ESSAYS IN REGIONAL AND URBAN ECONOMICS

By

DURBA CHAKRABARTY

Bachelor of Science in Economics
University of Calcutta
Kolkata, India
2004

Master of Science in Economics
University of Calcutta
Kolkata, India
2006

Master of Philosophy in Economics
University of Calcutta
Kolkata, India
2009

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
May, 2018

ESSAYS IN REGIONAL AND URBAN ECONOMICS

Dissertation Approved:

Dr. Dan S. Rickman

Dissertation Adviser

Dr. John V. Winters

Dr. Mehtabul Azam

Dr. Dave Shideler

ACKNOWLEDGEMENTS

First, I want to thank my advisor Dr. Dan S. Rickman for his time, dedication, and infinite patience during my Ph.D. pursuit. Besides, I would like to thank the rest of my committee members for their insightful comments. I would like to thank my fellow classmates in the economics department for the great times that we have shared together. I am particularly thankful to Michael, a good friend, with whom I had exciting discussions and debates that allowed me to thrive in my doctoral journey. I am grateful for time spent with friends and their continuous presence made my life in Stillwater enjoyable.

I dedicate this thesis to the memory of my Baba, who is my role model and remains a constant source of inspiration. I embarked on this journey due to my Ma and Baba. Lastly, this Ph.D. had been incomplete without the tremendous support, love and sacrifice of my patient husband, Rudra.

Name: DURBA CHAKRABARTY

Date of Degree: MAY 2018

Title of Study: ESSAYS IN REGIONAL AND URBAN ECONOMICS

Major Field: ECONOMICS

Abstract: My dissertation contains three applied essays. The first essay examines how proximity to higher-tiered centers in the urban hierarchy affects the population growth. The massive population growth in India is not driven by amenities. I find a negative and statistically significant effect of the nearest city distance variable on the growth rates of towns. In terms of distance penalty for towns, it is approximately 6.1% less population growth given the mean distance to its nearest city averaged at 53.57 km. Thus, my findings lend support to the hypothesis of urban hierarchical effects as evident in other countries like United States, Canada and China.

The second essay document assimilation patterns of broad race and Asian immigrants groups on attaining a STEM major by different ages of arrival. Among the child immigrants, the early arrivals (0-5 years) are less likely to specialize in STEM major compared to the late arrivals for white and Asians. The assimilation pattern in terms of attainment of STEM majors for the immigrant groups depends on the length of stay in the destination country. The later age of arrival groups (12-17 years) have already acquired a larger part of K-12 education in STEM driven countries of Asia like India and China and developed their math and science skills.

The third essay examine the spread and backwash effects of urban growth on the hinterlands of India. The dynamics of rural areas are unique in the context of India and needs to be addressed separately. The uniqueness stems from the definition of rural areas by the Census of India. A negative and statistically significant effect of the nearest town distance and incremental distance to a city variable on the growth rates of villages is found in this study. Thus, the villages experience spread effects for being in close proximity to a town and a city but backwash effects for being closer to a large city. The distance penalty or protection varies by the position of tier in the urban hierarchy.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

CHAPTER I

POPULATION GROWTH IN INDIA: DOES PROXIMITY TO URBAN HIERARCHY MATTERS

1. Introduction

There is a need to study the increasing role of distance in the urban hierarchy for India. The Urban India is rapidly growing and expanding in terms of boundaries with the level of urbanization increasing from 27.81% in the 2001 Census to 31.16% in the 2011 Census. The urban centers differ in size, density and in terms of characteristics. There are different tiers in the urban hierarchy, where the higher tiered centers offer a higher end of goods and services and economic opportunities compared to the lower-tiered centers. Studies find that the trajectory of the urban hierarchy in India is stable, as most of the growth in urban India has mainly occurred in large metropolitan areas compared to small and medium size cities (Schaffar & Dimou, 2012). There exists a large regional literature in the US on proximity to urban agglomerations (Partridge et al., 2008a; Partridge et al., 2008b; Partridge et al., 2009; Partridge et al., 2010). In addition, there are studies done in Canada (Partridge et al., 2007) and China (Chen & Partridge; 2013). Until now, no such study has been carried out in India. Thus, this paper aims to study the proximity of higher-tiered urban centers in explaining the population growth across towns from 2001 to 2011 and the role of amenities in spatial distribution of population across India. The spatial studies in India have focused on agglomerations economies for formal and informal sectors (Lall et al., 2004; Ghani et al., 2012;

Ghani et al., 2014; Mukim, 2015). Even various studies have been carried out in determining the growth rates based on initial population sizes and also on the relationship between ranks and size of cities (Soo, 2012; Luckstead & Devadoss, 2014 a; Luckstead & Devadoss, 2014 b; Chauvin et al., 2017). The growth in India among different states have been tested based on economic indicators like per-capita income and well-being (Chitke, 2011; Mallick, 2014; Arora & Ratnasiri; 2015). Recent study in India looks at the distance effect on urban transformation in the national capital region of India (Jain, 2017).

Another important aspect is the role of geography and amenities (weather, man-made amenities) in shaping up the spatial distribution of population. The importance of geography holds true for the US, as there is amenity-driven migration. Though most of the growth in the US and China took place along the coastal areas, no such conclusion can be made for China (Chen & Partridge, 2013). There exists a mixed literature on the contribution of weather to the growth of European cities. While Cheshire & Magrini (2006) find amenities matter only on a national scale for the European countries, Rodriguez-Pose & Ketterer (2012) observe amenities play an important role in explaining population patterns across Europe. In the case of Canada, the growth is urban centric and weather hardly plays any role as there is little variation in climate across Canada (Partridge et al., 2007).

The population growth in India has been urban centric like Canada and China. The development trajectory for most of the countries in the world including the US and even China has been a shift from agriculture to industries and then to the services sector. However, India's growth path has been unique because of its transition from the agriculture to the service sector. In addition, the growth driven by the services and to a lesser extent by the manufacturing has taken place in high-density clusters unlike the US and China where the medium density locations are the major drivers of growth (Desmet et al., 2015).

I employ a cross-sectional population growth equation to estimate the percentage change in population growth rate of Indian towns from 2001 to 2011. A key feature of this reduced form empirical model are all the initial period values of the explanatory variables are considered and additionally, the distance to different man-made amenities (railway station, hospital, college and school) have been reported rather than the actual number of amenities, which lessens the problems of endogeneity. There is also advantage in terms of the Indian census data being used. The disaggregated level of data on towns helps to capture a large degree of heterogeneity. Further, the distance variables are measured in terms of road distance in kilometers rather than straight-line distance. This reduces to some extent the measurement error created by using straight-line distances.

One can easily conclude that the massive population growth in Indian towns is not driven by natural amenities. The weather has an insignificant role to play in explaining the growth patterns of towns, as these low-income developing countries do not have the means to pay for nice weather. They care about other necessities in life. Even for man-made amenities, the distance to some of these infrastructure like railways, hospitals, colleges and schools have a negative but not statistically significant effect on population growth. Thus, no clear definite conclusion can be inferred that the growth is driven by amenities.

The geographical position of an area with respect to its urban hierarchy has an influence on population. The farther a town is away from the higher tiered urban centers, the lower is the growth rate of population. I find a negative and statistically significant effect of the nearest city[1] distance variable on the growth rates of towns. Every kilometer farther away from the city is associated with approximately 0.113% less population growth in towns for 2001-2011. In terms of distance penalty for towns, it is approximately 6.1% less population growth given the mean distance

[1] Towns with a population of 100,000 and more.

to its nearest city averaged at 53.57 km (Table 1). There is a cost for remoteness in terms of lower population growth. Each of the higher ordered urban tier offers access to additional amenities and services, and thus there is an incremental distance penalty. There is the growth penalty of approximately 0.05% per incremental kilometer to reach a large city[2], which are higher ordered city. As expected, the quadratic terms associated with the distance variables have a declining marginal effect. Thus, the penalty varies across different tiers in urban hierarchy. Irrespective of the size of the urban center, the agglomeration effects will be greatest in the highest tiered center. The results are robust using different specifications which proves the initial hypothesis that one must incur costs for remoteness in terms of lower population growth. The agglomeration effects have much wider reach in terms of geography. The findings indicate that the proximity to cities and higher tiered cities are important in explaining the urban hierarchical effects for towns in India. Thus, the results lend support to the hypothesis of urban hierarchical effects as evident in other countries like the US, Canada and China (Partridge et al., 2007; Partridge et al., 2008a; Partridge et al., 2008b; Chen & Partridge, 2013).

The remainder of the paper is organized as follows. Section 2 presents a brief background of India covering the pre, post-independence area, and the economic reforms in 1991 and effects of such economic reforms on regional growth in India. Section 3 discusses the geography and the population patterns of India. Section 4 discusses the spatial studies carried out in India followed by the conceptual framework and the related literature in section 5. Section 6 discusses the data and the definitions used. Section 7 specifies the empirical approach used and section 8 presents the main empirical findings. Finally, section 9 describes different robustness checks followed by conclusion in section 10.

[2] A large city is a town with a population of 500,000 and above.

2. Background of India

2.1 Pre-Independence Era

There is a long history to follow but I mention the ones that have relevance in shaping today's spatial India. The Mughal Empire ruled over large parts of India during the 16th and 17th centuries and centered its operations in Delhi, which is now the capital city of India. Different European countries colonized India at different periods starting with Portuguese and then the French, Danish, Dutch and British who had initial trade relationships with India. These European colonizers opened different trade stations along the coastal regions. However, the British expanded its operation through setting up the East India Company in 1612 (Winser, 2002; Berglee, 2012).

The East India Company developed the main port cities of Bombay, Calcutta, and Madras, now known as Mumbai, Kolkata and Chennai respectively. These were the nodal points of exchange of goods with the markets between India and Europe. These port cities are still of relevance today due to export and import relations with the rest of the world and serve as major industrial centers in India. Madras served as a port to southern India. Bombay was the financial center and became the largest city. Calcutta was the capital of Britain India for a large time. In 1912, the British shifted their capital from Calcutta to New Delhi, which is currently the capital of India too. Among the reasons stated, one was that Calcutta was located at the extreme eastern part of the country and was detached from the rest of the parts of the country. Delhi, on other hand was located in the northern part and the British could get better access to interior India (Berglee, 2012).

2.2 Post-Independence Era

Although India became independent in 1947, it faced several challenges. There were political problems of the integration of over 500 princely states[3] to form a united nation. Another was the

[3] The princely states were autonomous, ruled by a local monarch unlike the British provinces, and had an alliance with the powerful British Raj.

accommodation of mass migration of people from Pakistan[4] due to partitioning of colonial India into two separate nations along religious lines in 1947. Moreover, there were economic problems such as poverty, underdevelopment, inequality, illiteracy and so forth.

To liberate from these economic problems, India adopted the concept of economic planning along the socialistic lines of Russia. A series of Five-Year Plans have been undertaken since then, setting goals and targets according to the needs of the economy at that point in time. Despite these plans and policies, there was stagnation of economic development. Most of the earlier Five-Year plans in India laid emphasis on a growing public sector with massive investments in basic and heavy industries. The public sector was protected by various forms of licenses and tariffs, which led to poor performance of different sectors in the economy and constituted a major bottleneck on the conduct of business activity. Another primary drawback of such five-year plans was uniform targets set across the nation without paying attention to the social, as well as the economic, differences that exist in different states of India (Marshall, 2001; Bates, 2011; Chitke, 2011; Sharma, 2013).

2.3 Economic Reforms

India was lagging in terms of growth rates behind most of the developed and developing nations in the world due to rising fiscal deficits and a series of balance of payment crises. In 1990, the rise in oil prices due to the Gulf War further exacerbated the balance of payments problem. To improve economic performance and to solve the balance of payment crisis, a comprehensive set of economic reforms, known as economic liberalization was undertaken in 1991. It had three features of liberalization, privatization and globalization of the economy. There was the abolition of "License Raj" whereby the central planners had the ultimate authority to decide on products being manufactured by which sectors and also on the amounts of being manufactured. In addition, the

[4] Pakistan was formed to become a Muslim majority nation, and India, to be the Hindu majority nation.

import licensing was put to an end on all intermediate and capital goods and there has been a reduction in tariff rates. Moreover, it is still higher compared to other Asian countries.

In essence, these reforms freed the domestic economy from the state control and pushed the way for a market-oriented economy. To keep along the lines of reforms of 1991, from the Ninth Five-Year Plan (1997-2002) onwards, there has been less emphasis on the public sector and an increasing role for the private sector. Another crucial reform was to attract the foreign direct investment (FDI) in India. The New Industrial Policy in 1991 further improved the scope of FDI in priority industries and raising the percentage of foreign equity in infrastructure industries (Panagriya, 2001; Chitke, 2011; Statistical Book, 2017).

2.4 Post-Reform Period

The economic reforms of 1991 favored the rich states located along the southern and western coasts, particularly Gujarat, Maharashtra, and Tamil Nadu, which had strong industrial bases. The Western region is much ahead of the Eastern region because of the rich states of Maharashtra and Gujarat, which performed well even after the economic reforms. The state of Maharashtra is spearheaded by Mumbai having well-developed financial markets, high-tech firms (computer) and manufacturing industry (auto manufacturing and film industry). In addition, the location of ports provides easy access to the international markets. Gujarat benefits from having a large coastline and possesses oil and natural gas resources. Further, it made huge improvements in electricity and infrastructure. It also created an environment for attracting private and foreign investment, leading to large business opportunities. The Eastern region was naturally endowed with minerals in the states of Bihar and West Bengal, which led to the development of heavy industries. With the passage of time, these have reduced to rustbelt status with closing down of factories and industries. The Central region consists of the low-income state of Madhya Pradesh, which improved after

reforms. The states of North East region are the lagging states in terms of measures of welfare and development (Chitke, 2011; Berglee, 2012).

After liberalization, the southern region (Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu) performed much better than the northern regions. An important sector driving the economic growth is the booming ICT (Information and Communication Technology) and related services (Business Process Outsourcing) sector. India has been exporting mainly telecommunications and information technology services to most of the multinational companies across the world. Information Technology (IT) has been the largest contributor to GDP and generated the second largest employment just after agriculture. The state of Karnataka is known as the Silicon Valley of India as it is the leader of India's total software exports. Tamil Nadu ranks third in software exports and is home to the largest BPOs (business process outsourcing), such as the World Bank, Citibank and others. The development of Hyderabad in Andhra Pradesh is worth noting as it is a major IT hub and also has a large number of BPOs. Kerala's development is driven by high levels of literacy, and other measures of human development (World Bank, 2009; Chitke, 2011; Berglee, 2012).

Thus, this sudden spurt of services sector in last two decades has shaped up the spatial development in India. India's growth path has been unique because of its transition directly from the agriculture to the service sector. For most of the other countries in the world like the US, and even China, the path of development has been a shift from agriculture to industries and then to the services sector.

3. Geography and Population Patterns of India

India lies in the northern hemisphere, has a long coastline, extending to about 7,500 km. The entire nation is divided into twenty-eight states and seven union territories[5] based on the recent Census

[5] The difference between the states and union territories is that states have their own elected governments but union territories are ruled directly by the Central Government.

of 2011. There are mainly six broad regions in India, the North, the South, the East, the West, the Central and the Northeast. At the next level, are the districts[6], which are the second level of administrative division of the country after states and union territories. The lowest administrative unit is town in urban areas. One thing to note is the creation of three new states from original states in November 2000. These are Jharkhand, Chhattisgarh and Uttarakhand (formerly known as Uttaranchal) carved out of the southern part of state of Bihar, southeastern districts of state of Madhya Pradesh and Himalayan and adjoining northwestern districts of state of Uttar Pradesh. Since my data correspond to the periods from 2001 to 2011, there are no issues with formation of new states.

There is a wide variation of topography in India, as it comprises mountains, plateaus, plains, deserts and coasts. However, there is no such tendency to concentrate along coasts such as the US and China despite having a long coastline. Most of the concentrations in India are along the major rivers of Ganges and Godavari, and hence more inward-oriented. These places were home to ancient civilizations, because of fertile soils and easy availability of water for agriculture. Initially, the migration took place from the lagging states of Bihar, Orissa, Rajasthan, and Uttar Pradesh to agriculturally prosperous northwest states of Gujarat, Maharashtra and Punjab. There have been shifts in migration from these backward states to cities in search of jobs following the Second Five-Year Plan (1956-1961), whose focus was on rapid industrial development of the country. Even after economic reforms in 1990, there have been incentives for rural people to migrate to urban areas for work. Most of the migration is inter-district rather than inter-state. The highest level of migration is observed within the same district. (World Bank, 2009).

The current growth in India is geographically scattered but concentrated along few urban areas. Desmet et al. (2015) finds the large mega cities of India that are driving the spatial

[6] These districts are equivalent to counties in the US and China.

development unlike the US and China where medium density locations are the major drivers. The spatial concentration of industries depends on the age of the industry. The services industries (considered as young) are located in medium-sized locations like California's Silicon Valley, Boston's Route 128, and the North Carolina Research Triangle and others in US. However, in India, it is the "young" services industry (driven by strong agglomeration forces) and to a lesser extent the "less mature" manufacturing industry that are located in high-density locations. There is no clear explanation for the obstacles faced by these medium-density locations in their growth but find evidence of low levels of highly educated persons and the inadequate local infrastructure, in terms of poor access to telecommunication services.

In addition, given the geographical size, most of the densely populated cities in the interior are not as far away from coasts as in the cases of China and the US and thus have an added advantage. Other reasons for less concentration along the coasts can be attributed to the growth in the services sector. For its operation, one needs to have good access to the internet and telephones, rather than being close to a coast for exporting manufactured goods as in China (Sachs et al., 2002; World Bank, 2009; Soo, 2012; World Bank, 2013b).

The spatial aspect is an important feature especially for developing countries like India where economic geography is still shaping up. In explaining the growth in regions, both Gibrat's and Zipf's law have been tested for different sets of countries. While Gibrat's law states that growth rates of cities should be independent of initial population sizes, on the other hand, the sizes of cities are inversely proportional to ranks according to Zipf's law. Chauvin et al. (2017) observes that both these laws hold for the US and Brazil but the situation is different in case of India and China. However, Gibrat's law holds for the large cities in India. Zipf's Law does not hold for India, but holds for China and Brazil (Luckstead & Devadoss, 2014b). Initially, China had restrictive mobility of people from rural to urban areas imposed by hukou system. In recent decades, there has been relaxation of hukou requirements for China and thus the distribution of population in large cities

are close to Zipf's Law. Although there exists no such restrictions in India but there are mobility restrictions in terms of variations in ethnic groups, languages, culture and traditions all across India (Luckstead & Devadoss, 2014a; Soo 2012).

4. Spatial Studies in India

The spatial research on India is focused on industrial concentrations or agglomerations[7]. One of the reasons being easy access to the firm level data[8]. Lall et al. (2004) finds agglomeration economies hold in formal sectors such as the electronics, computer equipment, and machine tools. Among different factors, market access and proximity to transport hubs explains net benefits of such agglomerations. However, it aggravates spatial inequalities as the new firms locate near places having existing industrial clusters (Lall & Chakravorty, 2005). Mukim (2015) observes the patterns of coagglomeration of formal and informal manufacturing industries in India. The buyer-seller linkages between formal and informal sectors within the same industry and technological spillovers are the significant factors explaining such coagglomerations.

Ghani et al. (2014) observes evidence of agglomeration economies among the manufacturing industries. They further analyze the spatial determinants of entrepreneurship rates in the manufacturing and services sector in India and attribute them to the local education levels, physical infrastructure and to a lesser extent on banking conditions, and labor laws. Besides the services industry, the industrial agglomerations, like the National Capital Region of Delhi, Western Maharashtra, and Chennai-Bangalore, home to the automobile industry seen expansion of agglomerations. Among several factors are the rise in demand for automobile components,

[7] Agglomeration economies explain the economic benefits of concentration of economic activities by firms and households in cities.

[8] The Annual Survey of Industries (ASI) is conducted on the organized manufacturing sector and the National Sample Survey Organization (NSSO) collects information on the unorganized manufacturing sector and the services sector.

expansion of foreign automobile components, conducive environment promoted by the policies of Indian government (Tomozawa, 2016).

Another variant of agglomeration economies is the New Economic Geography (NEG). It explains the patterns of firms and workers location and the formation of agglomeration in geographical space. Further, it employs imperfect competition, increasing returns to scale and perfect mobility of key factors in general equilibrium framework settings (Fujita and Krugman, 2004). Koo & Lall (2007) finds significant association between production activities of manufacturing firms in India and NEG. The association is slightly reduced if conditional on firm location choice. NEG is applicable to the booming services sector in India, which in turn is creating demand linkages throughout the country (Shingal, 2014).

There have also been several studies about convergence for major states and districts in India after economic reforms to figure out whether the regional imbalances have increased or diminished after these reforms (Das et al., 2015). Lall (2007) finds that the regional growth in India can be attributed significantly to development of the public infrastructure. Infrastructure also plays a key role in positioning of industries in emerging economies of China and India, especially the transportation infrastructure. Ghani et al. (2014) finds that the organized manufacturing sector grew in very large proportions along the Golden Quadrilateral[9] highway and benefited the originally positioned industries along this network in terms of allocative efficiency.

The benefits of transportation and communication networks accrued more to the lagging states rather than the developed states. Chitke (2011) observes evidence of divergence in per capita NSDP (net state domestic product) across different states both in pre- and post-reform periods. Mallick (2014) finds convergence of per capita income of Indian states by conditioning private

[9] It is a large-scale highway construction and improvement project in India connecting the four major cities of Delhi, Mumbai, Chennai and Kolkata.

investment and public investment, along with other factors of economic growth such as population growth and human capital. Arora & Ratnasiri (2015) observes that the economic well-being, i.e., knowledge, health, income, technology and infrastructure have declined after the reforms. The reforms have benefited mostly the high well-being states compared to low well-being states. Further, it leads to widening inequalities between urban and rural areas.

