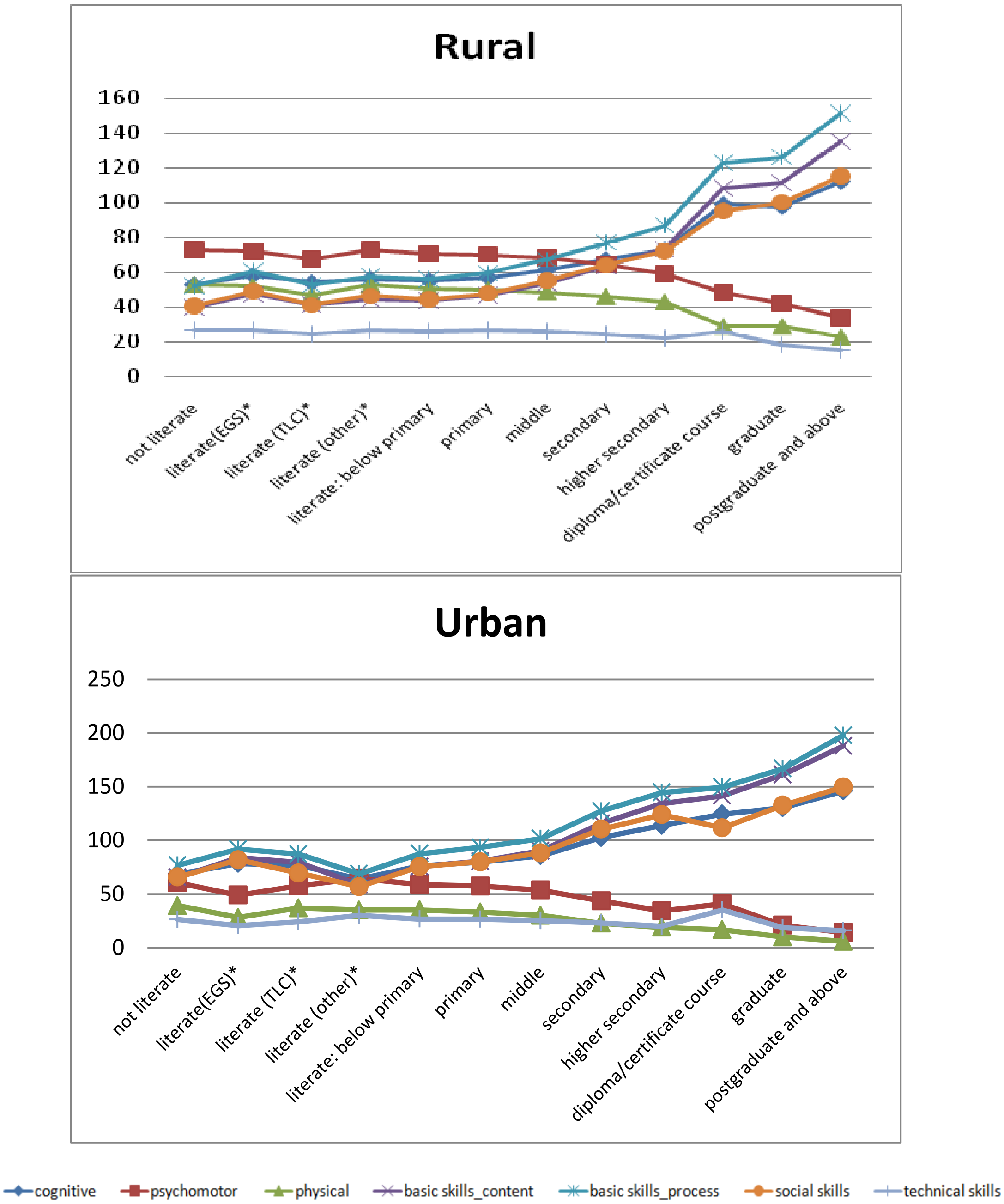
In sum, we find evidence for the fact that 'jobs follow skills'. This implies that net gains in employment in the past decade have been in skilled regions. Such employment gains include both self-employment (which could have been generated simply because of the expansion of population or migration) and new employment generated because of new firms and high growth (due to externalities) in these regions. We have estimated the population equation. Population changes are the sum of births net of deaths (which depends on demographic factors) and net migration. In Table 3, the skill variable for the population change equation takes a negative (but not significant) coefficient much in contrast to the employment change equations. This could either be because the natural expansion of population took place in less skilled areas or because higher migration was towards less skilled areas. This implies that employment growth in skilled regions have taken place despite the fact that population did not grow in these regions. This suggests that employment growth in skilled regions were not simply due to expansion of population but due to additional employment created in skilled regions due to new firms or high growth in these regions

