© The Pakistan Development Review
58:2 (2019) pp. 159–176
DOI: 10.30541/v58i2pp.159-176

The Effects of Agglomeration on Socio-economic Outcomes: A District Level Panel Study of Punjab

ANNUS AZHAR and SHAHID ADIL

This paper examines the variation of agglomeration across districts over time in Punjab and analyses the effects of agglomeration on socio-economic outcomes in terms of social inclusion and efficiency of firms at the district level in Punjab. Earlier studies in this regard faced multiple problems since they used cross-sectional data. To bridge the gap, a newly constructed panel data from CMI is used. Factor Analysis technique is used to analyse social-inclusion variable, in addition to some other control variables as well. Data Envelopment Analysis (DEA) with bootstrap technique (performed in R) is used to calculate district-wise firm efficiency. The study argues that agglomeration is a logical consequence of China Pakistan Economic Corridor (CPEC) through an increase in the economic activity in various districts of the province. The results show that district agglomeration has a positive effect on the average district-wise efficiency of firms and has a positive statistically significant relation with social inclusion. Interesting implications arise from results, setting up clusters in urbanised rather than highly urbanised areas under CPEC can be a game changer for the economy of Pakistan especially Punjab since it has significant potential positive effects on the economy of Punjab.

JEL Classification: D62, I38, L52, R13
Keywords: Agglomeration, CPEC, Social Inclusion, Factor Analysis, Data Envelopment Analysis, Efficiency

1. INTRODUCTION

Punjab is the biggest province of Pakistan with the total population of more than 100 million which is about 60 percent of the total population of the country. It is administratively divided into nine divisions and 36 districts. It has a long history of being overshadowed by agriculture sector which has resulted in the neglect of industrial sector. In the past, Punjab lacked a clear vision/policy for the industrial sector. The recent negative growth rate in the agriculture sector along with the positive trend of huge foreign direct investment from China has put the spotlight on the manufacturing sector.

Annus Azhar <annusazhar@hotmail.com> is Assistant Research Fellow at Punjab Economic Research Institute (PERI) with MS Economics from Lahore University of Management Sciences (LUMS). Shahid Adil is presently working as Senior Research Economist at Punjab Economic Research Institute (PERI) and Additional Secretary Planning and Development Department, Government of Punjab.

Authors' Note: This paper was presented at The 32nd Annual General Meeting and Conference of the Pakistan Society of Development Economists.

Manufacturing is the backbone of the industrial sector and large-scale manufacturing is the most pivotal subsector in manufacturing. It is the main source of tax proceeds for the government and also contributes significantly to the provision of job opportunities to the labour force. According to Pakistan Economic Survey 2015-16, the industrial sector of Pakistan contributes 20 percent to GDP. This sector has experienced dynamic changes over time.

Over the years clusters have been developed in Punjab due to geographical, social and historical reasons. Punjab has geographically divergent industrial clusters comprising Gujranwala, Sialkot, and Gujarat. In total, there are seven industrial zones/clusters in Punjab: Faisalabad, Lahore, Gujranwala, Sheikhupura, Sialkot, Rawalpindi and Wazirabad (Figure 1). One can clearly see that development in Punjab is only limited to industrial clusters present in North East and North West of Punjab (Figure 1). This has led to the uneven economic development in the province. Labour in Punjab is not very mobile that is why no new clusters have been able to develop over the years. Moreover labour is not skilled and mobile enough that it can switch within industries. That is why we do not see inter-industry spillovers in Punjab. Many studies [Glaeser, *et al.* (1992); Rizov, *et al.* (2012); Ciccone and Hall (1996) and Burki and Khan (2013)] have been conducted to examine the impact of such agglomeration (clusters) on firm efficiency/productivity. However, none of these studies have examined the welfare aspect of these clusters.

In the manufacturing sector, large producers manufacture high-quality output because of adoption of modern methods of production and employment of both skilled and unskilled labour that lead to the income generation and reduction of poverty in areas where these large businesses operate. This supports the hypothesis that industrialisation leads to social inclusion.[1] This idea is commonly known as *the trickle-down effect*, a phenomenon that has not yet been proven in the case of Pakistan. Only one study by Chaudhry and Haroon (2015) is available in the literature which examined the effect of entry of new firms on variables as diverse as employment, education, hospitals and schooling. Under China Pakistan Economic Corridor (CPEC) Pakistan will receive multi-billion-dollar investment which will be used to build infrastructure as well as industrial estates in various districts of Punjab. Setting up industrial estates will lead to clusters or agglomeration. Since clusters/districts are diverse in terms of industry type, average firm size, legal status, and geographical location, a "one-size-fits-all" industrial policy will not be suitable. Therefore, classifying constraints to industrial growth at the district level serves two important purposes: First, it helps policy-makers to classify and rank agglomeration constraints at the district rather than industry level. Second, this more detailed assessment can contribute to tailoring a policy for districts and sectors in order to spur industrial growth and productivity. That being said, it is equally important to look at the dynamics of industry as well. It needs to be seen what the trend in agglomeration is in overall industrial sector in Punjab. Thus looking at agglomeration at district[2] and industry[3] level may provide useful insights to policy-makers.

[1]Social inclusion is both an outcome and a process of improving the terms on which people take part in society. It is central to ending extreme poverty and fostering shared prosperity (World Bank).

[2]This will be measured through Lee and Lee Index.

[3]This will be measured through Ellison-Glaeser Index.

The second phase of CPEC is critically important that will emerge as an opportunity for the domestic and foreign investors to invest in industrial parks to seek the benefits of cheap labour. The designing of industrial parks by the provincial governments is in its initial phase. In this perspective, this study highlights the importance of social inclusion of labour force in the production process to seek the full benefits of CPEC. How clustering of business activities, that is, the development of industrial zones which is widely recognised as the agglomeration, will affect the social inclusion is an important point of concern among the civil society, academics and applied researchers but yet needed to explore through a strong micro-founded evidence. In this perspective, this study takes a lead over the existing literature. Further to it, under the assumption of the slow pace of change in social variables, the survey data of CMI and MICS are pooled.

