

Final report



Spatial analysis of small and cottage industries in Punjab, Pakistan



Syed M. Hasan
Attique ur Rehman
Ijaz Nabi
Naved Hamid
Usman Naeem

September 2017

When citing this paper, please
use the title and the following
reference number:
F-37310-PAK-2

DIRECTED BY



FUNDED BY



Spatial Analysis of Small and Cottage Industries in Punjab, Pakistan

Syed M. Hasan, Attique ur Rehman, Ijaz Nabi, Naved Hamid and Usman Naeem

Final Report

Contents

1. Introduction.....	4
2. Census of Small & Cottage Industry in Punjab 2011-13.....	8
2.1 Quality and Reliability of Data.....	9
3. Understanding Total Factor Productivity?	10
4. Reasons for Productivity Variation.....	11
5. Determinants of Productivity	13
5.1 Micro-level Factors.....	13
5.1.1 Managerial Skills.....	13
5.1.2 Quality of Labor and Capital	13
5.1.3 Learning-by-Doing	14
5.1.4 Product Innovation.....	14
5.1.5 Firm Structure Decisions	15
5.2 External Drivers of Productivity.....	15
5.2.1 Productivity Spillovers	15
5.2.2 Competition.....	17
5.2.3 Regulations	17
5.2.4 Infrastructure and Social Issues	18
6. Empirical Methodology: Estimating TFP.....	18
7. Econometric Issues and Proposed Solutions.....	20
7.1 Potential Sources of Bias.....	20
7.1.1 Simultaneity Bias.....	20
7.1.2 Other Possible Sources of Bias.....	21
7.2 Data Cleaning.....	23
8. Empirical Estimation and Discussion of Results	23
9. Conclusion & Policy Recommendations	30
Section B: Data Entry Component	59
1. Introduction.....	59
1.1 Description of Census of Small and Cottage Industry in Punjab: 2011-2013	59
1.2 Motivation for the Current Project.....	60
2. Project Implementation	60
2.1 Challenges Faced during Data Entry and how they were addressed	60

2.1.1	Data Entry Management System and Quality Assurance	61
3.	Project Output	61
3.1	Description of the Data Set	61
4.	Descriptive Data Analysis	62
5.	Future Plans	70
5.1	How to Access the Data Set?	70
5.2	Merging the Data Set with Other Data Sets	71

Determinants of Total Factor Productivity: An Analysis of Small and Cottage Industries of Punjab Province, Pakistan

Abstract

This paper uses estimates of total factor productivity of small and cottage enterprises located in Punjab province of Pakistan, to identify the reasons behind idiosyncratic variations. As the firm level productivity distribution indicate considerable differences, even within the same industry defined at 3-digit ISIC level, empirical analysis is required to identify the external and internal determinants as detailed in Syverson (2011) framework. The firm level data used for the study comes from a novel dataset developed through a census of cottage and small industries of Punjab Province. The universe of the census comprises of entire rural and urban areas of Punjab. As the robustness of empirical analysis hinges on reliable estimation of productivity, we correct the baseline OLS estimates for the simultaneity bias using instrumental variables and selectivity bias through Heckman correction as suggested in Wooldridge (2002). The productivity distributions for the cottage industry, small industry and the entire small manufacturing sector were separately estimated using residuals approach. The empirical findings are helpful in identifying significance of factors operating within the firms such as educational qualifications and professional training of employees for attaining higher levels of productivity. Results also signify the external drivers of productivity differences such as energy sources, marketing channels, access to finance and agglomeration economies. The results suggest direction of policy interventions towards provision of energy, appropriate marketing channels and selective professional training for improved efficiency and competitiveness.

1. Introduction

State support for growth and development of small scale industry is an important policy intervention aimed at addressing issues such as poverty and unemployment. Small scale industry, besides its several unique features such as frugal capital investment, resilience against recessions, ability to maintain growth over time, generation of agglomeration economies and support for entrepreneurial talent, is primarily recognized for its potential to generate employment opportunities. However, the direction of public policy to support small scaled

industry needs to be carefully determined so that impact of public intervention leads to a positive productivity shock for the efficient enterprise instead of providing protection to inefficient organizations. For optimal policy design, it is very important to identify the factors that control productivity variation among firms operating either at a specified level of production such as small or cottage or those operating within narrowly defined industrial classification. Accordingly, these factors can be broadly categorized as either internal or external depending upon if the firm itself or external economic environment determines them. Once determinants of productivity are segregated among these two categories it becomes simpler to identify the direction and source of required intervention aimed at improved efficiency.

Although, small and cottage industries may sound trivial when we compare their contribution to the country's GDP as shown in figure 1, but they actually play a very important part in the overall growth of an economy, characterized by their unique feature of labor intensiveness. The total number of people employed in this industry has been calculated to be around 80 percent of the industrial labor force in Pakistan (Noman, 1989). Another unique feature of small and cottage industry is their geographic spread across the entire province. Spatial map given in figure 2 shows that unlike large manufacturing sector which is limited mainly to few districts, small manufacturing is spread across all districts. As such small manufacturing, can efficiently act as agent for overcoming regional disparities. Developing countries like Pakistan, which are characterized by their high population growth along with shrinking agriculture and manufacturing sectors, are apprehensive of unemployment issues and therefore lay especial emphasis on building networks of small scale industries. It has been empirically proved that small scale and cottage industries are adept in distributing national income in more efficient and equitable manner, among the various participants involved in the economic process, than their medium or larger counterparts (Singh et.al. 2011). It is imperative to focus on the small and cottage industry so that the productivity of the economy can be improved as 'small scale manufacturing industries in Punjab contribute almost 58 percent to the country's overall industrial production, and account for about 60 percent of value-added in its manufacturing sector (Turab et al., 2012).

For the purpose of empirical analysis, we focus here on the Punjab province of Pakistan. Punjab is the largest province of Pakistan both in terms of population and its contribution towards the national economy. With an estimated area of 205,344 square kms, the region is divided into 36 districts and is a hub to more than 48 thousand industrial units (SMEDA, 2011). With a contribution of almost 60 percent towards the national GDP, the average growth rate of the province's GDP from 1999-2014 is almost 3.6 percent which is higher than the national average (PBIT, 2016). The region is copious in terms of its small scaled cottage industries with almost 79 thousand cottage and 93 thousand small scaled identified units (PSIC, 2013).

The industrial establishments in Punjab comprises of a multitude of spatial clusters that are diverse not only in terms of their industrial classification, historical backgrounds and competitive labor skills, but also in geographical location. Stretching across urban centers; from Lahore to Sialkot, Faisalabad to Sargodha and Multan to Rahim Yar Khan these clusters deal in the production of low tech cutlery products hi-tech auto parts, agriculture or processed food stuff, textile products, sports goods, traditional carved or modern aesthetic design furniture etc.

According to the officially notified definition, cottage industry is an enterprise in which an owner is both investor and laborer, employed capital is not more than one hundred thousand rupees in a given year, the number of workers in the enterprise is not more than 15 (on a single shift basis) at any time in a year whether working part or full time, and the owner doesn't own any other cottage industry or any enterprise but if he or she does, then the capital employed together should not exceed PKR:100,000. Similarly, small scale industry is a unit with fixed capital investment not exceeding PKR: 10 Million. No official permission is required for setting up an enterprise in small-scale industry.