VI.SUMMARY AND CONCLUDING REMARKS

There are large disparities in employment growth across Indian districts. Given the national level skill-biased demand shifts, we explore the impact of the spatial distribution of skills on spatial employment growth differences. We measure regional skills by combining regional occupational distribution and the skill-content of occupations. Using a spatial simultaneous equation model, we find a robust positive relationship between skills and employment growth for male workers. However, the relationship is negative for female workers which shows that the observed de-feminisation of the Indian workforce has been particularly stronger in skilled regions.The results, we believe, are relevant for public policy, especially policies dealing with labour market issues, skill development and urban issues.

One implication of the positive relation is that skill development policies not only help in meeting the (future) demand for more productive employment but are themselves drivers of change. The regions that have a higher stock of skilled labour will have increased employment by increasing *both* skilled and unskilled labour to the extent that they complement each other. The positive relation is good news for skilled regions and for people acquiring skills. However, it also reflects the reality that the less skilled regions which are already poor had slower employment growth. If this situation continues, it could aggravate poverty in these regions further leading to a vicious cycle. The concentration of employment in already skilled regions will lead to more spatial divergence in terms of growth and employment on one hand and will also pose mounting pressures on urban policy on the other hand. This calls for targeted skill development policies. Alongside supply side measures that are taken to increase the (supply of) skills of the workforce, demand side measures should concentrate on not only creating the demand for skills but also on allocating these opportunities equally and efficiently across space. Less skilled areas should also create conducive environment and amenities to attract skilled labour. This would ensure that growth is inclusive in the spatial sense as well.

Figure A1: Skill indices for different categories of education

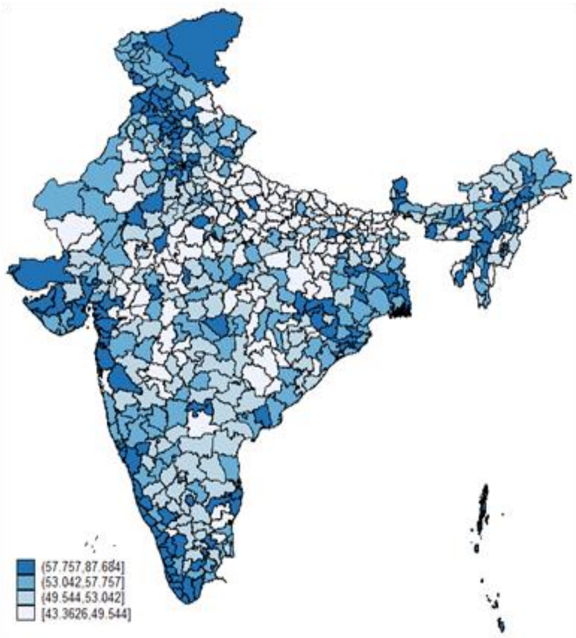


Source: Author's own calculations based on data from NSSO Employment-Unemployment Survey (round 66)