5. Literature Review and Conceptual Framework

5.1 Role of Amenities

The geography determines largely the spatial distribution of population. The US is such an example as the people are highly mobile compared to other parts of the world. Despite having abundant interior land, it is mostly a coastal nation as most of the population is in counties within eighty kilometers of an ocean or Great Lake (Rappaport & Sachs, 2003). The historical conditions drove the initial concentration of human settlements in coastal places. In recent times, coastal proximity has present day contributions to both productivity and quality of life.

Nice weather is valued as a consumption amenity contributing to quality of life due to rising income levels. According to the Rosen-Roback compensating differential methodology, weather is valued as the sum of the wages that a person is willing to forgo and the increased housing prices that she/he is willing to pay to enjoy it. A significant part of the migration in the twentieth century took place due to nice weather (Rappaport, 2007). Further, in some metropolitan areas with increasing numbers of high-income households combined with inelastic supply of land, rapid housing price appreciation occurred (Gyourko et al., 2013). Here, factors like inherent local productivity, amenities, or fiscal policies alone did not lead to the high value of house prices in these superstar cities as described by the standard compensating differential models in urban economics. However, the weather did not contribute to growth of Sunbelt areas since 1950s. The Sunbelt areas are places with warm winters and hot summers and so before the introduction of air

conditioning, these places were unpleasant to live. Even after the introduction of air conditioning, it has not grown because of household amenity attractiveness. Glaeser & Tobio (2008) have attributed the stronger population growth to the elastic growth in housing supply. In the US, the growth is amenity-driven leading to migration away from historically-created urban areas that is from Frostbelt to amenity-attractive areas, many in non-humid Sunbelt areas.

There exists a great deal of skepticism of amenity-related migration in Europe. Cheshire & Magrini (2006) find the importance of weather in explaining population growth but only within European Union countries. Besides economic factors, amenities play an important role in migration patterns across Europe according to Rodriguez-Pose & Ketterer (2012). The Canadian population growth was driven by the metropolitan areas from 1981-2001 supporting the hypothesis of cities being engines of growth. The role of amenities in explaining population growth in Canada is not important due to less climactic variation (Partridge et al., 2007). China too has urban centric growth patterns like Canada. In addition, the scale is much larger in China. Most of the growth in China is driven by megacities on coast (Chen & Partridge, 2013), but it is not clear whether the growth is amenity driven or not.

5.2 Urban Hierarchy and Distance

Urban places differ in size and it terms of characteristics. There are different tiers in the urban hierarchy that vary in terms of economic activity and range of services provided. For example, the higher tiered centers offer a higher end of goods and services, such as legal, accounting, and management services, operas, nice restaurants and higher order entertainment services compared to the lower tiered centers. It follows from Central Place theory (CPT) developed in 1933 by a German geographer, Walter Christaller, who studied the settlement patterns in southern Germany. He showed where and how the central places in the urban hierarchy would be functionally and spatially distributed.

According to the theory, the market town having a central location will provide goods and services to neighboring areas. Thus, a hierarchy forms based on provision of goods and services. The high-ordered central places are settlements, which provide a higher range of goods and services than other places. The number of such higher-order places are fewer and more widely distributed than lower-order places. Further, the settlements having the same hierarchical level will have the same types of business and all higher-order central places must contain all the business types already contained in lower-order places. However, Lösch modified CPT in 1954 to capture the real world. The settlements of the similar size may not have the same arrangement of business types, and the higher-order central places do not need to possess all the functions available in lower-order places (King, 1985).

The trajectory of the urban hierarchy in India is quite different from that of other countries, including China. Schaffar & Dimou (2012) finds that India urban hierarchies are more stable than Chinese urban hierarchies. Most of the growth in urban India have taken place in large metropolitan areas compared to small and medium size cities. In China, with the relaxation of hukou restrictions, there has been increase in the number of medium and small size cities. Also, there has been control over mega cities due to anti-megacities policies followed by China (Desmet et al., 2015).

With the innovations of technology, it is expected that the role of distance has diminished over time. Though there has been debate regarding the role of distance in regional economic growth and development, one cannot deny the importance of distance in accessing different tiers in urban hierarchy. Several regional studies look at the proximity to urban agglomeration as measured by distance to higher tiered areas. The proximity effects in terms of employment growth on the US counties holds for rural and nonmetropolitan samples in addition to the metropolitan sample between 1970 and 2000, particularly for the 1990s (Partridge et al., 2008a). Fallah et al. (2013) even finds the role of distance affecting specific sector employment like high-technology sector. As the distance to large cities in urban hierarchy increases, there is negative impact on high

technology sector employment. Partridge et al. (2009) studies capitalized penalties for remoteness into factor prices. They observe significant distance penalties on median earnings and housing costs both for rural and urban counties across the US urban hierarchy. Further, Partridge et al. (2010) assess the relative importance of proximity to urban consumer amenities and production spillovers in explaining the changing factor price penalties for remoteness.

Chen & Partridge (2013) uses both the CPT and NEG framework to study spread and backwash effects of cities on GDP and employment growth in counties of China. The NEG models explains regional development in terms of market potential differences and hence is suitable to Chinese economy due to its large markets on the coast. However, the heterogeneous effects across the Chinese urban hierarchy supports the CPT framework. Thus to explain the differential urban hierarchical effects in India, it will be suitable to employ the CPT to study the population growth in Indian towns.

5.3 Penalty Distance to Urban Hierarchy

The urban hierarchy in India will be different from the developed countries of the world. Instead of offering higher end services and amenities, the urban places will differ in terms of basic amenities such as hospital, schools, college and infrastructure like transportation facilities. I am looking at three tiers of urban hierarchy in India. For the urban sample, at the lowest tier are the towns. The next higher level of hierarchy are the cities, one whose population is 100,000 and more. At the top level, are the large cities, those with population 500,000 and above. As stated earlier, there exists considerable differences across these tiers in terms of quality of amenities and services being provided. Both the households and firms from a lower tiered town have to travel to the nearest urban center, which is a city to access amenities, services, jobs etc. There is an incremental distance penalty for travelling to a large city as each high tier city offers additional or high quality amenities.

I measure the distance penalty in the same manner as Partridge et al. (2009). It is referred as a penalty as one is incurring a cost in terms of less population growth due to increasing distance. Here penalty is defined as the sum of added penalties on population growth for a given town *i* in the *j*th tier of the urban hierarchy.

$$Penalty_{ij} = \sum_t d^t \varphi^t$$

Suppose an area is at the lowest tier in urban hierarchy, a town, then people have to travel distance d_1 to access the nearest urban center (city) and d_2 for travelling furthest to a higher ordered city (large city with a population of 500,000 and more), and thus an incremental distance of (d_2-d_1). At the same time, being farther away from a city and a higher tiered city will incur a penalty captured by φ. For example, Murud is a town (population of 12,552[10]) in the state of Maharashtra. The nearest city is Panvel (population of 104,058), which is at a distance of 102 kilometers (d_1) from Murud. The topmost large city is Mumbai (population of 11,978,450) and the incremental distance from Panvel to Mumbai is 16 kilometers (d_2-d_1). Thus, Mumbai is 118 kilometers away from the town of Murud. The location of tier in urban hierarchy will give different values of φ i.e., marginal penalties. Thus, the distance impacts are not the same and will have nonlinear distance effects in the urban hierarchy.

6. Data

The data come from the District Census Hand Book (DCHB), which forms part of the Indian Census data. The Office of the Registrar General and Census Commissioner of India under the Ministry of Home Affairs, Government of India conducts the decennial Census. The first population Census started in 1872 when India was under the British rule. The first Census of independent India began in 1951. The latest one is the 2011 census. The Census of India is the

[10] All the population figures are from 2001 Census data of India.

largest statistical database in providing information on demographic and other socio-economic variables of India.

The DCHB contains both the Census data and the non-Census data on various civic amenities and infrastructural facilities. I am using part-A of DCHB as it has data at the town level of the district, which constitute the town directory. The town directory consists of different natural amenities and infrastructural facilities. The natural amenities include information on rainfall in millimeters, maximum and minimum temperature in centigrade. These physical aspects have been recorded by taking the average for the last ten years, i.e., from the years 1990 to 1999 for the 2001 census. The infrastructural facilities include education, medical, drinking water, communication and transport, post and telegraph, electricity, recreational, banking, and other miscellaneous facilities. The incremental distance to a large city is computed as the distance from the city (population of 100,000 and more) to the nearest large city (population of 500,000 and above). The following information has been collected from Maps of India[11]. Each of this city corresponds to the nearest city associated with each of towns in the sample. I pick up the year 2001 and 2011 as the town directory for these years contain detailed information on distance variables, that is the distance to the nearest city and also the amenity distance variables.

One of the biggest advantages is the disaggregated data at town level. Most of the earlier literature considered states or districts as units of analysis as the data are of better quality and reliable at the state and district level. However, for detailed spatial analysis, it is better to have disaggregated level of data, which will capture a greater degree of spatial dimension. Further, the districts are wider geographic areas that include the rural population. Since our focus of analysis is on urban areas, it is better to have separate and demarcated data on towns. Further, it captures a larger degree of heterogeneity in terms of demographic and economic nature of variables.

[11] https://www.mapsofindia.com/aboutus.html

A town is defined by the Census of India as areas where population is 5,000 and above; the density of population is 400 per square kilometer (i.e. 1000 per sq. mile); and at least 75 per cent of the male main working population engaged in non-agricultural pursuits. There exist different categories of towns based on population size. These are as follows: (i) Class I towns are those with population of 100,000 and above (ii) Class II towns have population between 50,000 to 99,999; (iii) Class III towns are those with population of 20,000 to 49,999 (iv) Class IV towns have population between 10,000 to 19,999; and finally (v) Class V towns have population between 5,000 to 9,999. Class I towns are also known as cities. These cities and towns form the sample of urban areas. Another type of town are the district headquarters. These are home to administrative offices and serves as the seat of the state government. There are more than one headquarter town for each of the states.

There are 5161 towns according to 2001 Census. In 2011 Census, the total towns increased to 7,935. I matched 2001 with 2011 Census data by town names resulting in 5,177 towns as some of the new towns are not part of 2001 Census data. Thus, the omitted observations is due to matching the data of 2001 with 2011[12].One needs to keep in mind while considering the figures for urban population is the statistics suffers from downward bias. As stated by experts in India and internationally, the definition of urban areas misses the population growth occurring in urban areas just outside of the official city boundary, and hence is restrictive. Despite these problems, the Census definition of urban population is at best, just an approximation of reality (Cohen, 2004; Mckinsey Report, 2010; World Bank, 2013b).

7. Empirical Specification

This chapter examines the proximity to a nearest high-ordered urban center and incremental distances to higher ordered cities affects the population growth models for urban areas. The

[12] Overall, the omission of data was by 0.3% in urban sample.

proximity to a major center (such as city) for urban areas matters, as the cities are centers of jobs, education and household amenities. Additionally, the incremental distance to a higher ordered city means better access to potential markets, services and amenities.

The base specification is a cross-sectional growth equation given below which captures the total population growth for town *i*, located in state, *s*, for the time interval from 2001 to 2011. The dependent variable is the percentage change in population between periods 0, the initial period, and t, the end period.

$$\%\Delta\mathbf{POP}_{\mathrm{is(t-0)}} = \alpha + \beta\mathbf{DEMOG}_{iso} + \lambda\mathbf{AMENITY}_{\mathrm{is0}} + \gamma\mathbf{GEOG}_{\mathrm{is}} + \theta\mathbf{DIST}_{\mathbf{is}} + \delta_s + \varepsilon_{\mathrm{is(t-0)}} \quad (1)$$

DEMOG includes the initial period population density to account for the initial agglomeration including market potential and localization effects. **AMENITY** takes into account the natural amenities, like the rainfall, the maximum and the minimum temperature variables. One concern regarding the measurement of these variables is these have been recorded as averages for the last ten years. Hence, it might not be a good representation like the mean July and mean January temperatures which account for the summers and winters in the US. One of the reasons for recording in such a manner could be that India's variety of climate ranging from the tropical monsoon in south to the temperate in north and thus making overall generalization difficult.

DIST includes the man-made amenity distance variables. These are distance to the nearest rail station, distance to the nearest hospital, distance to a nearest college and distance to nearest school. These infrastructure facilities are important in determining concentrations of people. For example, the location of railways is important, as there are concentrations of people close to those networks (World Bank, 2013a). The initial development of railways was done by the British colonial India, connecting the interior to three main port cities of Kolkata, Mumbai, and Chennai, to transport raw materials from interior to coasts. After Independence, the Indian government

expanded their railway networks and connected cities with other cities and suburbs. According to the Guinness Book of World Records, the Indian Railways are biggest in the world in terms of size (World Bank, 2009).

GEOG is a vector of spatial distance variables reflecting proximity to urban areas and its location in urban hierarchy. The first one measures the distance to the nearest city. The second includes the incremental distance to a higher ordered city to account for added benefits such as spillovers and agglomeration effects for being closer to large cities. Further, the quadratic terms of the distance to the nearest city variable and incremental distance to a large city are considered to capture the non-linearity in distance. $\boldsymbol{\delta}_s$ are the state dummy variables capturing the fixed effects. These account for time-invariant state-related growth factors such as migration flows, geographic location, regulatory policies in a state, etc., and $\boldsymbol{\varepsilon}$ is the residual. The standard errors are clustered by districts to address issues of spatial autocorrelation and heteroscedasticity.

All the distance variables report the road distance in kilometers. Most of the spatial literature considers straight-line distance due to its ease of interpretation and use. However, it ignores roads, rails and other forms of transportation and their travel times. Further, the straight-line distance faces the classic measurement error which would bias the distance regression coefficients towards zero (Lall, 2004; Partridge et al., 2008b; Partridge et al., 2014).

8. Empirical Results

8.1 Summary Statistics

Table 1 reports the descriptive statistics for the urban sample consisting of towns and cities. Column 1 and Column 2 display the means and standard deviations and Column 3 and Column 4 represents the minimum and the maximum values of the variable. The mean population levels of 2011 are higher than 2001 implying there has been growth of towns. However, the mean growth rates of

urban areas is approximately 20%, which is much lower compared to earlier growth rates in previous years. The nearest urban center is a city and the second closest higher tier urban center is a large city. Thus, the nearest city distance variable averaged 54 km while the incremental distance to a large city stands at an average of 81 km. Thus, one has to travel more to access a higher ordered city.

For the man-made amenity distance variables, the mean distances to hospitals and schools are approximately 4 km and 2 km respectively. These are much lower compared to other distance amenity variables. It implies that the basic amenities like hospitals and schools are present in almost all towns and people do not have to travel much to access these health and school amenities. However, to avail higher education like colleges and to access railway infrastructure, one has to travel further.

8.2 Regression Analysis for Towns

Table 2 shows the regression results for the 2001-2011 period town sample. One of the concerns for these cross-sectional regressions is the problem of endogeneity, which is taken care by considering the initial period value of the explanatory variables. Besides, the natural amenity variables and density, all the other variables are distance variables. Even for man-made amenities, the distance to different amenities is taken into account rather than the actual number of amenities, which might be endogenous. All these distance variables measure the road distance in kilometers.

Another way to is to add the different category of variables at each stage to identify the causal effects of those added variables. The stepwise regression includes five models. Model 1 includes only the natural amenity variables as shown in column (1). Model 2 adds to the initial model the distance to the nearest city along with its square term[13] as represented in column (2). In

[13] I also checked the model fit comparing models with logarithmic distances and the one with cubic and quartic distance effects added. Based on AIC/BIC, the smaller values are observed for the logarithmic

Model 3, the incremental distance to a large higher ordered city is further added to model 2 in column (3). In column (4), model 4, which is the base model, adds further the demographic variable and the state fixed effects, and finally model 5 includes the amenity distance variables as shown in column (5).

Population density has a negative impact on the growth rates but is statically significant at 10% level implying that denser towns have lower population growth rates. The nearest city distance variable is negative and statistically significant across all models implying that proximity does matter for population growth. Thus, the further a town is located away from the city, the lower is the population growth. An increase in nearest city distance variable by one kilometer ceteris paribus reduces the population growth of towns by 0.113 percentage points as shown in the base model. In terms of distance penalty for towns, it is approximately 6.1% less population growth given the mean distance to its nearest city averaged at 53.57 km (Table 1). The distance penalty to reach the nearest city becomes less important at greater distances as evident by the positive coefficient for quadratic terms. The results suggest that positive marginal effects from proximity to nearest city extend out about 333 km for towns.

The next closest higher ordered urban center is a large city. All the coefficients for the incremental distance to a large city are negative and statistically significant again. Thus, one will be penalized even more if the current town is not only far away from a city but also farther away from a large city. There is an added penalty of 0.049% per incremental kilometer to reach a higher ordered large city. Also, the positive coefficient for quadratic terms points that the incremental distance penalty to reach a large city becomes less important at greater distances. Given the mean incremental distance to reach a large city from a city is 80.74 km (Table 1), the incremental distance penalty results in approximately 4% lesser population growth of towns. Adding up, the total

distance model. However, based on R2 and Adjusted R2, there are slight differences between the two models, though the model with cubic and quartic distances have slightly higher R2 and adjusted R2.

distance penalty for a town to be farther away from a higher ordered city is approximately 10% less population growth.

The natural amenity variables do not have much impact on growth of towns. The only significant result holds for model 1 (in column 1) when only the natural amenities are included and it is the negative effect of the average rainfall variable, which is significant at the 10% level. However, there exists mixed results for these variables on population growth as these are measured as averages for the last ten years, which does not give a clear picture about natural amenity and topographic variables as in the related US literature. All of the man-made amenity distance variables have a negative impact on the population growth of towns. The only significant result holds for the nearest school distance at 10% level. The nearest city distance variable might actually contain these effects, as most of these amenities are present in cities. The nearest distance for these amenities might coincide with distance to nearest city for some of them.

A consistent pattern is observed for both the distance to the nearest city and incremental distance to a large city and its respective quadratic terms across all the five models. It implies strong evidence for urban hierarchical effects. Thus, the proximity to a nearest high-ordered urban center and incremental distances to higher ordered cities affect the population growth of towns in India. Moreover, one can conclude that the given findings are not driven by one particular specification or by one single model as the results holds across all specifications.

Two things need to be noted. One could argue that the model is not able to capture the large variations in town population growth as given by low R^2. Despite having low R^2, the model can be a good fit given the cross sectional level of data which produces large variation across individual units of observation. Thus, R^2 alone cannot fully explain the suitability of the model. Another is the different number of observation being reported for different specifications. Most of it is due to the missing data on the amenity distance variables. However, I ran all the five models with the same

set of observations. The resulting significance levels are the same with slight difference in magnitudes but exhibit a similar pattern.

9. Sensitivity Analysis

Though our empirical model is in general exogenous due to the predetermined nature of the explanatory variables, several other specifications are estimated to check for robustness. Table 3 re-estimates the town level regression analysis considering the initial population levels of 2001 instead of the initial population density. The resulting coefficients and t-statistics on the distance to nearest city and incremental distance to a large city are approximately the same as for the population density model, indicating that the conclusions would be essentially unchanged regardless of using population density or initial population to measure the urban hierarchy effects.

Table 4 looks at the town level population growth considering the distance to nearest district headquarter. In this urban hierarchy, the nearest urban center is the district headquarters, and the second next urban center is the proximity to a nearest city and at the topmost level urban center is the distance to a large city. All the distance variables (related to agglomeration) are negative and statistically significant lending support to urban hierarchy claims. The distance penalty is much higher when a town is located farther away from a district headquarter than farther away from a city. However, the incremental distance to nearest city from the district headquarters is still negative and statistically significant. Even the incremental distance to a large city exhibits a similar pattern.

Jawaharlal Nehru National Urban Renewal Mission (JNNURM) is a huge mission undertaken by Government of India meant for urban development focusing on the Indian cities. It was launched in 2005 for a seven-year period (up to March 2012) to encourage cities to initiate steps for bringing phased improvements in their civic service levels. The government had extended the tenure of the mission for two years, i.e., from April 2012 to March 31 2014 for lagging projects.

Given the period of my sample, this large-scale program might have an effect on the population growth of the town. Table 5 presents the regression results where I exclude from my sample the JNNURM cities and towns. The negative and significant impact of distance to the nearest city and the incremental distance to a higher ordered city on town population growth holds even for this specification. Thus, one can easily conclude that the urban hierarchy effect holds and is not driven by the massive urbanization program.

As already mentioned above, a large part of GDP growth in India is driven by the Information Technology (IT) services sector and is the second largest generation of employment after agriculture. The IT cities are Bangalore, Gurgaon, Noida, Mumbai, Hyderabad, Pune, Ahmedabad, Gandhinagar, Chennai, Trivandrum and Delhi. Therefore, Table 6 reports the regression results for the non-IT sample, which excludes the 26 IT cities mentioned above. Here again, the negative and statistically significant results for the nearest city distance and incremental distance to a large city variables point to the urban hierarchical effect and is not driven by these IT cities.

An important question is whether the urban hierarchical effects differs across fast, medium and slow growing towns. Table 7 reports the quantile regression analysis for 25th percentile, 50th percentile and 95th percentile respectively. The urban hierarchical effects as shown by the negative and statistically significant coefficients for the nearest city variable and incremental distance to a higher ordered city variable holds for all categories of towns, that is slow, median and fast growing towns. The adverse distance impacts are highest for the fast growing towns as reflected by the coefficient in the nearest city distance and also incremental distance to a large city.