The small-scale industry is particularly specialized in the production of consumer commodities with labor mainly working as farmers, carpenters, blacksmiths, potters, craftsmen and cotton weavers. By acting as an essential medium for the efficient utilization of the skills as well as resources available locally they can be instrumental in creating opportunities in rural areas so as to check the unbridled trend of urbanization. As these industries, do not require

substantial capital investment, they mainly utilize the labor power for the production of goods. The notion of providing gainful employment to non-farm workforce who were not being employed by the large and capital intensive industries nicely fits in to develop a mechanism that can scale down the extent of unemployment as well as poverty. Such industries also support women economic empowerment. Traditionally, women are not allowed or encouraged to undertake outside work, however, cottage or small-scale industries like carpet weaving, candle making or handicrafts can be setup within homes and women can be profitably employed.

The extent of cottage industries is not limited to rural areas alone but they have expanded to urban centers also. There is great demand for hand knotted carpets, embroidery, brassware, rugs and traditional bangles, both in national level and the international market. These goods alone provide around 30 percent of the export incomes of the entire manufacturing sector (Noman, 1989). Perkins (1977), highlights the importance of small-scale and cottage industry acting as a catalyst for the economic development of the Chinese economy by reiterating the fact that, small sized industry has a potentially large number of effects on the social structure of the country and the countryside through the possibility of absorbing rural labor, which would otherwise be unemployed during parts of the year and strengthening the position of women in the society’.

This paper analyzes the total factor productivity distributions of firms to identify the economic drivers. We are constrained to work with cross-sectional data as no earlier similar data with firm identifiers is available. We have taken care of potential biases in estimation on this account by adopting econometric methodologies suggested in literature. We are also careful in interpreting empirical findings. In this case, we hope to find an appropriate way to estimate a production function using cross sectional data, but before we look into what productivity is and how is it measured, we will first talk about the survey that we will be using for analytical purposes.

2. Census of Small & Cottage Industry in Punjab 2011-13

The data used in the study was collected during a census carried out on the initiative of the Punjab Small Industries Corporation (PSIC). PSIC was formed in 1972 to “promote sustained industrial development through provision of market driven credit, infrastructure and technological support to contribute towards poverty alleviation through job creation and socioeconomic uplift of the province”. The census was conducted between January 2012 to June 2013, where the focus was to identify the issues related to availability of raw material and manpower, efficiency of production processes and technical and financial constraints faced by the small-scale manufacturing industries in Punjab. For data collection, the province was divided into 10 administrative units and a total of 132,224 units of cottage and small industries were interviewed using listing forms and detailed questionnaire. The census was carried out across the 36 districts of the region covering both rural and urban areas. The urban frame consists of small dense areas with well-defined boundaries categorized as Enumeration Blocks (EBs), each of which holds about 250 households. On the basis of population census of 1998, there are 14766 EBs in urban areas of Punjab. For the rural areas, the survey frame consists of villages (Mouzas). A Mouza is the smallest revenue unit recognized by its unique HADBAST number, within a tehsil; an administrative sub-division of a district. There are total 25914 Mouzas in Punjab.

The entire census data comprises of total 164,860 enterprises; 71,913 cottage enterprises and 92,947 small enterprises at the census cluster level. The dataset contains industrial information at 3-digit, 4-digit and 6-digit Pakistan Industrial Classification which is consistent with the International Standard Industrial Classification Rev 3.1. The census was divided into two categories (i) cottage industries, and (ii) small-scaled industries, where cottage industries are mostly run by the artisans in their own cottages along with their own family members and the capital does not exceed more than one hundred thousand rupees. On the other hand, small-scaled industries are located in urban and semi-urban areas with the capital less than rupees 10 million (PSIC, 2013). Summary statistics for entire small manufacturing, small industry and cottage industry are given in tables 1-3. Table 4 indicates top ten revenue

generators at provincial level. Also concentration of cottage and small industry in each district are shown in figure 2.

2.1 Quality and Reliability of Data

A review about the census data is presented here to apprise the reader about the quality and reliability of data. Whereas PSIC undertook a gigantic task of collecting statistics of all enterprises working in the small and cottage industry through detailed interviews, it appears that the final dataset still needs a lot to be desired. The PSIC census report indicates that detailed interviews based on questionnaire were completed for 132,224 enterprises. The detailed review of the dataset indicates that several important variables have incomplete information.

Two such variables that are required for our empirical analysis, specifically the TFP estimation are labor wages and raw material values. Thus, wage information is available for 46,405 and raw material values are available for 53,582 enterprises only. The non-availability of data thus limits estimation of production function and unbiased TFP using instrumental variables approach to just 27,641 enterprises. Although this number is much less than the entire universe of small and cottage enterprises, yet it is around 20% of the entire enterprise population. Moreover, in order to ensure that the estimates are unbiased on spatial basis we confirmed that the sample is geographically spread across the districts as shown in table 7.

Having pointed out the above limitation, it is still possible to utilize the larger dataset in several beneficial ways. A few of these are indicated here and are as follows. First, the gross numbers of cottage and industry enterprises can be used to figure out presence of clusters. Figure 2 shows district level clusters of small and cottage industry. Second, in more detailed terms using industry level employment statistics regional diversity of small manufacturing can be estimated. We estimate this for districts in Punjab using the Henderson Lee and Lee (HLL) index. Third, spatial concentration of select industry can be estimated using Ellison Glaeser (EG) index. The methodology to calculate these indices and interpret their values is discussed in detail in section 5. Both these indices, HLL and EG are calculated for districts in Punjab and their top industries and reported in table 6. A quick glance at the HLL index indicates that Lahore,

Kasur, Rajanpur and Okara are few of the most diverse districts in terms of small manufacturing industry. Focusing on specific industries, EG index indicates that spinning and weaving of textiles, sports goods, insulated cable and wire manufacture, stone cutting etc. have high spatial concentration. Finally, the available data can be used to calculate Location Quotient for specific industries to determine their competitiveness and exporting potential.

Knowing that due to data limitations our estimates of TFP are limited to enterprises with complete information on all variables, we were curious to know if our empirical findings are generally true for the entire universe. In order to have some handle on this we plotted some of our explanatory variables such as age of firm and education of entrepreneurs at district level as shown in figure 8. The plots indicate that districts with high level of TFP have more educated entrepreneurs and firms that benefit from learning by doing.

3. Understanding Total Factor Productivity?

The word 'productivity' has gained popularity in recent times but its roots can be traced back to 1957 when Robert Solow introduced his Neoclassical Growth Model as an extension to the Harrod-Domar Model and talked about how productivity is linked to the production function i.e. technology, capital accumulation and labor force. According to Griliches (1994), economists explain productivity through estimating the existing state of production technology and therefore interpret the impact of any 'technological change by outward shifts in the production possibilities frontier (Griliches, 1994). In another instance, Law (2000) concludes that when economists speak of productivity growth, they essentially mean that the growth rate of the economy's ability to produce output from a given stock of inputs. Productivity therefore refers to output increase without any change in the scale of operations. Thus, the total factor productivity of an economy only increases if it obtains more output from a given supply of inputs (Law, 2000). On the other hand, Jorgenson (1995) explains the characteristic of productivity as being external to the economic activities. Thus, his focus is on spillovers associated with scale economies.