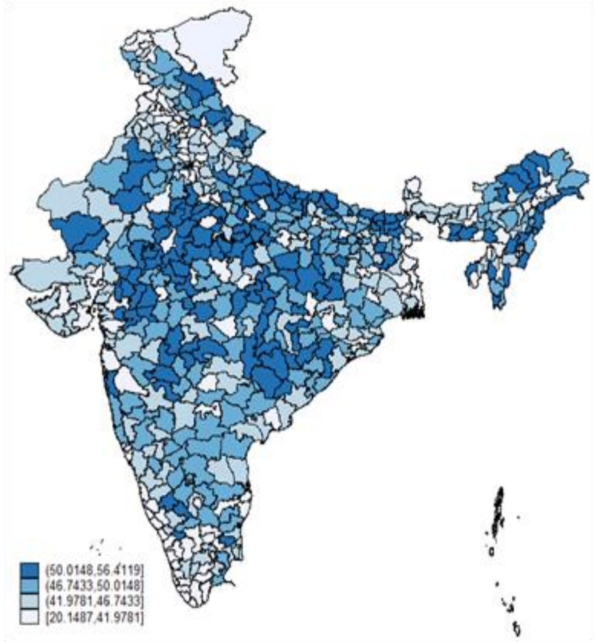
Notes: The indices for each occupation are calculated as explained in section 3 and then combined with the information on the occupation of each worker as given in NSS. The graphed values of each index are the average values of that index in each education group.
* Literate without formal schooling

Figure A2: Distribution of Skill Index Across Districts

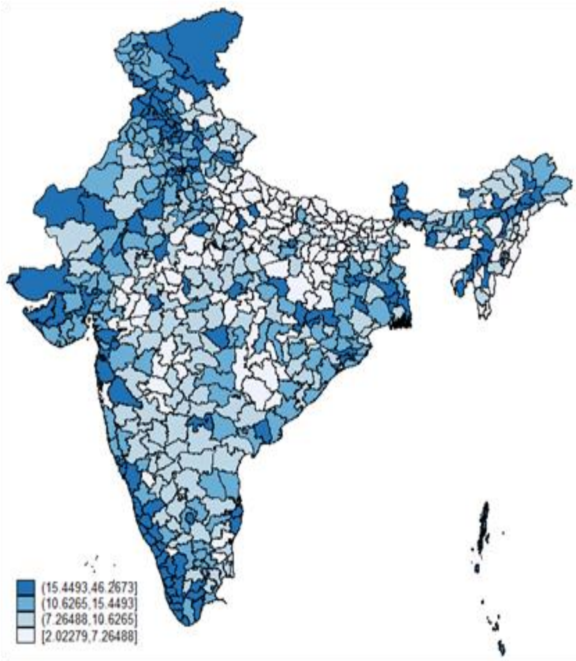
(a) Entire Workforce (Cognitive Skills)



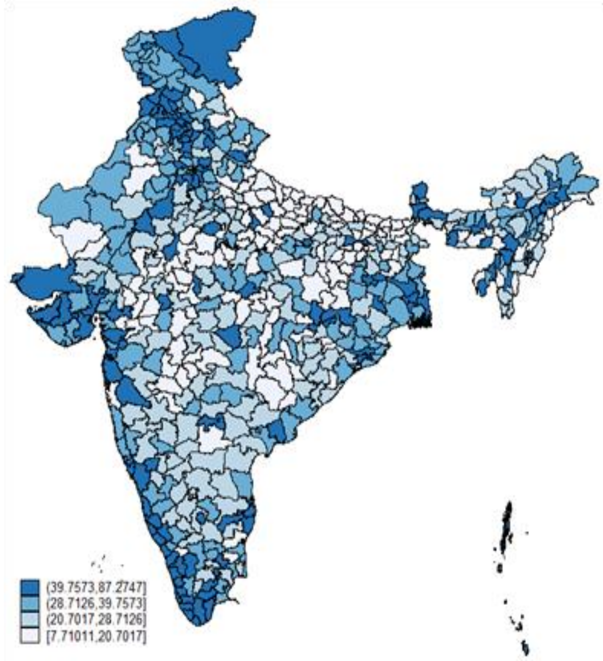
(b) Entire Workforce (Physical Skills)



(c) Non-farm Workforce (Cognitive Skills)