Table 8 reports the town level analysis of population growth by different town sizes. The small town sample includes towns with a population of 5,000 and less than 100,000. The large town consist of towns whose populations is more than 100,0000. For both small and large towns as

differentiated by population levels, the resulting coefficients and t-statistics on the distance to nearest city and incremental distance to a large city are negative and statistically significant. However, the distance penalty to reach to a nearest city and further to reach a high ordered city is maximum for the large town sample.

Finally, Table 9 incorporates deeper lags by taking into consideration the 1991 population level. The urban hierarchical effects holds across most of the specifications. The base model which includes the natural amenity variables, vector of spatial variables, population of 1991 and the state fixed effects points out to urban hierarchical effects as evident by distance to the nearest city and incremental distance to a large city. Thus, all sensitivity analyses supports the claim that the proximity to a nearest and higher ordered urban center matters and the distance impacts on population growth differs by its status in urban hierarchy.

10. Summary and Conclusions

Although India is a part of the South East Asian countries, it possesses distinctive traits. It is the second most populated country in the world and is the world's largest democracy. The large population caters to the world market as consumer, producer and investor. It is projected that the population of India is expected to surpass that of China within a decade or two (Mckinsey Report, 2010). It is a point of concern given that such a large population will use resources for food, clothing, housing, medical care etc. However, India will have young and working population unlike other countries like China, which will face a burden of older population.

The uniqueness is seen in the population growth patterns of India. Regional population growth in India is quite different from other countries as it has been in the interior rather than along the coasts. One could argue that the interior places in India have an added advantage compared to interior places in the US and China in that these are not far away from the coasts. Another reason is that the current growth in India is mostly driven by the high-tech services, such as Information

Technology (IT), financial, and telecommunications. This led to the development of high tech cities like Bangalore, Gurgaon, Noida, Mumbai, Hyderabad, Pune and Chennai, which are located in the interior of India, except Mumbai, because exporting these services does not require access to a coast.

However, India has experienced similar trends of urban centric growth like Canada and China. There was only one city Kolkata (then Calcutta) having a population of one million in the beginning of the twentieth century. By 1991, the number of cities had increased and also the share of urban population. There are six megacities (Kolkata, Mumbai, Delhi, Chennai, Bangalore, and Hyderabad) with population of 5 million and above in 2001. The policy reforms in the 1980s and 1990s favored the coastal provinces due to export-driven growth in China leading to a divergence between the coastal and interior regions. However, India had no such preferential policies. The economic reforms in 1991 set up the path for a market-oriented economy in India. Thus, the urban agglomerations in India are market- oriented and policy-driven by schemes like Jawaharlal Nehru National Urban Renewal Mission (Sridhar, 2010).

Based on a set of countries, World Bank (2013a) finds that China, India, Indonesia, and Vietnam are in the intermediate stage of urbanization. Thus, it is expected that cities will matter, as this will generate 70 percent of new jobs given the recent trends of urbanization. People will settle for locations close to cities to enjoy improved access to jobs, markets, and urban infrastructure or amenities (Mckinsey Report, 2010; Bloom 2011). The findings also provide such insights to these projections. A strong inverse relationship exists between town population growth and the two important distance variables; the distance to the nearest city and incremental distance to a large city. Further, the positive quadratic term implies these adverse effects weaken at greater distances. In other words, a town is penalized for being located away from nearest urban center that is a city. In additional, an incremental penalty is incurred from being farther away from large cities. Thus, people have to pay a price for remoteness in terms of lower growth rate of population in town.

There is insignificant impact of weather on urban growth and no clear definite conclusion can be inferred about the man-made amenities. Hence, one can conclude that the growth is not driven by amenities unlike the US, where growth is amenity-driven leading to migration away from historically created urban areas in the Frostbelt to Sunbelt areas.

Thus, the Indian policymakers need to keep in mind the role of urban centers and its differential effects due to its position in urban hierarchy. The adverse distance impacts on population growth is evident and thus policies need to be framed keeping in mind the importance of distance penalty. Thus, one should encourage public investment to reduce transportation time and costs. The place-based policies like JNNURM has been successful in terms of inputs and processes but not much can be concluded regarding the impacts (World bank, 2013b). One should focus more on the broad based policy bringing improvement in economic indicators and quality of life considerations such as environment, health, and infrastructure. Policies should target the cities given these will be highest generator of jobs in future.

Recently, India's metropolitan cities have been stagnant and have not shown any improvement in economic indicators (World Bank, 2013a). The urbanization pattern of India is following a different path. Although urban India is rapidly growing and expanding in terms of boundaries, it is not significant. The level of urbanization in recent years (2001-2011) was lesser in extent compared to the last twenty years (Sridhar, 2010 ; Sharma, 2013; Tripathi, 2013; Das et al., 2015; Mukim, 2015). Most of the concentrations have been taking place in suburbs rather than metropolitan areas, which is termed as suburbanization. It can be attributable to poor land management practices, insufficient modes of transport etc. (World Bank, 2013b). This could also be driven by restrictive definitions of urban areas whereby there exists grey zones of large villages which have urban traits (Cohen, 2004; Urban India, 2011). The towns should be given more importance as this could reduce the load from the cities that are already dense and congested. Policies such as investment in infrastructure, like schools, colleges, hospitals etc. should be taken

for towns. Given the growth differences across the urban hierarchy exist and will continue to attract the attention of researchers and policymakers, future research can be carried out to decompose these growth differentials into productivity and amenity effects through wages and housing prices.

CHAPTER II

MATTERS ASSIMILATION PATTERN OF STEM MAJORS: ANALYSIS BY AGE OF ARRIVAL AND IMMIGRANT GROUPS

1. Introduction

Immigration plays an important role in the United States and continues to attract the attention of both policy makers and academics. To meet the demand for increased STEM workforce, there were changes in immigration laws in 1990, such as the introduction of the H-1B visa which led to a steady rise in skilled immigrants, especially the foreign-born STEM (Science, Technology, Engineering and Mathematics) workforce in the US (Hunt 2011; Peri et al., 2015). However, it may have had unintended consequences in terms of displacement of natives out of the STEM fields, and more so among minorities and women. It also may have had adverse labor market outcomes such as displacement from STEM jobs for natives (Ransom & Winters, 2016). There also exists a large body of literature on STEM workers in the US. Most of these studies have looked at different labor market outcomes like wages, employment and other factors such as filing of patents (Hanson & Slaughter, 2016; Peri et al., 2015; Winters, 2015; Ransom & Winters, 2016).

This paper use the 1% Integrated Public Use Microdata Series (IPUMS, Ruggles et al., 2017) American Community Survey (ACS) sample to examine assimilation patterns of broad race and Asian child immigrants groups on attaining a STEM major by different ages of arrival. Less is

known about the path of STEM assimilation among the first generation of immigrants and more so of child arrivals based on different ages of entry. One of the best ways of measuring assimilation is to compare the educational outcomes of the foreign-born population with that of the natives of the US. The decision of a college major is crucial for one's future. It is lucrative to study STEM fields to reap the benefits later in their careers. Most of the STEM occupations seem to be the most financially rewarding jobs. In addition, studies find that college graduates and even STEM graduates create large human capital externalities (Moretti, 2004; Winters, 2014).

The percentage of the foreign-born population attaining a STEM major conditional on having a bachelor's degree and more is high across all the broad race immigrant groups compared to natives in Table 1. It is also evident that Asians represent a higher percentage of STEM population in terms of immigrants and natives. The reasons for higher STEM attainment rates among the Asians could be driven by the curriculum, career focus and job opportunities in their respective origin countries (Jia, 2017). For detailed Asian immigrant groups, the gap between the foreign-born and native populations in attaining a STEM field is not that large except for Indians. Thus, to capture a better analysis of heterogeneity among the Asian groups, it of interest to study the STEM patterns of child immigrants belonging to different Asian groups by ages of arrival to the US.

The STEM assimilation pattern holds for white and Asian immigrants whereby the late child arrivals (12-17 years) have higher probability of attaining a STEM major conditioned on the education levels[14]. For the detailed Asian groups, it is applicable to Chinese, Other Asians, Vietnamese and Koreans. In other words, the later child arrivals have higher STEM attainment rates than the early child (0-5 years) and the middle child (6-11 years) arrivals. The early child arrivals have lower conditional probability of attaining a STEM major compared to middle child

[14] Here it refers to those having a bachelor's degree and more. The additional degrees are having a Master's degree, Professional beyond a bachelor's degree and a doctoral degree.

arrivals as these arrive in the US very earlier, stay for the longest duration, and hence absorb the American culture and education system.

This paper provides insights regarding the pattern of attainment of STEM majors among child immigrants arriving at different ages and examines the extent to which the observed patterns corroborate or contradict various assimilation theories. The current study expands on the existing research in several ways. First, the education outcomes in terms of STEM majors across immigrant race groups are studied. Previous literature considers different college outcomes like college attainment and performance in college calculus, enrollment in two and four-year public colleges and universities (Hagy & Staniec, 2002; Feliciano, 2005; Barnett et al., 2012). The average years of schooling (Rong & Brown, 2001; Gonzalez, 2003; Chiswick & DebBurman, 2004; Feliciano, 2005) and different test scores such as reading, mathematics and science (Cortes, 2006; Ohinata & Ours, 2012) have been studied. Second, this study focuses on different ages of arrival among child immigrants. These children are brought in the US by their parents and hence do not suffer from individual self-selection. Third, the different age of immigration helps in explaining the differences in choice of educational attainment between foreign-born and natives depending on the length of stay in the host country and explains the assimilation patterns of attaining a STEM major. Most of the earlier assimilation studies are on immigrant generations (Rong & Brown, 2001; Chiswick & DebBurman, 2004; Cortes, 2006; Barnett et al., 2012).

Based on different ages of arrival for child immigrants, it is clear that the late child arrivals have higher STEM attainment rates than the early child arrivals. One should weigh both the positive and negative externalities generated by foreign-born STEM graduates. The policy makers need to keep in mind the different age of arrival categories while framing immigration policies. The documentation of such assimilation trends will also help in carving out an informed higher education policy in the US. Most of these late child arrivals are supposed to acquire a major part of

their K-12 education from their origin countries. These STEM driven countries curriculum focus more on mathematics and science oriented fields compared to other fields of study.

The rest of the paper is organized as follows. Section 2 presents the conceptual framework and related literature. Section 3 describes the data and empirical specification. Section 4 describes the summary statistics and empirical results and finally Section 5 concludes the paper.

2. Conceptual Framework and Related Literature

2.1 Assimilation Theory

The classical theory explains assimilation as the integration of immigrants in the American culture and lifestyle over time. Scholars mostly from the fields of sociology and anthropology have carried out several assimilation studies. Immigrant face difficulties in adjusting to the host country as they are at a disadvantage compared to natives because of their language, culture and social skills. However, they become integrated to the culture and society, the longer is their duration of stay and thus they get assimilated. At the same time, social scientists have also stressed immigrant advantage theory whereby foreign students are at an advantage over the natives. These relate to the culture of their country like commitment to their family and emphasis on value of educational achievement (Carter & Suegra, 1979; Perlmann, 1988; Greenman & Xie, 2008).

A good way of measuring assimilation is through attainment of educational outcomes. According to the assimilation process, the newly arrived immigrants will have different education patterns than those of the host country as they have studied in their respective countries. With the passage of time and generations, they will pick up the curriculum, language and other specific skills of the host country and become similar to natives in terms of educational attainment.

2.2 Age of Immigration

Economists also have conducted several assimilation studies taking into account the age of immigration. The magnitude of assimilation depends on whether you arrive as adults or children. Chiswick & DebBurman (2004) look at the education attainment of adults by different generations (age at immigration) of foreign-born in the US. Gonzalez (2003) examines the effect of age at arrival on education and wages of the immigrant population in the US. Those that enter the US at a young age will have a relative advantage in the classroom in terms of curriculum, language and culture of the US over the immigrants who arrive at an older age. The early arrivals attend the same years of schooling as native and acquire American-specific education like developing soft skills etc. Thus, the age of arrival is an important factor explaining the assimilation patterns.

There are similar assimilation studies in other countries. Schaafsma & Sweetman (2001) considers the effect of age of immigration both on educational attainment and on earnings for Canada. They find that the immigrants arriving between ages 15–18 acquire less total education and have lower earnings than those who immigrate at a younger or older age. Ours & Veenman (2006) compare the educational attainment of young immigrants to second-generation immigrants by age of arrival in the Netherlands. Reading literacy, mathematical skills and science skills of young immigrant children in the Netherlands depend on the age at immigration and whether one of the parents is native Dutch (Ohinata & Ours, 2012).

2.3 Immigration Generations

Most of these assimilation studies focus on the educational outcomes of immigrant generations. Rong & Brown (2001) observes the educational attainment of immigrant Black youth across generations. Feliciano (2005) analyzed patterns of educational selection among immigrants to the US and educational attainment outcomes among children of immigrants. They found as immigrants' educational selectivity increases, the college attainment of the second generation also

increases. The assimilation patterns of first and second-generation child immigrants are studied especially for the enclave schools (Cortes, 2006). The education gap exists even when they arrive early. Barnett et al. (2012) figure out the effect of various immigrant generations on college student's mathematic performances. They find that both foreign-born and early immigrant generation had a higher college calculus grade than the later immigrant generations and non-immigrant students.

2.4 Race Groups & Other Outcomes

The magnitude of assimilation might vary across race groups. One can define race groups based on country of origin, race/ethnicity, and culture (Chiswick & DebBurman, 2004; Gonzalez, 2003; Feliciano, 2005). There exists considerable differences by country of origin as these countries differ by culture, language, ethnicity, and their curriculum focus and education levels. Thus, it accounts for unobserved cultural influences on the assimilation patterns of immigrants. The studies on race groups describe four broad categories, white, black, Asian/Pacific Islander, and Hispanic which are defined on lines of race/ethnicity (Hagy & Staniec, 2002; Barnett et al., 2012).

Hagy & Staniec (2002) looks at the effect of immigrant status on the college choice behavior of recent high school graduates based on immigrant generational status and race/ethnicity. In their study, Asian immigrants are more likely to attend public four-year schools and both Asian and Hispanic first-generation students exhibiting a greater likelihood of attending public two-year colleges. Stiefel et al. (2010) studies the performance in public schools of immigrant students who come to large cities such as New York. The performance of students who immigrate during high school (teen immigrants) is better compared to native-born students and the student who immigrated during middle school (tween immigrants) have an added foreign-born advantage. The immigrants on temporary work or on student/trainee visas in the US have an advantage over natives in terms of wages, patenting, commercializing and licensing patents, publishing books and papers

and writing papers for presentation at major conferences and starting successful companies (Hunt, 2011).

2.5 STEM Outcomes

Most of the STEM literature looks at labor market integration in terms of wage and occupation outcomes. Prior to 1991, there is issuance of temporary work visas (H-1 visa) in specialty occupations, which require a specialized knowledge and a bachelor's degree or equivalent. Most of these occupations are STEM related. As opposed to the earlier H-1 program, the new H-1B program reduced barriers for foreign-born persons having this temporary work visa to get permanent residency. Peri et al. (2015) find that an inflow of foreign-born STEM workers increases the wage growth for both native college educated labor and native non-college educated workers. It also boosts the economic productivity especially for college-educated workers. Ransom & Winters (2016) however, observes differing effects. The foreign-born STEM workers harm the native working population by crowding natives out of STEM fields and of STEM occupations later and the effect is stronger more for women and minority native population. Hanson & Slaughter (2016) explain patterns of high skilled immigration in STEM occupations among workers with at least a college degree.

Unlike other studies, which concentrated only on STEM workers, Winters (2014) finds that an increase in the local stock of both STEM graduates and non-STEM graduates creates positive externalities by increasing the wages of other workers in the labor market. The effect is more pronounced for STEM graduates. Boyd & Tian (2017) compare the performance of STEM educated immigrants with the native counterparts in the Canadian labor market. The Canadian STEM educated natives earn higher wages in STEM occupations compared to non-STEM occupations. Chen & Skuterud (2017) studies the labor market performance of former international

students (FISs) entering the Canadian labor market during the first decade of the 2000s to their Canadian-born-and-educated (CBE) graduates and foreign-born-and-educated (FBE) immigrants.

The choice of college major is an important decision in life as it paves the way for future careers and jobs. Liu et al. (2017) explains how the choice of college major responds to external shocks, like the Great Recession. They find that there is an increase in STEM fields majors like computer and information sciences and a decrease in business majors especially among finance and management. There is immigrant STEM attainment advantage largely among Asian and white students. The first-generation Asian and white immigrant students have an advantage over the natives in terms of attainment of STEM majors mainly due to better academic preparation in math and science in their respective schools for K-12 education (Jia, 2017). Further from the economy's point of view, the STEM fields are the major drivers of innovation and boost economic productivity. Winters (2015) finds that STEM graduates have greater effects on innovation as measured by patents compared to non-STEM graduates.

3. Data and Empirical Specification

This paper examines the assimilation patterns of child immigrants on attaining a STEM major by ages of arrival and different race groups. The data comes from 2009-2016 American Community Survey (ACS) extracted from the IPUMS (Ruggles et al. 2017). The ACS is an annual survey of one percent of the US population that provides basic information on age, sex, race, marital status, citizenship status, highest educational attainment and others. These data are suitable for our analysis as it reports the major field of study in attaining a bachelor's degree and hence crucial in defining the STEM (Science, Technology, Engineering and Mathematics) major[15]. The reason for picking

[15] There is no information on college major for persons with less than a bachelor's degree.

up the year 2009 is the ACS began reporting the field in which the person attained the Bachelor's degree and 2016 is the latest data available.

I classify STEM majors as those having a college major whose either first or second field is in STEM. The detailed list of STEM fields in terms of ACS code are in Appendix Table A1. Here, I follow the definition used by Winters (2015), where STEM college majors are based on the lists used by U.S. Immigration and Customs Enforcement. The country of birth and citizenship status determines the status of the immigrants. An immigrant is a foreign-born person who was born outside the US, and is either a non-citizen or naturalized citizen. Thus, the natives are those born in the US and include people born abroad of American parents and those born in outlying areas/ territories[16] of the US. The sample includes individuals aged between 22-60 years old. I construct the age of arrival variable as the difference between a person's actual age and number of years in the US for foreign-born persons[17].

All of the adult immigrants have attained their STEM majors in their own country and their reasons of higher STEM attainment rates will be different from those entering in the US as child immigrants. Hence, the focus of analysis is child immigrants based on different age of arrival and their attaining of a STEM major in the US. The age of arrival continuous variable is divided into three different categories. The first category comprise of child immigrants who arrived in the US at the ages between 0-5 years of age. These early child arrivals will start primary school at the same time as natives, and might behave similar to natives. The second category are the middle child immigrants who entered in the US at ages 6-11. The second group might attend some elementary and middle school in the US. Finally, the last category are the late child arrivals stepping in the US at ages 12-17. These late child arrival groups likely acquire some secondary education in the US,

[16] It includes American Samoa, Guam, Puerto Rico, US Virgin Islands, and other US possessions.
[17] Such calculation resulted in some negative values, like -1, -2, which is recoded to zero implying that these foreign-born immigrated at a very small age.

but attended primary education in their respective countries before coming to the US. The different age intervals will likely have a differential effect on attaining a STEM major due to differences in language, culture, commitment towards family emphasis on value of education and motivation towards different fields.

Next, I divide immigrants into two major groupings based on ethnicity and race. One is the broad race groups of immigrants and another is the detailed Asian immigrant groups. The broad race groups include the white non-Hispanic immigrants, black non-Hispanic Immigrants, Asian non-Hispanic immigrants, Hispanic immigrants and Other Races[18]. The detailed Asian groups consist of Chinese non-Hispanic immigrants, Japanese non-Hispanic immigrants, Filipino non-Hispanic immigrants, Indian non-Hispanic immigrants, Korean non-Hispanic immigrants, Vietnamese non-Hispanic immigrants, and Other Asian[19] non-Hispanic immigrants. This group captures a better analysis of heterogeneity among the Asian groups, as not all Asian race groups contribute in the same manner to STEM majors in the US. For all these groups, the white non-Hispanic natives is the omitted category and reference group to which it is compared to all the other race groups.

To observe the assimilation patterns of child immigrants on attaining a STEM major by ages of arrival and race groups, I use a linear probability model (LPM) of the following form:

$$P(Y_{ijt} = 1) = \theta_j Group_j + X_{ijt}\beta + \Gamma_t \qquad (2)$$

where Y is a binary dependent variable indicating STEM Major for immigrant *i*, from race group *j*, and observed in survey year *t*. $Group_j$ is a group dummy variable taking a value of 1 if an

[18] This includes the remaining races besides white non-Hispanic, black non-Hispanic , Asian non-Hispanic immigrants and Hispanic immigrants.

[19] This includes the Asian races besides Chinese non-Hispanic immigrants, Japanese non-Hispanic immigrants, Filipino non-Hispanic immigrants, Indian non-Hispanic immigrants, Korean non-Hispanic immigrants, Vietnamese non-Hispanic immigrants.

immigrant belongs to race group j and zero otherwise. As mentioned earlier, it could be broad set of race groups or the detailed Asian groups. In either of these groups, it is the white non-Hispanic natives, which is the reference group. The model controls for a vector of demographic characteristics like gender and age of person as contained in X_{ijt}. It also includes the survey year fixed effects, Γ_t. Our coefficient of interest is θ_j, which captures differences in attainment of STEM major rates between each group and the omitted group across three different age of arrival groups. For example, in case of broad race groups, the Hispanic/Asian/white/black/Other races child immigrants are compared to white non-Hispanic child natives for each of the three age categorical variables that is for ages 0-5, 6-11 and 12-17 years of age respectively. Similarly, for detailed Asian groups, the Indian/Chinese/Filipino/Other Asian/Vietnamese/Korean/Japanese child immigrants are compared again to white non-Hispanic child natives for each of the three age categorical variables. Thus, these child immigrants not only differ from natives but there might exist substantial heterogeneity even among these child immigrants. They arrived in the US at different points in time and also came from different countries bringing in their host culture.