While there are diverging views on the interpretation of productivity, Lipsey (2000) sums it up by concluding that the conventional group holds that changes in total factor

productivity measure the rate of technical change. (Law, Statscan, Krugman, Young.) and the other (J&G) group holds that TFP measures just the free lunch of technical change, which is mainly associated with externalities and scale effects (Jorgenson, and Griliches). In simple words, one can conclude that productivity represents efficiency of production process i.e. for a given inputs how much output can be obtained. Graphically, TFP variations are represented by shifts in the firm's isoquants as production carried out with high productivity process will produce greater output with the same set of inputs. Shifts along isoquants caused by factor price variation does not influence TFP as it does not induce any shifts in isoquants. An easy way to estimate TFP is through a production function where output is the product of a function of inputs and Hicks-neutral shift given as follows:

$$Y = A F(K, L, M, E) \quad (1)$$

where Y is output, $F(\cdot)$ is a function of observable inputs capital K, labor L, intermediate materials M, energy E and A is the Hicks neutral shift which also gives the value of TFP.

Using this approach, TFP is, essentially a residual. Being residuals, it is fair to say that it is a measure of our ignorance (Abramovitz, 1956); it is the change in output produced that cannot be explained based on inputs consumed. Following Syverson (2011), this study is an attempt to “put a face on” the residual. Given the multiple determinants of productivity variation (Syverson 2011) actually identifies various features of the face.

4. Reasons for Productivity Variation

Before we move on to identify the determinants of productivity, it is important to take into account the reasons for variations in productivity within same industry. A recent research by Melitz (2003) develops a model to explain reallocations of factors of production and hence economic activity toward relatively higher productivity firms, both through entry and exit and expansion of business in local and foreign markets. In this context, Hsieh and Klenow (2009) raise the question about potential of growth in Chinese and Indian economies if they attained efficiency close to that of United States.

Availability of firm level micro panel data in recent times has helped in understanding the dynamics of long-run growth while controlling for the unobservable. This was not possible with aggregated industry level data and thus, models of economic changes caused by productivity shocks are estimated using firm-level data (see for example Fisher, 2006). In industrial organization literature, models connect productivity levels to factors such as production technology, consumer demand, and market structure. Similarly, other models estimate the effect of competition (Syverson 2004), the magnitude of sunk costs (Allan Collard-Wexler 2010), etc. Theoretical frameworks explaining heterogeneity through idiosyncratic productivity of firm used by Eaton and Kortum (2002) and Melitz (2003) provide the latest approaches to observe export market participation decision. In these models, intensive and extensive margins of trade depend on firms' productivity levels. Although this literature does not talk much about the determinants of productivity yet it is mostly expected that aggregate productivity gains comes from trade induced selection and competition.

Given the significant fact about existence of productivity differences, this raises important question: why some firms exhibit higher productivity than others? A simple way to find an answer is to analyze productivity empirically. The idiosyncratic productivity estimate for the firm when plotted for all incumbents of a region or sector give the complete productivity distribution. For a region or a sector performing at high productivity level will have a distribution that has higher mean; that is shifted to the right compared to a low productivity region or sector (Combes et al, 2012). Increases in the average productivity level across plants will thus expectedly translate into higher aggregate industry or regional productivity. One possible explanation of dispersion of productivity distribution comes from the reallocation mechanisms- resources namely factors of production move towards most efficient producer. A productivity shock in case of a particular sector may not mean the same for all members of the sector. Those receiving favorable shocks will observe expansion, while those hit by negative shocks will recede either from foreign to local markets or may even exit. Even some may be forced to exit the market. Thus following the shock the economy will readjust so that reallocation of inputs takes place.

5. Determinants of Productivity

Taking the conventional definition of productivity forward, the determinants of productivity and hence explanation of within industry differences across firms can be estimated both at micro level i.e. at firm level as well at macro level, i.e. elements of the industry or business or economic environment. However, it must be kept in mind that at times some of these forces can overlap and multiple factors can act in combination. Hence, we will divide our discussion into two parts, and try to connect the dots between productivity and internal/external factors affecting it.

5.1 Micro-level Factors

5.1.1 Managerial Skills

Research believes that managers drive productivity differences. Managers perform multitude of duties. They organize the efficient utilization of labor, capital, and intermediate inputs and hence one might expect poor management to lead to inefficient production operations. However, measuring the performance of the manager is a complicated task. Recent studies have made considerable efforts to bridge this estimation gap. Bloom et al., (2007) offer comprehensive research relating management practices to productivity in which they document that improved management practices are significantly correlated with various tangible measures of firm productivity and efficiency, such as labor productivity, TFP, sales growth, probability of survival etc.

A reasonable proxy to gauge the managerial skills in our case is the age of proprietor. This is argued on the basis that in case of cottage and small industry the proprietor is the manager also and accordingly his experience plays a major role in the productivity of the firm in terms of his ability to make smart and intelligent decisions. The empirical analysis in this study therefore uses the age of the proprietor as proxy for their managerial skills.

5.1.2 Quality of Labor and Capital

Another theoretical yet important factor determining the differences in productivity is the input quality differences between labor and capital. Factors like labor's level of education, specialized training, work experience and longevity of tenure at a firm are some of the reasons

for these differences in input quality. Pekka et al (2004), for example, use Finnish plant-level data to show that productivity increases with workers' education and age. Capital variation measures must take into account technological progress as accounting book-value figures may not capture these. Keeping these factors in consideration, our empirical results show similar trends; firms with educated workforce and professionally trained workforce production turn out to be more productive.

5.1.3 Learning-by-Doing

In human resource development firms require their employees to undertake induction training whereby, the new employees learn how to work by looking at more experienced and trained employees i.e. 'learning by doing'. The very act of operating under some experienced personnel can increase productivity and our results show a likewise trend; experienced and aged proprietors and firms were more productive than other. This not only allows producers to determine opportunities for production process improvements but also improve productivity.

5.1.4 Product Innovation

Man's imagination produces a constant process of technological progress and therefore induces a firm to innovate. These innovations may not necessarily increase the quantity of the product being produced, but rather raise its quality, thereby increasing the product price and, hence, the firm's revenue. If one thinks about productivity in terms of improved quality, product innovation is the way to improve productivity. As market price of a product signals its quality, this can be captured using standard revenue-based productivity measures. The effect on productivity using product innovation has been studied in several papers. One such paper written by Bartel et. al., (2007) deals with IT-based productivity growth that led to an improved ability to customize products. As mentioned before, research and development and higher-quality employees is one other input that can trigger innovation. In another paper, Bernard et. al., (2010) prove that a firm's TFP is positively correlated with the number of products it produces. However, the available empirical evidence in this regard is mixed. Thus, in case of our research, number of products produced by the enterprise are negatively correlated with their

total factor productivity. Thus innovation in small manufacturing comes with certain additional costs and hence lower revenues which is quite in line with Schumpeterian hypothesis.

5.1.5 Firm Structure Decisions

Last but not the least, the organizational structure of a firm, its industrial classification and vertical and horizontal linkages within the industry, etc. also plays a major role in its productivity. Studies have also suggested that decentralization and the flexibility of a firm i.e. and how easily new technologies are adopted also play a major role. One such study carried out by Bloom, et al (2009) explains European firms slacked productivity growth. In another study, Schoar (2002) concludes that plants that are diversified in terms of the products they are producing or belong to the conglomerate end of the structure have higher permanent productivity levels.

Using our questionnaire results, if we link the theory with the stats, it is safe to conclude that firm's structure does play some part in the productivity of a firm. This is established by the fact that our categorical variable about the organizational set-up of establishment has turned out to be significant both jointly for small and cottage industries as well as individually for small-scaled industries with a positive coefficient; thereby meaning that as we go from sole trader to a partnership, the productivity increases. This makes sense because a larger firm structure would mean there are more ideas, more capital to invest and work can be divided equally among partners. The marketing channels adopted by high productivity firms also play statistically significant role in their productivity levels.