CPEC especially special economic zones are in their infancy stage. However, researchers have used available datasets to draw an important conclusion regarding CPEC. For example, Chaudhry, *et al.* (2017) analysed Pak-China Free Trade Agreement (FTA) of 2007 to draw important lessons from CPEC. Malik, *et al.* (2017) did a similar thing but also used international trade data between Pakistan and China which was taken from IMF (Direction of Trade Statistics). Under CPEC many special economic zones will be established (see Table A1 for a list of proposed special economic zones). These zones will cluster business activities in pockets of geographical areas thus leading to agglomeration. We cannot presently see the impact of such proposed zones on socio-economic outcomes but we can see how much these economic zones have had the impact in the past. We will use historical data to draw an inference regarding the impact of CPEC. Our study finds out that agglomeration leads to social inclusion as seen by positive sign of agglomeration coefficient in Table 5. Since we have used agglomeration as a proxy for development of special economic zones we can safely say that CPEC will, in fact, lead to social-inclusion in future.

Punjab could ensure balanced development by developing different clusters located all over the districts. A smaller investment could be sufficient for establishing an assembling unit in a cluster where all backward as well as forward linkage industries are available. Therefore, cluster development could be a powerful tool for the inclusive and sustainable growth of Punjab as well as Pakistan. Provincial and federal governments can play a role in cluster development. Cluster initiatives alone are less effective if they are not part of an overarching approach to improve competitiveness on the national and/or regional level. There is a need to focus on cross-cluster issues that affect the whole economy. A sound macroeconomic, political, legal, and social context creates the potential for competitiveness but is not sufficient. Competitiveness ultimately depends on improving the microeconomic capability of the economy and the sophistication of local companies and local competition. The government may follow an approach to cluster development aimed at addressing the main causes of cluster stagnation and help unleash their growth potential. Hundreds of enterprises share few common problems in a cluster and it is worthwhile to solve a problem for hundred enterprises than that of a smaller group or few scattered entities. This type of agglomeration policy will be more inclusive and lead to better socio-economic outcomes for society as a whole.

In urban economics, and more recently in the international economics literature, agglomeration has been considered as a principal determinant of new investment

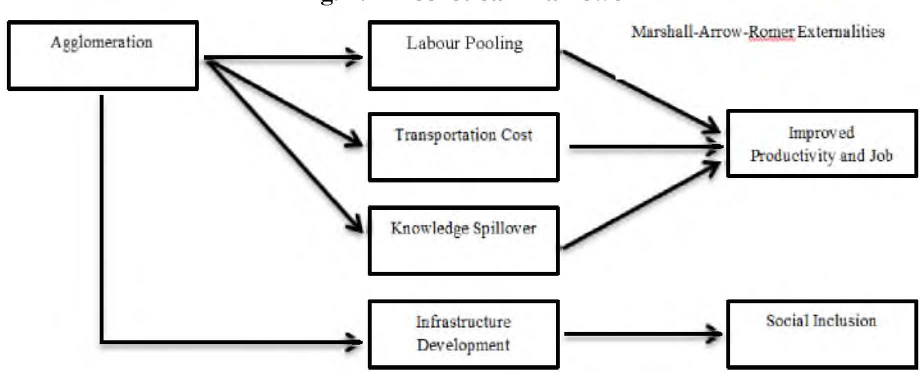
[Guimaraes, *et al.* (2000)]. Bronzini (2004) found strong evidence that specialised geographical areas attract FDI. This paper addresses the important question of how agglomeration economies affect socio-economic variables. It also studies the impact of agglomeration on average firm efficiency at the district level. The results provide evidence to support the hypothesis that agglomeration leads to social inclusion or that growth of the industrial sector has *trickle-down* effect by creating jobs and promoting income for the poor (Figure 2) and that if infrastructure is also provided with cluster development then binding social and economic constraints will also be removed.

Fig. 1. Industrial Zones and Their Major Industries

Source: Authors Illustration using GIS.
Note: Highlighted districts show major industrial centres of Punjab.

Districts	Specialisation
Rawalpindi	Food, Garment, Textile
Sialkot	Leather and Leather Products, Garment, Machinery and Sports
Gujranwala and Wazirabad	Textile, Machinery and Equipment and Electronics
Faisalabad	Textiles, Garments, Machinery and Equipment
Shiekupura	Textile, Food, and Machinery and Equipment
Lahore	Food, Garments, Textiles

Fig. 2. Theoretical Framework



Source: Authors' preparation.

Figure 2 shows the link between firm agglomeration and income/poverty. Agglomeration leads to positive externalities like labour pooling, reduction in transport cost, information and knowledge spillover etc. This leads to a rise in productivity of firms and more job opportunities in the area which in turn lead to rising income and reduction in poverty. In Solow's Model, productivity is a key link between the performance of firms, economic growth and improving the welfare of people. This productivity can only be gained if private sector takes charge of economy and government sets up industrial estates to help the private sector.

The main objective of the study is to examine the effects of agglomeration on socioeconomic outcomes at the district level in Punjab. The specific objectives of the study are to: examine whether agglomeration leads to social inclusion; find out the determinants of social inclusion; analyse the link between agglomeration and average district efficiency; find out socio-economic benefits of CPEC; and provide policy guidelines for government on how to improve efficiency and social inclusion. The rest of the paper is organised as follows. Section 2 presents literature review. Section 3 and 4 discusses data and econometric specification respectively. Empirical results are presented in Section 5 which is followed by conclusion and policy recommendations in Section 6.