The LPM does a good job in explaining the partial effects of the explanatory variables especially when values of the independent variables are near the averages in the sample[20]. Some of the predicted probabilities might fall outside the {0, 1} interval, but it is a matter of concern only if we have extreme values of independent variables in our model. Most of our independent variables are discrete in nature and hence, the fitted probability are simply the average y within each cell defined by the different values of x. Also, the empirical research uses it more because of the ease of estimation, and is much easier to interpret than probit or logit models, and, once the proper scaling is done, the estimated effects are often similar near the mean or median values of the explanatory variables.

[20] Wooldridge, p.455 (2002), Econometric Analysis of Cross Section and Panel Data.

4. Summary Statistics and Empirical Results

Tables 2 reports the weighted[21] STEM major rates by different ages of arrival and immigrant race groups conditional on having a bachelor's degree and more. Conditioning on earning a degree beyond a bachelor's degree helps in explaining clearly the STEM advantage among the immigrants. The seven ages of arrival groups are (1) 0-5 years; (2) 6-11 years; (3) 12-17 years; (4) 18-21years; (5) 22-24 years; and (6) 25-40 years. The first three categories are the child immigrant groups as described earlier. The fourth group contains those who will likely attain their college degree in the US. The fifth group are probably the ones who might have attained their college degrees in their respective countries of origin. Finally, the last group are those who have likely served the workforce in their respective origin countries, have gained experience, and have arrived in the US late. The last three groups in general represent the adult immigrants. There is a higher percentage of having a STEM major in adult immigrants compared to child immigrants for all the broad race groups. Among the detailed Asian groups, the Indians, Chinese, and the Other Asian groups exhibit a stronger and similar STEM attainment rates for the adult immigrants. All these adult immigrants have attained their STEM majors in their own country. Therefore, my focus of interest lies on the child immigrants who attain their college degree in the US. Among the child immigrants, the percentage of those attaining a STEM major in the US is higher for the late arrivals among all race groups except Filipinos and Japanese.

4.1 Broad Race Groups

Table 3 presents the regression results of age of arrival on STEM attainment rates among broad race immigrant groups. These child immigrant groups affect the conditional probability of

[21] The current sample might not be represent a true sample with respect to all the variables captured in the survey and there are problems of self-selection. In order to rectify, a weighting scheme is followed based on population levels.

individuals obtaining a STEM degree, conditional on the having a bachelor's degree and more[22]. The positive signs across the broad race groups, except the Hispanics (0-5 years) point to the fact that all these immigrants are doing well in terms of STEM major compared to white non-Hispanic natives. In other words, the comparison across the broad race groups is done with reference to the white non-Hispanic natives. Across all age of arrival groups, the Asian child immigrants represent a higher conditional probability of STEM major rates compared to all broad race groups. For example, in case of white immigrants, the STEM major rates are higher for children who arrive later in the US. In other words, the white immigrants of 12-17 age of arrival group with reference to white non-Hispanic natives have a higher conditional probability of attaining a STEM major than the 6-11 years and 0-5 years of age category again reference to the white non-Hispanic natives of similar age. Also, the 6-11 years of age category has a higher conditional probability of STEM attainment rates than the early arrival groups (0-5 years) again. A similar assimilation pattern holds for Asians and Other races.

As described earlier, assimilation refers to the adjustment of immigrants to the host culture. The longer a person stays in the US, he/she picks up the curriculum of the US and has lower probability of attaining a STEM major. Those who arrive late as child immigrants (12-17 years) in the US have higher STEM major rates compared to earlier arrivals, 0-5 years and 6-11 years of age category. These foreign-born populations coming from their respective countries have better skills in science and mathematics (Jia, 2017) and thus the late arrivals will have higher STEM major rates compared to earlier arrivals. If perfect assimilation holds, then the coefficients should be zero for the early arrival age group (0-5 years) as these early arrivals will behave similarly to natives.

[22] The conditional probability estimates of attaining a STEM major might suffer from upward bias arising from positive selection. The positive selection is due to conditioning on having a Bachelor's degree or more.

However, the coefficients are still positive. It could be driven by cultural differences where in children are motivated by parents to take up STEM fields.

4.2 Detailed Asian Groups

As already observed in Table 3, the STEM assimilation pattern is strongest for Asians compared to other broad race immigrant groups. To get a clear idea which Asian countries are driving the STEM majors in the US, it is important to decompose the STEM patterns by major Asian countries and examine how it differs across three age of arrival categories. In Table 4, the positive STEM attainment rates hold for every Asian child immigrant group. However, there exists heterogeneity among these detailed Asian groups. The differences in STEM attainment rates between Indians and the white non-Hispanic natives are higher for the early age of arrival group (0-5 years) compared to any other Asian immigrant groups with respect to white non-Hispanic natives. For the later age of arrival category, 6-11 years and 12-17 years, the higher STEM attainment rates are observed for the Vietnamese followed by Indians.

A clear STEM assimilation pattern holds for Chinese, Other Asians, Vietnamese and Koreans whereby the earlier arrival groups stay for the longest time in the US and assimilate with the American culture in terms of a lower conditional probability of attaining STEM majors. In the case of Indians and Filipinos, the late child immigrants have higher STEM major rates than early arrivals (0-5 years) but no clear assimilation pattern holds as the middle arrivals (6-11 years) should have lower STEM attainment rates than the early arrivals (0-5 years). However, the STEM attainments are much higher for Indians than Filipinos.

4.3 Sub Fields of STEM

An important dimension to consider is the STEM assimilation pattern for specific STEM majors. The computer science and engineering field is one of the highest paying for immigrants in the US and thus there is an increased probability of majoring in such STEM sub fields (Liu et al, 2017).

Panel A of Table 5 reports the conditional computer science STEM major estimates. The conditional probability of majoring in a STEM sub field of computer science is statistically significant and higher among the Indians, Chinese, Filipinos, Other Asian and Vietnamese immigrant groups compared to white non-Hispanic native. A clear STEM assimilation pattern holds for the Asian immigrant groups mentioned above including Indians and Filipinos. But, there was no such consistent pattern for Indians and Filipinos when I considered the overall STEM major.

For the engineering sub field, all of the Asian immigrant groups have higher STEM attainment rates for all ages of arrival groups compared to white non-Hispanic natives as observed in Panel B of Table 5. The STEM assimilation pattern holds for all of these Asian immigrant groups except the Japanese. In other words, the later age of the arrival group (12-17 years) has higher conditional probability of attaining an engineering sub field STEM major compared to the earlier age of arrival groups, 6-11 years and 0-5 years. Thus, there exists differences among the different Asian immigrant groups in driving the STEM majors depending on the sub fields of STEM major. Nevertheless, there exists a clear STEM assimilation patterns for the majority of the Asian immigrants groups.

4.4 Detailed Asian Groups by Sex

Another important factor driving the decision of majoring in STEM are the preferences and tastes in addition to monetary gains. Thus, the college major decisions varies by sex as they have different aspirations and goals in their lives. Zafar (2013) explains the gender gap in science and engineering, driven by preferences and tastes and not due to academic ability. A similar gender gap exists even in attainment of STEM majors in Table 6.

Despite the gap, even for the female sample, the STEM attainment rates of detailed Asian immigrant females are higher compared to white non-Hispanic female natives of similar age of arrival groups in Panel A of Table 6. However, the STEM assimilation pattern is evident for females

in Chinese, Other Asians, Vietnamese and Koreans. There is a higher conditional probability of attaining a STEM major for males across all the age of arrival categories for most of the Asian immigrant groups when compared to white non-Hispanic natives in Panel B. The STEM assimilation pattern holds for Chinese, Other Asians, and Vietnamese and Korean males similar to females. In addition, it also holds for Indian and Filipino male counterparts. Thus, the findings support the potential heterogeneity of STEM major decision across sex (Ransom & Winters, 2016).

4.5 Comparing Detailed Asian Groups with their Own Natives

The STEM assimilation process for Asian immigrant groups holds with reference to the white non-Hispanic natives group. However, it will be interesting to figure out whether the different Asian immigrant groups fare better or worse in terms of STEM attainment rates when compared to their respective native groups. The positive statistically significant coefficient for the Other Asian groups point to the fact that the Other Asian immigrants are performing better than the Other Asian natives across the three different age of arrival categories. However, for the remaining Asian groups, the immigrants are at a disadvantage in terms of STEM major compared to their respective natives as in Table 7. The Indian, Chinese, Vietnamese and Korean immigrants in the later age of arrival group (12-17) are performing better in terms of STEM attainment rates than the earlier arrivals i.e., 0-5 years and 6-11 years.

No clear STEM assimilation pattern holds for the Asian immigrant groups except the Other Asians when compared to their respective natives. Thus, these Asian immigrants fare better when compared to white-non Hispanic natives but no such clear inference can be drawn when being compared to their own native group. The role of K-12 education in these Asian countries could be the driving force for higher STEM attainment rates for the late arrivals (12-17 years). These later arrivals develop their math and science skills completely and thus their performance are better

whether compared to their respective natives or the white non-Hispanic natives who are getting their K-12 education in the US.

4.6 Other STEM Outcomes

There exists previous literature related to STEM workers and outcomes like wages, employment and others. It will be interesting to look at the patterns among Asian immigrant groups who are likely to have attained their college degree and have joined the labor force by the same age of arrival groups and whether it follows the same path as in attainment of STEM majors. The STEM occupation is defined on the basis of 2010 Census Bureau occupational classification scheme[23].

In table 8, the coefficients for the Indians, Chinese, Filipino, Other Asian and Vietnamese immigrant groups are positive and statistically significant implying a higher conditional probability of being engaged in STEM occupation compared to non-Hispanic white natives. A similar pattern is observed for those having STEM occupations like the STEM assimilation pattern. The later age of arrival group (12-17 years) has a higher conditional probability of being employed in a STEM occupation than the early age of arrival categories, (6-11 years) and (0-5 years). Even the middle child arrivals have higher STEM occupation rates than the early child arrivals. Thus, the foreign-born STEM workers who arrive in the US after having their entire or a part of their K-12 education in their respective countries will have a higher conditional probability of having STEM jobs.

4.7 Other Educational Outcomes

In terms of educational attainment, for example having a bachelor's degree and more, one finds a different assimilation path than the STEM major. The longer a person stays in the US, the rates of such educational attainment are higher. Thus, the education infrastructure in the US is suitable for such kind of education. Table 9 examines which Asian countries are driving the assimilation

[23] The entire list of STEM occupations is described in Appendix Table A2.

patterns for attainment of such educational degrees. The Indian, Chinese, Korean and Japanese have positive and statistically significant attainment rates of Bachelor's degree and more compared to white non-Hispanic natives. For Filipino and Other Asian late child immigrants, they have a lower probability of attaining a bachelor's degree and more compared to white non-Hispanic natives. Thus, it could be the other factors like culture, preferences and motivation of these immigrant groups in addition to education infrastructure that drives the results.

For Koreans and Japanese, one can find that the early child arrivals (0-5) have higher conditional probability of obtaining a bachelor's degree and more than the late child arrivals (12-17 years). However, it is the Indians and Chinese that follow clear assimilation patterns. The early child arrivals (0-5 years) have higher attainment rates than the middle child arrivals (6-11 years) and the later child arrivals (12-17 years) and also the middle child immigrants (6-11 years) have higher attainment rates than the later child immigrants (12-17 years) when compared to white non-Hispanic natives.

4.8 Other Robustness Checks

Next, I consider the marginal effect of different child age of arrival groups on attaining a STEM major by using interactions of these ages of arrival and immigrants groups to see whether the STEM assimilation pattern still holds. Table A3 does it for broad race immigrant groups and table A4 repeats it for detailed Asian immigrant groups. Thus, for broad race immigrant groups, the Hispanics, Asians, Whites, Blacks and Other Races of similar ages of arrival are compared to the same reference group non-Hispanic white natives. This makes the immigrants groups more comparable to each other. In a similar manner, the Indian, Chinese, Filipino, Other Asians, Vietnamese, Korean and Japanese are compared to white non-Hispanic natives for the detailed Asian immigrant groups. The results are almost the same and statistically significant as reported in Tables A3 and A4.

All the earlier results are conditional on individuals with a bachelor's degree and more. I re-estimate the results in Tables 3-8 unconditional on education levels for the sake of completeness. Thus, the sample is restricted to individuals aged 22-60 in the 2009-2016 ACS regardless of their educational levels. Tables A5-A10 overall reinforces the STEM assimilation story. Table A11 reports the unconditional estimates of STEM attainment rates by different ages of arrival among the broad race immigrant groups. The STEM assimilation pattern holds for white and Asian immigrants.

5. Summary & Conclusion:

I study the heterogeneous effect of child immigrants arriving in the US at different ages on STEM attainment rates by different race groups. In other words, this chapter examines how the patterns of STEM majors who immigrate in the 0-5 year age of arrival group differ from those who immigrate during 6-11 years (going to attend elementary or middle school) or 12-17 years age of arrival group (will attend high school).

There is a higher probability of attaining a STEM major for late child arrivals compared to early child arrivals for some broad race immigrant groups and some detailed Asian immigrant groups. This path of assimilation is even found for different sub fields of STEM like computer science and engineering. Though it might vary by sex, the STEM assimilation pattern is evident for males in all Asian race groups except the Japanese and holds for Chinese, Other Asians, Vietnamese and Korean females. Similar STEM assimilation patterns hold for STEM occupation rates. However, the assimilation rates for having a bachelor's degree and more is different in the sense the early arrivals have an advantage over the late arrivals as they pick up the curriculum of the US and have higher probability of such education attainment rates.

The assimilation studies have been carried out in different disciplines. Most of the assimilation studies in the field of economics have focused on immigrant generations. These

assimilation patterns vary by race groups and ages of arrival. As expected, the early arrivals come to the US at an age where it is easier to pick up the habits, culture and education of the host country compared to late arrivals. These earlier ones gets integrated in American society gradually.

The assimilation patterns might vary depending on the type of outcome. There is an immigrant STEM advantage in terms of high attainment rates of STEM fields mainly due to better preparation in high school math and science (Jia, 2017). Most of the late child arrivals (12-17 years) from the STEM driven countries are likely to receive a part or whole of their K-12 education in respective origin countries where there is curriculum focusing more on mathematics and science developing skills and less on the English language. However, the longer a person stays in the US he/she picks up the curriculum, culture and even the skills of natives. Thus, the early child arrivals behave similarly to natives and may specialize in other majors besides STEM compared to the late child arrivals from STEM driven Asian countries like India, China. As a result, it is expected that there will be rapid assimilation of child immigrants compared to the adult immigrant groups.

One can infer the STEM assimilation pattern is unique and needs further study for broader policy perspectives. The STEM driven countries, mostly Asian countries like India, China contribute to such fields in the US. Further, these foreign-born STEM graduates make the future labor force of the US. Previous research has already shown that both STEM and non-STEM graduates create positive wage externalities but STEM graduates generate additional benefits (Winters, 2014). The higher wages in STEM jobs to some extent explains the inverse relationship between STEM education and self-employment for immigrants (Cai & Winters, 2017). There is a debate whether these foreign-born STEM graduates will have adverse impacts for the natives similar to the foreign-born STEM workforce as these STEM graduates will form a considerable portion of the foreign-born STEM workforce in the US. Thus, there is a need for informed policy to encourage the natives to become part of such fields as they are the major drivers of innovation and boost economic productivity. There have been policies to increase the stock of STEM graduates

especially among the natives given the disproportionate share of foreign-born in STEM fields but more needs to be done. Researchers have pointed out the importance of K-12 education in these STEM driven countries. It will be useful if future research can be carried out on the effect of role of K-12 education on attainment of STEM majors.

CHAPTER III

RURAL POPULATION GROWTH: SPREAD OR BACKWASH EFFECTS?

1. Introduction

Although the level of urbanization has increased over the past decades, a large percentage of India's population still resides in villages. According to the recent 2011 Census of India, the size of the rural population is 68.84% of the total population and agriculture is the source of livelihood for these two-thirds of the population in India. One also needs to analyze the various linkages that exist between urban and rural areas as both of them are interdependent.

There have been several studies across the world focusing on the interrelationship between metropolitan and non-metropolitan areas. One of the commonly used concepts being applied in regional studies is the spread and backwash effects (Myrdal, 1957). Some of the earlier studies concentrated on specific regions (Barkley et al., 1997; Henry et al., 1999). However, the later literature (Partridge et. al, 2007b; Partridge et. al, 2008a; Partridge et. al, 2008b; Partridge et. al, 2010) have been applied to a broader national context and include the importance of distance in explaining spillovers of urban areas on rural areas.

Most of the spatial studies in India have focused on only urban areas and little importance has been given to rural areas. In recent years, there have been few studies on urban-rural linkages. Kumar et al (2014) observes the negative effects of distance to health facility on in-facility birth in

rural India. One recent study by Sharma (2016) considers the effect of urban proximity on land use patterns and economic development in rural India. Asher et al (2017) finds the effect of remoteness on mean monthly earnings, non-farm employment and literacy rates in the villages.

The aim of this paper is to find the spread and backwash effects of urban growth on the hinterlands of India. Besides distance to the nearest town[24], I am using different tiers in the urban hierarchy to study the proximity effects of urban centers for the villages. Most of the earlier literature focusses only on the relationship between core and periphery, where the core is the town and village is the periphery. However, such a measure might not capture the exact magnitude of the spread and backwash effect as the nearest urban center is low in its hierarchy and higher tiered centers offer more services and goods and a greater variety of jobs. Thus, the dynamics of the relationship between rural and urban areas might vary by the position of urban centers in hierarchy.

For example, it is not only the location of towns but also the cities[25] and the large cities[26], which matter for development of villages. Given the urban hierarchy, the nearest urban tier for villages are towns, followed by cities which is the next higher tier in terms of provision of goods and services and also centers of education, employment and at the topmost level are the large cities. Further, this study uses the rich Indian Census data, which provide information on the distance variables like distance to the nearest town, incremental distance to a city and the incremental distance to a large city. It also has detailed information on different man-made amenity distance variables including hospitals, schools, colleges, railways and banks.

The spread and backwash literature looks at the metropolitan and non-metropolitan samples together. For example, Partridge et al (2007a) examines Canadian population growth to

[24] Areas having a population above 5,000; the population density is more than 400 persons per square kilometer; and at least 75% of the males of the working population are engaged in non-agricultural pursuits.
[25] Towns having a population of 100,000 and more.
[26] Towns having a population of 500,000 and above.

find whether urban centers acts as engines of growth. They consider separate models for rural and urban areas as these areas differ in terms of population, with rural areas being less populated. However, the dynamics of rural areas are unique in the context of India and quite different from that of an urban area and needs to be addressed separately. The uniqueness stems from the definition of rural areas by the Census of India. It is not only the difference in population figures that distinguishes rural from urban areas, but also their main occupation. A major distinction between towns and villages is at least 75% of the males of the working population are engaged in agricultural pursuits for villages. Thus, the urban hierarchy effects on growth of towns are quite different from the growth of villages.

A negative and statistically significant effect of the nearest town distance and incremental distance to a city variable on the growth rates of villages is found in this study. Thus, the villages experience spread effects for being in close proximity to a town and a city but backwash effects for being closer to a large city. There might be limited labor mobility due to low skill sets of people in the rural areas. Thus, the distance penalty or protection varies by the position of tier in the urban hierarchy. The results are robust when initial population levels rather than population density is used. Irrespective of the size of the villages or even different distances to towns, the villages overall experience a spread effect from the growth of towns and cities. No clear inference can be drawn about the man-made amenity distances as there is a mixture of results.

The rest of the paper is organized as follows. Section 2 provides the conceptual framework and literature review. Section 3 describes the dataset and the empirical strategy. The empirical results are discussed in Section 4. Section 5 concludes with summary and discussion.

2. Conceptual Framework and Literature Review

There exists interlinkages between towns and villages in India. Neither towns nor villages are self-sufficient. Both of them cater to each other through demand and supply of goods and labor.

However, there is greater dependence of villages on towns in India. The towns are higher in the hierarchy than villages in terms of the order of goods and services and access to markets. The towns also serve as centers of jobs and education. Thus, the location of towns holds relevance for villages.

Following Hirschman (1958) and Myrdal (1957), the concept of spread and backwash effects can help in in explaining the regional growth in villages of India as it has been used to describe the effects of urban growth on the hinterlands (Gaile, 1980). The spread effect refers to the positive effects of growth in a major urban core (town) for adjoining villages through agglomeration and spillovers. On the other hand, there might be adverse effect called the backwash effect due to the growth in major core. The towns might attract people from nearby areas or villages because of access to markets, amenities etc. and thus, people will migrate from village to towns.

Barkley et al. (1997) tested the spread–backwash effects for specific regions in North Carolina, South Carolina, and Georgia in the United States. They find evidence of backwash effects for most of the rural areas. A later study (Henry et al., 1999) considers the Danish rural communities and the rural communities of France in addition to the rural communities of South Carolina in the United States. They find a mixture of spread and backwash effects from growth in urban core and fringe areas on the growth in hinterlands.

Different approaches and variables have been employed to measure the spread and backwash effects. Some of the commonly used variables are population, employment and income growth (Partridge et. al, 2007; Partridge et. al, 2008, Partridge et. al, 2010).Henry et al. (1997) compares population density functions for eight functional economic areas in North Carolina, South Carolina, and Georgia. Most of the literature in China uses Gross Domestic Product (GDP) or per-capita GDP due to hukou restrictions. Ying (2000) observes significant spillovers to provinces for selected areas in China. Ke and Feser (2010) uses non-agricultural GDP and employment to find mixed evidence of spread and backwash effects of Chinese urban centers on adjoining cities and

counties in Greater Central China. However, there are backwash effects on rural counties. Chen & Partridge (2013) tested whether urban centers act as engines of growth by considering the entire country rather than focusing just on provinces or a few counties in a region.