5.2 External Drivers of Productivity

5.2.1 Productivity Spillovers

Sometimes when firms locate in the vicinity of each other, their decisions and how they operate have a major impact on nearby firms. These producer practices may result in spillovers i.e. externalities or third party effects which can affect the productivity. According to Marshal (1890), these externalities accrue in form of agglomeration mechanisms like thick-input-market

effects and knowledge transfers. However, one must take a note that knowledge spillovers need not to be tied to any single geographic or input market. This is true because most producers tend to emulate each other, following the leader-follower behavior where the larger firms with financial resources research, while the smaller firms try to imitate their behavior, regardless of whether they share a common input market.

In the past, research has proved that there is a positive correlation between productivity and spillovers. Yet the large and persistent productivity differences within narrowly defined industries suggests that any such spillover process does not lead perfect imitation and hence we do not observe productivity convergence. According to Syverson (2011), constraints do not allow followers from fully replicating industry leaders. In this framework, it is important to identify the extent of knowledge transfers, factors that influence this size, and the channels through which such spillovers operate. To quantify the agglomeration economies, we use Ellison Glaeser index for localization and Henderson Lee and Lee index for urbanization economies respectively.

The Ellison and Glaeser (EG) index, given as:

$$\theta_j = \frac{\sum_{i=1}^M (s_{ij} - x_i)^2 - (1 - \sum_i x_i^2) H_j}{(1 - \sum_i x_i^2)(1 - H_j)}$$

where s_{ij} is the share of industry j 's employment located in region i ; x_i is the share of overall manufacturing employment in region i ; $\sum_i (s_{ij} - x_i)^2$ is an index of region specific employment concentration given by the sum of squared deviations of employment shares of the industry j known as Gini-coefficient; $H_j = \sum_k M_{kj}^2$ is a Herfindahl-style measure of the industry j 's enterprise level concentration of employment, where M_{kj} is the k th enterprise's share in industry j 's employment. Typically θ scores above 0.05 indicate highly concentrated industry; and scores below 0.02 show geographically dispersed industry. The Ellison and Glaeser's index of geographical concentration of various industries computes industry and region specific share of employment and is also suitable for survey data where information up to plant level is available. Ellison and Glaeser (1997) assume that location choice of firms is motivated by profit maximizing behavior from internal and external economies specific to a particular location.

The Henderson, Lee and Lee (HLL) index, given as:

$$g_i^s = \sum_{i=1}^n \left[\frac{E_{in}}{E_i} - \frac{E_n}{E} \right]^2$$

where E is total national manufacturing employment and E_j total national employment in industry j . E_{ij} and E_i are the corresponding local employment figures. g_i^s Has a minimum value of zero, where, in a district, industrial employment shares are identical to the national share. The maximum value of g_i^s approaches two when a district is completely specialized in industry j . Therefore as g_i^s increases, diversity falls, that is it implies greater spatial specialization.

5.2.2 Competition

Although, spillovers play an important part in the productivity differences but another essential external driver is market competition. Competition acts as a pressure, threatening firms to improve constantly and to enhance their productivity. Thus, competition drives productivity through two key mechanisms. The first one draws its link from the Darwinian theory of natural selection i.e. survival of the fittest where heterogeneous compete with each other for a higher market share by increasing efficiency. The firm with the highest efficiency level would be able to produce at a lower cost and would be able to pass on these cost savings to the consumers in form of lower prices. This not only pushes existing firms to improve their efficiency levels but also raises the success bar for potential entrants.

The second mechanism becomes active as aggregate efficiency increase in an area. Due to high competition firms are motivated to take expensive productivity improving measures that they may avoid or delay otherwise. Such actions besides raising enterprise's own productivity levels, also leads to aggregate productivity growth at industry and regional level.

5.2.3 Regulations

While competition and spillovers acts as external agents, government's role is no less important when it comes to the differences in productivity among firms. Governments interfere in the productivity decisions of firms through policies either by incentivizing them or regulating them. Where inefficiently designed regulations can create disincentives that may reduce

productivity, deregulating or privatizing may reverse this. Bridgman, et al (2009), show that regulations in the U.S. sugar market discouraged the producers' incentive to raise productivity. Along with the U.S. Sugar Act of 1934 aimed at uplifting the agricultural sector out of the depression, the government funded a subsidy to sugar beet farmers with a tax on downstream sugar refining. Finally, when in 1974 the Act was repealed, the yields began to climb again immediately (Syverson 2011). Financial regulations such as those determining eligibility for business loans to customers with collateral or credit history are a major hurdle for small manufacturing sector.

5.2.4 Infrastructure and Social Issues

Although shared infrastructure such as road network, sea port or airport, high speed internet availability and vocational training institute do fall under the non-traded input sharing aspect of agglomeration economies, yet it is important to highlight their role on productivity levels attained by firms. Similarly, frequency of crimes and more specifically terrorist activities need to be examined to understand locational choices and industrial productivity.

6. Empirical Methodology: Estimating TFP

As discussed in the section above, studies have conventionally used the Cobb–Douglas production function to estimate productivity whereby the production function can be defined as:

$$Y_i = A_i K_i^{\beta_k} L_i^{\beta_l} M_i^{\beta_m} \quad (2)$$

The difference between this simple equation and the one discussed in the previous section is that now the powers of the parameters have beta coefficients where these beta coefficients measure output elasticities. The estimate of TFP obtained as the residual can then be used either to evaluate the impact of various policy measures, such as the extent of foreign ownership (e.g. Javorcik, 2004), trade. (Pavcnik, 2002; De Loecker, 2007) or its determinants (Syverson, 2011). Since 'A' i.e. the Hicksian neutral efficiency level of a firm is unobservable, we take the natural log of the Cobb-Douglas production function to convert the non-linear parameters in a linear form so that we can apply the OLS properties to our model and try to minimize the residuals. Taking natural log of (2) yields a linear production function:

$$y_i = \beta_0 + \beta_k k_i + \beta_l l_i + \beta_m m_i + \varepsilon_i \quad (3)$$

where lower-case letters represented logarithmic quantities and

$$\ln(A_i) = \beta_0 + \varepsilon_i \quad (4)$$

where β_0 measures the mean productivity level across firms and ε_i is the firm specific deviation from the mean. This last term can be further decomposed into an observable and unobservable component which yields the following equation,

$$y_i = \beta_0 + \beta_k k_i + \beta_l l_i + \beta_m m_i + v_i + u_i^q \quad (5)$$

where $\omega_i = \beta_0 + v_i$ represents firm-level productivity and u_i^q is an i.i.d. component, representing random deviations from the mean. Typically, empirical researchers estimate (3) and solve for ω_i .

Estimated productivity can then be calculated as follows:

$$w_i = v_i + \beta_0 = y_i - \beta_k k_i - \beta_l l_i - \beta_m m_i \quad (6)$$

and productivity in levels can be obtained as the exponential of w_i , *i. e.* $\Omega_i = \exp(w_i)$

The productivity measure resulting from equation (6) can be used to determine its various determinants at the firm level. Also, firm-level TFP can be aggregated to the industry level by using weights etc. Our model based on (6) minimizes the residual squares using OLS under the Gauss-Markov assumptions to estimate the log-TFP.