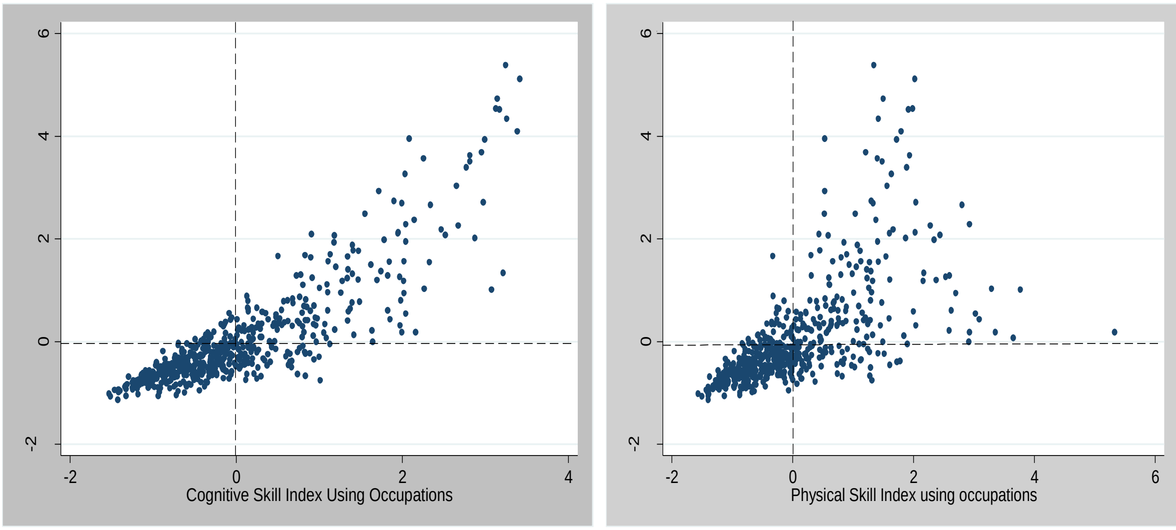
(d) Non-farm Workforce (Physical Skills)



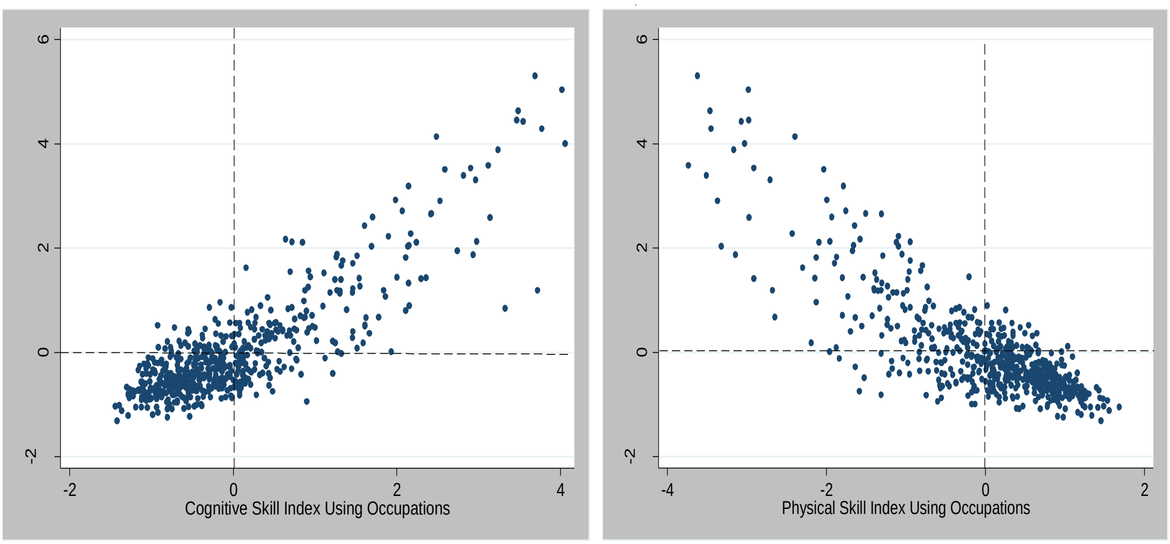
Notes: The above maps are quantile maps .i.e., each class includes approximately equal number of observation (districts in this case).

Figure A1: Comparison of Skill Index Calculated Using Education Profile and Occupational Profile of Districts

(a) Non-Farm Workforce Only



(b) Entire Workforce



Source: Author's own calculations using data from Census of India 2001

Notes: N=593; In the left panel of figure (a) and (b), the cognitive skill index across districts is plotted against the education index across districts (see text for details). In the right panel, the physical skill index across districts is plotted against the education index across districts; The farm workforce (Cultivators and Agricultural labourers) is excluded while calculating the skill index in figure(a) and included in figure(b).

