Zagreb International Review of Economics & Business, Vol. 28, No. 1, pp. 103-124, 2025

Faculty of Economics and Business, University of Zagreb and Sciendo. Printed in Croatia.
ISSN 1331-5609; UDC: 33+65
DOI: 10.2478/zireb-2025-0005

Inter and Intra-Regional Income Inequalities Attributable to Spatial Concentration in Pakistan

Uzma Tabassum[*+]
Munazah Nazeer[**]

Abstract: *Reducing inequality is essential to sustainable growth in regions. Due to varying geographic locations and locally formed development strategies, agglomeration disparities fluctuate among urban areas. This study is uniquely designed to measure the extent to which spatial agglomeration impairs inter and intra inequalities using two distinct techniques, Geographic Information System (GIS) and Propensity Score Matching (PSM). The results obtained from both analyses are in line with the theoretical framework established in the study. The results show that income growth is significantly impacted by the geographical concentration of industries. After matching, agglomerated regions have 22.5% higher average income than less agglomerated areas, which upsurges inter-regional disparities. Additionally, as income growth is unevenly distributed among inhabitants of the same region, inequalities in a treated regions are estimated to increase by approximately 2.5% more in comparison to the untreated regions that are relatively less concentrated, exacerbating intra-regional inequality.*

Keywords: Income inequality; urban regions; Propensity Score Matching; Arc GIS

JEL Classification: C31, D30, R12

[*] Applied Economics Research Centre, University of Karachi, Karachi, Pakistan.
[+] Corresponding Author E-Mail: uzma.tabasum@aerc.edu.pk
[**] Applied Economics Research Centre, University of Karachi, Karachi, Pakistan.

Introduction

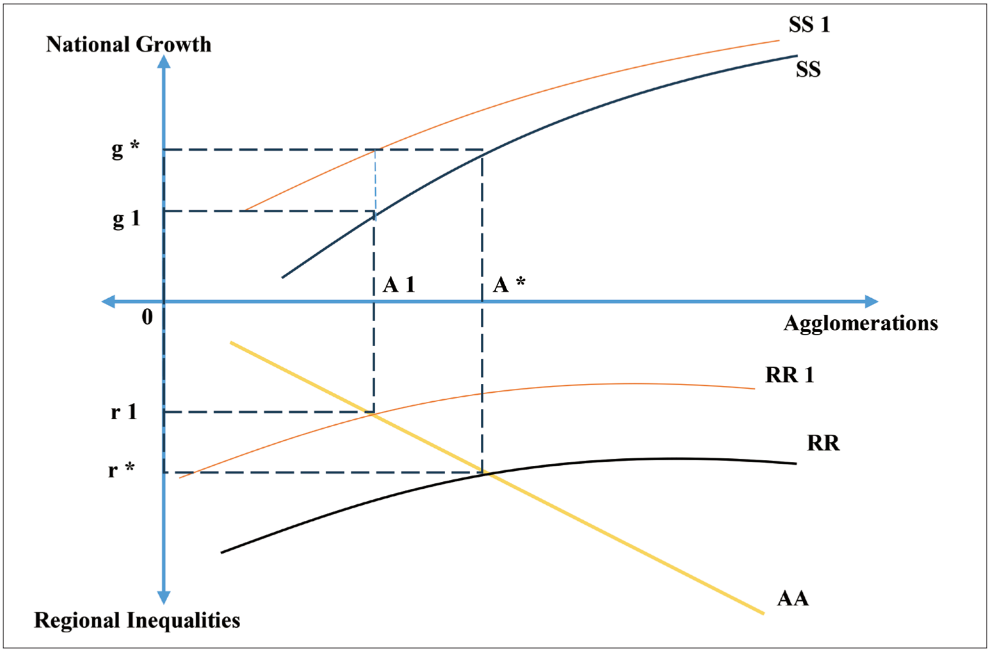
World patterns during the last several epochs have drawn attention to two distinct characteristics of economic expansion. The first is a rise in inequality, while the second is a growing spatial concentration of economic activities across areas. According to several theories, more openness of a region and the removal of barriers to free factor migration will likely result in higher levels of industry concentration and, as a result, greater specialization of the region. This could have ramifications in terms of widening wealth inequality between regions. Such expansion may also be seen in Pakistan, where income gaps are increasing as regions are becoming more concentrated and specialised (Tabassum, 2016). The basic idea is that spatial concentration amplifies inequalities across regions, as exogenous factors such as region characteristics, labour force skill composition, and agglomeration economies are attributed to productivity differences, which are then assumed to be reflected in income distribution. The factors of income inequalities have been the subject of a recent collection of empirical studies conducted both domestically and globally. Inequalities are influenced by socioeconomic, political, and open economy aspects, according to these studies (Sial et al, 2018., Fambon, 2017., Naseer & Ahmed, 2016., Davtyan; 2016., Burki & Khan, 2012., Jamal, 2006). The research evaluates the role of spatial concentration in exacerbating disparities across various locations, by examining the likelihood that affluent regions would grow even more affluent because of the concentration of economic activities and expanding labour demand. To extract the impact of agglomeration on spatial inequalities, it is crucial to understand how the concentration of industries affects the growth process of a particular location. Depending on the agglomeration behaviour, the various regions would have distinct physical and economic structures and sizes. The intensity and diversity of industrial agglomeration in a region determine its economic circumstances.

On the theoretical side, the concept of Agglomeration economies is defined as the process of clustering of firms in close proximity to one another to exploit the benefits of being together with an intention to minimise per unit cost of production or achieved economies of scale (Misra, 2011). The concept was first acknowledged by Weber (1909) while the detailed explanation of the sources of these economies was given by Marshall (1890). Marshall (1890) provided three bases through which economies of scale can be achieved. According to him, agglomeration enables firms to economise on their unit cost by sharing knowledge regarding efficient production technologies and management strategies. Further Firms may agglomerate to share the supply of intermediate inputs. The demand for intermediate input is greater when there are many firms located in the same place the average cost of providing such intermediate inputs to each firm will be low as high demand enables the producer of intermediate input to achieve scale economies. Lastly sharing a labour pool benefit firms when the variation in firm product demand is greater than the industry-level demand. The labour fired from a firm that is facing low product demand can be hired by a firm that is facing high product demand. For doing so firms have to bear no cost for searching and training labour thus declining unit cost for two reasons: one by firing unwanted labour by firms facing low product demand and second by reducing search and training costs by high product demand firms. In a nutshell, the productivity and income growth of a region are positively affected by agglomeration.