Nevertheless, numerous methodological issues arise when TFP is estimated using ordinary least squares (OLS). Estimating (6) using OLS leads to biased productivity estimates, caused by the endogeneity of input choices as productivity is unobservable to the econometrician but known to firm, and hence, input choices are likely to be correlated. Further allowance is to be made for entry and exit, to avoid selection bias. Besides the simultaneity and selection bias, various other methodological issues have been pointed out lately. In case there is imperfect competition in either output or input markets or both, an omitted variable bias will arise in standard TFP estimation if data on physical quantities and their corresponding firm-specific prices is unavailable and the industry-level deflators are used to proxy for firm-level

prices (Katayama et al., 2009). Finally, if firms produce multiple products, with significant difference in their production technology, not estimating the production function at the particular product level, rather than aggregating at the firm level, will introduce other methodological issues. Bernard et al. (2009) have recently questioned the relevant level of analysis for the estimation of a production function, in case of multi-product manufacturing sector firms in the US. The available empirical evidence regarding impact of multi-product output on TFP is generally mixed (Beveren 2010).

7. Econometric Issues and Proposed Solutions

In this section we focus on the potential biases in estimation of TFP, the econometric solutions to obtain reliable estimates and limitations of available data.

7.1 Potential Sources of Bias

7.1.1 Simultaneity Bias

The first and foremost important limitation to our TPF estimation is that of simultaneity because the inputs used in the production function are jointly determined leading to the problem of endogeneity. The issue of simultaneity arises when an independent variable, is not truly exogenous (i.e., it is a function of other variables). The non-fulfilment of this exogeneity condition or ‘endogeneity of inputs’ is defined as the correlation between the level of inputs chosen and unobserved productivity shocks (De Loecker, 2007).

In general, ordinary least squares (OLS) regression applied to a system infested with reverse causality will produce biased parameter estimates and this bias will not decrease as the sample size increases. Estimating parameters from a simultaneous equation model requires advanced methods such as two-stage least squares (2SLS) approach with instrumental variables (IV).

For industries, dependent heavily on flexible production factors such as labor, a failure to correct for endogeneity of input choice will increase the likelihood of a downward bias in estimated TFP, although the opposite is true for sectors making more intensive use of capital (Levinsohn and Petrin 2003). This issue of simultaneity has been the fundamental emphasis of methodological literature dealing with TFP estimation because the issue was first raised more

than 70 years ago by Marschak and Andrews (1944). Conventional methods dealing with the endogeneity of inputs issue include fixed effects and instrumental variables (Griliches and Mairesse, 1995). Lately methodologies like those introduced by Olley and Pakes (1996), Blundell and Bond (1999) and Levinsohn and Petrin (2003) have provided us with some viable solutions to the simultaneity issue but due to data limitations our model will use the conventional IV method to overcome this endogeneity caused by simultaneity.

The instrumental variables we have used in this model to avoid endogeneity are logarithms of establishment area and local raw-material quantity for the labor input in the production function. The Durbin Wu Hausman test confirms endogeneity in selecting labor quantity. To confirm the validity of our IVs, we used the Sargan Test to check for the over-identification restriction. The hypothesis being tested with the Sargan test is that the instrumental variables are uncorrelated to residuals, and therefore they are valid instruments. Furthermore, our use of Two-Stage Least Squares (2SLS) approach is in accordance with Wooldridge (2009) that asserts that the (2SLS) estimator is the most efficient IV estimator.

7.1.2 Other Possible Sources of Bias

Three other potential sources of bias are possible in our estimation of TFP. These are omitted variable bias, multi-product bias and selectivity bias. The omitted variable bias arises if the data does not provide quantities of firm level prices of inputs and outputs associated with production process. Thus if firm level price variation is correlated with input choices, this will result in biased input coefficients (De Loecker, 2007). However, as small and cottage firms are usually price takers and do not have the market power in either selling or buying markets so this limitation is not likely to bias our estimates.

Another potential source of bias could arise if firms produce multiple products. This is an issue because if firms produce multiple products within the same industry using different production process this will cause TFP estimates to be biased. Ideally, consistent estimation of TFP in case of multi-product firms requires information on the product mix, product level output, inputs, as well as prices.

A possible solution as proposed by Beveren (2010) is that ‘in the absence of information on inputs and outputs at the product level, it is possible to cluster firms into groups that make a single product to obtain estimates of product-specific factor elasticities and TFP levels. On the other hand, it may be noted that focusing only on single product firms could possibly lead to lower than actual estimates of TFP values, because it fails to take into account the potential synergies in the production process. Alternatively, if the researcher has knowledge of the number and type of products produced by each firm, consistent estimates of productivity can be obtained by allowing the parameters of the production technology to vary across firms producing a different product mix (Bernard, 2009)’. In case of our data 70% of firms produce one product only. Another 20% produce two additional but closely related products, yet the product mix is highly tilted in favor of one product. Furthermore our estimates are not likely to be biased on this account as we control for number of products in our model.

One of the major issues faced by our model is that since we are working with a cross-sectional data, we are unable to control for the exit of firms during that period of census. Empirically, it has been tested that entry and exit decisions are systematically related to differences in productivity. As firms’ exit patterns reflect initial productivity differences, leading to the prediction that higher productivity will lower the exit probability at the firm level (Beveren, 2010).

Intuitively, the bias emerges because the firms’ choice of inputs in a particular period are conditional on its survival. If firms have some knowledge about their productivity level ω_i prior to their exit, this will generate correlation between ε_i and the fixed input capital, conditional on being in the data set (ABBP, 2007). This correlation has its origin in the fact that firms with a higher capital supply will (*ceteris paribus*) be able to survive with lower ω_i relative to firms with a lower capital stock. In order to control selectivity bias in production function estimation, we use Heckman correction model with two equations; selection and outcome. As the econometric analysis for estimation of TFP potentially suffers from endogeneity and selectivity biases we use the methodology given in Wooldridge (2002, Chapter 19), where the selection equation follows a probit model. In order to correct for the bias due to endogenous

explanatory variables in such models, inverse Mills ratio is inserted in the IV-2SLS regression equation.

7.2 Data Cleaning

The data digitized by the PSIC had several issues which we had to be dealt with. First, several firms violated definition of cottage or small enterprise, they were dropped from analysis. The total number of dropped observations total to 32 enterprises.

Second, the raw material and the total output sections needed maximum cleaning. The per-unit prices of raw materials were not uniform and also the values of the variables were jumbled up and were entered in wrong cells. Third, for a lot of raw materials, zero-unit value was shown in the data. These were then assigned the average values based on identical firms

Fourth, to calculate the agglomeration indexes, we used 4-digit industrial classification and calculated the variables of interest at Tehsil level. However, due to incomplete information on labor and raw material only about 26000 establishments with complete information on labor, capital, raw materials and energy usage could be used in the analysis.

8. Empirical Estimation and Discussion of Results

Once reliable estimates of TFP are obtained that are free from simultaneity and selectivity bias, we proceed to perform econometric analysis to identify various internal and external determinants of enterprise productivity, as discussed in detail in section 5. For this, the following logarithmic model as given by (7) is estimated. The model is estimated separately for cottage industry, small industry and entire small manufacturing sector (combined cottage and small industries).

$$\omega_i = \beta_0 + \beta_{ext}X + \beta_{int}Z + \beta_{control}W + \mu_i \quad (7)$$

where X refers to factors external to the firm and Z to factors internal to the firm, both expectedly influencing the idiosyncratic TFP ω . Also in the model W refers to the control variables such as size of the firm and the area of its location whether urban or rural.