2. REVIEW OF LITERATURE

There is wide literature on the benefits of urban economics in terms of growth of cities through expansion of industries. Cities grow initially because of geography, history and then by their industrial structures based on the extent of specialisation or diversity of business. With industrial growth, firms get benefit from other businesses or overall level of economic activity around them e.g., accessibility of infrastructure, access to financial establishments and publishing and marketing. These externalities are known as Jacob externalities which echo the diversity in the area which in the case of present study is a district. Localisation economies exist when firm gains value from within the industry or firms which are involved in matching activity. Firms benefit from knowledge spillover due to the collaboration of agents, availability of particular labour, availability of non-tradable intermediate goods and low transportation cost due to access to a market. These externalities are also known as Marshall-Arrow-Romer (MAR) externalities in dynamic form. Many benefits arise due to both of these agglomeration economies. Thus the location of a firm may depend on closeness to target market to reduce transportation cost or because the nature of the product is perishable and thus requires speedy delivery [Marshall (1890); Myrdal (1957) and LaFountain (2005)]. However, some firms may be constrained to locate near the source of raw material [Hirschman (1958)].

Firms locating closer to each other may have significant potential benefits at different levels of economic activity. Hazledine, *et al.* (2013) summarised that the benefits of agglomeration can occur at four different levels: (i) Internal to individuals/households—individuals gain from wider job opportunities and better amenity; (ii) Internal to firms—firms gain from larger labour markets, and from economies of scale generated by access to effectively larger accessible output markets; (iii) Internal to industries—technological (knowledge) spillovers; a better choice of intermediate inputs; larger skilled labour pool; (iv) Internal to the city—scale of local markets and more efficient provision of infrastructure, public administration, and amenities.

Additionally, agglomeration also has direct benefits as well. Giang, *et al.* (2015) found a linkage between agglomeration and poverty reduction in the case of Vietnam. This effect was greater for houses with male younger and more educated household heads. Firms can improve household welfare and reduce poverty by having a positive effect on employment and wages. Chaudhry and Haroon (2015) observed that in case of the manufacturing sector of Pakistan, firm entry has a significant impact on socio-economic outcomes and that these outcomes normally materialise with a lag. They recommended that policy-makers should recognise that different type of firms have a different type of impact which warrants the need for a customised approach to industrial development. Thus agglomeration can lead to social uplift of people. Confirming these findings, Quintana and Royuela (2014) showed that agglomeration processes can be associated with economic growth, at least in countries at early stages of development.

Apart from affecting the community, agglomeration contributes positively to firm-level variables as well. Albert and Maudos (2002) found that investment in the physical capital also positively relates to business efficiency. Beeson and Husted (1989) in a cross-state study for the US observed that a substantial part of the difference of efficiency can be credited to regional dissimilarities of the labour force features, the intensity of urbanisation and industrial structure. The New Economic Geography literature points out that transport cost explains agglomeration. [Fujita, *et al.* (2001)].

Agglomeration if unchecked may lead to diseconomies as well. According to [Lall, *et al.* (2004)] agglomeration may be associated with negative consequences as well. Krugman (1991a) argues that when transport cost of a region decreases then it begins to invite industries towards it hence increasing the concentration of industry and eventually increasing the population of the region. Fujita and Thisse (2002) found that when the concentration of industry in a specific area crosses a certain level it begins to raise the cost of functioning in that area due to greater labour wages, greater land prices and rent, overpopulation, congestion cost, higher transportation cost and communication costs. According to Kim (2008), while negative spillovers result from an increased cluster of industry, it will eventually raise the cost of production and it is known as "Thin Market Effect" by Cohen and Paul (2005). Rising costs due to agglomeration shrink additional concentration of industry in the nearby areas and disperse economic activities in the region [Fujita and Thisse (1999); Kim (2008)]. The equilibrium between two positive and negative forces—centripetal and centrifugal—leads to stability. For example, Mitra (1999) studied the connection between agglomeration economies and technical efficiency of electrical machinery and cotton textile sector through firm-level data. The outcomes indicate that agglomeration raises the efficiency of firms but the effect starts to diminish for cities which are very bigger in size.

3. DATA AND METHODOLOGY

In the previous studies, due to data constraints, industry level firm efficiency was measured using cross-sectional data. Lall, *et al.* (2004) in his study on agglomeration in India mentioned similar data constraints. To understand the true impact of the independent variable on the dependent variable we have to follow the same units over time. Lall, *et al.* (2004) thus mentioned that ideally for work on agglomeration panel data should be used.

For this study, we use panel data constructed from CMI (2001, 2005 and 2010). Since the data has same *i's* for each *t*. We merge the district level panel data with that of MICS (2003, 2007 and 2011). Here we assume that social level variables change slowly over time, thus allowing us to merge two different datasets that were collected at most two years apart. This allowed us to merge two unique datasets at the district level. However, this may be seen as a limitation in data collection by government agencies where different surveys are done irrespective of timings of each other. It is expected that unobserved effects might be correlated with the independent variables. If this is indeed the case, pooled OLS will lead to biased results. Hausman test was run to check if Fixed Effects (FE) or Random Effects (RE) technique is appropriate.[4] Result yielded p value of 0.1194; thus we failed to reject the null hypothesis that both FE and RE are consistent.[5] We believe that heterogeneity of districts is an important issue and thus requires controlling for it in regression; hence FE model is used. The significance of the model can be judged from F test whose p value is 0.073, which means that the estimated model is significant at 10 percent significance level.

Table 1

Variable and Their Data Sources

Variables	Methodology	Data Sources
District Agglomeration	Lee and Lee Agglomeration Index	CMI (2001, 2005, 2010)
Sectoral Agglomeration	Ellison-Glaeser	CMI (2001, 2005, 2010)
Efficiency	DEA Bootstrap	Calculated in R-Software
Social Inclusion	Factor Analysis/ Principal Component Analysis	MICS, Punjab Development Statistics
Road Density	Ratio Road Length to Total Area of District	Punjab Development Statistics
Education Index	Factor Analysis	Punjab Development Statistics
Investment	Taken as reported in the source	Directory of Industries
Employment Cost	Taken as reported in the source	Directory of Industries
Number of Factories	Taken as reported in the source	Directory of Industries

Note: Panel data and variables used in efficiency model mentioned in Appendix were constructed by Ahmad (2016). Efficiency was calculated in R using DEA Bootstrap technique.