Subsequently, the growth of agglomerated urban regions, the spatial placement of population concentrations, and the clustering of economic activities within these population concentrations have all been the focus of extensive research in urban and regional economics. Concentration of economic activities may be the outcome of first-nature geography or second-nature geography, and thus the ensuing consequences may be addressed by focusing on its attributes. Basic geographic problems like the real topography of coastlines, mountains, and natural resources are addressed in the first nature geography. In contrast, second-nature geography is focused on the physical aspects of human behaviour, or the "geography of interactions between economic agents (Venables, 2006). The construction of highways, railroads, and rivers as well as the growth of regions, for instance, changed the expenses of economic contact across space. In contrast to first-nature geography, which is primarily external, second-nature geography is typically endogenous and may, in principle, be affected by policy.

With the rise of the "new economic geography" writing in the wake of Krugman (1991), the causes of the uneven distribution of economic activity across space have once again come to light. The New Economic Geography (NEG) paradigm states that spatial disparity is the combined result of two opposing forces: centrifugal force and centripetal force. Due to imperfect competition, agglomeration or aggregation of economic activities is determined by centripetal pressures, which also increase returns to scale and mobility of factors. Centrifugal forces, on the other hand, such as traffic congestion, high transportation costs, and restricted movement of factors, encourage economic activity dispersion. Businesses expand throughout immobile marketplaces when it is costly to transport goods to remote areas. Contrarily, when transit costs are minimal, immobile markets can readily access various locations, which motivates businesses to cluster at a particular location to reap the benefits from economies of scale and externalities. Increased economic activity in the region resulting from growing agglomeration leads to geographic disparities that demonstrate a positive interrelationship (Curve AA). Higher levels of agglomeration lead to greater congestion and undesirable externalities and thus counteract the positive benefits of agglomeration. As a result, a new connection between agglomeration and spatial disparity emerged that is in contrast to the earlier one (Curve RR) in Figure 1.

Figure 1: Connections between regional growth and inequalities



Source: Gardiner et al. 2010.

The spatial concentration of economic activities boosts the growth of the country's economy through localised spillovers. The SS curve shows how agglomeration boosts productivity and real output by labour pooling and disseminating knowledge. The degree of agglomeration, inequalities, and growth at equilibrium are determined by the intersection of the two curves, RR and AA. There exists a trade-off between national growth and agglomeration-induced spatial inequalities. Effective government action that breaks up agglomeration shifts the RR curve to RR1 and lowers geographic disparity by transferring resources to impoverished areas, but it also erases the spill-over effects that result in lower national growth. A comparable trade-off emerged within regions, as a consequence of the relatively faster increase in income in areas where economic activities concentrate.

The structure of agglomeration and its effect on spatial disparities in developing countries like Pakistan is vital to investigate. This research seeks to shed some light on the process of agglomeration and its effect on the income of regions that experiencing these agglomerations. The national literature on the spatial disparities between urban regions based on agglomeration is rather scarce. This study attempts to start bridging this gap by investigating the link between agglomeration, defined as location-specific economies of scale, and the incidence of income inequalities. In the study, large cities and other urban areas have been taken as spatial units. For estimating the agglomeration impact, this study used Labour Force Survey data, conducted by the Pakistan Bureau of Statistics (PBS), for the period of 2017-18. Further to identify the urban regions that are experiencing persistent inter-temporal spatial concentration in the case of Pakistan, the Herfindahl indices are computed for the years 2010 and 2018. The regions which have above-average levels of agglomeration in the most recent period and positive growth from the base year are marked as treated regions. This research is a valuable addition to the existing literature on Pakistan not only by analyzing the effect of agglomeration on regional income and inequalities but also for employing two distinct approaches for this evaluation. First, it uses a geographic information system (GIS) to geographically portray the connections between agglomeration, income and inequalities across treated and untreated regions to arrive at some valuable conclusions. Second, it employs propensity score matching for quantifying and validation of these results. This is the first study that employed propensity score matching to study agglomeration linkages with spatial inequalities arising from the unbalanced growth of regions.

Materials and Method

First, the Herfindahl indices will be computed to measure the inter-temporal spatial concentration of the sector. Using these indices values study will identify regions that experience higher than average initial levels of concentration and positive growth of spatial concentration. The Herfindahl index of spatial concentration (${HHC_j}^c$)is calculated as the sum of the regional shares in national employment in the particular industry both by industry and region. Symbolically

$$ {HHC_j}^c = \sum_{\mathrm{i}} CR_{ij}^{c^2} $$

Where ${CR_{ij}}^c$ symbolize the concentration ratio, which is computed as

$$ {CR_{ij}}^c = \frac{E_{ij}}{E_i} = \frac{E_{ij}}{\sum_i E_{ij}} $$

In the concentration ratio equation, E_{ij} represents the weight of employment in sector i from region j in the overall employment of sector i (E_i). The Herfindahl index of spatial concentration (*HHC*) ranges from 0 to 1. The more a region's Herfindahl index is closer to 0 the less spatially concentrated the region is, and the more it is away from 0 or closer to 1 the more it is said to be spatially concentrated.

After identifying treated regions using the Herfindahl index Arc GIS is used to visualize inequalities spatially. For this various dummies were generated for visualizing and categorizing the regions as per various criteria. Table 1 provides the details of these dummies along with their symbolic representations.

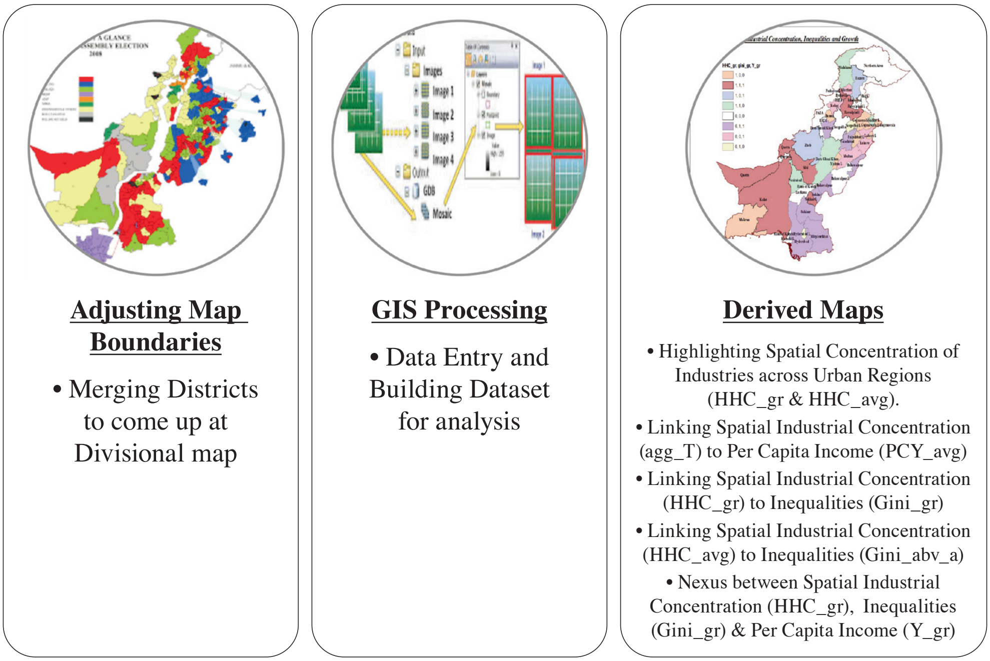
Table 1: description of variables and their Symbols used in GIS Maps