Table A1: Descriptive Statistics

Variable	Description	Min	Max	Mean	SD
DEPENDENT VARIABLES					
dE (Main workers)	Log ratio of total(male/female) main workers in 2011 to that in 2001	-0.37	0.59	0.13	0.13
(males)		-0.42	0.61	0.11	0.12
(females)		-1.20	0.95	0.18	0.26
(Non-farm workers)	Log ratio of total(male/female) non-farm workers.i.e., main workers excluding cultivators and agricultural labourers in 2011 to that in 2001	-0.34	0.75	0.20	0.13
(males)		-0.47	0.74	0.17	0.13
(females)		-1.28	1.42	0.41	0.32
dP (all)	Log Ratio of total (male/female) population above 6 years in 2011 to that in 2001	-0.31	0.61	0.19	0.09
(males)		-0.33	0.59	0.18	0.09
(females)		-0.28	0.64	0.19	0.09
INITIAL VALUES					
EMP01 (Main workers)	Natural log of total(male/female) main workers in the year 2001	9.04	14.86	12.81	0.99
(males)		8.87	14.70	12.52	1.03
(females)		7.19	13.54	11.27	1.03
(Non-farm workers)	Natural log of total(male/female) non-farm workers in the year 2001	8.57	14.86	11.87	1.12
(males)		8.29	14.70	11.68	1.13
(females)		6.67	13.06	10.02	1.17
POP01 (all)	Natural log of Total Population in the year 2001	10.20	15.92	13.82	1.03
(males)		9.52	15.25	13.16	1.02
(females)		9.40	15.21	13.09	1.03
ENDOGENOUS VARIABLES AND SPATIAL VARIABLES					
dP_HAT	Predicted values of endogenous population growth ($\widehat{dP}$)				
WPOP01	Population(log) of the residential zone in the initial year 2001.i.e., (I+W)*POP01				
WdP_HAT	Predicted values of endogenous population growth of the residential zone.i.e.,$\widehat{(I+W)*dP}$				
dE_HAT	Predicted values of endogenous employment growth($\widehat{dE}$)				
WEMP01	Employment (log) of the labour market zone in the initial year 2001.i.e., (I+W)*EMP01				
WdE_HAT	Predicted values of endogenous employment growth of the labour market zone.i.e.,$\widehat{(I+W)*dE}$				
EXOGENOUS VARIABLES					
URBAN	Urbanization Rate - Proportion of urban population in total population for the year 2001	0	1	0.24	0.20
WPR	Workforce Participation Rate- Proportion of total(male/female) main workers in total(male/female) population for the year(2001)	0.17	0.58	0.31	0.06
(for males)		0.29	0.70	0.45	0.06
(for females)		0.02	0.50	0.16	0.09
PROP_SC	Proportion of Scheduled Caste(SC) Population in total population in the year 2001	0	0.50	0.15	0.09
PROP_ST	Proportion of Scheduled Tribe(ST) Population in total population in the year 2001	0	0.98	0.16	0.26
ELEC	Proportion of Households with access to electricity in the year 2001	0.03	1	0.55	0.28
HH_MANU	Proportion of total main workers employed in Household Manufacturing in the year 2001	0	0.31	0.04	0.03
NHH_MANU	Proportion of total main workers employed in Non household manufacturing in the year 2001	0	0.56	0.07	0.06
EGW	Proportion of total main workers employed in section E (Electricity,Gas and Water Supply) in the year 2001	0	0.05	0.01	0.01
CONSTR	Proportion of total main workers employed in section F (Construction) in the year 2001	0	0.17	0.04	0.02
TRADE	Proportion of total main workers employed in section G (Wholesale and Retail Trade) in the year 2001	0	0.28	0.07	0.04
HOTEL	Proportion of total main workers employed in section H (Hotels and Restaurants) in the year 2001	0	0.05	0.01	0.01
TSC	Proportion of total main workers employed in section I	0	0.13	0.03	0.02