In estimation of (7) endogeneity issue can potentially bias the coefficients due to reverse causality. This eventuality is more likely to happen in case of internal drivers of

productivity. In all such cases, we performed the Durbin Wu Hausman test and in case the test confirmed the presence of endogeneity we used instrumental variables (IV) approach to correct for that.

The table 5 provides us with the detailed regression results reported along with their robust standard errors clustered on industries classified at 3-digit level. The three columns refer to results of cottage industry, small industry and the entire small manufacturing sector. The dependent variable is the firm level TFP for the relevant category. The explanatory variables can predominantly be classified into two categories i.e. internal variables that is factors which are controlled by the enterprise and external variables that is factors determined in consequence of a policy or by the overall economic environment. We will now talk about these variables in detail and discuss why they are important with respect to productivity of small manufacturing enterprises in Punjab.

The first firm level variable of interest is the education level of the entrepreneur. The results indicate that in case of cottage industry education of 8 years (middle level) to 12 years (Inter) is significantly correlated with TFP. Thus, enterprises managed by entrepreneurs with these educational levels and even without any professional training, are likely to have higher productivity in the range from 5-16%. However, in case of small industry, education of 12 years and beyond is found to be statistically significant with enterprises likely to exhibit productivity higher by 14-20%. The two results make a lot of sense in terms of educational qualification of the entrepreneur and the scope of the business activity. The results not only dispel the myth that education does not contribute much to the productivity of small enterprises but also provide direction for policy in terms of quantum of impact of educational intervention and resultant productivity improvement.

As discussed in section 5, the entrepreneur's age is used as a proxy for managerial experience as in (Luong and Hebert 2009). The empirical results show that managerial experience counts for its contribution towards enterprise productivity in case of small industries but not in case of even smaller cottage industries. We can intuitively interpret that as a startup manager of cottage scale business, an entrepreneur need not to worry about professional experience.

However, his early years' experience is likely to contribute positively when the business is scaled up to the level of small industry. Egression results show a positive relationship between the age of the entrepreneur and the firm productivity.. Econometrics show that 1% increase in age of the entrepreneurs will result in 24% higher productivity in case of small industry. For example, by the time an entrepreneur aged 40 years turns 44 year, the enterprise productivity shall increase by 24%. However, we need to be cautious while interpreting this as age does not have a linear relationship over the entire lifespan This finding is interesting as the impact of age increase in absolute number of years is diminishing over time. Various policy measures aim at providing venture support and often interested entrepreneurs self-select into these. However, in order to ensure higher productivity a screening based on managerial experience may yield better results. However, entry screening to ensure that entrepreneur's age is still in the positively related portion may help attain higher productivity gains.

Besides the level of managerial experience another important variable that contributes towards enterprise productivity is the age of the firm. The age of the firm signifies the "learning by doing" component of productivity and represents the improvement brought about by operating in a sector over time. Again, this component is empirically found to be significant for small industries only where 1% increase in age will yield a productivity higher by 5%. Thus, age does not matter very much in case of cottage industry but in case of small industry continuing business over years may positively help in attaining higher productivity. Also as individuals matter a lot in case of small scale informal businesses their experience has higher impact on productivity compared the age of the enterprise itself.

Along with educational qualification and experience of entrepreneur, another important firm level variable is the provision of professional training. Although our knowledge in this regard for the enterprises under study is limited as the dataset just has one binary variable on this yet we try to extract more information from this. The professional training variable is statistically significant for small industry only when estimated separately. Leaving aside missing responses, around similar proportion (0.53 to 0.055) of entrepreneurs in cottage and small industry reported to have received relevant professional trainings. Also, when we interact this

with the educational level it appears that in case of small industry, imparting training to entrepreneurs with 12 years of education or beyond has a statistically significant impact on the enterprise productivity in the range of 22-46%. For this analysis we excluded the firms which were older than 40 years or inherited by current proprietors.

Two inter-related variables that pertain to market of firm's products and the channel of marketing are also important to study. These variables are unique as they have an overlap in terms of firm's decision and also the policy support provided to them. For small industry, it can be seen that only high productivity enterprises are able to sell their products in international markets. This result is in line with the findings of the seminal work of Melitz (2003) which concludes that only high productivity enterprises are able to enter and compete in export markets. This finding points out that policy support either in form of technological upgradation or access to credit may help enterprises to cross the export market costs threshold and thus allow firms to compete in international markets.

Regarding sales and marketing channel, the commonly used channels are wholesalers, retailers, distributors, exports, online, contract and direct sales. The results indicate that in case of cottage industry sub-contracting as compared to direct one is associated with higher productivity firms through a higher value coefficient and level of significance. In case of small industry both sub-contracting and consignment basis marketing is associated with high productivity firms. These results signify that in case of small manufacturing sector, the use of sub-contracting channels is linked with higher level of enterprise productivity. Usually sub-contracting agreements involve performance of a part of a large contract and is somewhat financially supported through mobilization advances etc. The takeaway from this finding is that small and cottage enterprises need financial support during the production and marketing processes. The coefficient on distribution channel is negative as it is likely to deprive enterprise of its revenue share. The intermediary channels although ease the sales process as they are in direct contact with the customers but often charge high commissions. Provision of direct sales channel to firms shall help them attain higher productivity.

About the impact of organizational setup, the results indicate that enterprise based on partnerships have higher productivity. Specifically, by setting up the firm as a partnership, around 24% higher productivity level may be attained. This result makes sense on two accounts. First partnerships cause pooling of resources and ideas which results in productivity growth. Also partnerships cause businesses to operate on socially beneficial level whenever they share common resources. Policy implication of this result suggests that in certain sectors partnerships on the pattern of unitization schemes may help enterprises attain higher productivity.

Whereas collective diversity as in case of partnerships yields higher productivity, diversity in final products does not help in attaining higher productivity. The regression results indicate that in case of entire small manufacturing sector enterprises manufacturing more than one final output have lower productivity levels which is line with the findings of Dhyne et al (2016). Thus diverting production facilities to produce more than one products leads to loss of scale economies and hence lower the profitability of the firm. Diversifying the product base lowers the level of firm productivity, thus it appears that specialization on a single product may help improve enterprise efficiency.

The energy requirement of an enterprise plays a very important role in its efficiency and use of high-tech and mechanized production processes. The regression results indicate that in case of cottage as well as small industry, use of electricity is associated with high productivity level. In fact the use of electricity is an indicator that enterprise relies less on manual production techniques and employs electric machinery. Further use of electricity and its correlation with productivity may imply use of better illumination and office equipment by the enterprise. In this variable two interesting results pertain to use of generators and coal as alternate sources of energy. The use of electric generator is found to be highly significant. However, this raises doubts of potential endogeneity because of reverse causality; high productivity firms tend to use electric generators to overcome energy shortages. To check the robustness of results, we used instrumental variables approach and find that our doubts are not unfounded. Regarding use of coal, the results indicate that it being a less efficient source of

energy, it negatively impacts the productivity levels. The findings here have both private and public implications. Whereas the enterprise may consider switching to electricity or generators as source of energy instead of coal, smooth provision of electricity by the state may help in attaining higher productivity. In terms of policy significance, it is important to highlight on the role of various sources of energy. The biggest problem faced by manufacturing sector in Pakistan is power shortage. As a result firms have to resort to whatever alternate means are available to them, may it be coal, generator, electricity, wood, gas or oil. Empirically, all these sources except coal (coal being an inefficient source of energy which hampers productivity) have a positive coefficient in our model which is an indication of the positive correlation between energy and productivity, with the grid electricity having the largest numerical value for the parameter.