Variable Symbol	Description
HHC_gr	Being 1 if the growth of Hirschman Herfindahl concentration index (HHC) in the region is above average growth ($HHC\_gr_i > 1/n[\sum HHC\_gr_i]$) and is zero otherwise
HHC_avg	Being 1 if the HHC index value for a region is above average of all regions ($HHC_i > 1/n [\sum HHC_i]$) and zero otherwise.
agg_T	Is 1 for regions having above average concentration index value and having positive concentration index growth while its zero for regions failing to meet either of these criteria.
Y_gr	Being one if the real per capita income growth of the region is increasing and above average ($Y\_gr_i > 1/n [\sum Y\_gr_i]$) and zero otherwise.
PcY_avg	Is 1 if a region's income is higher than the national average ($PcY_i > 1/n [\sum PcY_i]$) and 0 otherwise.
Gini_gr	Is1 if the growth of the Gini index over time in a regions has positive growth ($Gini\_gr > 0$) and zero otherwise ($Gini\_gr < 0$).
Gini_abv_a	Being 1 if the Gini index value for a region is above average gini value of all regions ($Gini_i > 1/n [\sum Gini_i]$) and zero otherwise.

Source: Author's illustration

Using GIS the objective of this research would be presented in such a way that is both simple to comprehend and much more interesting to analyse. The steps involved in using GIS to analyse the objective are shown in Figure 2.

Figure 2: Step-wise GIS analysis.

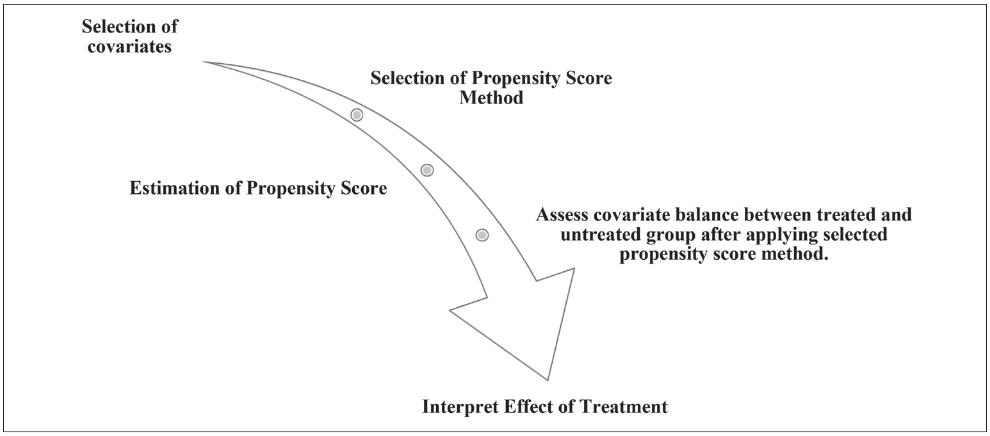


Source: Author's illustration

As the data from the microdata set is for large cities and other urban regions while the GIS map file is accessible with a distinct district map, the analysis using GIS starts by adjusting district and divisional borders. Therefore, according to the Pakistan Bureau of Statistics (PBS), all other districts within a region are combined, with the exception of large cities. To extract maps for the study, additional data for different important variables is then added. In maps, agg_T stands for agglomeration, with agg_T equal to 1 if the region meets both agglomeration criteria and zero otherwise. Per capita income is denoted by PCY, and PCY_avg = 1 if a region's income is higher than the national average and 0 otherwise. Gini_gr is the growth of the Gini index over time, which is 1 for regions with positive growth and zero otherwise.

Next to estimate the causal treatment effects of agglomeration in exaggerating income inequalities across various regions Propensity-Score Matching (PSM) technique is used. Rosenbaum and Rubin's (1985) propensity score analysis ensures that selection bias is minimised by equating the distribution of indicated significant traits (cofactors) between the treated and control groups. PSM enables one to obtain reliable estimates from observational analysis. Based on the similarity of their predicted probabilities of having treatment, or Propensity Scores, PSM balances the distribution of observed cofactors among the treated and untreated population. To determine mean effects, the PSM does not require subject-specific parametric modelling or any arbitrary assumptions about functional shape or error distribution. The method of carrying out Propensity Score Matching engrosses a series of six steps. Every step required decisions regarding the selection of covariates, models for generating propensity scores, matching distances and algorithms, the estimation of treatment effects, and finally the diagnoses of matching. Figure 3 illustrates the typical steps involved in the propensity score matching process.

Figure 3: Steps in Propensity Matching Score (PSM)

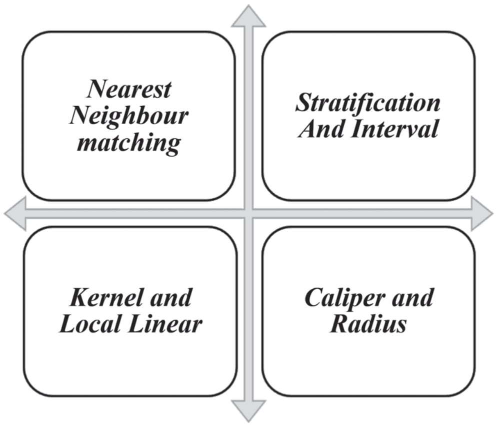


Source: Author's representation

The first step involved in-depth Knowledge of the problem and previous research to decide which variables are accounted to be confounders. While selecting the variable to be included in the estimation of the propensity score, it is important to keep in mind that the propensity score balances the covariates across treatment status. Executing matching therefore requires choosing covariates that can reasonably fulfil this requirement. Leaving out crucial variables can significantly increase bias in the results that are produced (Heckman et al; 1998). Only those variables that affect the treatment choice and the outcome variable at the same time should be included. Additionally, it should be assured that variables may not be impacted by treatment or anticipation of it. The second step is about the selection of the model to be used for estimating propensities. In contrast to the linear probability model, which makes forecasts outside of the [0, 1] limits for dichotomous dependent variables or binary treatment, both the logit and probit models present comparable bounded results.

The selection of a matching algorithm came next. Matching units based on their propensity scores can be done using a number of methods. The methods are summarised in Figure 4.