	(Transport,Storage and Communications) in the year 2001				
FINANCE	Proportion of total main workers employed in section J and K(Financial Intermediation, Real Estate, Renting and Business activities) in the year 2001	0	0.16	0.03	0.02
OTH_SERV	Proportion of total main workers employed in section L to Q (Public Administration, Defence, Education, Health, Social Work, Other community and social services)in the year 2001	0.03	0.60	0.11	0.07
SECONDARY	Proportion of total main workers with matric/secondary level education(but below graduate) in the year 2001	0.02	0.33	0.11	0.06
DIPLOMA	Proportion of total main workers with a technical diploma(not equal to degree) in the year 2001	0	0.06	0.01	0.01
DEGREE	Proportion of total main workers with a technical degree or diploma equal to degree in the year 2001	0.01	0.26	0.05	0.04
GRADUATE	Proportion of total main workers with a graduate degree and above(other than technical degree) in the year 2001	0	0.06	0.01	0.01
TOILET	Proportion of households having toilet facility within the house in the year 2001	0.05	0.97	0.37	0.23
WATER	Proportion of households having access to drinking water from any source within premises(within 100mts for urban areas and 500 mts for rural areas)	0.02	0.93	0.35	0.21
HOSP	No of hospitals(allopathic, ayurvedic, unani and homeopathic)	0	462	43	51
SCHOOL	No of schools(primary, middle, secondary and senior secondary)	39	10285	1962	1279
COLLEGE	No of colleges (includes arts, science, commerce, law, medical, engineering, polytechnics and other colleges)	0	246	23	25
CINEMA	No of cinema theatres and cinema halls	0	517	36	52
BANK	No of commercial and co-operative banks	3	1787	157	167
CREDIT_SOC	No of agricultural, non-agricultural and other credit societies	0	4010	331	498
Cognitive Skills	Log of the index I_j created as per eqn (1). To create $S_{i,}$we take a mean of all the variables under cognitive abilities (See Table A1)	2.04	4.47	3.36	0.47
Motor Skills	Log of the index I_j created as per eqn (1). To create $S_{i,}$we take a mean of all the variables under psychomotor abilities (See Table A1)	1.29	4.23	2.88	0.52
Physical Skills	Log of the index I_j created as per eqn (1). To create $S_{i,}$we take a mean of all the variables under physical abilities (See Table A1)	0.70	3.83	2.35	0.53
Content Skills	Log of the index I_j created as per eqn (1). To create $S_{i,}$we take a mean of all the variables under basic content skills(See Table A1)	2.15	4.58	3.42	0.47
Process Skills	Log of the index I_j created as per eqn (1). To create $S_{i,}$we take a mean of all the variables under basic process skills(See Table A1)	2.25	4.64	3.51	0.46
Social Skills	Log of the index I_j created as per eqn (1). To create $S_{i,}$we take a mean of all the variables under social skills(See Table A1)	2.06	4.49	3.33	0.46
Technical Skills	Log of the index I_j created as per eqn (1). To create $S_{i,}$we take a mean of all the variables under technical skills(See Table A1)	0.55	3.37	2.10	0.54

Notes: The source for all the data is Census of India (2001 and 2011). All the data except for the growth rates(dependent variable) is for the year 2001. The number of observation is 592 excluding one district which didnot have spatial co-ordinates.

Notes:

[1]This framework has been adopted in over 50 studies mostly in the US (see Hoogstra, 2005).

[2]Felstead et al (2007) discusses the advantages and disadvantages of these two methods and other alternative approaches such as occupational classification, performance in tests and self-assessment methods.

[3]In UK, British Skills Survey (BSS) is used as the main source for job-requirements. A major difference between the two datasets is the unit of analysis. O-NET provides data at the level of occupation whereas BSS provides data at the level of individual. While individual level data can account for within-occupation heterogeneity, it cannot be used outside UK.

[4]Even if it is, as in the case of BSS, it can be used only for that sample.

[5]We use version 17.0http://www.onetcenter.org/database.html.

[6]It is important to note that the skill index calculated for each occupation S_i is a cardinal index computed from ordinal scores of occupational skill requirements. That is, the absolute value of the index for occupation A Vs B captures the extent to which these occupations are different in terms of skill requirements. The district skill index $I_j$which is calculated using this index as a weight, is also therefore a cardinal index.