Independent Variable of interest is Agglomeration which has been measured in literature in a variety of ways [Chaudhry and Haroon (2015); Ahmad (2016)]. For example, some studies used the number of firms in a geographic area while others used location Gini Coefficient as a formalisation of agglomeration [Aiginger, *et al.* (1999)]. The latter measure had the benefit of providing for the concentration of industry but it did not show how firms are distributed among regions [Capello, *et al.* (2010)]. This led to the development of a measure for regional specialisation and was popularised by Lee and Lee who defined this index as the share of industry i's employment relative to total industry employment in a specific region j by contrast to the share of region j's employment relative to total (provincial in our case) employment in industry. This Lee and Lee index has been used in international literature for the case of India by Lal, *et al.*

[4]For details on the need to use FE or RE please see Wooldridge (2009) and Gujarati (2003).

[5]Consistency property also holds in our case since degree of freedom is greater than thirty.

(2004) and it has been used in the case of Pakistan by Burki and Khan (2013). To measure agglomeration albeit, in a different context, we, however, apply it for the first time in the context of Punjab with special district level focus. The formula for Lee and Lee Agglomeration (Diversity Index) is:

$$g_i^s = \sum_{n=1}^{N} \left[\frac{E_{ij}}{E_i} - \frac{E_j}{E}\right]^2$$

The above formula shows the agglomeration index used in this paper where *i* signifies district and *j* signifies industry, g_i^s represents the extent of localisation and urbanisation in the *i*th district, E_{ij} is employment in *the i*th district in *the j*th industry, *Ei* is employment in the *i*th district, E_j is employment in industry *j,* and *E* signifies total manufacturing sector employment. A lesser value of the index signifies high diversity which means urbanisation economies are stronger while a higher value represents that firms are specialising which indicates localisation economies are stronger. The index varies from 0 to 2 with zero meaning zero specialisation (high diversity) and two representing complete specialisation (zero diversity). In order to measure the extent and effects of localisation economies and urbanisation economies on technical efficiency, the diversity index has been used [as proposed by Henderson, *et al.* (2001)]. The index is calculated at the district level where district boundaries are frozen at 2000-01 level. The 29 districts that existed in Punjab at that time are used for the index.

In order to measure the agglomeration of industries, the Ellison Glaeser Index proposed by Ellison and Glaeser (1997) is adopted. A value of zero for this index means no agglomeration. We have computed this index at 3 digit level under Pakistan Standard Industrial Classification (PSIC) and industry codes are fixed at 2000-01 level to ensure uniformity across industry classification codes.

The formula for Ellison Glaeser is:

$$\emptyset_i = \frac{\sum_{j=1}^{n}(s_{ji} - x_j)^2 - (1-\sum_j x_j^2)H_i}{[1-H_i]\left[1-\sum_j x_j^2\right]}$$

Were s_{ji} is share of industry *i's* employment which is located in district *j,* x_j is share of industry's employment in district *j* as compared to the overall manufacturing sector employment. $\sum_{j=1}^{n}(s_{ji} - x_j)$^2 is referred to as Gini coefficient which shows raw geographical concentration of industry *i*. H_i is Herfindahl-index which measures plant share of employment in industry *i's* overall employment. The scores are allocated to each industry in Ellison Glaeser index. Average result of this index is presented in empirical results part of the paper.

Social Inclusion is a multi-dimensional concept. Its general definition is: *"Social inclusion is central to ending extreme poverty and fostering shared prosperity. It is both an outcome and a process of improving the terms on which people take part in society."* According to the World Bank, variables to include in this concept may be altered according to country specificity. In Pakistan researchers like Cheema, *et al.* (2008) have used health-related indicators as in its social variable formulation. These health level variables have cross-cutting importance since they also incorporate education/literacy levels as well. Eide and Showalter (2011) stated that there are numerous health benefits associated with education. For example, education can play a positive role in the ability to manage health care [Kaplan, *et al.* (2015)]. Since education affects health, it leads to social inclusion. Thus we have taken only health variables in the definition of social inclusion. Additionally, we have education as an independent variable in our main regression so our model incorporates the effects of education as well.

Dependent Variable (Social Inclusion) in this paper is constructed by conducting Factor Analysis using four variables taken from MICS[6] (2003, 2007, and 2011) namely: infant mortality rate[7] (IMR); antenatal care;[8] improved water sources[9] and improved Sanitation.[10] The data from MICS and CMI have been merged district wise by assuming that social variables change slowly over time.

The Principal Component Analysis (PCA) and Factor Analysis are used to transforms a number of (probably) correlated variables into a (lesser) number of uncorrelated variables called *Principal Components*. The first principal component accounts for maximum variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The goal of the principal components analysis is to explain the maximum amount of variance with the fewest number of principal components. Factor Analysis is also used which is similar to PCA technique. The principal component with the smallest eigenvalue contributes the least variance and so is least informative and is thus discarded.

In order to control for infrastructure, we have taken road density as a suitable proxy. Employment cost and investment have been taken to see if firm efficiency is sensitive to cost and investment changes respectively. The results are robust as seen by minor changes in coefficients even if we add/drop few variables. Factors for education are calculated using variables such as primary enrollment of boys and girls, high school enrollment of boys and girls, staff number, enrollment in poly-technology institute etc. Factors for crime include murder, attempted murder, kidnapping, and burglary.

Table 2

Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
Lee and Lee Index	.2458	.2031	.0204	.9621
Ellison-Glaeser	0.142	0.200	–0.340	1.112
District Efficiency Score	.8508	.0533	.7424	.9571
Road Density	.3614	.2062	.0562	1.0291
Investment	21.9	33.3	0.389182	157
Employment Cost	0.014049.5	0.033365.15	0.000003	0.214907
No. of Factories	120	214	2	1170

Notes: Investment and employment costs are in millions.