Figure 4: Ways to perform matching



Source: Author's visualization

Nearest Neighbor Matching (Rosenbaum & Rubin; 1985) is a widespread technique for propensity score matching that matches each unit in the treatment group with a unit in the untreated group based on the closets' absolute distance between their propensities. As an alternative, caliper matching pairs each unit in the treatment group with a unit in the control group within a specific caliper band "b" (Cochran & Rubin, 1973). According to Rosenbaum and Rubin (1985), the caliper band should be less than or equal to 0.25 of the propensity score's standard deviation. Later, Austin (2011) proposed that 0.20 for b would be the ideal number. Radius matching, which matches each unit in the treated group with numerous units in the untreated group within a specified band, was first suggested by Dehejia & Wahba in 2002. Nearest Neighbor (NN) matching is used in this research, which is the most frequently used matching strategy. Differences are obtained after matching the complete treated unit to the control units and then the Average Treatment effect of Treated (ATT) is obtained by averaging these differences. The general formula to calculate ATT is defined as

$$\text{ATT}=(O_1 - O_0 |\, C = 1) = E(O_1|C = 1) - E(O_0|C = 1)$$

Here O specifies the potential outcome. C =1 indicates that the subject unit belongs to the treatment group with subsequent potential outcome O_1 and C =0 shows that the subject unit is in a counterfactual group with potential outcome O_0. Using Nearest Neighbor (NN) matching ATT is estimated as

$$\text{ATT}^{\text{NN}} = \frac{1}{N^C} \sum\nolimits_{\text{i:wi=1}} \left[O_i{}^{obs} - \sum\nolimits_{j \in C(i)M} W_{ij}\, O_{ij}{}^{obs} \right]$$

$$= \frac{1}{N^C} \sum\nolimits_{i:wi=1} O_i{}^{obs} - \frac{1}{N^T} \sum\nolimits_{j \in C(i)M} W_j\, O_{ij}{}^{obs}$$

N^C indicate the number of observations in the treated group
N_i^U indicate the number of controls or untreated matched with treated observations i
$W_{ij} = \frac{1}{N^U}$ if j is a control unit of i and 0 otherwise and $W_j = \sum_i W_{ij}$
$C(i)_M$ indicate the set of first M matches for unit i

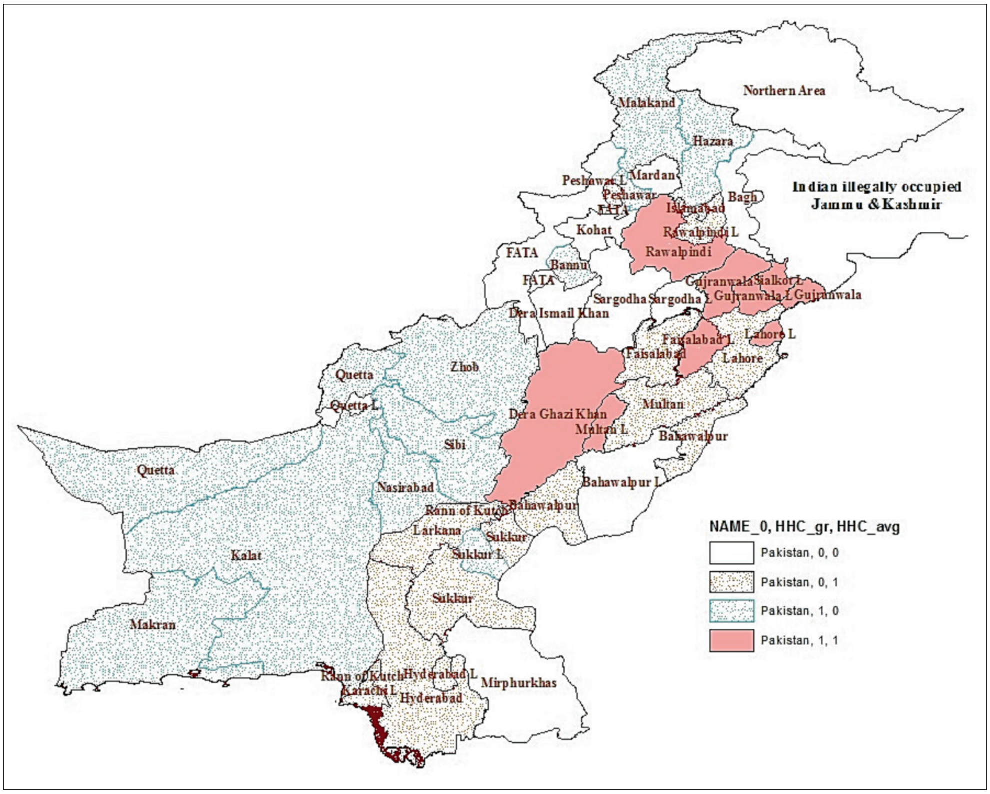
Results and Discussion

The core objective of this research is to analyze the impact of agglomeration in exaggerating inequalities via increased income levels in regions experiencing industrial concentration. Two distinct approaches are used to accomplish the said objective. First, the connections between spatial agglomeration and regional disparities are depicted geographically using a Geographic Information System (GIS) and then the Propensity Score Matching (PSM) is used to quantify and validate the outcomes. The quantification analysis is performed in two steps. To begin, we need to identify the urban regions in Pakistan that are more agglomerated. These areas are designated as treated regions if agglomeration to them has been above average in the most recent period and has positively grown since 2010. Then, considering these regions as treated their effect on regional income and income disparities are examined.

GIS Estimation Results

Spatial visualization of the regions experiencing concentration of industries with positive concentration growth (HHC_gr) over time and above average concentration (HHC_avg) in the most recent period under consideration have been represented in Figure 5.