The locations decisions occur simultaneously by the household utility maximization and firm cost minimization as stated in Roback's (1982) static general equilibrium model and later incorporating the dynamic aspect by Rapport (2004). Following the same model, Partridge et al. (2007a) use 1981-2001 Canadian data to examine the spread and the backwash effects. An important inclusion of this paper is to capture the spillover effects through the urban distance discount (UDD). While considering spread effects of the urban centers on rural communities, an important role is played by people commuting from rural areas to urban for jobs. Partridge et al. (2010) explains Canadian agglomeration spillovers and job growth spread effects on rural areas through commuting.

There have been similar studies in the United States to study the effects of distance on the population growth in hinterlands. The proximity to urban agglomeration as measured by distance to higher tiered areas and proximity to market potential have negative effects on population growth in hinterland areas (Partridge et al., 2008a). Other studies (Partridge & Rickman, 2008b) consider the impact of distance as differentiated by urban hierarchy on rural poverty. The more distant is a rural place from a metropolitan area, the higher is rural poverty. A recent study by Ganning et al. (2013) estimate the impact spread and backwash effects of metropolitan growth on population growth in nonmetropolitan communities.

Mitra & Mehta (2011) estimates the urban (city) domestic product for selected states in India to find cities being engines of growth and thus focusing only on agglomeration economies. Ghani et al. (2012) observes the manufacturing sector moving from urban to rural areas across districts from 1989-2005 period. These rural areas have strong education and infrastructure levels

to support industries with high capital and land intensity. Recent studies have taken into account the remoteness of villages or rural areas and measured the distance effects on different outcome variables. Kumar et al (2014) captures the effect of distance to the nearest health facility on the place of delivery after controlling for socio economic factors. There is the role of linkage between urban proximity and rural land use and on rural development pattern (Sharma, 2016). Asher et al. (2017) observes negative relationship between distance and rural living standards. Thus, the villages have to incur the cost of remoteness in terms of lower mean monthly earnings, reduction in non-farm employment, and decrease in literacy rates.

3. Data and Empirical Specification

This paper assesses the spread and backwash effects of major urban centers on the population growth of the hinterlands. The data come from the Part –A of District Census Hand Book (DCHB)[27] that constitute both the town and the village directory. It has information on several demographic and socioeconomic characteristics that is part of Census data and contains non-Census data on various civic amenities and infrastructural facilities. The reason for using Census data is the disaggregated level of data at the village level. Most of the earlier studies had data limitations and used states or districts as unit of analysis.

All of the variables used in the analysis come from the village directory except the two distance variables, incremental distance to a city and incremental distance to a large city. These distance variables are obtained from the town directory. The other distance variable, distance to the nearest town is already there in the village directory. Given those town names, I obtained the information on the other distance variables from the town directory and matched the data sets by using the town names. There are 638,588 villages in India out of which 593,731 are inhabited villages according to 2001 Census of India. In 2011, there is a slight increase in villages to 640,932

[27] It is part of the Indian Census data that is conducted every after 10 years.

out of which inhabited villages are 597,608. I was able to match approximately villages up to 91% between 2001 and 2011 Census. Further merging of village data with town data resulted in matching of data approximately to 90%.

Based on the definition of Census, an area is defined as rural if the population is below 5,000; the density of population is less than 400 per square kilometer; and further in such areas at least 75% of the males of the working population are engaged in agricultural pursuits. The reason for picking up year 2001-2011 is the advantage of village data containing detailed information on distance variables and the range of man-made amenity distance variables. The following model is implemented by estimating a cross-sectional growth equation as described below.

$$\%\Delta\mathbf{POP}_{\mathrm{ir(t-0)}} = \alpha + \beta\mathbf{DEMOG}_{iro} + \ \gamma\mathbf{SPATIAL}_{\mathrm{ir}} + \emptyset\mathbf{DIST}_{ir} + \delta_{\mathrm{r}} + \varepsilon_{\mathrm{ir(t-0)}} \qquad (3)$$

The dependent variable is the percentage change in population between initial period, 0, and final period, t, for rural areas, *i*, located in region *r*. The initial period in our data is 2001 and the final period is 2011. The population density is captured in **DEMOG** to account for the initial agglomeration, congestion and localization effects.

The best way to capture the spread and backwash effect is through distances between areas as urban growth spills over to its nearby area. **SPATIAL** consists of a list of distance variables measuring the spread and the backwash effects. The distance to the nearest town captures the road distance in kilometers from a village to the nearest town. Next, is the incremental distance to city which measures the additional road distance in kilometers from the nearest town corresponding to the village and to a nearest city. Similarly, there is the incremental distance to a large city capturing the additional road distance in kilometers from a city to a large city. The quadratic term of the each of these distance variables are considered to capture the non-linearity in distance effects. The impact of these distance variables will vary depending on the position in urban hierarchy. For the village sample, the nearest town distance variable is reported rather than the nearest city distance

variable. The nearest town is the next higher level in hierarchy in terms of provision of goods, services and even jobs. The second higher level in urban hierarchy are the cities and at the topmost level are the large cities.

There is no information on natural amenity variables as these people in rural areas care about basic man-made amenities. The man-made amenity distance variables are contained in **DIST.** It is recorded as a categorical variable, where it is coded as "0" if that amenity is present in the village, "1" if that amenity is in the range of less than 5 kilometers, "2" if that amenity is available within 5-10 kilometers, and "3" if that amenity is in the range greater than 10 kilometers. This variable is suitable to capture the hierarchical effects as it is measured in road distance in kilometers. These amenity distance variables include access to different types of infrastructure like hospitals, colleges, schools, railways, and banks. An advantage of having distance amenity variables is the exogeneity of such variables. Most of the models having large numbers of amenity measures suffer from endogeneity issues as there might be effects of population growth on man-made amenities leading to problems of simultaneity.

In addition to the role played by demographic attributes, spatial measure and man-made amenity distance variables, there might be time invariant unobserved socio-economic and geographic factors which might impact the population growth of the rural areas. $\boldsymbol{\delta}_r$ are the region fixed effects controlling for other region-specific factors including, language, culture, migration flows, tax and expenditure policies, regulatory differences, geographic location with respect to coasts and $\boldsymbol{\varepsilon}$ is the residual. Although the villages in India fall under the political arrangement of the state, the state fixed effects might not give a true picture. The state fixed effects are expected to absorb all the across state variation. However, our data points out the fact that there is much variation in growth rates between villages of different states than within states. Hence, the region fixed effects are considered.

4. Empirical Results

4.1 Summary Statistics

The summary statistics are reported in Table 1. The mean population levels of 2011 are higher compared to 2001. The mean population density is 586 people per square kilometer[28]. The nearest urban center for villages are towns and the next higher level of urban center is a city and at the topmost level are the large cities. The nearest town distance variable averaged 25 km while the incremental distance to a city averages 77 km. Further, the mean incremental distance to a large city stands at 97 km. Thus, to avail the services of higher ordered urban centers, one has to travel even more as evident by the mean incremental distances. For the amenity distance variables, the mean school range is approximately within the range of 5 kilometers. However, for other amenities like hospitals and commercial banks, the mean range is within 5-10 kilometers. For amenities like college and railways, one has to travel farther, as it is in the range greater than 10 kilometers.

4.2 Regression Analysis for Villages

Table 2 presents the regression results for villages between 2001 and 2011. Column (1) considers population density and includes the region fixed effects. Column (2) adds to the initial model the spatial variables, such as distance to the nearest town and its square term. Column (3) further adds additional distance variables that is incremental distance to a city and its respective quadratic term. Column (4) includes another set of distance variables that is incremental distance to a large city and the quadratic term. Finally, column (5) includes the amenity distance variables along with demographic variables and a list of spatial variables. The population density has a negative effect on growth rates but none of them across models is statistically significant. There is a negative statistical impact of the nearest town distance variable on population growth of the villages. It

[28] According to Census definition of village areas, the population density should be less than 400 persons per square kilometer (i.e. 1000 per sq. mile).

implies that proximity of towns matters for villages. Thus, the further a village is located from a town, the more adverse the impacts on population growth. However, the quadratic term on distance to the nearest town is not significant. Thus, it is not clear whether the distance penalty becomes more or less important if the distance to the nearest town increases. Nothing much can be inferred about the incremental distance variables, that is incremental distance to a city and incremental distance to a large city. As expected, spillover effects of the urban areas will spread to rural areas closest to the urban core. Based on distance to nearest town variable, there are spread effects of urban growth (towns) on the growth rate of villages.

For the amenity distance variables, only some of the variables are significant and have opposite effects on the growth rates of villages. One could argue that the following rural sample consists of villages, which fall outside the definition of rural areas by the Census of India and hence might not capture the true effects.

4.3 Regression Analysis for Villages with restrictions

Therefore, I ran the regressions for villages with restrictions. The rural sample is restricted to only those villages whose population lies between 100 and less than 5000 for both time periods of 2000 and 2011 and having population density of less than 400 persons per square kilometer. First, the definition of villages by the Census is followed. Further, most of the rural population live in villages between 500 and 5,000 inhabitants (Cohen, 2004). Second, there exists a certain fraction of rural people residing in villages having population more than 5,000, but are referred to as grey zones as these possess urban characteristics (Urban India, 2011). It is difficult to categorize such zones as these fall neither into urban areas nor into rural areas based on the Census definition. Also, the villages having population less than 100 is too small to capture any meaningful growth.

The density variable is negative and statistically significant across all specifications. Thus, the denser is the village, the lower is the population growth. The two distance variables, distance

to the nearest town and the incremental distance to a city have a negative statistical impact on population growth of the villages. Thus, the location of both towns and cities matter for the development of villages. A village being farther way from a town is at a disadvantage as evident in lower population growth rates. An increase in nearest town distance variable by one kilometer reduces the population growth by approximately 0.05 percentage points. Further, the positive coefficient across the quadratic terms point out that the distance exerts a negative influence on growth rates at a decreasing rate. In terms of distance penalty for villages, it is approximately 1.3% less population growth given the mean distance to its nearest city averaged at 24.73 km (Table 1).

A village is penalized for being not only farther away from a town but also from a city as given by the negative and statistically significant coefficient of the incremental distance to a city variable. The penalties differ depending on the position of tier in the urban hierarchy following Central Place Theory (Christaller, 1933), which states that urban areas differ in terms of provision of goods and services. The positive quadratic term implies that increased distance has negative effects on growth rates but at a diminishing rate. The added penalty of incremental distance to a city is approximately 1.5% lesser population growth of villages. Adding up, the total penalty for a village to be farther away from a higher ordered city is approximately 2.8% less population growth. Thus, in terms of spread effect, the growth in an urban area spills to the nearest rural areas. The rural area closest to town and city will benefit from proximity. Given positive square term for each of the distance variables mentioned above, the distance penalty to reach to the nearest town and incremental distance to reach nearest city becomes less important at greater distances

However, the coefficient of incremental distance to a large city is positive and significant pointing to backwash effects of these top most tiers of urban centers on villages. The negative square terms indicate these effects do not attenuate with distance. The people in these remote villages are protected from being distant to a large city called the “distance protection” effect (Polese & Shearmur, 2004). It could be such an area even being remote have minimum necessities

to support them. Further, people in these areas have a skill set which might not be useful to large cities, which mostly have high skilled jobs. This is true for villagers in India who are engaged in agriculture. It could be that these people cannot commute to these large cities due to lack of proper infrastructure like roads, railways and long hours of commuting times. It could be that rural villages closer to larges cities experience more out-migration. A similar distance effect is seen on rural out-commuting within the urban hierarchy in Canada (Partridge et al., 2010).

A mixture of results is again observed for the amenity distance variables and no clear conclusion can be made as there might be omitted influences. One thing to note is the man-made amenity variables are reported as categorical variables where the reference category is having that amenity present in the village. There is a negative statistical effect of college amenity distance on growth of villages. Thus, the villages are penalized for being farther away from a college education. In terms of education infrastructure, the location of college matters for villages. For other amenities like hospitals, rail stations and commercial banks, some of the coefficients have opposite signs and are not significant. All of the coefficients for school are positive, and thus one is willing to travel more.

4.4 Regression Analysis for Villages using Initial Population

To check for robustness, Table 4 estimates the village regressions by considering the initial population levels of 2001 rather than using initial population density. The signs on the coefficients and t-statistics on the distance to nearest town and incremental distance to city variables are approximately the same as for the population density model except the coefficients are lesser in magnitude. Thus, a village has to incur a cost for being remote not only from a town but also from a city and the cost is in terms of less population growth. Again, there is spillover effects of close proximity of village to urban centers (towns and cities). However, for the incremental distance to the topmost city, the villages are not penalized as given by the positive coefficients. These villages

are so remote that it is completely insulated from these higher tiered urban centric growth effects and experience backwash effects from growth of such large cities. Thus, the relation between population growth in villages and its proximity to urban centers depends on the geographic location in urban hierarchy.

4.5 Regression Analysis for Villages based on Different Distance Tiers to Towns

The difference in distance to town might affect the impact of spread and backwash effects of urban centers on the growth rates of villages. I look at two distance rings, one in which the distance of a village from the nearest town is less than 50 kilometers and one where it exceeds 50 kilometers in Table 5. The choice of distance rings follow the previous literature (Henry et al., 1997; Partridge et al., 2007a). The population density has a negative impact on growth rates for villages having different distances to towns. The magnitude of the impact of the distance to nearest town on the village population growth rate for those less than 50 kilometer range points out to the fact that spillover matters for villages in close proximity to towns. The villages are further penalized being farther away from a city as shown by the coefficients of the incremental distance to a city variable for both distance sizes. The backwash effects holds for villages being farther away from a large city.

4.6 Regression Analysis for Villages based on Size

The size of the spread and the backwash effects may vary depending on the size of villages. I have considered three different sets of villages, small villages (population of 100 and less than 500), medium (population of 500 and less than 1000) and large villages (having a population of 1000 and less than 5,000). For all village sizes, population density has a negative statistical effect on the growth rate of villages from 2001 to 2011 as seen in Table 6. Irrespective of village size, all these villages are penalized for being far away from a town and thus the urban spread effects holds for all villages. The distance penalty to reach a town becomes less important at greater distances as

shown by positive significant quadratic terms for small and large villages. Only the small and the medium villages are further penalized for greater distance away from city. However, the spread effects are stronger for small villages as reflected by negative statistical magnitudes of both the distance variables on the growth rate of villages. There exist fewer spread effects for allsizes of villages for being far away from a large city.

4.7 Regression Analysis for Villages based on Different Growth Rates

Besides the size of village, it is important to estimate the spread-backwash effects based on the growth rate of villages. One might differ across fast, medium and slow growing villages. Table 7 reports the quantile regression analysis for 25th percentile, 50th percentile and 95th percentile respectively. The negative statistical coefficients on the distance to the nearest town point out the spread effects for all types of villages. The negative effects of incremental distance to a city on rural population growth holds for fast and slow growing villages. The adverse distance impacts seems stronger for fast growing villages as reflected by the distance to the nearest town and incremental distance to a city. Thus, the spread effects of urban centers (towns and villages) is seen as greatest for the fast growing villages. There is no distance penalty for such remote villages rather these are distance protected in terms of incremental distance to the top most tier in urban hierarchy. Thus, the small and medium growing villages experience backwash effects from being farther away from a large city.

5. Conclusion & Discussion

There exists a symbiotic relationship between towns and villages. However, one cannot deny the role of urban areas in explaining growth patterns of villages. The best way to capture the rural-urban integration is through proximity to urban centers. The present study looks at the spread and backwash effects of urban centers on the population growth of villages. There exists heterogeneity of effects depending on the position of such urban center in the urban hierarchy. The

empirical findings indicate that the urban centers like towns and cities have spread effects on the villages. However, for remote villages which are farther way from a large city experience backwash effects. There are no additional benefits being closer to large cities as commuting from village to topmost urban center takes time. Thus, growth spreads out to neighboring rural areas at least those within commuting distance to urban centers.

The urban-rural linkage also depends on the status of the agricultural sector as the people in villages are directly engaged in agriculture. The major policy reforms related to agriculture was the Green Revolution[29], which continued until the economic liberalization in 1991. The 1991 era and thereafter mostly implemented macroeconomic reforms in industry, the exchange rate and foreign trade and investments and not much for the agricultural sector. However, the Uruguay Round Agreement on Agriculture (URAA) under the WTO in 1994 officially extended the reform wave to agriculture, but mostly aimed at gradually decontrolling agricultural trade flows. Thus, there have been lagging agricultural reforms in the domestic sector in the current scenario (Gulati & Fan, 2007).

Further, there exists a large share of workers in villages who commute to urban areas for non-agricultural jobs[30]. One needs to consider this while considering the spread and backwash effects. Thus, the dichotomous definition of urban and rural areas misses not only people involved in non-agricultural pursuits but also the peri urban areas or the grey areas where most of the positive spillover effects take place.

Several place-based policies have been undertaken in rural areas. The Pradhan Mantri Gram Sadak Yojana (PMGSY) was launched in 2000 to build a paved road connecting every village

[29] Adoption of high yielding variety seeds, fertilizers, irrigation facilities and other infrastructure like tractors in agriculture in 1965 which led to increase in food grain production in India.
[30] Chandrasekhar (2011) showed a total of 8.05 million workers commute from rural to urban area based on 2009-10 NSSO data.

in India. The impacts were observed in terms of educational attainment, like staying in school longer and better performance on standardized exams (Adukia et al., 2017) and also affecting rural employment and economic outcomes (Asher & Novosad, 2016). The National Rural Employment Guarantee Act (NREGA) aims at proving better rural infrastructure like water, roads, electricity etc and is expected to improve the poverty scenario in rural areas (Chakraborty & Guha, 2009). Even the 11th Five Year Plan (2007-2012) focuses on bridging the gap between rural and urban areas in terms of basic services, jobs and development measures. However, the villages incur costs of remoteness for being farther away from a town and city. One needs to keep in mind the urban areas play a key role in functioning of villages. Thus, the future rural development policy rather than just focusing on place-based policy should think of developing the urban cores as these have positive spillover effects on villages and needs to be implemented.

REFERENCES

Adukia, A., Asher, S., & Novosad, P. (2017). Educational investment responses to economic opportunity: Evidence from Indian road construction. Working Paper.

Arora, R. U., & Ratnasiri, S. (2015). Economic reforms, growth and well-being: evidence from India. Journal of Economic Policy Reform, 18(1), 16-33.

Asher, S., Nagpal, K., & Novosad, P. (2017). The cost of remoteness: Evidence from 600,000 Indian villages. Working Paper.

Asher, S., & Novosad, P. (2016). Market access and structural transformation: Evidence from rural roads in India. Manuscript: Department of Economics, University of Oxford.

Barkley, D. L., Henry, M. S., & Bao, S. (1996). Identifying" spread" versus" backwash" effects in regional economic areas: A density functions approach. Land economics, 336-357.

Barnett, M. D., Sonnert, G., & Sadler, P. M. (2012). More like us: The effect of immigrant generation on college success in mathematics. International Migration Review, 46(4), 891-918.

Bates, C. (2011). The hidden story of partition and its legacies. BBC History, 3.

Berglee, R. (2012). World Regional Geography: People, Places and Globalization. The Saylor Foundation.

Bloom, D. E. (2011). Population dynamics in India and implications for economic growth. Program

on the Gobal Demography of Aging (PGDA) Working Paper No. 65.

Boyd, M., & Tian, S. (2017). STEM education and STEM work: Nativity inequalities in occupations and earnings. International Migration, 55(1), 75-98.

Cai, Z., & Winters, J. V. (2017). Self-employment differentials among foreign-born STEM and non-STEM workers. Journal of Business Venturing, 32(4), 371-384.

Carter, T. P., & Segura, R. D. (1979). Mexican Americans in school: A decade of change. New York: College Entrance Examination Board.

Chakraborty, D., & Guha, A. (2009). Infrastructure and economic growth in India: Analysing the village-level connectivity scenario of the states. Journal of Infrastructure Development, 1(1), 67-86.

Chandrasekhar, S. (2011). Workers commuting between the rural and urban: Estimates from NSSO data. Economic and Political Weekly, 22-25.

Chauvin, J. P., Glaeser, E., Ma, Y., & Tobio, K. (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. Journal of Urban Economics, 98, 17-49.

Chen, A., & Partridge, M. D. (2013). When are cities engines of growth in China? Spread and backwash effects across the urban hierarchy. Regional Studies, 47(8), 1313-1331.

Chen, Z., & Skuterud, M. (2017). The relative labour market performance of former international students: Evidence from the Canadian National Graduates Survey. IZA DP No. 10699.

Cheshire, P. C., & Magrini, S. (2006). Population growth in European cities: Weather matters – but only nationally. Regional Studies, 40(1), 23-37.

Chikte, R. P. (2011). Income convergence and regional growth in India: Before and after the

economic liberalization. South Asia Economic Journal, 12(2), 239-269.

Chiswick, B. R., & DebBurman, N. (2004). Educational attainment: analysis by immigrant generation. Economics of Education Review, 23(4), 361-379.

Cohen, B. (2004). Urban growth in developing countries: a review of current trends and a caution regarding existing forecasts. World Development, 32(1), 23-51.

Cortes, K. E. (2006). The effects of age at arrival and enclave schools on the academic performance of immigrant children. Economics of Education Review, 25(2), 121-132.

Das, S., Ghate, C., & Robertson, P. E. (2015). Remoteness, urbanization, and India's unbalanced growth. World Development, 66, 572-587.