[6]MICS collects important information on socio-economic variables.

[7]Probability of dying between birth and the first birthday.

[8]Skilled person providing antenatal care to women aged 15-49 who gave birth during preceding two years in Punjab.

[9]Percent distribution of household population according to the main source of drinking water and percentage of household population using improved drinking water sources.

[10]Percent distribution of the household population according to the type of toilet facility used by the household, and the percentage of the household population using sanitary means of excreta disposal.

Summary Statistics (see Table A2) show that the district with most agglomeration are Layah, Rajanpur, Mianwali, and Rahim Yar Khan whereas districts with most diversity include Lahore, Khanewal, Multan, Kasur, Attock and Shiekupura. Additionally, districts with highest average efficiency of firms are Sargodha, Jhangh, Kasur, Rahim Yar Khan, Sheikhupura and Faisalabad. Contrary to this, districts with lowest average efficiency[11] include Rawalpindi, Lahore, Sahiwal and Gujranwala.

Due to data constraints, the efficiency model could not be estimated with full robustness due to degrees of freedom problem. This issue was expected since we have used district-level data. The aforementioned problem could have been avoided had the regression was run at firm level but that would not have added anything substantial to the already dense literature on agglomeration. Perhaps future studies could address this degree of freedom limitation.

4. ECONOMETRIC SPECIFICATION

Following Buki and Khan (2011), we have estimated the following equation:

$$Y_{it} = B_1X_{it} + B_2X_{2it} + B_3X_{3it} + \alpha_1 + \alpha_2 + \dots \alpha_n \epsilon_{it} \quad \dots \quad \dots \quad \dots \quad (1)$$

The use of above equation (Fixed Effects) addresses the problem of endogeneity by ensuring that the assumption of $Cov\ (X, U) = 0$ is not violated. Table 5 describes the data sources for different variables. The dependent variable is social inclusion and the independent variable of interest is agglomeration. Other control variables are road density, total education, crime factors, number of reporting factories and employment cost.

It is necessary to check whether data is normally distributed or not. Therefore, we use Cameron and Trivedi's (1990) decomposition of IM-test in Stata. Overall there is no skewness[12] or kurtosis[13] in the data. Individually the variables for social inclusion, agglomeration, road density and crime factors follow a normal distribution. Ramsey (1969) *reset* test checks for misspecification in a model and also omitted variable bias. P value is 0.4118 which is greater than 0.05. Therefore we failed to reject the null hypothesis. This implies the model is correctly specified and that it has no omitted variable bias.

Group-wise heteroscedasticity is checked by running Modified Wald test in the Fixed Effect regression model using Stata. The same results were obtained in terms of Breusch-Pagan (1979) and Cook-Weisberg (1983) tests for heteroscedasticity.[14] Therefore the null hypothesis is rejected which implies that heteroscedasticity exits. To counter this problem, we have used heteroscedastic robust standard errors.

Multicollinearity diagnostic criteria are given below:

[11]Larger districts e.g. Lahore etc. may have low average efficiency due to huge variation in the operations of firms.

[12]P value was 0.9908.

[13]P value was 0.1994.

[14]P value was 0.000.

Table 4

Multicollinearity Diagnostic Criteria

Variables	Eigenvalues	VIF	1/VIF
Agglomeration	2.0155	1.0847	0.9219
Road Density	0.8975	1.4904	0.6710
Crime	0.5899	1.3554	.7378
Total Education	0.4971	1.3867	0.7211

Source: Authors own Calculations.

The variance inflation factor (VIF) is most commonly used criteria to identify the problem of multicollinearity in regression analysis. According to Gujarati (2003), if VIF is above 10, then a severe problem of multicollinearity exists among the predictors. However, VIF calculated shows no issue of multicollinearity as all the values for VIF are lower than 10. If the Eigenvalues are close to zero then the chances are that multicollinearity exists, but none of the Eigenvalues is zero, so there is no issue of multicollinearity. The 1/VIF is called the tolerance test and if its value is less than 0.10 than there is multicollinearity but none of the explanatory variables has tolerance value less than 0.10 [Gujarati (2003)]. Since no multicollinearity exits, therefore, it shows that t values are robust.

5. EMPIRICAL RESULTS

The Agglomeration index indicates the specialisation/diversity. If the value of agglomeration index increases, then it means that specialisation is increasing and if its value falls then it means that diversity is increasing. Social Inclusion is the dependent variable of the the reported regressions given in Table 5. The slope parameter of agglomeration index is statistically significant at 1 percent level of significance. Thus benefits of industrial development in Punjab are being enjoyed by lower segment of the population as well. These positive effects of specialisation rather than diversity are supported by many empirical findings [Henderson, *et al.* (2001); Ciccone and Hall (1996) and Henderson (1990)].

As for the Ellison and Glaeser index we have estimated a score of greater than 0.05 for this index which indicates that industries in Punjab are very agglomerated. If the score is in between 0.02 and 0.05 it shows that the industry is reasonably agglomerated and a score of less than 0.02 shows very weak agglomeration. As shown in Table 2, the mean value of this index is 0.142 which according to aforementioned range shows that on average industries of Punjab are highly agglomerated.

Road density, agglomeration and total education have a positive relationship with social inclusion whereas crime has a negative association with social inclusion. All the signs are as expected. The main variable of interest is statistically significant at 10 percent level of significance.