[7]We can also use the total labour force. But, using labour force will capture the underemployment and unemployment as well. Hence, the skill index will be lower not only if the employed population is unskilled but also if there is huge unemployment and underemployment.

[8] For each occupation, the respondent is either an incumbent (who is currently involved in this occupation) or an occupational analyst. For example, all questions related to the occupation 'journalist' are asked to a few journalists or occupation analysts.

[9] For each variable, the description of three level scales is given in the questionnaire to elicit better response from the respondents. The descriptions of level scales for each variable can be found in the ONET database. For example, the following are the level scale descriptions for the skill 'Writing':Level 2 - Take a telephone message ;Level 4- Write a memo to staff outlining new directives ;Level6- Write a novel for publication.

[10] For example, consider two occupations A (Novelist) and B (stenographer) and the skill 'Writing'. Say, both occupations score a high value (say 5) on the *importance* of writing skill for the occupation. However, these occupations are different in the *level* of writing skill needed on which occupation A and B could score a lower value (say 3) and a higher value(say 7). By taking the product of importance and level, occupation A (with score 15) becomes quite distinct from occupation B (with score 35).

[11] ONET 17.0 follows SOC-2010 codes which has 840 detailed occupation titles under 23 major and 97 minor groups. However, ONET has data only for 758 titles. For some titles, ONET has data at a finer disaggregation. Accordingly, there are 903 titles in ONET.

[12]Census 2001 has 434 occupations. ISCO 88 has 389 occupations. There are 7 occupations in ISCO that are absent in census and 36 occupations in census that are absent in ISCO. There are 4 ISCO codes for which a detailed occupational breakup is given in census. Essentially, 20 census occupations correspond to 4 ISCO occupations.

[13] Concordance tables are available for SOC-2010 to ISCO-2008 and ISCO-2008 to ISC0-1988.

[14]For example, occupations like hand pedal drivers, astrologers, occupations in the 'others' category cannot be matched. Some occupations like 'religious workers' found matches but were later removed due to inconsistencies.

[15]Excluding cultivators and agricultural labourers. We have calculated the skill index for cultivators and agricultural labourers also. If we match the cultivators of India to that occupation code in SOC that matches by definition, we would substantially overestimate the skills of cultivators. Hence, we match it according to the tasks involved.

[16] We divide the sample into ten deciles based on the values taken by 30 day per capita expenditure.

[17]The data for these variables have been taken from the village amenities and town directory information in the 2001 census. The values of these variables are the sum of data at the village level and the town level.

[18]Primary Census Abstract, Census of India 2011

[19]Out of 593 districts, 540 were unchanged in 2011. 35 districts were neatly divided to form 71 districts in 2011. The figures for these districts had to be added to correspond to the original district. Remaining 18 districts underwent more complicated changes to form 29 districts in 2011. For these districts, the population proportions of the corresponding districts in both the years were used to adjust the employment figures.

REFERENCES

Abraham ,V. (2009) "Employment Growth in Rural India:Distress Driven?", *Economic and Political Weekly*, April 18.

Aedo , C., Hentschel ,J., Luque,J. and Moreno,M. (2013) "From Occupations to Embedded Skills: A Cross-Country Comparison", Background paper for the World Development Report 2013

Autor,D.H. and Handel,M.J. (2013) "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages", *Journal of Labor Economics*, Vol. 31, No. 2, The Princeton Data Improvement Initiative, pp. S59-S96

Bacolod, M., Blum, B.S. and Strange, W. (2009) "Skills in the City", *Journal of Urban Economics*,65:136-153

Berry,C.R. and Glaeser,E.L. (2005) "The Divergence of Human Capital Levels Across Cities", *Papers in Regional Science*, 84: 407-444.

Boarnet, M. G. (1994) "An Empirical Model of Intrametropolitan Population and Employment Growth", *Papers in Regional Science*, 73, 135-153.

Carlino, G.A. and Mills,E.S. (1987) "The determinants of county growth", *Journal of Regional Science*, 27, pp. 39-54.

Chaudhuri, S and Gupta, N. (2009) "Levels of Living and Poverty Patterns: A District-Wise Analysis for India, *Economic and Political Weekly*, Feb 28.