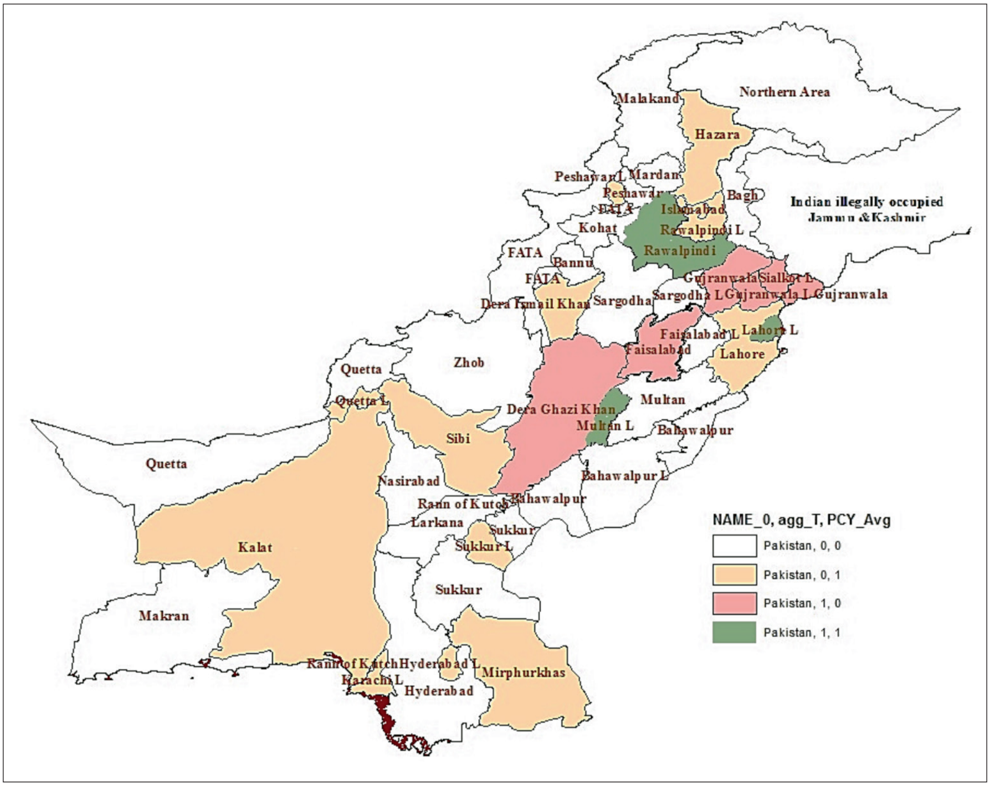
Figure 5: Highlighting Spatial Concentration of Industries (HHC_gr & HHC_avg) across Urban Regions



Source: Author's depiction using GIS.

Figure 5 indicates that there were 19 regions with above-average values of HHC_i including the mega-cities Karachi & Lahore while 20 regions were with positive agglomeration growth as highlighted by the legends (1,0) and (1,1). 8 regions qualify for both criteria that have above average agglomeration and were experiencing positive agglomeration growth. These regions include large cities Lahore, Faisalabad, Sialkot, Gujranwala, Multan, and other urban areas DG Khan, Rawalpindi and Gujranwala. Interestingly mega city Karachi despite having above average industrial agglomeration fall out because of having negative agglomeration growth indicating that industries are being crowded out from the city in response to the socio-political distress majorly along with electricity shortages too (Hasan et al; 2012 and Hasan & Raza; 2015).

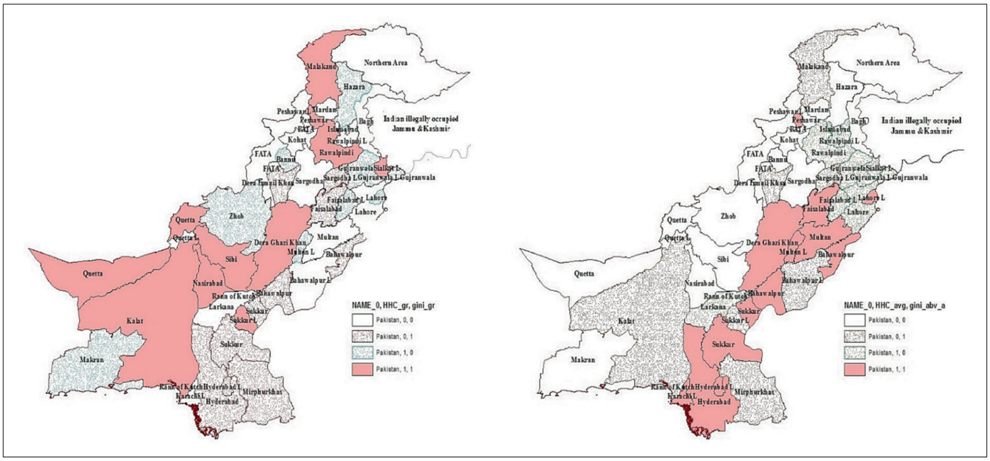
Figure 6: Linking Spatial Industrial Concentration (agg_T) to Per Capita Income (PCY_avg)



Source: Author's depiction using GIS.

Next agglomeration and income are linked geographically in Figure 6, with agg_T being 1 if the region qualifies for both agglomeration criterion and otherwise zero. Regions such as Kalat, Sibbi, Peshawar, Hazara, and DG Khan have above-average income because of the increasing concentration of economic activities (HHC_gr in previous map) as fruits of the China-Pakistan Economic Corridor (CPEC). Agglomeration and inequalities are examined spatially using the Gini index in Figure 7.

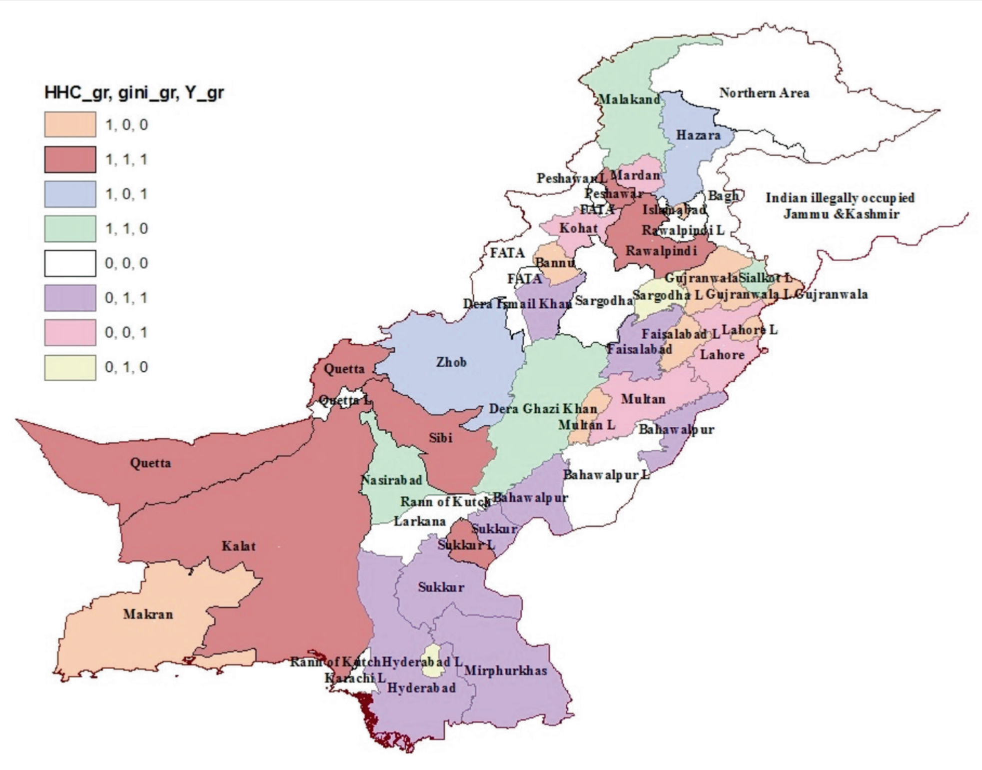
Figure 7: Linking Spatial Industrial Concentration & inequalities (HHC_avg) & (gini_ab g) and their growth (HHC_gr) & (gini_gr)



Source: Author's depiction using GIS.