Table 5

FE Estimates of Agglomeration Model

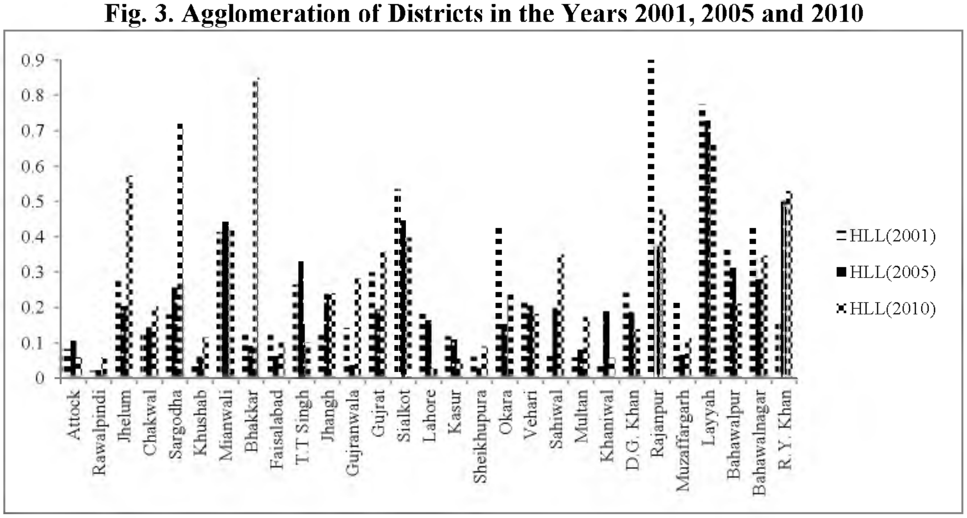
Variables	(1) Social Inclusion	(2) Social Inclusion	(3) Social Inclusion	(4) Social Inclusion
Agglomeration	0.457***	0.488**	0.494**	0.485**
	(0.165)	(0.179)	(0.180)	(0.177)
Road Density	0.260	0.542**	0.539**	0.535**
	(0.205)	(0.211)	(0.214)	(0.216)
Total Education	0.0806	0.101	0.109	0.117
	(0.157)	(0.144)	(0.143)	(0.140)
Crime Factors	–0.101	–0.0706	–0.0666	–0.0823
	(0.164)	(0.155)	(0.155)	(0.155)
No of Reporting Factories			0.000109	0.000153
			(0.000179)	(0.000187)
Employment Cost				8.18e-07
				(9.84e-07)
Constant	–0.206***	–0.250***	–0.260***	–0.261***
	(0.0703)	(0.0667)	(0.0677)	(0.0677)
District Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	No	Yes	Yes	Yes
Observations	87	87	87	87
R-squared	0.097	0.123	0.125	0.131

Source: Authors' own Calculations.

Robust Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 (Std. Err. adjusted for 29 clusters in districts).

Correlation between district efficiency and agglomeration is positive. There are only 42 observations and if fixed effects are used this number falls to 28. With *n* less than 30 the OLS assumptions of normality will be violated. There is a vast literature that supports the hypothesis that agglomeration increases the efficiency of firms. Thus sign and significance may be checked without going into the details of robustness of results (see Table A3).

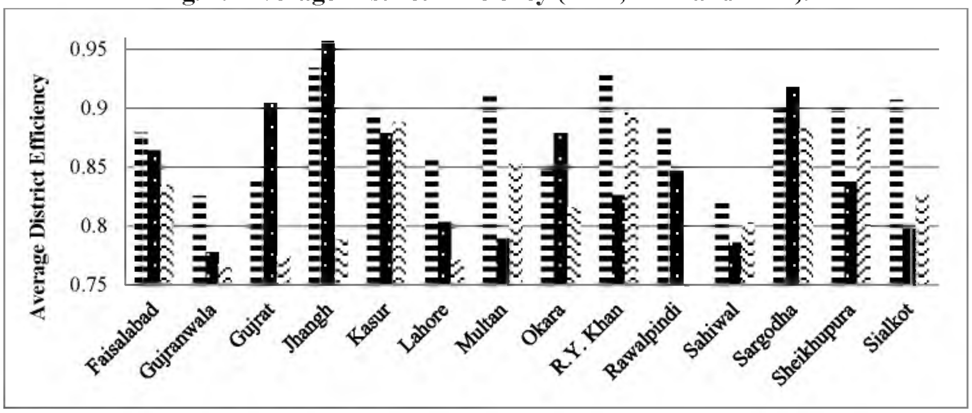
Fig. 3. Agglomeration of Districts in the Years 2001, 2005 and 2010



Source: Authors' Own rendering using Panel CMI data.

Agglomeration of each of the district in the sample is illustrated above. The x-axis shows time period (2001, 2005 and 2010) whereas y-axis shows agglomeration level. Districts show a considerable change in the level of agglomeration. A mixed trend of change in agglomeration levels is observed: there is a rise in agglomeration in Bhakkar, Sargodha and Jehlum whereas Rajanpur and Layyah show a fall in agglomeration level. On the other hand, Rajanpur, Layah, Sialkot, and Okara show the most level of agglomeration in the year 2001 whereas Layyah, Rahim Yar Khan, Sialkot, and Mianwali show the most level of agglomeration in the year 2005 and in the year 2010 the most level of agglomeration is shown by Layyah, Bhakkar, Sargodha, Jehlum.

Fig. 4. Average District Efficiency (2001, 2005 and 2010).



Source: Authors' Own rendering using Panel CMI data.

There has been a consistent fall in efficiency in Faisalabad, Gujranwala, and Lahore from 2001 till 2010 whereas no district has shown a consistent rise in average firm efficiency over the same period. For the most districts like Gujrat, Jhangh, Okara, and Sargodha, there has been a rise in average efficiency from 2001 till 2005 but for the next half-decade, we see a falling trend in efficiency level. The inputs used to calculate the efficiency are capital, labour, and materials and energy, whereas output is taken as value added by firms. The fall in efficiency from 2001 till 2005 can be attributed to the fact that due to opening up of Pakistan's economy it faced fierce competition from international firms. However, the fall in efficiency following this period was due to lack of business-friendly policies of the government.

As stated earlier, results for regression of district efficiency on agglomeration are not robust due to the degree of freedom problem. This problem arises because our regressions are run on the district basis and not on the basis of individual firm. Therefore, this paper utilises the trend of efficiency (Figure 4) over the years (2001-2010). This ensures robustness of results as well since same firms are followed over time to measure efficiency.