Amongst the external variables, the benefits to productivity that accrue from external economies of scale are quantified through EG and HLL index. As shown earlier, the EG index indicates the intensity of agglomeration externalities due to localization economies. Our results show that the EG index is statistically significant for cottage and small industries with a positive coefficient which means that the more geographically close the small manufacturing industries are located, the more productive they will be due to knowledge spillovers, labor matching and input sharing as outlined in the relevant literature (see for example; Rosenthal and Strange, 2004). In case of small manufacturing industries, the significance of EG index indicates that concentration of own industry employment has positive impact on the level of productivity. The regression results thus clearly suggest that building same industry clusters shall have positive impact on the individual firm's productivity. As small manufacturing is often labor intensive, building clusters shall also help in more elastic supply of labor at the cluster level and hence firms shall have a higher expected profit.

On the other hand, the HLL index represents benefits arising due to urbanization economies and it takes a negative value for the small industry. The HLL index measures the level of diversity and the index measures for each industry how much the local production mix differs from the national mix. If all industries mimic the national mix the index has a value zero

i.e. perfectly diverse while a highly specialized one would have an index approaching two. As the HLL index is statistically significant with respect to productivity and have a negative coefficient, this implies that in case of small manufacturing sector less regional diversity harms productivity. Reversing the argument, we can say that the more diverse the industrial structure is at a specific area, the more productive the small manufacturing there is. The geographical proximity of diverse industries helps in improving productivity through provision of varied needs of enterprises and generating greater opportunities. Also diverse regions help absorb negative shocks through inter-industry adjustment of labor and generation of more creative ideas and innovative designs. The policy implication of this result is that small industries be provided infrastructure support so that they are housed in diverse clusters.

Following the interesting results obtained for localization and urbanization economies, we calculate EG index for top revenue generating industries of Punjab in each district. We also identify the district level dominant clusters of cottage and small industry across the province and calculate HLL for each region. The EG and HLL indices are listed in Table 6. The HLL indices for these districts shows that Chiniot is the most specialized cluster of small manufacturing industry. The spatial mapping of the clusters is shown in figure 4 which may be looked in conjunction with table 6.

Lastly, the two remaining variables that we have not discussed so far but which externally affect the productivity of firms are the crime rate and the public infrastructure. Our two district level variables road length and crime also have some impact on the productivity. It is important to talk about these two variables because in a country like Pakistan where war on terror has reduced direct foreign investment in the country, the level of crime has equally reduced business confidence. This diminution in business confidence reduces business optimism, making businessmen skeptical of investment decisions. The negative but significant correlation between road and productivity for cottage industry indicates that road network development was not undertaken with due consideration to provide access to the small manufacturing enterprises.

The district level total factor productivity estimates are also obtained for the entire small manufacturing sector. This gives an idea about efficiency level or productivity level attained by each district. The spatial mapping of this is shown in figure 5. Along with mean TFP levels we also obtained dispersal in productivity distribution which is also shown in figure 5. This exercise was done to find out districts which have higher than average productivity. Also the dispersal of productivity distribution signifies if the region is supporting both low and high productivity enterprises. A review of figure 5 reveals that districts having higher TFP have also greater dispersion of the productivity distribution. This finding is line with Forslid and Okubo (2014) who show that both type of firms, high and low productivity firms with different capital intensive production functions may self-select in certain core regions. As cottage and small industry significantly differ from each other in terms of scope and scale of business, we draw separate maps, on the pattern of figure 5 for both as shown in figures 6 & 7 respectively. The trend of with high productivity with large dispersion in productivity distribution can be observed in case of cottage industry. However, similar locational trend is not observed in case of small industry due to forces of competition; the presence of high productivity enterprises forces less productive enterprises to exit the district market.

In conclusion, it can be said that our results provide us with a clear picture of the situation and problems faced by the small and cottage industries in the province. The productivity distributions (figure 3) and district level spatial maps shown in figure 5, are good tools to identify productivity trends and patterns across the province. Using such tools and empirical findings for pointing ways to improve productivity, we have added some policy recommendations in the conclusion section of the paper.

9. Conclusion & Policy Recommendations

The research investigating productivity variation across enterprises has come a long way since Bartelsman and Doms (2000) surveyed the literature. Extant literature is much more informed about what causes the measured differences in productivity, and how factors both internal and external to the plant or firm shape the distribution. The research done in this context has not only progressed a lot on theoretical frontier but has also stood the test of empirical analysis. On the theoretical side, it has been quite challenging to obtain bias free TFP estimates. However,

controlling for potential simultaneity and selectivity bias firm-level TFP estimates indicated significant within industry variations. These observed variations raised questions about the determinants of productivity; hence search for the contribution of various factors towards enterprise' productivity while controlling for its size etc.

The recently renewed interest in firm level productivity stems from another strand of literature also. Trade and industry models such as those by Melitz (2003) explain firm heterogeneity and hence its market survival in terms of its idiosyncratic productivity. According to these models, firm's ability to enter export market depends on its placement on the productivity spectrum. As this finding has important policy implications it becomes all the more necessary to segregate high productivity firms from the low efficiency firms for a more targeted and better designed public intervention.

While thinking about designing public interventions, it is extremely important that specific areas be identified and resources be directed towards them. As the empirical literature indicates a diverse range of factors that impact firm's productivity it is necessary to segregate these into internal and external ones. Syeverson (2011) surveys literature pertaining to firm's productivity and accordingly places various determinants in the two categories. One potential use of this categorization is to identify areas where a firm or enterprise can itself intervene and improve its efficiency, or the state may provide some requisite support.

The purpose of the current project is to apply this knowledge on the novel data set of small and cottage enterprises in the Punjab province. To the best of our knowledge no such study has been undertaken for the small manufacturing sector of Punjab, which just to reiterate is the largest job contributor to the national economy. Moreover, the study looks at the issue of improving total factor productivity more comprehensively. By segregating the factors in the two broad categories helps pin down exactly the who should initiate the intervention that would trigger the requisite productivity shock; the firm itself or the state. In contrast the focus of earlier studies is mostly towards evaluation of public policies without giving any serious thought to the measures that can be taken by the enterprise themselves and improve their efficiency.

A review of past policies in Pakistan indicates that the focus of industrial growth and development either on the pretext of import substitution or export promotion has largely been limited to the large enterprises. Thus small manufacturing sector continued to suffer from a range of issues. The current study is an attempt to identify the issues that are inflicting productivity loss on enterprises operating in the small manufacturing sector. By identifying the internal and external determinants of productivity the study can be utilized private sector and public policy makers to develop appropriate strategy for growth of small manufacturing enterprises in Punjab.

Following the discussion of results in the previous section it can be safely concluded that education plays a significant role in improving efficiency of the firm. Whereas some contribution for this may come from the enterprise itself, the significance of public role should not be under estimated as there is a known market failure in this sector. Several studies in case of US have analyzed the role of human capital measured through share of population with college degrees in urban growth. In the context of developing countries, a study by Mody and Wang (1997) indicates that secondary school enrollment has significant impact on average city wages. Our results also indicate that education up to grade 8 and beyond has positive impact on firm productivity. This also implies that besides regular schools, specialized programs for primary school dropouts may be helpful in this regard.