Felstead, A., Gallie , D., Green, F. and Zhou,Y. (2007) "Skills At Work, 1986 to 2006", SKOPE, Oxford University

Feser, E. (2003) "What regions do rather than make: A proposed set of knowledge-based occupation clusters", *Urban Studies*, 40 (10): 1937-1958.

Gleaser, E.L., Scheinkman, J.A. and Shleifer, A. (1995) "Economic Growth in a Cross-Section of Cities", *Journal of Monetary Economics*, 36: 117-143.

Glaeser, E., A. Saiz, 2004. "The Rise of the Skilled City", *Brooking-Warton Papers on Urban Affairs* 5, 47-94.

Henry, M.S., Schmitt, B., Kristensen. K, Barkley . D.L. and Bao, S. (1999) " Extending Carlino-Mills Models to Examine Urban Size and Growth Impacts on Proximate Rural Areas", *Growth and Change*, Vol 30, pp 526-548.

Henry, M.S., Schmitt, B., and Piguet, V. (2001) "Spatial econometric models for simultaneous systems: Application to rural community growth in France", *International Regional Science Review*, 24, pp. 171-193.

Hoogstra, G.J., Florax, R.J.G.M., and van Dijk, J. (2005) "Do 'Jobs follow People' or 'People follow Jobs'? A meta-analysis of Carlino-Mills Studies", Paper prepared for the 45th Congress of the European Regional Science Association 23-27 August 2005, Amsterdam, the Netherlands.

IMF (2014), "Report for Selected Countries and Subjects", World Economic Outlook Database .

Kelejian, H.H., andPrucha.I.R.(2004) “Estimation of simultaneous systems of spatially interrelated cross sectional equations”, *Journal of Econometrics* 118, 27–50.

Lucas, R. (1988) “On the mechanics of economic development”, *Journal of Monetary Economics*,22, pp. 1–42.

Mehrotra, S., Gandhi,A.,Sahoo,B.K. and Saha ,P.(2012) “Creating Employment in the Twelvth Five Year Plan”, *Economic and Political Weekly*, May12

OECD (2012), “India: Sustaining High and Inclusive Growth”, OECD Better Policies Series

Poelhekke, S. (2009) “Human Capital and Employment Growth in German Metropolitan Areas: New Evidence”, DNB Working paper no 209

Ramaswamy. K.V., (2007) “Regional Dimension of Growth and Employment”, *Economic and Political Weekly*, Vol. 42, No. 49 (Dec. 8 - 14, 2007), pp. 47-56

Rey, S.J., and Boarnet.M.G.(2004) “A taxonomy of spatial econometric models for simultaneous equations systems”. In: L. Anselin, R.J.G.M. Florax& S.J. Rey (eds.), *Advances in spatial econometrics. methodology, tools and applications.* Berlin: Springer, 99–119.

Romer, P. (1986) “Increasing returns and longrun growth”, *Journal of Political Economy*, 94,pp. 1002–1037.

Scott, A.J. (2009) “Human capital resources and requirements across the metropolitan hierarchy of the USA”, *Journal of Economic Geography* 9: 207-226

Shapiro, J. (2006): “Smart Cities – Quality of Life, Productivity, and the Growth Effects of Human Capital”, *The Review of Economics and Statistics* 88 (2): 324-335.

Simon, C.J.(1998) “Human capital and metropolitan employment growth”, *Journal of Urban Economics*, 43:223—243, March 1998.

Simon, C. J., and Nardinelli, C. (2002) “Human capital and the rise of American cities, 1900-1990”, *Regional Science and Urban Economics*, 32:59—96, January 2002.

Steinnes, D., and Fisher, W. (1974) “An econometric model of intraurban location”, *Journal of Regional Science*. 14: 65-80.

Suedekum, J. (2006) “Human Capital Externalities and the Growth of High and Low Skilled Jobs”, IZA Discussion Paper No 1969

Thomas, J.J. (2014) “The Demographic Challenge and Employment Growth in India”, *Economic and Political Weekly*, Feb 8.

Wheeler ,C.H. (2006) “Human Capital Growth in a Cross Section of U.S. Metropolitan Areas” Federal Reserve Bank of St. Louis *Review*, March/April 2006, *88*(2), pp. 113-32.