6. CONCLUSION AND POLICY IMPLICATIONS

This paper investigated the district level agglomeration economies in the manufacturing sector of Punjab. The DEA bootstrap analysis which incorporated technical efficiency model was applied. Plant level panel data constructed from CMI dataset for the years 2000-01, 2005-06 and 2010-11 were used. The Agglomeration Index (Diversity index) was then calculated which measured local scale externalities at the district level while the mean agglomeration level of industries was also calculated. This study found that social inclusion and firm efficiency is positively related to agglomeration in districts.

The results indicate that industries in Punjab are agglomerated[15] thus showing intra industry spillovers in Punjab and that this agglomeration is positively associated with efficiency of firms. Agglomeration at district level is also positively associated with social inclusion in districts. Thus both firms and districts in Punjab are benefitting from positive externalities of agglomeration economies. Further, the results show that better infrastructure in districts also allows for more social inclusion. This means that government may focus on the provision of better infrastructure facilities such as better road network which will lead to greater connectivity and better social inclusion in districts of Punjab.

This study uses past information to draw inference about the potential future positive consequences of CPEC. The results show that agglomeration which will be a natural consequence of industrial development as a consequence of CPEC will yield social inclusion.

[15]As shown by high mean value of Ellison-Glaeser Index in Table 2.

APPENDIX A

Table A1

CPEC Special Economic Zones (SEZs)

Serial Number	Project Name
1.	Rashakai Economic Zone on M-1
2.	Special Economic Zone Dhabeji
3.	Bostan Industrial Zone
4.	Punjab - China Economic Zone, M-2 District Sheikhupura
5.	ICT Model Industrial Zone, Islamabad
6.	Development of Industrial Park on Pakistan Steel Mills Land at Port Qasim near Karachi

Source: Official CPEC website. Government of Pakistan.

Table A2

Summary Statistic (Mean) from 2001-2010

Districts	Agglomeration	Road Density	Social Inclusion
Attock	0.0816	0.233	–0.125
Rawalpindi	0.0335	0.537	1.098
Jhelum	0.351	0.291	0.651
Chakwal	0.156	0.292	–0.224
Sargodha	0.385	0.399	0.655
Khushab	0.0733	0.228	0.576
Mianwali	0.429	0.225	0.659
Bhakkar	0.357	0.197	–0.975
Faisalabad	0.0944	0.508	1.827
T.T Singh	0.232	0.456	0.819
Jhangh	0.201	0.333	–1.43
Gujranwala	0.154	0.598	0.577
Gujrat	0.283	0.49	1.031
Sialkot	0.459	0.579	0.733
Lahore	0.124	0.629	2.171
Kasur	0.0932	0.366	0.261
Sheikhupura	0.0584	0.35	0.946
Okara	0.272	0.501	–0.662
Vehari	0.199	0.481	–0.101
Sahiwal	0.203	0.645	–0.183
Multan	0.103	0.565	–0.46
Khaniwal	0.0974	0.401	–0.452
D.G. Khan	0.189	0.113	–1.776
Rajanpur	0.603	0.0965	–2.337
Muzaffargarh	0.131	0.236	–1.127
Layyah	0.725	0.175	–0.693
Bahawalpur	0.296	0.082	0.0157
Bahawalnagar	0.35	0.23	–0.583
R.Y. Khan	0.396	0.244	–0.893

Source: Authors' own Calculation.

Table A3

Regression of District Efficiency on Agglomeration

Variables	District Efficiency
Employment Cost	–1.47e-06***
	(4.90e-07)
No. of Reporting Factories	–0.000325***
	(0.000104)
No. of Reporting Factories Squared	1.80e-07*
	(9.09e-08)
Investment	7.06e-07***
	(2.51e-07)
Total Education	0.000530
	(0.0101)
Constant	0.881***
	(0.0138)
Observations	42
R-squared	0.383

Source: Authors' own Calculation.
Note: Standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1.

Agglomeration (number of reporting factories), investment, and the number of reporting factories squared have a positive relationship with average district efficiency of firms whereas employment cost and the number of reporting factories in level form have a negative relation with average district efficiency of firms. The number of firms is taken as a proxy for agglomeration as taken in [Barry, Gorg, and Strobl (2003)].

REFERENCES

Ahmad, Taimoor (2016) Agglomeration Economies and their Effects on Technical Inefficiency of Manufacturing Firms in Punjab, Pakistan: Panel Data Analysis from 2000-01 to 2010-11(MS Thesis). Retrieved from the Author.

Aiginger, K., M. Boeheim, K. Gugler, M. Pfaffermayer, and Y. Wolfmayr-Schnitzer (1999) Specialisation and (Geographic) Concentration of European Manufacturing, Background Paper for 'The competitiveness of European Industry: 1999 Report' of DG Enterprise.

Barry, F., H. Görg, and E. Strobl (2003) Foreign Direct Investment, Agglomerations, and Demonstration Effects: An Empirical Investigation. *Review of World Economics* 139:4, 583–600.

Beeson, P. E. and S. Husted (1989) Patterns and Determinants of Productive Efficiency in State Manufacturing. *Journal of Regional Science* 29:1, 15–28.

Barkley, D. L. and M. S. Henry (1997) Rural Industrial Development: To Cluster or not to Cluster? *Review of Agricultural Economics* 1:2, 308–325.

Braunerhjelm, P. and B. Borgman (2004) Geographical Concentration, Entrepreneurship, and Regional Growth: Evidence from Regional Data in Sweden, 1975–99. *Regional Studies* 38:8, 929–947.

Breusch, T. S. and A. R. Pagan (1979) A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica: Journal of the Econometric Society* 1287–1294.

Bronzini, R. (2004) Foreign Direct Investment and Agglomeration: Evidence from Italy, Roma, Banca d'Italia, Tema di Discussione 526.

Burki, A. A. and M. A. Khan (2013) Agglomeration Economies and Their Effects on Technical Inefficiency of Manufacturing Firms: Evidence from Pakistan.