Desmet, K., Ghani, E., O'Connell, S., & Rossi-Hansberg, E. (2015). The spatial development of India. Journal of Regional Science, 55(1), 10-30.

Fallah, B., Partridge, M. D., & Rickman, D. S. (2013). Geography and High-Tech Employment Growth in US Counties . Journal of Economic Geography, 14(4), 683-720.

Feliciano, C. (2005). Does selective migration matter? Explaining ethnic disparities in educational attainment among immigrants' children. International Migration Review, 39(4), 841-871.

Fujita, M., & Krugman, P. (2004). The new economic geography: Past, present and the future. Papers in Regional Science, 83(1), 139-164.

Gaile, G. L. (1980). The spread-backwash concept. Regional Studies, 14(1), 15-25.

Ganning, J. P., Baylis, K., & Lee, B. (2013). Spread and backwash effects for nonmetropolitan communities in the US. Journal of Regional Science, 53(3), 464-480.

Ghani, E., Goswami, A. G., & Kerr, W. R. (2012). Is India's manufacturing sector moving away

from cities? National Bureau of Economic Research (NBER) Working Paper No.17992.

Ghani, E., Goswami, A. G., & Kerr, W. R. (2014). Highway to success: The impact of the golden quadrilateral project for the location and performance of Indian manufacturing. The Economic Journal, 126(591), 317-357.

Ghani, E., Kerr, W. R., & O'Connell, S. (2014). Spatial determinants of entrepreneurship in India. Regional Studies, 48(6), 1071-1089.

Glaeser, E. L., & Tobiot, K. (2008). The rise of the Sunbelt. Southern Economic Journal, 74(3), 609-643.

Gonzalez, A. (2003). The education and wages of immigrant children: the impact of age at arrival. Economics of Education Review, 22(2), 203-212.

Greenman, E., & Xie, Y. (2008). Is assimilation theory dead? The effect of assimilation on adolescent well-being. Social Science Research, 37(1), 109-137.

Gulati, A., & Fan, S. (2007). The dragon and the elephant: agricultural and rural reforms in China and India: International Food Policy Research Institute (IFPRI).

Gyourko, J., Mayer, C., & Sinai, T. (2013). Superstar Cities. American Economic Journal. Economic Policy, 5(4), 167-199.

Hagy, A. P., & Staniec, J. F. O. (2002). Immigrant status, race, and institutional choice in higher education. Economics of Education Review, 21(4), 381-392.

Hanson, G. H., & Slaughter, M. J. (2016). High-skilled immigration and the rise of STEM occupations in US employment. NBER Working Paper 22623.

Henry, M. S., Barkley, D. L., & Bao, S. (1997). The hinterland's stake in metropolitan growth: Evidence from selected southern regions. Journal of Regional Science, 37(3), 479-501.

Henry, M. S., Schmitt, B., Kristensen, K., Barkley, D. L., & Bao, S. (1999). Extending Carlino-Mills models to examine urban size and growth impacts on proximate rural areas. Growth and Change, 30(4), 526-548.

Hirschman, A. O. (1958). The Strategy of Economic Development. Yale University Press, New Haven, CT.

Hunt, J. (2011). Which immigrants are most innovative and entrepreneurial? Distinctions by entry visa. Journal of Labor Economics, 29(3), 417-457.

Jain, M. (2017). The effect of distance on urban transformation in the Capital Region, India. International Planning Studies, 1-14.

Jia, N. (2017). Laying the tracks for successful science, technology, engineering and mathematics education: what can we learn from comparisons of immigrant–native achievement in the USA? Pacific Economic Review, 1-24.

Ke, S., & Feser, E. (2010). Count on the growth pole strategy for regional economic growth? Spread–backwash effects in Greater Central China. Regional Studies, 44(9), 1131-1147.

King, L. J. (1985). Central place theory. Regional Research Institute, West Virginia University Book Chapters.

Koo, J., & Lall, S. (2007). New economic geography: Real or hype? International Regional Science Review, 30(1), 3-19.

Lall, S. V. (2007). Infrastructure and regional growth, growth dynamics and policy relevance for India. The Annals of Regional Science, 41(3), 581-599.

Lall, S. V., & Chakravorty, S. (2005). Industrial location and spatial inequality: Theory and evidence from India. Review of Development Economics, 9(1), 47-68.

Lall, S. V., Shalizi, Z., & Deichmann, U. (2004). Agglomeration economies and productivity in Indian industry. Journal of Development Economics, 73(2), 643-673.

Liu, S., Sun, W., & Winters, J.V. (2017). The Great Recession and college major choice. Working paper.

Luckstead, J., & Devadoss, S. (2014a). A comparison of city size distributions for China and India from 1950 to 2010. Economics Letters, 124(2), 290-295.

Luckstead, J., & Devadoss, S. (2014b). A nonparametric analysis of the growth process of Indian cities. Economics Letters, 124(3), 516-519.

Mallick, J. (2014). Regional convergence of economic growth during post-reform period in India. The Singapore Economic Review, 59(2), 1450012.

Marshall, P. J. (2001). The Cambridge illustrated history of the British Empire: Cambridge University Press.

Mitra, A., & Mehta, B. (2011). Cities as the engine of growth: Evidence from India. Journal of Urban Planning and Development, 137(2), 171-183.

Moretti, E. (2004). Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. Journal of Econometrics, 121(1–2), 175-212.

Mukim, M. (2015). Coagglomeration of formal and informal industry: evidence from India. Journal of Economic Geography, 15(2), 329-351.

Myrdal, G. (1957). Economic theory and under-developed regions. Harper & Row, New York, NY.

Ohinata, A., & van Ours, J. C. (2012). Young immigrant children and their educational attainment. Economics Letters, 116(3), 288-290.

Panagariya, A. (2001). India's economic reforms what has been accomplished? What remains to be done? Asian Development Bank (ADB), Economics Research and Development Policy (ERDC) Briefs No.2.

Partridge, M., Bollman, R. D., Olfert, M. R., & Alasia, A. (2007). Riding the wave of urban growth in the countryside: Spread, backwash, or stagnation? Land economics, 83(2), 128-152.

Partridge, M., Olfert, M. R., & Alasia, A. (2007). Canadian cities as regional engines of growth: agglomeration and amenities, Canadian Journal of Economics, 40(1), 39-68.

Partridge, M. D., Ali, K., & Olfert, M. (2010). Rural-to-urban commuting: Three degrees of integration. Growth and Change, 41(2), 303-335.

Partridge, M. D., Olfert, M. R., & Ali, K. (2009). Towards a rural development policy: lessons from the United States and Canada. Journal of Regional Analysis and Policy, 39(2), 109-125.

Partridge, M. D., & Rickman, D. S. (2008). Distance from urban agglomeration economies and rural poverty. Journal of Regional Science, 48(2), 285-310.

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2008a). Employment growth in the American urban hierarchy: long live distance. The BE Journal of Macroeconomics, 8(1), 1-36.

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2008b). Lost in space: population growth in the American hinterlands and small cities. Journal of Economic Geography, 8(6), 727-757.

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2009). Agglomeration spillovers and wage and housing cost gradients across the urban hierarchy. Journal of International Economics, 78(1), 126-140.

Partridge, M. D., Rickman, D. S., Ali, K., & Olfert, M. R. (2010). Recent spatial growth dynamics in wages and housing costs: Proximity to urban production externalities and consumer amenities.

Regional Science and Urban Economics, 40(6), 440-452.

Peri, G., Shih, K., & Sparber, C. (2015). STEM workers, H-1B visas, and productivity in US cities. Journal of Labor Economics, 33(S1), S225-S255.

Perlmann, J. (1989). Ethnic differences: Schooling and social structure among the Irish, Italians, Jews, and Blacks in an American city, 1880-1935: Cambridge University Press.

Polèse, M., & Shearmur, R. (2004). Is distance really dead? Comparing industrial location patterns over time in Canada. International Regional Science Review, 27(4), 431-457.

Ransom, T., & Winters, J. V. (2016). Do foreigners crowd natives out of STEM degrees and occupations? Evidence from the US Immigration Act of 1990. IZA DP No. 9920.

Rappaport, J. (2007). Moving to nice weather. Regional Science and Urban Economics, 37(3), 375-398.

Rappaport, J., & Sachs, J. D. (2003). The United States as a coastal nation. Journal of Economic Growth, 8(1), 5-46.

Revi, A., Idicheria, C., Jain, G., Anand, G., Sudhira, H., Seddon, J., & Srinivasan, S. (2011). Urban India 2011: Evidence. Paper presented at the Indian Institute of Human settlements India Urban Conference: Evidence and Experience (IUC 2011).

Roback, J. (1982). Wages, rents, and the quality of life. Journal of Political Economy, 90(6), 1257-1278.

Rodríguez-Pose, A., & Ketterer, T. D. (2012). Do local amenities affect the appeal of regions in Europe for migrants. Journal of Regional Science, 52(4), 535-561.

Rong, X. L., & Brown, F. (2001). The effects of immigrant generation and ethnicity on educational attainment among young African and Caribbean Blacks in the United States. Harvard Educational

Review, 71(3), 536-566.

Rosen, S. (1979). Wage-based indexes of urban quality of life. Current issues in urban economics, 3, 324-345.

Ruggles, S. J., Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek (2017). Integrated Public Use Microdata Series: Version 6.0 [dataset]. Minneapolis: University of Minnesota.

Sachs, J. D., Bajpai, N., & Ramiah, A. (2002). Understanding regional economic growth in India. Asian Economic Papers, 1(3), 32-62.

Sankhe, S., & Vittal, I. (2010). India's urban awakening: building inclusive cities. sustaining economic growth, McKinsey Global Institute Report.

Schaafsma, J., & Sweetman, A. (2001). Immigrant earnings: age at immigration matters. Canadian Journal of Economics, 1066-1099.

Schaffar, A., & Dimou, M. (2012). Rank-size city dynamics in China and India, 1981–2004. Regional Studies, 46(6), 707-721.

Sharma, A. (2016). Urban proximity and spatial pattern of land use and development in rural India. The Journal of Development Studies, 52(11), 1593-1611.

Sharma, R. (2013). The Rise of the Rest of India: How States Have Become the Engines of Growth. Foreign Affairs, 92, 75.

Shingal, A. (2014). Services globalization and sub-national demand linkages: The case of India. Journal of International Commerce, Economics and Policy, 5(01), 1440002.

Soo, K. T. (2014). Zipf, Gibrat and geography: evidence from China, India and Brazil. Papers in Regional Science, 93(1), 159-181.

Sridhar, K. S. (2010). Determinants of city growth and output in India. Review of Urban & Regional Development Studies, 22(1), 22-38.

Statistical Book (2017). Five Year Plans, Chapter 7, 1-13.

Stiefel, L., Schwartz, A. E., & Conger, D. (2010). Age of entry and the high school performance of immigrant youth. Journal of Urban Economics, 67(3), 303-314.

Tomozawa, K. (2016). Development and spatial dynamism of the automobile component industry in India. Journal of Urban and Regional Studies on Contemporary India, 3(1), 11-27.

Tripathi, S. (2013). Do large agglomerations lead to economic growth? Evidence from urban India. Review of Urban & Regional Development Studies, 25(3), 176-200.

van Ours, J. C., & Veenman, J. (2006). Age at immigration and educational attainment of young immigrants. Economics Letters, 90(3), 310-316.

Winser, S. (2002). Vasco da Gama. History BBC,17.

Winters, J. V. (2014). STEM graduates, human capital externalities, and wages in the US. Regional Science and Urban Economics, 48, 190-198.

Winters, J. V. (2015). Do STEM and Non-STEM college graduates increase innovation? Evidence from U.S. Metropolitan Areas. Working paper.

World Bank. (2009). World Development Report 2009: Reshaping Economic Geography. World Bank.

World Bank. (2013a). Planning, Connecting, and Financing Cities—Now: Priorities for City Leaders, Washington, DC, World Bank.

World Bank. (2013b). Urbanization beyond municipal boundaries: nurturing metropolitan

economies and connecting peri-urban areas in India. Directions in development. Washington, DC, World Bank.

Ying, L. G. (2000). Measuring the spillover effects: Some Chinese evidence. Papers in Regional Science, 79(1), 75-89.

Zafar, B. (2013). College major choice and the gender gap. Journal of Human Resources, 48(3), 545-595.

Table 1.1: Descriptive Statistics

Variable	(1) Mean	(2) Std. Dev.	(3) Min	(4) Max
Growth rate of population (2001-2011)	20.31	74.09	-98.91	3909.88
Growth rate of population (1991-2001)	30.29	60.08	-97.63	1846.28
Population 2001	55413.21	253411	338	12000000
Population 2011	66482.67	307584	110	12400000
Population density	4426.21	5921.17	41	104267
Nearest city distance	**53.57**	**64.46**	**0**	**1232**
(Nearest city distance)^2	7187.66	41386.04	0	1517824
Nearest district headquarter distance	**37.25**	**32.06**	**0**	**296**
(Nearest district headquarter distance)^2	2414.92	4267.95	0	87616
Incremental distance to a city	16.70	65.65	-203	1190
(Incremental distance to a city)^2	4588.22	39244.38	0	1416100
Incremental distance to a large city	**80.74**	**99.45**	**0**	**551**
(Incremental distance to a large city)^2	16407.77	36073	0	303601
Average rainfall	1144.84	774.02	16	10270
Maximum temperature	36.78	5.58	9	88.4
Minimum temperature	14.64	7.13	-14.4	35
Nearest railway distance	21.85	56.14	0	1232
Nearest hospital distance	4.07	11.20	0	129
Nearest college distance	63.17	71.41	0	1232
Nearest school distance	1.78	6.56	0	150

Notes: The sample consists of towns and cities. Cities are often referred as Class I towns, population of 100,000 and above. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city (population of 100,000 and more) and is equal to zero if the town itself is a city. Nearest district headquarter distance considers the distance from a town to the nearest district. Incremental distance to a large city measures the additional distance from a district headquarters to a city with population of 100,000 and more. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. Amenity distance variables are zero if that amenity is present in that town. There are 5,177 towns according to 2001 Census of India. In 2011 Census, there were additional towns, which were not part of our 2001 Census data set. Thus, due to missing data, definitional problems and matching data of 2001 with 2011 resulted in 5,161 towns. Overall, the omission of data was by 0.3% in town sample.

Table 1.2: Town-level Analysis of Population Growth 2001-2011

Variable	(1) Natural Amenity	(2) Distance to nearest city	(3) Inc dist city (500,000+)	(4) Base Model (+Demog)	(5) Amenity distance
Average Rainfall	-0.002*	-0.002	-0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Max Temperature	-0.084	-0.188	-0.069	0.069	0.026
	(0.170)	(0.162)	(0.168)	(0.166)	(0.170)
Min Temperature	-0.157	-0.219	-0.216	0.159	0.030
	(0.153)	(0.145)	(0.145)	(0.206)	(0.214)
Nearest city dis		**-0.080*** **	**-0.081*** **	**-0.113*** **	**-0.073*** **
		(0.018)	(0.019)	(0.025)	(0.028)
(Nearest city dis)^2		0.009***	0.008***	0.017***	0.010
		(0.003)	(0.003)	(0.006)	(0.006)
Inc dis large city			**-0.051*** **	**-0.049*** **	**-0.038** **
			(0.016)	(0.016)	(0.015)
(Inc dis large city)^2			0.008**	0.010***	0.007**
			(0.004)	(0.004)	(0.003)
Density				-0.017*	-0.021*
				(0.009)	(0.011)
Nearest rail dis					-0.031
					(0.047)
Nearest hosp dis					-0.051
					(0.100)
Nearest college dis					-0.015
					(0.014)
Nearest school dis					-0.365*
					(0.195)
Observations	4,779	4,779	4,779	4,761	3,922
R-squared	0.003	0.009	0.014	0.075	0.092
State Fixed Effects	No	No	No	Yes	Yes

Notes: The sample consists of towns and cities. Class I towns are often referred as cities, population of 100,000 and above. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city and is equal to zero if the town itself is a city. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. All the distance square terms are expressed as hundreds of square kilometers. Population Density is also reported as hundreds of square kilometers. Amenity distance variables are zero if that amenity is present in that town. Standard Errors are clustered by districts. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 1.3: Town-level Analysis of Population Growth 2001-2011 using initial population

Variable	(1) Natural Amenity	(2) Distance to nearest city	(3) Inc dist city (500,000+)	(4) Base Model (+Demog)	(5) Amenity distance
Average Rainfall	-0.002*	-0.002	-0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Max Temperature	-0.084	-0.188	-0.069	0.074	0.037
	(0.170)	(0.162)	(0.168)	(0.164)	(0.169)
Min Temperature	-0.157	-0.219	-0.216	0.155	0.026
	(0.153)	(0.145)	(0.145)	(0.204)	(0.211)
Nearest city dis		**-0.080*** **	**-0.081*** **	**-0.110*** **	**-0.067** **
		(0.018)	(0.019)	(0.025)	(0.028)
(Nearest city dis)^2		0.009***	0.008***	0.017***	0.009
		(0.003)	(0.003)	(0.006)	(0.006)
Incr dist large city			**-0.051*** **	**-0.048*** **	**-0.036** **
			(0.016)	(0.016)	(0.015)
(Incr dist large city)^2			0.008**	0.010***	0.007**
			(0.004)	(0.004)	(0.003)
Initial Population				-0.012	0.070
				(0.162)	(0.190)
Nearest rail dis					-0.029
					(0.047)
Nearest hospital dis					-0.046
					(0.099)
Nearest college dis					-0.016
					(0.014)
Nearest school dis					-0.358*
					(0.195)
Observations	4,779	4,779	4,779	4,779	3,938
R-squared	0.003	0.009	0.014	0.081	0.100
State Fixed Effects	No	No	No	Yes	Yes

Notes: The sample consists of towns and cities. Class I towns are often referred as cities, population of 100,000 and above. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city and is equal to zero if the town itself is a city. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. All the distance square terms are expressed as hundreds of square kilometers. Initial Population (2001) is reported as hundreds of thousands unit. Amenity distance variables are zero if that amenity is present in that town. Standard Errors are clustered by districts. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 1.4: Town-level Analysis of Population Growth 2001-2011 considering distance to the nearest district headquarter

Variable	(1) Natural Amenity	(2) Distance to nearest city	(3) Inc dist city (100,000+)	(4) Inc dist city (500,000+)	(5) Base Model (+Demog)	(6) Amenity distance
Average Rainfall	-0.002***	-0.003***	-0.003***	-0.003***	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Max Temperature	-0.086	-0.052	-0.162	-0.074	-0.013	-0.020
	(0.110)	(0.100)	(0.102)	(0.104)	(0.126)	(0.133)
Min Temperature	-0.158*	-0.092	-0.159**	-0.161**	0.173	0.128
	(0.084)	(0.076)	(0.077)	(0.077)	(0.131)	(0.141)
Nearest district dis		**-0.146*** **	**-0.166*** **	**-0.153*** **	**-0.120*** **	**-0.147*** **
		(0.035)	(0.035)	(0.035)	(0.036)	(0.039)
(Nearest dist dis)^2		0.056**	0.050*	0.044*	0.017	0.032
		(0.026)	(0.026)	(0.026)	(0.028)	(0.029)
Incr city distance			**-0.060*** **	**-0.064*** **	**-0.088*** **	**-0.067*** **
			(0.012)	(0.012)	(0.015)	(0.017)
(Incr city distance)^2			0.007***	0.007***	0.017***	0.009
			(0.003)	(0.003)	(0.006)	(0.006)
Incr dist large city				**-0.045*** **	**-0.047*** **	**-0.038*** **
				(0.011)	(0.012)	(0.012)
(Incr dis large city)^2				0.008***	0.010***	0.007**
				(0.003)	(0.003)	(0.003)
Density					-0.017*	-0.019**
					(0.009)	(0.010)
Nearest rail dis						0.032
						(0.021)
Nearest hospital dis						0.060
						(0.051)
Nearest college dis						-0.002
						(0.011)
Nearest school dis						-0.192**
						(0.090)
Observations	4,780	4,766	4,766	4,766	4,748	3,911
R-squared	0.003	0.011	0.016	0.020	0.073	0.090
State Fixed Effects	No	No	No	No	Yes	Yes

Notes: The sample consist of towns and cities. Class I towns are often referred as cities, population of 100,000 and above. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest district headquarter distance captures the distance from a town to the nearest district. Incremental distance to a large city measures the additional distance from a district headquarter to a city with a population of 100,000 and more. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. All the distance square terms are expressed as hundreds of square kilometers. Population Density is also reported as hundreds of square kilometers. Amenity distance variables are zero if that amenity is present in that town. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 1.5: Town-level Analysis of Population Growth 2001-2011 using non-JNNURM sample

Variable	(1) Natural Amenity	(2) Distance to nearest city	(3) Inc dist city (500,000+)	(4) Base Model (+Demog)	(5) Amenity distance
Average Rainfall	-0.002**	-0.002	-0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Max Temperature	-0.074	-0.182	-0.069	0.102	0.065
	(0.171)	(0.161)	(0.168)	(0.157)	(0.158)
Min Temperature	-0.138	-0.202	-0.216	0.161	0.027
	(0.156)	(0.146)	(0.145)	(0.207)	(0.214)
Nearest city dis		**-0.078*** **	**-0.081*** **	**-0.111*** **	**-0.070*** **
		(0.018)	(0.019)	(0.025)	(0.027)
(Nearest city dis)^2		0.009***	0.008***	0.016***	0.009
		(0.003)	(0.003)	(0.006)	(0.006)
Inc dis large city			**-0.051*** **	**-0.050*** **	**-0.039** **
			(0.016)	(0.016)	(0.015)
(Inc dis large city)^2			0.008**	0.011***	0.008**
			(0.004)	(0.004)	(0.003)
Density				-0.019**	-0.024**
				(0.009)	(0.011)
Nearest rail dis					-0.041
					(0.046)
Nearest hosp dis					-0.060
					(0.106)
Nearest college dis					-0.013
					(0.014)
Nearest school dis					-0.346*
					(0.205)
Observations	4,715	4,715	4,779	4,699	3,863
R-squared	0.003	0.010	0.014	0.080	0.099
State Fixed Effects	No	No	No	Yes	Yes