The maps in Figure 7 conclude that in general growing inequalities in regions are attributable to the growing concentration of industries in them. Compellingly, it can be observed that the growth of agglomeration and inequalities are rising in Balochistan and the KPK region along the CPEC route supporting the new economic geography model. The regions that are already experiencing above-average agglomeration are also associated with above-average inequalities. Mostly these include regions from Sindh and Punjab Provinces such as Karachi, Lahore, Hyderabad & Multan cities, and other urban areas including Sukkur, Bahawalpur, Multan, Faisalabad and DG Khan. Also, Peshawar city shows similar results. Finally, the nexus between industrial concentration, income and inequalities is simultaneously analysed in Figure 8.

Figure 8: Nexus between Spatial Industrial Concentration (HHC_gr), Inequalities (gini_gr) & Per Capita Income (Y_gr)



Source: Author's depiction using GIS.

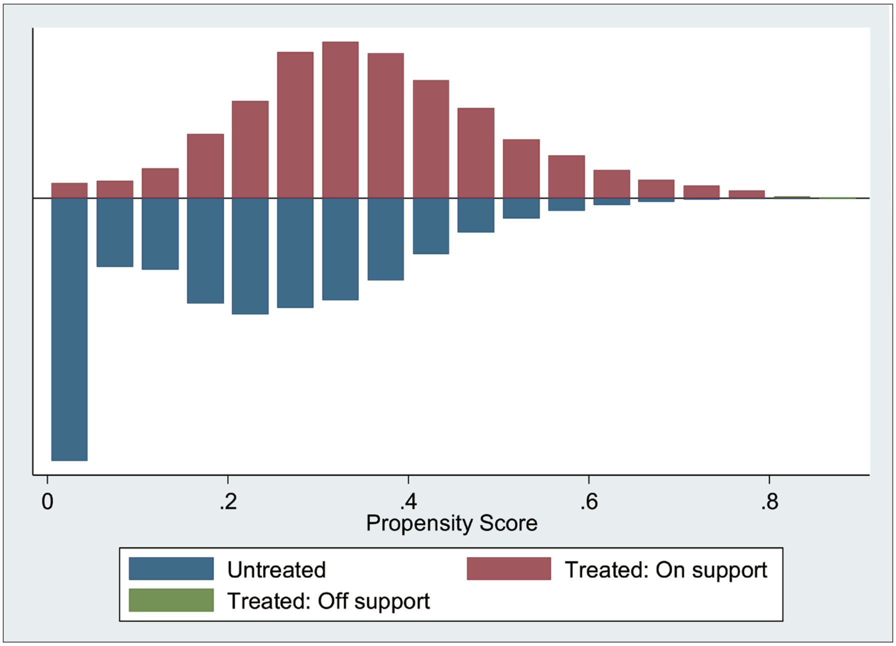
The concentration of economic activities in regions positively contributes towards increasing regional income but at the cost of increasing disparities within and across regions as proposed by the literature reviewed above. The regions that are confronting growth-enhancing agglomeration also encounter inequalities. This acknowledges the empirical considerations that economic growth is by nature uneven and this unevenness is connected with the spatial concentration of economic activities (Harvey, 2009). The China-Pakistan Economic Corridor (CPEC) is viewed as a big push to the economic development of Pakistan. CPEC is a long route passing probably from various backward areas of relatively less developed provinces of our country. The concentration of economic activities in the concerned regions due to improved physical infrastructure and road connectivity has raised the income level of the respective regions but these benefits of concentration are achieved at the cost of rising inequalities in these regions as apparent from the nexus present in figure 8.

PSM Estimation Results

To formulate policy recommendations, it is vital to measure the extent of the growing regional disparities caused by increasing agglomeration in some particular areas. ArcGIS contributed to rendering the phenomenon quite apparent, but generalisation needed significance and impact size. To confirm the reliability of the conclusion derived through GIS analysis, the PSM method is applied. Agglomeration of industries (agg_T) is regarded as a treatment variable that exhibits similar meaning for this analysis as well. Propensity scores are created using variables that seem significant for this purpose to infer a causative treatment impact of agglomeration on disparities in the distribution of income. These determinants include factors related to demographics, learning, and the labour market. All of these factors are anticipated to have an impact on the household's earnings. To estimate the propensities, the binary treatment variable -spatial concentration is run on the variables chosen; however, the probit model estimates are only to predict the propensities and thus do not provide any meaningful full explanation. The predictions are then used to generate propensity scores, which are then used for matching purposes. The above-mentioned findings are annexed (see annexe table A1). Overall, the model is statistically fit based on the likelihood chi-square.

It is crucial to evaluate the common support area in both the treated and untreated groups. It would be troublesome if the distribution did not overlap, but the issue could be solved using the minimum and maximum comparison. To accomplish this, all data with propensity scores less than the minimum or greater than the maximum would be removed. As a result, all observations that fall outside of the common support area [0.002, 0.853] are removed from the analysis. Figure 9 illustrates a depiction of the density distribution of propensity scores. The upper portion of Figure 9 depicts the propensity score distribution for the treated, while the lower section of Figure 9 depicts the untreated. The symmetric distribution of the variables after matching shows that the treated and non-treated groups coincide, implying that the common support condition is fulfilled.

Figure 9: Distribution of Propensity Score to test Quality of Matching among Treated and Untreated Groups



Source: Author's estimation using LFS 2017-18

To conclude whether the concentration of industries in a particular region positively contributes towards region income growth or is becoming an obstacle towards it by enhancing inequalities between and within regions two outcome variables log of income and Gini index are used. Matching results provide strong evidence of statistically significant income disparities between the treated and untreated regions. Treated regions are those which are experiencing above-average and positive growth of industrial concentration over time. In Table 2 the unmatched difference of 0.374 for log income indicates that the regions experiencing concentration exhibit about 45.4% [($e^{0.374}$– 1)*100 =45.4 %] higher income compared to the regions not qualifying as treated because of either declining growth or below average concentration. However, after matching the gap closes to 0.203 in log income which is approximately 22.5% [($e^{0.20335}$– 1) = 22.55%]. The unmatched difference for the Gini index is 0.02309, which indicates greater inequality in agglomerated regions. After matching, this difference marginally rises to 0.02452, revealing a persistent association between inequality and spatial concentration. The statistical significance of these effects is indicated by the high values of the t-statistics for both outcomes, which show that agglomeration increases income inequalities both within and between agglomerated regions by raising income levels.