Cameron, A. C. and P. K. Trivedi (1990) Regression-based Tests for Overdispersion in the Poisson Model. *Journal of Econometrics* 46:3, 347–364.

Capello, Roberta and Peter Nijkamp (eds.) (2010) Handbook of Regional Growth and Development Theories. Edward Elgar Publishing.

Castells-Quintana, D. and V. Royuela (2014) Agglomeration, Inequality and Economic Growth. *The Annals of Regional Science* 52:2, 343–366.

Chaudhry, A. and M. Haroon (2015) The Economic Impact of New Firms in Punjab. *The Lahore Journal of Economics* 20: SE (September 2015), 143–182.

Chaudhry, T., N. Jamil, and A. Chaudhry (2017) Pakistan's Experience with the Pakistan—China FTA: Lessons for CPEC. *Lahore Journal of Economics* 22:(Special Edition), 1–24.

Cheema, A., L. Khalid, and M. Patnam (2008) The Geography of Poverty: Evidence from the Punjab. *The Lahore Journal of Economics* (Special edition) 163–188.

Ciccone, A. (2002) Agglomeration Effects in Europe. *European Economic Review* 46:2, 213–228.

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

Cohen, J. P. and C. J. M. Paul (2005) Agglomeration Economies and Industry Location Decisions: The Impacts of Spatial and Industrial Spillovers. *Regional Science and Urban Economics* 35:3, 215–237.

Cook, R. D. and S. Weisberg (1983) Diagnostics for Heteroscedasticity in Regression. *Biometrika* 70:1, 1–10.

Eric R. Eide, Mark H. Showalter (2011) Estimating the Relation between Health and Education: What do we know and what do we Need to Know? *Economics of Education Review* 30:5, 778–791.

Fujita, M., P. R. Krugman, and A. J. Venables (2001) *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, Massachusetts: MIT Press.

Fujita, M. and J. F. Thisse (2002) *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth.* Cambridge: Cambridge University Press.

Giang, L. T., C. V. Nguyen, and T. Q. Tran (2015) A Linkage between Firm Agglomeration and Poverty Reduction First evidence in Vietnam (No. 2015–617).

Glaeser, E., H. Kallal, J. Scheinkman, and A. Schleifer (1992) Growth in Cities. *Journal of Political Economy* 100, 1126–1152.

Guimaraes, P., O. Figueiredo, and D. Woodward (2000) Agglomeration and the Location of Foreign Direct Investment in Portugal. *Journal of Urban Economics* 47:1, 115–135.

Gujarati, D. N. (2003) *Basic Econometrics*. Tata McGraw-Hill Education.

Gumbau-Albert, M. and J. Maudos (2002) The Determinants of Efficiency: The Case of the Spanish Industry. *Applied Economics* 34:15, 1941–1948.

Hamid, N. and M. Khan (2015) Pakistan: A Case of Premature Deindustrialisation? *The Lahore Journal of Economics* 20, 107.

Hazledine, T, S. Donovan, and J. Bolland (2013) The Contribution of Public Transport to Economic Productivity. (NZ Transport Agency Research Report 514). 63pp.

Helsley, R. W. and W. C. Strange (1990) Matching and Agglomeration Economies in a System of Cities. *Regional Science and Urban Economics* 20:2, 189–212.

Henderson, J. V., T. Lee, and Y. J. Lee (2001) Scale Externalities in Korea. *Journal of Urban Economics* 49, 479–504.

Hirschman, A. O. (1958) *The Strategy of Economic Development* (No. HD82 H49).

Kaplan, R. M., M. L. Spittel, and D. H. David (eds.) (2015) *Population Health: Behavioural and Social Science Insights.* Government Printing Office.

Kathuria, V. and A. S. George (2005) Spatial Location of Industries—Factors Influencing Locational Choice. Conference 2nd Annual Conference on Economic Growth and Development.

Kim, S. (2008) Spatial Inequality and Economic Development: Theories, Facts and Policies. Commission on Growth and Development, World Bank. (Working Paper No.16).

Krugman, P. (1991a) *Geography and Trade*. Cambridge: MIT Press.

LaFountain, C. (2005) Where do Firms Locate? Testing Competing Models of Agglomeration. *Journal of Urban Economics* 58:2, 338–366.

Lall, S. V., Z. Shalizi, and U. Deichmann (2004) Agglomeration Economies and Productivity in Indian Industry. *Journal of Development Economics* 73, 643–673.

Malik, A., E. Ghani, and Musleh ud Din (2017) An Assessment of Pakistan's Export Performance and the Way Forward. Pakistan Institute of Development Economics, Islamabad.

Marshall, Alfred (1920) *Principles of Economics.* London: Mac-Millan (1890).

Mitra, A. (1999) Agglomeration Economies as Manifested in Technical Efficiency at the Firm Level. *Journal of Urban Economics* 45:3, 490–500.

Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions*. London, Duckworth.

Podnieks, E. (2006) Social Inclusion: An Interplay of the Determinants of Health—New Insights into Elder Abuse. *Journal of Gerontological Social Work* 46:(3-4), 57–79.

Quintana, D. C. and V. R. Mora (2012) Agglomeration, Inequality and Economic Growth: Cross-section and Panel Data Analysis.

Ramsey, J. B. (1969) Tests for Specification Errors in Classical Least-squares Regression Analysis. *Journal of the Royal Statistical Society*, Series B, 31, 350–71.

Rizov, M., A. Oskam, and P. Walsh (2012) Is there a Limit to Agglomeration? Evidence from Productivity of Dutch Firms. *Regional Science and Urban Economics* 42:4, 595–606.

Wing, P. E. A. S. (2005) *Pakistan Economic Survey (2015-16).* Finance Division.

Wooldridge, J. M. (2009) *Econometrics*. Cengage Learning.

World Bank (2013) Retrieved from https://www.worldbank.org/en/topic/social development/brief/social-inclusion