Notes: The sample consist of towns and cities are not covered by JNNURM (Jawaharlal Nehru National Urban Renewal Mission). Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city and is equal to zero if the town itself is a city. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. Amenity distance variables are zero if that amenity is present in that town. Standard Errors are clustered by districts. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 1.6: Town-level Analysis of Population Growth 2001-2011 using non-IT sample

Variable	(1) Natural Amenity	(2) Distance to nearest city	(3) Inc dist city (500,000+)	(4) Base Model (+Demog)	(5) Amenity distance
Average Rainfall	-0.002*	-0.002	-0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Max Temperature	-0.100	-0.201	-0.083	0.063	0.019
	(0.169)	(0.161)	(0.167)	(0.165)	(0.170)
Min Temperature	-0.142	-0.202	-0.200	0.169	0.043
	(0.152)	(0.143)	(0.144)	(0.206)	(0.214)
Nearest city dis		-0.077***	-0.078***	-0.110***	-0.069**
		(0.018)	(0.019)	(0.025)	(0.028)
(Nearest city dis)^2		0.008***	0.008***	0.016***	0.009
		(0.003)	(0.003)	(0.006)	(0.006)
Inc dis large city			-0.051***	-0.047***	-0.035**
			(0.017)	(0.016)	(0.015)
(Inc dis large city)^2			0.008**	0.010***	0.007**
			(0.004)	(0.004)	(0.003)
Density				-0.017*	-0.021*
				(0.009)	(0.011)
Nearest rail dis					-0.031
					(0.047)
Nearest hosp dis					-0.042
					(0.100)
Nearest college dis					-0.015
					(0.014)
Nearest school dis					-0.366*
					(0.195)
Observations	4,762	4,762	4,762	4,744	3,906
R-squared	0.002	0.009	0.014	0.075	0.092
State Fixed Effects	No	No	No	Yes	Yes

Notes: The sample consist of towns and cities which are not IT (information technology) driven. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city and is equal to zero if the town itself is a city. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. All the distance square terms are expressed as hundreds of square kilometers. Population Density is also reported as hundreds of square kilometers. Amenity distance variables are zero if that amenity is present in that town. Standard Errors are clustered by districts. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 1.7: Quantile Town-level Analysis of Population Growth 2001-2011

	(1)	(2)	(3)
Variable	.25	.50	.95
Average Rainfall	-0.001**	-0.001	0.004
	(0.000)	(0.000)	(0.003)
Max Temperature	0.037	0.008	-0.230
	(0.057)	(0.050)	(0.417)
Min Temperature	0.021	-0.096*	0.282
	(0.060)	(0.053)	(0.437)
Nearest city dis	-0.018**	-0.032***	-0.345***
	(0.008)	(0.007)	(0.060)
(Nearest city dis)^2	-0.002	0.003	0.100***
	(0.003)	(0.002)	(0.020)
Inc dist large city	-0.016***	-0.015***	-0.101***
	(0.005)	(0.005)	(0.039)
(Inc dist large city)^2	0.004***	0.004***	0.023**
	(0.001)	(0.001)	(0.011)
Density	-0.010**	-0.005	-0.026
	(0.004)	(0.003)	(0.029)
Observations	4,762	4,762	4,762
State Fixed Effects	Yes	Yes	Yes

Notes: The sample consist of towns and cities. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city and is equal to zero if the town itself is a city. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. All the distance square terms are expressed as hundreds of square kilometers. Population Density is also reported as hundreds of square kilometers. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 1.8: Town-level Analysis of Population Growth 2001-2011 by different town size

Variable	(1) Small town sample	(2) Large town sample
Average Rainfall	0.000	-0.001
	(0.001)	(0.004)
Max Temperature	-0.061	0.584
	(0.138)	(0.787)
Min Temperature	0.003	1.805*
	(0.216)	(1.016)
Nearest city dis	-0.106***	-1.873*
	(0.025)	(1.132)
(Nearest city dis)^2	0.016***	4.268
	(0.006)	(2.699)
Inc distance city	-0.034**	-0.158**
	(0.014)	(0.073)
(Inc distance city)^2	0.007**	0.037**
	(0.003)	(0.018)
Density	-0.016*	0.008
	(0.009)	(0.029)
Observations	4,199	407
R-squared	0.106	0.136
State Fixed Effects	Yes	Yes

Notes: The small town sample consist of towns whose population is greater than equal to 5,000 and less than 100,000 and large town sample include those towns whose populations is more than 100,0000. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city and is equal to zero if the town itself is a city. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. All the distance square terms are expressed as hundreds of square kilometers. Population Density is also reported as hundreds of square kilometers. Standard Errors are clustered by districts. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 1.9: Town-level Analysis of Population Growth 2001-2011 using 1991 population

Variable	(1) Natural Amenity	(2) Distance to nearest city	(3) Inc dist city (500,000+)	(4) Base Model (+Demog)	(5) Amenity distance
Average Rainfall	-0.002*	-0.002	-0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Max Temperature	-0.084	-0.188	-0.069	0.170	0.120
	(0.170)	(0.162)	(0.168)	(0.171)	(0.172)
Min Temperature	-0.157	-0.219	-0.216	0.264	0.121
	(0.153)	(0.145)	(0.145)	(0.184)	(0.186)
Nearest city dis		-0.080***	-0.081***	-0.091***	-0.042
		(0.018)	(0.019)	(0.023)	(0.027)
(Nearest city dis)^2		0.009***	0.008***	0.015**	0.006
		(0.003)	(0.003)	(0.006)	(0.007)
Inc dist large city			-0.051***	-0.037***	-0.026*
			(0.016)	(0.014)	(0.014)
(Inc dist large city)^2			0.008**	0.007**	0.005
			(0.004)	(0.003)	(0.003)
1991 Population				-0.059	0.051
				(0.188)	(0.219)
Nearest rail dis					-0.053
					(0.058)
Nearest hospital dis					-0.035
					(0.126)
Nearest college dis					-0.016
					(0.014)
Nearest school dis					-0.413*
					(0.225)
Observations	4,779	4,779	4,779	4,057	3,352
R-squared	0.003	0.009	0.014	0.067	0.084
State Fixed Effects	No	No	No	Yes	Yes

Notes: The sample consist of towns and cities. Class I towns are often referred as cities, population of 100,000 and above. Temperatures are measured in centigrade and rainfall is recorded in millimeters. All the distance variables are measured in terms of road distance in kilometers. Nearest city distance captures the road distance in kilometers from a town to the nearest city and is equal to zero if the town itself is a city. Incremental distance to a large city measures the additional distance from a city with population of 100,000 and more to a higher ordered city with population of 500,000 and more. All the distance square terms are expressed as hundreds of square kilometers. Population Density is also reported as hundreds of square kilometers. Amenity distance variables are zero if that amenity is present in that town. Standard Errors are clustered by districts. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 2.1: Weighted STEM major rates of different race groups conditional on having a bachelor's degree and more

	STEM major rates		
	Full Sample	**Natives**	**Immigrants**
Groups			
White	22.594	21.618	38.130
Black	19.200	17.274	28.550
Asian	45.075	35.121	47.741
Hispanic	22.445	19.239	27.492
Other Races	26.108	22.519	41.332
Detailed Asian Groups			
Chinese	28.623	28.621	28.624
Japanese	16.674	16.350	17.051
Filipino	12.172	12.159	12.177
Indian	46.685	33.293	48.354
Korean	17.882	18.816	17.682
Vietnamese	14.748	19.656	13.892
Other Asians	13.519	9.998	14.995

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 American Community Survey (ACS). STEM major is a college major whose either first or second field is a STEM (Science, Technology, Engineering and Mathematics). All of these numbers represent percentages. Immigrant is a foreign-born person who was born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are defined by race and ethnicity.

Table 2.2: Weighted STEM major rates of broad race and detailed Asian immigrant groups by age of arrival conditional on having a bachelor's degree and more

	0-5	**6-11**	**12-17**	**18-21**	**22-24**	**25-40**
Groups						
Hispanic	0.199	0.219	0.261	0.266	0.278	0.311
Asian	0.356	0.403	0.461	0.517	0.559	0.490
White	0.277	0.289	0.370	0.367	0.371	0.420
Black	0.236	0.222	0.283	0.323	0.301	0.297
Other Races	0.273	0.310	0.401	0.486	0.513	0.425
Asian Groups						
Indian	0.467	0.457	0.527	0.682	0.707	0.649
Chinese	0.399	0.429	0.468	0.500	0.581	0.555
Filipino	0.279	0.308	0.290	0.221	0.197	0.241
Other Asians	0.341	0.371	0.452	0.475	0.419	0.389
Vietnamese	0.406	0.478	0.611	0.635	0.551	0.417
Korean	0.258	0.326	0.378	0.319	0.285	0.325
Japanese	0.278	0.367	0.266	0.204	0.164	0.298

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. STEM major is a college major whose either first or second field is a STEM (Science, Technology, Engineering and Mathematics). Immigrant is a foreign-born person who was born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are defined by race and ethnicity and are arranged in order of having the largest weighted population. Age of arrival, captures the difference between a person's actual age and number of years in the US for immigrant persons.

Table 2.3: Effects of age of arrival on attainment of STEM major by broad race immigrant groups

Age of Arrival	(1) Hispanic	(2) Asian	(3) White	(4) Black	(5) Other Races
0-5 years	-0.013***	0.144***	0.059***	0.031***	0.066***
	(0.004)	(0.004)	(0.004)	(0.011)	(0.014)
6-11 years	0.004	0.184***	0.070***	0.019**	0.095***
	(0.005)	(0.004)	(0.005)	(0.008)	(0.017)
12-17 years	0.042***	0.241***	0.148***	0.077***	0.177***
	(0.005)	(0.004)	(0.005)	(0.007)	(0.017)
N	2,794,307	2,827,494	2,799,740	2,771,957	2,763,595

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 American Community Survey (ACS). Dependent variable is an indicator for completing a STEM major, conditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM (Science, Technology, Engineering and Mathematics). Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

Table 2.4: Effects of age of arrival on attainment of STEM major by detailed Asian immigrant groups

Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.245***	0.180***	0.063***	0.130***	0.193***	0.062***	0.066***
	(0.010)	(0.008)	(0.010)	(0.011)	(0.012)	(0.008)	(0.024)
6-11 years	0.235***	0.210***	0.094***	0.150***	0.260***	0.107***	0.155***
	(0.011)	(0.007)	(0.010)	(0.012)	(0.010)	(0.010)	(0.034)
12-17 years	0.299***	0.250***	0.086***	0.227***	0.379***	0.159***	0.063**
	(0.009)	(0.007)	(0.010)	(0.011)	(0.010)	(0.010)	(0.029)
N	2,770,558	2,778,532	2,768,032	2,767,777	2,769,443	2,770,448	2,760,778

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, conditional on having a bachelor's degree or more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *$p<0.1$; ** $p<0.05$; *** $p<0.01$.

Table 2.5: Effects of age of arrival on attainment of STEM major by sub fields

Panel A: Computer Science							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age of Arrival	Indian	Chinese	Filipino	Other Asian	Vietnamese	Korean	Japanese
0-5 years	0.036***	0.040***	0.021***	0.045***	0.036***	-0.005	-0.006
	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)	(0.003)	(0.007)
6-11 years	0.049***	0.055***	0.036***	0.049***	0.047***	0.007	0.010
	(0.007)	(0.005)	(0.006)	(0.008)	(0.006)	(0.005)	(0.013)
12-17 years	0.087***	0.075***	0.038***	0.051***	0.104***	0.022***	0.017
	(0.007)	(0.004)	(0.006)	(0.006)	(0.008)	(0.005)	(0.016)
N	2,411,367	2,418,459	2,409,431	2,409,140	2,410,666	2,411,505	2,403,212

Panel B: Engineering							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age of Arrival	Indian	Chinese	Filipino	Other Asian	Vietnamese	Korean	Japanese
0-5 years	0.067***	0.053***	0.024***	0.020***	0.024***	0.009**	0.038**
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.004)	(0.017)
6-11 years	0.070***	0.082***	0.028***	0.041***	0.088***	0.022***	0.079***
	(0.008)	(0.006)	(0.007)	(0.008)	(0.008)	(0.006)	(0.028)
12-17 years	0.120***	0.122***	0.033***	0.110***	0.192***	0.065***	0.047**
	(0.008)	(0.006)	(0.007)	(0.010)	(0.009)	(0.008)	(0.020)
N	2,411,367	2,418,459	2,409,431	2,409,140	2,410,666	2,411,505	2,403,212

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, conditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table 2.6: Sex-Specific Effects of age of arrival on attainment of STEM major

Panel A: Female Sample							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age of Arrival	Indian	Chinese	Filipino	Other Asian	Vietnamese	Korean	Japanese
0-5 years	0.245***	0.164***	0.060***	0.100***	0.179***	0.060***	0.045*
	(0.014)	(0.011)	(0.012)	(0.013)	(0.015)	(0.009)	(0.027)
6-11 years	0.206***	0.187***	0.091***	0.117***	0.212***	0.097***	0.147***
	(0.015)	(0.010)	(0.013)	(0.015)	(0.014)	(0.013)	(0.042)
12-17 years	0.258***	0.200***	0.063***	0.190***	0.319***	0.133***	0.072**
	(0.014)	(0.009)	(0.011)	(0.016)	(0.015)	(0.013)	(0.035)
N	1,492,681	1,497,172	1,491,926	1,491,477	1,492,207	1,493,406	1,487,906

Panel B: Male Sample							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age of Arrival	Indian	Chinese	Filipino	Other Asian	Vietnamese	Korean	Japanese
0-5 years	0.246***	0.199***	0.067***	0.166***	0.210***	0.061***	0.091**
	(0.015)	(0.013)	(0.016)	(0.018)	(0.018)	(0.014)	(0.042)
6-11 years	0.264***	0.236***	0.097***	0.185***	0.310***	0.117***	0.165***
	(0.016)	(0.011)	(0.017)	(0.018)	(0.015)	(0.016)	(0.054)
12-17 years	0.339***	0.306***	0.121***	0.262***	0.429***	0.185***	0.051
	(0.013)	(0.010)	(0.017)	(0.016)	(0.012)	(0.016)	(0.048)
N	1,277,877	1,281,360	1,276,106	1,276,300	1,277,236	1,277,042	1,272,872

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, conditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table 2.7: Effects of age of arrival on attainment of STEM major relative to native Asian groups

Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.020*	0.007	-0.008	0.021*	-0.025	-0.019*	0.001
	(0.012)	(0.009)	(0.011)	(0.012)	(0.016)	(0.011)	(0.025)
6-11 years	0.011	0.035***	0.025**	0.042***	0.028*	0.019	0.084**
	(0.012)	(0.008)	(0.012)	(0.013)	(0.015)	(0.013)	(0.034)
12-17 years	0.077***	0.075***	0.017	0.125***	0.124***	0.071***	0.002
	(0.011)	(0.008)	(0.011)	(0.013)	(0.016)	(0.013)	(0.029)
N	21,649	41,186	19,699	18,439	14,291	17,172	11,322

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, conditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. For each immigrant group, their corresponding natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table 2.8: Effects of age of arrival on STEM Occupation

Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.042***	0.065***	0.015***	0.036***	0.051***	0.005	0.020
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.004)	(0.016)
6-11 years	0.068***	0.077***	0.026***	0.039***	0.075***	0.013**	0.051**
	(0.008)	(0.005)	(0.007)	(0.007)	(0.007)	(0.006)	(0.021)
12-17 years	0.090***	0.099***	0.032***	0.072***	0.163***	0.023***	0.032**
	(0.007)	(0.005)	(0.006)	(0.008)	(0.009)	(0.006)	(0.016)
N	2,770,558	2,778,532	2,768,032	2,767,777	2,769,443	2,770,448	2,760,778

Notes: The sample includes individuals employed in STEM occupations from 2009-2016 ACS. Dependent variable is an indicator for having a STEM occupation, conditional on having a bachelor's degree and more. STEM occupations are defined on the basis of the 2010 Census Bureau occupational classification scheme. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. * p<0.1; ** p<0.05; *** p<0.01.

Table 2.9: Effects of age of arrival on attainment of Bachelor's degree and more

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age of Arrival	Indian	Chinese	Filipino	Other Asian	Vietnamese	Korean	Japanese
0-5 years	0.367***	0.355***	0.105***	0.007	0.188***	0.265***	0.257***
	(0.008)	(0.007)	(0.008)	(0.007)	(0.009)	(0.007)	(0.022)
6-11 years	0.290***	0.303***	0.033***	-0.006	0.147***	0.297***	0.257***
	(0.009)	(0.006)	(0.007)	(0.007)	(0.008)	(0.009)	(0.027)
12-17 years	0.201***	0.187***	-0.060***	-0.045***	0.008	0.208***	0.216***
	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)	(0.024)
N	8,031,251	8,044,259	8,036,455	8,036,479	8,035,150	8,031,556	8,016,117

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a bachelor's degree and more. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

Table 3.1: Summary Statistics

Variable	(1) Mean	(2) Std. Dev.	(3) Min	(4) Max
Growth rate of population (2001-2011)	26.28	387.94	-100	163000
Population 2001	1226.80	1730.44	1	68205
Population 2011	1420.60	1983.39	0	66062
Population density	586.21	6770.02	0.01	3170000
Nearest town distance	**24.73**	**25.16**	**1**	**1717**
(Nearest town distance)^2	**1244.57**	**7176.37**	**1**	**2948089**
Incremental distance to a city	**77.49**	**72.18**	**0**	**1232**
(Incremental distance to a city)^2	**11215.22**	**46071.36**	**0**	**1517824**
Incremental distance to a large city	**96.88**	**108.04**	**0**	**551**
(Incremental distance to a large city)^2	**21058.64**	**38371.43**	**0**	**303601**
Distance range of school	0.96	0.83	0	3
Distance range of college	2.59	0.69	0	3
Distance range of hospital	2.42	0.80	0	3
Distance range of railway	2.61	0.73	0	3
Distance range of bank	1.87	0.91	0	3

Notes: The rural sample consists of villages. All the distance variable are measured in terms of road distance in kilometers. Nearest town distance captures the road distance in kilometers from a village to the nearest town. Incremental distance to a city distance measures the additional distance in kilometers from a town to the nearest city, with a population of 100,000 and more. Incremental distance to a large city measures the additional distance in kilometers from a city with a population of 100,000 and more to a higher ordered city with a population of 500,000 and above. For rural sample, there are range codes for the different amenity variables. Range Code is “0” if that amenity is present in that village, “1” if that amenity is in the range of less than 5 kilometers, “2” if that amenity is available within 5-10 kilometers, and “3” if that amenity is in the range greater than 10 kilometers.