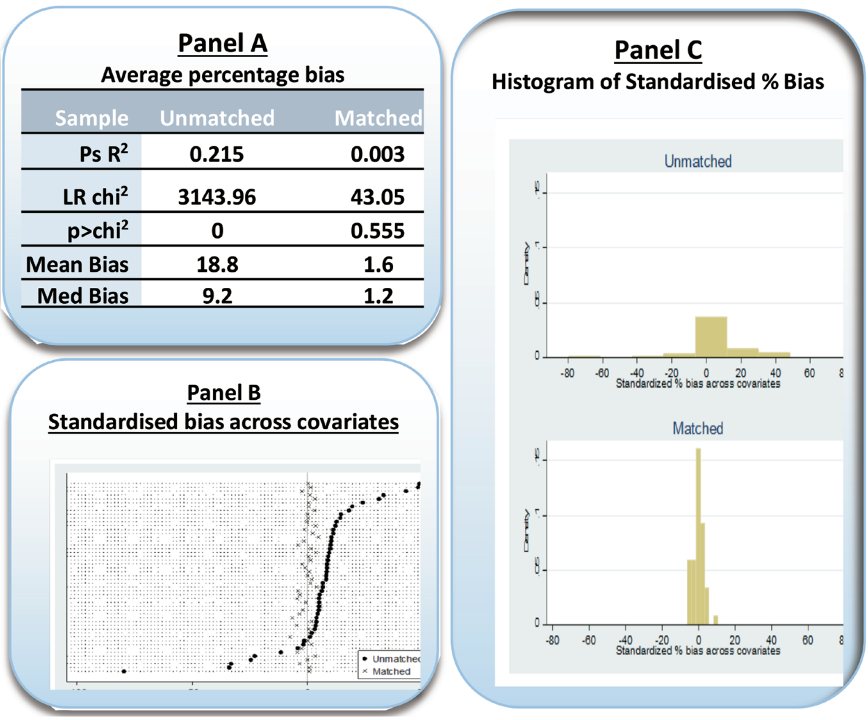
Table 2: Average Treatment Effects of Agglomeration on Outcome variables

Outcome Variable	Log Income		Gini Index	
Matched/ Unmatched	Unmatched	ATT	Unmatched	ATT
Treated	12.75	12.76	0.39859	0.39856
Controls	12.38	12.55	0.37549	0.37405
Difference	0.374	0.203	0.02309	0.02452
T-stat	27.07	9.09	18.37	13.89

Source: Author's estimation using LFS 2017-18

To assess the soundness of the matching quality the study carried out t-tests for the comparison of household characteristics within the treatment group to the corresponding untreated group. The results specify that before matching most of the variables indicate statistically significant differences but after matching has been done almost all covariates were found balanced and no significant statistical differences have been found (see annex table A2). The percentage bias for each of the covariates is also reduced significantly after performing matching.

Figure 10: Quality of Matching (PS-Test)



Source: Author estimation using LFS-2017-18

It can be observed from panel A of Figure 10 that the absolute mean bias has reduced from 18.8% to 1.6% and the absolute median bias reduced from 9.2% to 1.2% after performing the PSM. The graph in panels B & C of Figure 10 indicates how individual variables balance after matching. The x-axis of the graph presented in panel B displays the standardised bias, which is the percentage difference of the sample means in the treated and non-treated as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Rosenbaum and Rubin; 1985). It can be visualised from the figure that the unmatched sample exhibits large imbalances with standardised bias being present across many covariates but once the matching is done the standardised differences diminished significantly. Further, the same results can also be seen through the histogram (panel-C of Figure 10). It can be observed that the standardised percentage bias has reduced significantly after the matching.

Conclusion and Policy Implications

This study makes a novel contribution to the existing literature by conducting an in-depth analysis of inter and intra-urban income disparities within Pakistan through the lens of agglomeration dynamics which is rarely explored specifically in the case of Pakistan. The dual approach leveraging spatial visualization through ArcGIS and applying propensity score matching to quantify the influence of agglomeration on regional income inequality provides a robust understanding of the consequences of spatial concentration. The key findings of the analysis are as follows.

- Industry agglomeration boosts regional income, while it simultaneously intensifies regional inequalities.
- Using the values of Herfindahl indices urban regions are identified that have positive and above average level of concentration. Karachi and Lahore are among 19 urban areas with above-average agglomeration while positive agglomeration growth is evident in twenty regions.
- Eight regions exhibit both above-average concentration of economic activities as well as positive agglomeration growth. These eight regions comprise four large cities and four other urban areas, namely Gujranwala L, Sialkot L, Lahore L, Multan L, Rawalpindi, Faisalabad, Gujranwala, and DG Khan. These eight regions were considered treated regions while the rest that could not qualify both criteria were taken as controlled or untreated regions.
- Before matching, the average income level of agglomerated areas is on average 45.4% higher compared to regions considered untreated due to either below-average agglomeration or negative agglomeration growth. The variation narrows to 22.5% after matching, demonstrating the significant income-boosting impact of agglomeration and its ability to spur regional income growth and raising inter-regional disparities.

- Furthermore, even after controlling for matching, agglomeration is linked to considerably larger inequality, suggesting that agglomeration might have a significant role in heightening intra-regional disparities in incomes.

The unveiled results carry significant policy implications for the region, suggesting that curbing the agglomeration in the pursuit of fostering equality may be detrimental to productivity. Focusing on the growth of regional agglomerations of activities rather than distributing resources to every area appears to be the best course of action. Agglomeration is considered essential for promoting sustainable development in an area, even if it tends to create income inequality because it concentrates economic activity in some regions resulting in increased innovation, productivity, and total economic dynamism that all are critical components of long-term sustainability. Governments must thus develop sophisticated policies that strike a careful balance between encouraging agglomeration and reducing economic inequality. Targeted investments in technology, training, education, and infrastructure in concentrated areas might be part of strategies, along with steps to provide fair access to opportunities within the wider range. Additionally, the detrimental effects of income disparity both within and across areas can be lessened by the establishment of social welfare programs, progressive tax legislation, and inclusive economic initiatives.

The spillover effects from agglomerated regions can have a transformative impact on less-developed areas. Agglomerated regions can contribute positively to the general economic upliftment of the surrounding areas, and this effect can be further amplified by strategic government policies that promote connectivity between disadvantaged and agglomerated regions, such as targeted investment in communication, transportation, and infrastructure. Moreover, to provide local workers the knowledge and skills they need to be more marketable to industries and firms from agglomerated areas, the government should implement education and skill development programmes in underprivileged regions. By adopting a multifaceted approach that combines these measures, governments can foster a more equitable distribution of resources and opportunities, fostering sustainable growth across diverse regions within the country without compromising the benefits of agglomeration.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest/Competing interests

There is no conflict of interest/Competing interests

Availability of data and material

The data that support the findings of this study are openly available in the website of Pakistan Bureau of Statistics (PBS).