Table 3.2: Village-level Analysis of Population Growth 2001-2011

Variable	(1) Demog	(2) Near town distance	(3) Incr dist city	(4) Incr dist large city	(5) Amenity distance
Pop Density 2001	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dist to nearest town		-0.084***	-0.075**	-0.084***	-0.096***
		(0.029)	(0.030)	(0.031)	(0.032)
(Dist to near town)^2		0.011	0.011	0.011	0.011
		(0.010)	(0.010)	(0.010)	(0.010)
Inc dist to a city			-0.006	-0.003	-0.003
			(0.014)	(0.014)	(0.014)
(Inc dist to a city)^2			-0.001	-0.001	-0.000
			(0.003)	(0.003)	(0.003)
Inc dist large city				0.013	0.016
				(0.012)	(0.012)
(Inc dist large city)^2				-0.002	-0.003
				(0.003)	(0.003)
N	520,626	520,626	518,026	511,998	511,998
R-squared	0.000	0.000	0.000	0.000	0.001
Amenity Distance	No	No	No	No	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: The sample consists of villages. All the distance variable are measured in terms of road distance in kilometers. Nearest town distance captures the road distance in kilometers from a village to the nearest town. Incremental distance to a city distance measures the additional distance in kilometers from a town to the nearest city, with a population of 100,000 and more. Incremental distance to a large city measures the additional distance in kilometers from a city with a population of 100,000 and more to a higher ordered city with a population of 500,000 and above. Amenity distance variables include distance to nearest school, college, hospital, rail station and bank. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 3.3: Village-level Analysis of Population Growth 2001-2011 using restricted definition

Variable	(1) Demog	(2) Near town distance	(3) Incr dist city	(4) Incr dist large city	(5) Amenity distance
Pop Density 2001	-0.033***	-0.036***	-0.037***	-0.036***	-0.035***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Dist to nearest town		-0.067***	-0.054***	-0.054***	-0.048***
		(0.004)	(0.004)	(0.004)	(0.004)
(Dist to near town)^2		0.007***	0.005***	0.005***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
Inc dist to a city			-0.023***	-0.021***	-0.018***
			(0.002)	(0.002)	(0.002)
(Inc dist to a city)^2			0.002***	0.001***	0.001***
			(0.000)	(0.000)	(0.000)
Inc dist to a large city				0.008***	0.009***
				(0.002)	(0.002)
(Inc dist to a large city)^2				-0.002***	-0.002***
				(0.001)	(0.001)
N	286,873	286,873	285,593	282,629	282,629
R-squared	0.021	0.022	0.022	0.023	0.027
Amenity Distance	No	No	No	No	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: The sample consists of villages, which have population of 100 and more, and less than 5,000 for both 2001 and 2011 and having population density of less than 400 per square kilometer according to Census definition. All the distance variable are measured in terms of road distance in kilometers. Nearest town distance captures the road distance in kilometers from a village to the nearest town. Incremental distance to a city distance measures the additional distance in kilometers from a town to the nearest city, with a population of 100,000 and more. Incremental distance to a large city measures the additional distance in kilometers from a city with a population of 100,000 and more to a higher ordered city with a population of 500,000 and above. All the distance square terms are expressed as hundreds of square kilometers. Amenity distance variables include distance to nearest school, college, hospital, rail station and bank. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 3.4: Village-level Analysis of Population Growth 2001-2011 using initial population

Variable	(1) Demog	(2) Near town distance	(3) Incr dist city	(4) Incr dist large city	(5) Amenity distance
Population 2001	-0.004***	-0.004***	-0.004***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Dist to nearest town		-0.050***	-0.040***	-0.040***	-0.038***
		(0.004)	(0.004)	(0.004)	(0.004)
(Dist to near town)^2		0.004***	0.003***	0.003***	0.003***
		(0.001)	(0.001)	(0.001)	(0.001)
Inc dist to a city			-0.016***	-0.014***	-0.013***
			(0.002)	(0.002)	(0.002)
(Inc dist to a city)^2			0.001***	0.001***	0.001***
			(0.000)	(0.000)	(0.000)
Inc dist to a large city				0.012***	0.012***
				(0.002)	(0.002)
(Inc dist to a large city)^2				-0.002***	-0.003***
				(0.001)	(0.001)
N	286,873	286,873	285,593	282,629	282,629
R-squared	0.021	0.022	0.022	0.023	0.024
Amenity Distance	No	No	No	No	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: The sample consists of villages, which have population of 100 and more, and less than 5,000 for both 2001 and 2011 and having population density of less than 400 per square kilometer according to Census definition. All the distance variable are measured in terms of road distance in kilometers. Nearest town distance captures the road distance in kilometers from a village to the nearest town. Incremental distance to a city distance measures the additional distance in kilometers from a town to the nearest city, with a population of 100,000 and more. Incremental distance to a large city measures the additional distance in kilometers from a city with a population of 100,000 and more to a higher ordered city with a population of 500,000 and above. All the distance square terms are expressed as hundreds of square kilometers. Amenity distance variables include distance to nearest school, college, hospital, rail station and bank. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 3.5: Village-level Analysis of Population Growth 2001-2011 based on different distance rings

Variable	(1) Dist < 50 km	(2) Dist >= 50 km
Pop Density 2001	-0.036***	-0.027***
	(0.001)	(0.002)
Dist to nearest town	-0.240***	-0.018**
	(0.028)	(0.008)
(Dist to near town)^2	0.348***	0.001
	(0.057)	(0.001)
Inc dist to a city	-0.018***	-0.011***
	(0.002)	(0.003)
(Inc dist to a city)^2	0.001***	0.000
	(0.000)	(0.000)
Inc dist to a large city	0.009***	0.012***
	(0.002)	(0.005)
(Inc dist to a large city)^2	-0.002***	-0.002*
	(0.001)	(0.001)
N	248,731	33,898
R-squared	0.027	0.032
Amenity Distance	Yes	Yes
Region Fixed Effects	Yes	Yes

Notes: The sample consists of villages, which have population of 100 and more, and less than 5,000 for both 2001 and 2011 and having population density of less than 400 per square kilometer according to Census definition. All the distance variable are measured in terms of road distance in kilometers. Nearest town distance captures the road distance in kilometers from a village to the nearest town. Incremental distance to a city distance measures the additional distance in kilometers from a town to the nearest city, with a population of 100,000 and more. Incremental distance to a large city measures the additional distance in kilometers from a city with a population of 100,000 and more to a higher ordered city with a population of 500,000 and above. All the distance square terms are expressed as hundreds of square kilometers. Amenity distance variables include distance to nearest school, college, hospital, rail station and bank. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 3.6: Village-level Analysis of Population Growth 2001-2011 based on different village sizes

Variables	(1) 100-500	(2) 500-1000	(3) 1000-5000
Pop Density 2001	-0.043***	-0.020***	-0.013***
	(0.002)	(0.001)	(0.001)
Dist to nearest town	-0.068***	-0.026***	-0.033***
	(0.009)	(0.006)	(0.006)
(Dist to near town)^2	0.006***	0.002	0.009**
	(0.002)	(0.002)	(0.004)
Inc dist to a city	-0.040***	-0.009***	0.003*
	(0.005)	(0.002)	(0.002)
(Inc dist to a city)^2	0.000***	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Inc dist to a large city	0.013***	0.010***	0.004***
	(0.005)	(0.002)	(0.002)
(Inc dist to a large city)^2	-0.000	-0.003***	-0.001**
	(0.001)	(0.001)	(0.000)
N	113,038	82,269	91,348
R-squared	0.011	0.027	0.042
Amenity Distance	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes

Notes: The sample consists of villages, which have population density of less than 400 per square kilometer according to Census definition. All the distance variable are measured in terms of road distance in kilometers. Nearest town distance captures the road distance in kilometers from a village to the nearest town. Incremental distance to a city distance measures the additional distance in kilometers from a town to the nearest city, with a population of 100,000 and more. Incremental distance to a large city measures the additional distance in kilometers from a city with a population of 100,000 and more to a higher ordered city with a population of 500,000 and above. All the distance square terms are expressed as hundreds of square kilometers. Amenity distance variables include distance to nearest school, college, hospital, rail station and bank. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

Table 3.7: Quantile Village-level Analysis of Population Growth 2001-2011

Variables	(1) 25th percentile	(2) 50th percentile	(3) 95th percentile
Pop Density 2001	-0.005***	-0.014***	-0.082***
	(0.000)	(0.000)	(0.002)
Dist to nearest town	-0.021***	-0.010***	-0.066***
	(0.002)	(0.002)	(0.012)
(Dist to near town)^2	0.001**	0.000	0.014***
	(0.001)	(0.001)	(0.000)
Inc dist to a city	-0.009***	-0.001	-0.028***
	(0.001)	(0.001)	(0.005)
(Inc dist to a city)^2	0.000	-0.000***	0.003***
	(0.000)	(0.000)	(0.001)
Inc dist to a large city	0.017***	0.012***	0.006
	(0.001)	(0.001)	(0.005)
(Inc dist to a large city)^2	-0.004***	-0.003***	-0.004
	(0.000)	(0.000)	(0.002)
N	282,629	282,629	282,629
Amenity Distance	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes

Notes: The sample consists of villages, which have population density of less than 400 per square kilometer according to Census definition. All the distance variable are measured in terms of road distance in kilometers. Nearest town distance captures the road distance in kilometers from a village to the nearest town. Incremental distance to a city distance measures the additional distance in kilometers from a town to the nearest city, with a population of 100,000 and more. Incremental distance to a large city measures the additional distance in kilometers from a city with a population of 100,000 and more to a higher ordered city with a population of 500,000 and above. All the distance square terms are expressed as hundreds of square kilometers. Amenity distance variables include distance to nearest school, college, hospital, rail station and bank. *Significant at 10% level; **Significant at 5% level; ***Significant at 1% level.

APPENDICES

Table A1: List of STEM majors based on ACS codes in IPUMS

ACS	STEM major description
1103	Animal Sciences
1104	Food Science
1105	Plant Science and Agronomy
1106	Soil Science
1301	Environmental Science
1302	Forestry
2001	Communication Technologies
2100	Computer and Information Systems
2101	Computer Programming and Data Processing
2102	Computer Science
2105	Information Sciences
2106	Computer Information Management and Security
2107	Computer Networking and Telecommunications
2400	General Engineering
2401	Aerospace Engineering
2402	Biological Engineering
2403	Architectural Engineering
2404	Biomedical Engineering
2405	Chemical Engineering
2406	Civil Engineering
2407	Computer Engineering
2408	Electrical Engineering

2409 Engineering Mechanics, Physics and Science
2410 Environmental Engineering
2411 Geological and Geophysical Engineering
2412 Industrial and Manufacturing Engineering
2413 Materials Engineering and Materials Science
2414 Mechanical Engineering
2415 Metallurgical Engineering
2416 Mining and Mineral Engineering
2417 Naval Architecture and Marine Engineering
2418 Nuclear Engineering
2419 Petroleum Engineering
2499 Miscellaneous Engineering
2500 Engineering Technologies
2501 Engineering and Industrial Management
2502 Electrical Engineering Technology
2503 Industrial Production Technologies
2504 Mechanical Engineering Related Technologies
2599 Miscellaneous Engineering Technologies
3600 Biology
3601 Biochemical Sciences
3602 Botany
3603 Molecular Biology
3604 Ecology
3605 Genetics
3606 Microbiology
3607 Pharmacology
3608 Physiology
3609 Zoology
3611 Neuroscience
3699 Miscellaneous Biology
3700 Mathematics

3701 Applied Mathematics

3702 Statistics and Decision Science

3801 Military Technologies

4002 Nutrition Sciences

4003 Neuroscience

4005 Mathematics and Computer Science

4006 Cognitive Science and Biopsychology

5000 Physical Sciences

5001 Astronomy and Astrophysics

5002 Atmospheric Sciences and Meteorology

5003 Chemistry

5004 Geology and Earth Science

5005 Geosciences

5006 Oceanography

5007 Physics

5008 Materials Science

5098 Multi-disciplinary or General Science

5102 Nuclear, Industrial Radiology and Biological Technologies

5901 Transportation Sciences and Technologies

6106 Health and Medical Preparatory Programs

6108 Pharmacy, Pharmaceutical Sciences and Administration

6202 Actuarial Science

6212 Management Information Systems and Statistics

Table A2: List of STEM occupations based on Occupation 2010 codes in IPUMS

Occ2010	STEM occupation description
1000	Computer Scientists and Systems Analysts
1020	Software Developers, Applications and Systems Software
1200	Actuaries
1240	Mathematical science occupations, nec
1320	Aerospace engineer
1350	Chemical engineers
1360	Civil engineers
1410	Electrical and Electronics Engineers
1430	Industrial engineers, including health and safety
1450	Materials engineers
1460	Mechanical engineers
1520	Petroleum, mining and geological engineers, including mining safety engineers
1530	Engineers, nec
1600	Agricultural and Food Scientists
1610	Biological Scientists
1640	Conservation Scientists and Foresters
1650	Medical Scientists, and Life Scientists
1700	Astronomers and Physicists
1710	Atmospheric and Space Scientists
1720	Chemists and Materials Scientists
1740	Environmental Scientists and Geoscientists
1760	Physical Scientists, nec

Table A3: Interactions for age of arrival and broad race groups on attainment of STEM major

Variables	STEM Major
(0-5 years)*Hispanic	-0.013***
	(0.004)
(6-11 years)*Hispanic	0.004
	(0.005)
(12-17 years)*Hispanic	0.042***
	(0.005)
(0-5 years)*Asian	0.144***
	(0.004)
(6-11 years)*Asian	0.184***
	(0.004)
(12-17 years)*Asian	0.241***
	(0.004)
(0-5 years)*White	0.059***
	(0.004)
(0-5 years)*White	0.070***
	(0.005)
(0-5 years)*White	0.148***
	(0.005)
(0-5 years)*Black	0.031***
	(0.011)
(0-5 years)*Black	0.019**
	(0.008)
(0-5 years)*Black	0.077***
	(0.007)
(0-5 years)*Other Races	0.067***
	(0.014)
(0-5 years)*Other Races	0.095***
	(0.017)
(0-5 years)*Other Races	0.177***
	(0.017)
N	2,918,377

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, conditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table A4: Interactions for age of arrival and detailed Asian immigrant groups on attainment of STEM major

Variables	STEM Major
(0-5 years)*Indian	0.245***
	(0.010)
(6-11 years)*Indian	0.235***
	(0.011)
(12-17 years)*Indian	0.300***
	(0.009)
(0-5 years)*Chinese	0.180***
	(0.008)
(6-11 years)*Chinese	0.210***
	(0.007)
(12-17 years)*Chinese	0.250***
	(0.007)
(0-5 years)*Filipino	0.063***
	(0.010)
(6-11 years)*Filipino	0.094***
	(0.010)
(12-17 years)*Filipino	0.086***
	(0.010)
(0-5 years)*Other Asian	0.130***
	(0.011)
(6-11 years)*Other Asian	0.150***
	(0.012)
(12-17 years)*Other Asian	0.227***
	(0.011)
(0-5 years)*Vietnamese	0.193***
	(0.012)
(6-11 years)*Vietnamese	0.260***
	(0.010)
(12-17 years)*Vietnamese	0.379***
	(0.010)
(0-5 years)*Korean	0.062***
	(0.008)
(6-11 years)*Korean	0.107***
	(0.010)
(12-17 years)*Korean	0.159***
	(0.010)
(0-5 years)*Japanese	0.066***
	(0.024)
(6-11 years)*Japanese	0.155***
	(0.034)
(12-17 years)*Japanese	0.063**
	(0.029)
N	2,827,494

Notes:. Dependent variable is an indicator for completing a STEM major, conditional on having a bachelor's degree and more. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table A5: Effects of age of arrival on attainment of STEM major among broad race immigrant groups unconditional on education

Age of Arrival	(1) Hispanic	(2) Asian	(3) White	(4) Black	(5) Other Races
0-5 years	-0.043***	0.127***	0.043***	0.010**	0.037***
	(0.001)	(0.003)	(0.002)	(0.004)	(0.007)
6-11 years	-0.046***	0.136***	0.047***	0.000	0.032***
	(0.001)	(0.003)	(0.002)	(0.003)	(0.007)
12-17 years	-0.063***	0.123***	0.068***	0.011***	0.049***
	(0.000)	(0.002)	(0.002)	(0.003)	(0.006)
N	8,319,602	8,145,377	8,108,989	8,051,148	8,025,191

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, unconditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. * p<0.1; ** p<0.05; *** p<0.01.

Table A6: Effects of age of arrival on attainment of STEM major among detailed Asian immigrant groups unconditional on education

Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.261***	0.208***	0.052***	0.050***	0.147***	0.092***	0.097***
	(0.008)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)	(0.016)
6-11 years	0.218***	0.206***	0.043***	0.055***	0.165***	0.137***	0.152***
	(0.008)	(0.006)	(0.005)	(0.005)	(0.006)	(0.007)	(0.023)
12-17 years	0.216***	0.176***	0.010***	0.062***	0.140***	0.136***	0.081***
	(0.006)	(0.004)	(0.003)	(0.004)	(0.005)	(0.006)	(0.017)
N	8,031,251	8,044,259	8,036,455	8,036,479	8,035,150	8,031,556	8,016,117

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, unconditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table A7: Effects of age of arrival on attainment of STEM major by sub fields unconditional on education

Panel A: Computer Science							
Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.039***	0.038***	0.012***	0.016***	0.025***	0.004**	0.008*
	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.005)
6-11 years	0.042***	0.046***	0.015***	0.017***	0.029***	0.014***	0.013*
	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.007)
12-17 years	0.058***	0.046***	0.010***	0.016***	0.038***	0.021***	0.014
	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.009)
N	8,031,251	8,044,259	8,036,455	8,036,479	8,035,150	8,031,556	8,016,117

Panel B: Engineering							
Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.073***	0.062***	0.016***	0.007***	0.023***	0.019***	0.039***
	(0.005)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)	(0.010)
6-11 years	0.061***	0.072***	0.013***	0.015***	0.053***	0.032***	0.070***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.017)
12-17 years	0.080***	0.074***	0.006***	0.031***	0.069***	0.049***	0.036***
	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.011)
N	8,031,251	8,044,259	8,036,455	8,036,479	8,035,150	8,031,556	8,016,117

Notes: The sample includes individuals between ages 22-60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, unconditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table A8: Sex-Specific Effects of age of arrival on attainment of STEM major unconditional on education

Panel A: Female Sample							
Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.232***	0.171***	0.044***	0.042***	0.135***	0.076***	0.058***
	(0.011)	(0.009)	(0.007)	(0.006)	(0.010)	(0.006)	(0.017)
6-11 years	0.176***	0.174***	0.044***	0.037***	0.138***	0.105***	0.115***
	(0.011)	(0.007)	(0.006)	(0.006)	(0.008)	(0.009)	(0.027)
12-17 years	0.164***	0.139***	0.016***	0.047***	0.110***	0.102***	0.070***
	(0.008)	(0.006)	(0.004)	(0.005)	(0.006)	(0.008)	(0.022)
N	4,052,140	4,058,795	4,054,984	4,054,866	4,053,694	4,053,369	4,044,888

Panel B: Male Sample							
Age of Arrival	(1) Indian	(2) Chinese	(3) Filipino	(4) Other Asian	(5) Vietnamese	(6) Korean	(7) Japanese
0-5 years	0.289***	0.244***	0.059***	0.059***	0.159***	0.116***	0.146***
	(0.012)	(0.010)	(0.008)	(0.008)	(0.011)	(0.009)	(0.029)
6-11 years	0.258***	0.236***	0.043***	0.074***	0.189***	0.167***	0.204***
	(0.012)	(0.008)	(0.007)	(0.008)	(0.010)	(0.011)	(0.039)
12-17 years	0.264***	0.211***	0.004	0.076***	0.165***	0.171***	0.094***
	(0.010)	(0.007)	(0.005)	(0.007)	(0.007)	(0.010)	(0.028)
N	3,979,111	3,985,464	3,981,471	3,981,613	3,981,456	3,978,187	3,971,229

Notes: The sample includes individuals between ages 22 and 60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, unconditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table A9: Effects of age of arrival on attainment of STEM major relative to native Asian groups unconditional on education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age of Arrival	Indian	Chinese	Filipino	Other Asian	Vietnamese	Korean	Japanese
0-5 years	-0.008	-0.010	-0.002	0.016***	0.004	-0.034***	0.008
	(0.010)	(0.007)	(0.006)	(0.005)	(0.010)	(0.007)	(0.016)
6-11 years	-0.048***	-0.013**	-0.010*	0.024***	0.015	-0.002	0.064***
	(0.010)	(0.006)	(0.005)	(0.006)	(0.009)	(0.009)	(0.023)
12-17 years	-0.052***	-0.041***	-0.043***	0.037***	-0.024***	0.002	-0.007
	(0.008)	(0.005)	(0.004)	(0.005)	(0.008)	(0.008)	(0.018)
N	30,931	60,046	46,570	53,637	29,617	26,677	19,148

Notes: The sample includes individuals between ages 22 and 60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a STEM major, unconditional on having a bachelor's degree and more. STEM major is a college major whose either first or second field is a STEM. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. For each immigrant group, their corresponding natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. *p<0.1; ** p<0.05; *** p<0.01.

Table A10: Effects of age of arrival on STEM Occupation unconditional on education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age of Arrival	Indian	Chinese	Filipino	Other Asian	Vietnamese	Korean	Japanese
0-5 years	0.057***	0.069***	0.017***	0.016***	0.040***	0.020***	0.032***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.011)
6-11 years	0.068***	0.072***	0.020***	0.017***	0.050***	0.029***	0.052***
	(0.005)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.014)
12-17 years	0.065***	0.064***	0.009***	0.020***	0.062***	0.027***	0.035***
	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.012)
N	8,031,251	8,044,259	8,036,455	8,036,479	8,035,150	8,031,556	8,016,117

Notes: The sample includes individuals employed in STEM occupations from 2009-2016 ACS. Dependent variable is an indicator for having a STEM occupation, unconditional on having a bachelor's degree and more. STEM occupations are defined on the basis of the 2010 Census Bureau occupational classification scheme. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. * p<0.1; ** p<0.05; *** p<0.01.

Table A11: Effects of age of arrival on attainment of Bachelor's degree or more

Age of Arrival	(1) Hispanic	(2) Asian	(3) White	(4) Black	(5) Other Races
0-5 years	-0.186***	0.207***	0.080***	0.002	0.053***
	(0.002)	(0.003)	(0.003)	(0.008)	(0.010)
6-11 years	-0.215***	0.171***	0.077***	-0.020***	-0.005
	(0.002)	(0.003)	(0.004)	(0.005)	(0.010)
12-17 years	-0.289***	0.080***	0.041***	-0.055***	-0.038***
	(0.001)	(0.003)	(0.003)	(0.004)	(0.009)
N	8,319,602	8,145,377	8,108,989	8,051,148	8,025,191

Notes: The sample includes individuals between ages 22 and 60 years in the 2009-2016 ACS. Dependent variable is an indicator for completing a bachelor's degree or more. Immigrant is a foreign-born person, born outside the US, and is either a non-citizen or naturalized citizen. Immigrant groups are arranged in order of having the largest weighted population. White non-Hispanic natives is the reference group. Regression include dummy variable controls for age, survey year and gender. Standard errors in parentheses are robust to heteroscedasticity. * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

VITA

Durba Chakrabarty

Candidate for the Degree of

Doctor of Philosophy

Thesis: ESSAYS IN REGIONAL AND URBAN ECONOMICS

Major Field: Economics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Economics at Oklahoma State University, Stillwater, Oklahoma in May, 2018.

Completed the requirements for the Master of Philosophy in Economics at Oklahoma State University, Stillwater, Oklahoma in 2009.

Completed the requirements for the Master of Science in Economics at Oklahoma State University, Stillwater, Oklahoma in 2006.

Completed the requirements for the Bachelor of Science in Economics at Oklahoma State University, Stillwater, Oklahoma in 2004.

Experience:

Graduate Student Instructor, Money and Banking, Oklahoma State University, 2015-2017.

Graduate Teaching Assistant, Oklahoma State University, 2013-2015, 2017, 2018.

Research Assistant, Georgia Institute of Technology, 2011-2013.

Project Fellow, University of Calcutta, India, 2008-2009.

Part-time Lecturer, India, 2007-2008.

Graduate and Professional Student Government Association (GPSGA), Representative August 2015 - December 2017.