Code Availability

The results generated using Stata and GIS software are presented in the form of tables and graph within the manuscript.

Authors' Contributions

Uzma Tabassum was primarily responsible for conceptualizing and designing of the study and leading the writing of the manuscript. Munazah Nazeer contributed significantly by conducting the Geographic Information System (GIS) analysis and providing technical support for the spatial data interpretation. Both authors reviewed and approved the final version of the manuscript.

REFERENCES

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioural research, 46(3), 399-424.

Burki, A. A & Khan, A. M. (2012). Agglomeration Economies and their Effects on Technical Inefficiency of Manufacturing Firms: Evidence from Pakistan, International Growth Centre Working paper

Cochran, W. G., & Rubin, D. B. (1973). Controlling bias in observational studies: A review. Sankhyā: The Indian Journal of Statistics, Series A, 417-446.

Davtyan, K (2016). Interrelation among Economic Growth, Income Inequality, and Fiscal Performance: Evidence from Anglo-Saxon Countries, Review of Public Economics, 217(2), 37-66.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for non-experimental causal studies. Review of Economics and statistics, 84(1), 151-161.

Fambon, S. (2017) "The Determinants of Inequality and Income Gap between Urban and Rural Areas in Cameroon: Evidence from the ECAM3 Household Survey". Advances in Economics and Business, 5(7), 394 - 410.

Gardiner, B., Martin, R., & Tyler, P. (2010). Does spatial agglomeration increase national growth? Some evidence from Europe. Journal of Economic Geography, 11(6), 979-100.

Hasan, S. A., Subhani, M. I., & Osman, M. (2012). The energy short fall and it's after effects (a case study for Karachi city in context to Karachi electric supply corporation). MPRA Paper No. 37663 Pg. 42-49

Hasan, A., & Raza, M. (2015). Impacts of economic crises and reform on the informal textile industry in Karachi.

Harvey, D. (2009). Reshaping economic geography: the world development report 2009. Development and change, 40(6), 1269.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. The review of economic studies, 65(2), 261-294.

Jamal, H (2006). “Does inequality matter for poverty reduction? Evidence from Pakistan’s poverty trends” The Pakistan Development Review 45(3), pp. 439-459

Krugman, P. (1991). Increasing returns and economic geography. Journal of political economy, 99(3), 483-499.

Marshall, A. (1890) “Principles of Economics” London.

Misra, M. (2011). Planning Sustainable Cities: Global Report on Human Settlements 2009.

Naseer, S. & Ahmed, A. M. (2016) “Determinants of Income Inequality among the Earners in Pakistan” S3H Working Paper Series, National University of Sciences and Technology (NUST).

Rosenbaum, Paul R., and Donald B. Rubin (1985) “Constructing a control propensity score.” The American Statistician 39(1), pp. 33-38

Sial, M.S., Zheng, C., Khuong, N.V., Khan, T., Usman, M. (2018), does firm performance influence corporate social responsibility reporting of chinese listed companies? Sustainability, 10(7), 1-12.

Tabassum, U. (2016). Regional Specialization Trade and economic size of cities” MPhil dissertation, University of Karachi.

Venables, A. J. (2006). Shifts in economic geography and their causes (No. 767). Centre for Economic Performance, London School of Economics and Political Science.

Weber A. (1909). Über den Standort der Industrien (Mohr, Tübingen). Translated by C. J. Friedrich, 1929, as Alfred Weber’s Theory of the Location of Industries (University of Chicago Press, Chicago).

Annexure

Table A1: Estimates of Probit Model (If Agglomeration=1)

Variables	Coefficient	Standard Error
Average schooling	0.075***	0.00517
Age of highest earner	0.009***	0.00092
Education of highest earner	-0.012***	0.0044
Age of head of household	0.005***	0.00093
Dependency Ratio	0.0184**	0.00802
Migrant household	0.394***	0.0440
Technical/Vocational training	0.266***	0.0410
Working-age members in household	0.0224***	0.00589
Gender of HOH	0.298***	0.0445
Work hours	0.0099***	0.00118
Female Participation Rate	-0.282***	0.0351
Constant	-1.617***	0.246
Occupational dummies	Most of the dummies appear statistically significant	
Provincial dummies		
Observations	18,722	
LR chi2(46)	3143.96***	
Pseudo R2	0.215	

*** p<0.01, ** p<0.05, * p<0.1

Author's estimations based on LFS 2017-18

Table-A2. Covariates Balancing Test (PS-test)

Variables	Unmatched / Matched	Mean		%Reduction		t-test	V(T)/ V(C1)
		Treated	Control	%bias	Bias	T	
Female participation Rate	U	0.138	0.213	-21.8	74	-30.26	0.63*
	M	0.258	0.239	5.7		2.58	1.03
Gender of the Head – Male	U	1.084	1.059	9.8	77.3	16.08	1.39*
	M	1.077	1.071	2.2		1.06	1.07*
Working age members in household	U	3.841	3.713	6.2	70.7	9.35	0.94*
	M	4.472	4.254	10.6		5.23	1.12*
Education of the Highest Earner	U	5.631	4.897	19.5	93.1	23.33	1.09*
	M	5.731	5.781	-1.3		-0.66	1.08*
Age of the Highest Earner	U	37.07	35.59	12.6	97.8	15.07	1.06*
	M	36.27	36.30	-0.3		-0.12	0.95
Technical/Vocational training	U	0.253	0.166	30.9	99.4	50.06	1.32*
	M	0.285	0.285	0.2		0.09	1.05
Average hours work	U	52.061	47.177	42.8	96.2	62.69	0.95*
	M	49.668	49.851	-1.6		-0.86	0.79*
Average schooling	U	8.203	6.806	48.5	99	71.9	1
	M	8.263	8.278	-0.5		-0.23	1.01
Migrant HOH	U	1.166	1.089	33	99.3	56.51	1.68*
	M	1.178	1.179	-0.2		-0.1	1.05
Dependency ratio	U	3.189	3.448	-11.5	45.1	-15.82	0.67*
	M	2.091	2.234	-6.3		-4.26	0.94*
Age of HOH	U	47.15	45.826	10.3	70.3	15.62	0.95
	M	48.863	48.47	3.1		1.52	1.03

*if variance ratio outside [0.98; 1.02] for U and [0.94; 1.06] for M

Source: Author's estimation using LFS-2017-18