Besides education, professional training and financial assistance (access to credit) relate to two important public policy areas. The results obtained in this context indicate that policy intervention in this regard is not very helpful in improving enterprise efficiency unless the recipient have received certain educational qualification. The results obtained in this context indicate that policy intervention in this regard may be more helpful in improving enterprise efficiency if the recipient have received certain educational qualification. The results therefore suggest that participants in professional training programs or soft loans may be screened for education to ensure efficient outcome. Access to credit through soft loans has significance on multiple accounts. Loans may help enterprise improve their production process through improved technology, incorporate innovative production processes or produce more innovative

products. Such loans may also help firms to overcome the financial threshold to access foreign markets.

Another policy area that needs immediate response is the provision of clean and uninterrupted energy supply to the small manufacturers. The empirical results have shown that small industry firms have higher productivity if they use electric energy. On the other hand, coal as source of energy is negatively correlated with firm productivity. Hence use of coal as source of energy leads to increased inefficiency. Similarly, firms with ability to use generators have higher productivity. This result nevertheless, highlights the significance of uninterrupted power supply towards attaining higher productivity.

Another important contribution towards improved productivity can potentially come from provision of infrastructure. Our empirical findings show a strange result whereby one major indicator of infrastructure that is roads shows no correlation with productivity. Our discussion with stakeholders done for another related project pointed out that current road network does not provide significant access to small manufacturing enterprises. It is needless to emphasize the role of roads and similar infrastructure in improving firm productivity through better provision of cheaper inputs, easier commute for the labor and swift access to output markets. Planning of future road networks may consider their role in the context of small and cottage industry besides other factors.

In order to improve their productivity, the enterprise can focus on educated workforce with appropriate professional trainings, specializing on single product, selecting location to gain from agglomeration economies etc. On the other hand the public policy intervention can help enhance productivity levels through provision of sufficient and reliable energy supply, removal of financial constraints, building infrastructure to provide land access to small manufacturing, development of direct marketing channels and export market openings.

References

- Abramovitz, Moses. (1956). "Resource and Output Trends in the U.S. since 1870." *American Economic Review*. Papers and Proceedings, 46(2), pp. 5-23.
- Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu. (2008). "R&D Investments, Exporting, and the Evolution of Firm Productivity." *American Economic Review*, 98(2): 451–56.
- Bartel, Ann, Casey Ichniowski, and Kathryn Shaw. (2007). "How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills." *Quarterly Journal of Economics*, 122(4): 1721–58.
- Bartelsman, Eric and Mark Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata," *J. Econ. Lit.* 38:3, pp. 569-94
- Bernard, A.B., Redding, S.J. and Schott, P.K. (2009). "Products and Productivity." *Scandinavian Journal of Economics* 111(4): 681–709.
- Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. (2010). "Multiple-Product Firms and Product Switching." *American Economic Review*, 100(1): 70–97.
- Beveren, Ilke Van. (2010). "Total Factor Productivity Estimation: A Practical Review." *Journal of Economic Surveys* 26.1: 98-128. Web.
- Bloom, Nicholas, and John Van Reenen. (2007). "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics*, 122(4): 1351–1408.
- Bloom, Nicholas, Mark Schankerman, and John Van Reenen. (2007). "Identifying Technology Spillovers and Product Market Rivalry." *National Bureau of Economic Research Working Paper* 13060.
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. (2009). "The Organization of Firms across Countries." *National Bureau of Economic Research Working Paper* 15129.
- Blundell, R. and Bond, S. (1999). "GMM estimation with persistent panel data: an application to production functions." *IFS Working Paper* W99/4.
- Bridgman, Benjamin, Shi Qi, and James A. Schmitz. (2009). "The Economic Performance of Cartels: Evidence from the New Deal U.S. Sugar Manufacturing."
- Campbell, Jeffrey R., and Jonas D. M. Fisher. (2004). "Idiosyncratic Risk and Aggregate Employment Dynamics." *Review of Economic Dynamics*, 7(2): 331–53
- Cartel, 1(934–74). "Federal Reserve Bank of Minneapolis Staff Report 437."
- Collard-Wexler, Allan. (2010). "Demand Fluctuations in the Ready-Mix Concrete Industry." Unpublished.

- Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, Puga, Diego, and Roux, Sébastien. (2012). "The productivity advantages of large markets: Distinguishing agglomeration from firm selection." *Econometrica*, 80(6):2543–2594.
- Crespi, Gustavo, Chiara Criscuolo, Jonathan E. Haskel, and Matthew Slaughter. (2008). "Productivity Growth, Knowledge Flows, and Spillovers." National Bureau of Economic Research Working Paper 13959.
- Dhyne, Emmanuel, Amil Petrin, Valerie Smeets, and Frederic Warzynski. 2016. "Multi-Product Firms, Import Competition, and the Evolution of Firm-product Technical Efficiencies." Unpublished, University of Aarhus.
- De Loecker, J. (2007). "Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity." National Bureau of Economic Research Working Paper Series 13155.
- Doraszelski, Ulrich, and Jordi Jaumandreu. (2009). "R&D and Productivity: Estimating Endogenous Productivity." Unpublished.
- Eaton, Jonathan, and Samuel Kortum. (2002). "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741–79.
- Eslava, M., Haltiwanger, J., Kugler, A. and Kugler, M. (2004). "The effects of structural reforms on productivity and profitability enhancing reallocation: evidence from Colombia." *Journal of Development Economics* 75: 333–371.
- Fisher, Jonas D. M. (2006). "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks." *Journal of Political Economy*. 114 (June): 413.
- Forbes, Silke J., and Mara Lederman. (2011). "Does Vertical Integration Affect Firm Performance? Evidence from the Airline Industry." *Rand Journal of Economics*, 41(4): 765–90.
- Forslid, R., & Okubo, T. (2014). Spatial sorting with heterogeneous firms and heterogeneous sectors. *Regional Science and Urban Economics*, 46, 42-56.
- Foster, Lucia, John Haltiwanger, and C. J. Krizan. (2001). "Aggregate Productivity Growth: Lessons from Microeconomic Evidence." In *New Developments in Productivity Analysis*, ed.
- Griliches, S. (1994), "Productivity, R&D, and the Data Constraint", *AER*, 84(1).
- Griliches, S. (1995), "The Discovery of the Residual". NBER. Working Paper #5348.
- Griliches, Z. and Mairesse, J. (1995). "Production functions: the search for identification." National Bureau of Economic Research Working Paper Series 5067.
- Hamilton, Barton, Jack Nickerson, and Hideo Owan. (2004). "Diversity and Productivity in Production Teams." Working paper. N.p.: Washington U in St. Louis. Print.

- Hsieh, Chang-Tai, and Peter J. Klenow. (2009). "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–48.
- Jorgenson. (1995). "Productivity: Postwar U.S. Economic Growth" (Cambridge: MIT Press).
- Katayama, H., Lu, S. and Tybout, J.R. (2009). Firm-level productivity studies: illusions and a solution. *International Journal of Industrial Organization* 27: 403–413.
- Kellogg, Ryan. (2009). "Learning by Drilling: Inter-firm Learning and Relationship Persistence in the Texas Oilpatch." National Bureau of Economic Research Working Paper 15060.
- Lall, S. (1984). "India's Technological Capacity: Effects of Trade, Industrial Science and Technology Policies in Technology Capability in the Third World". Fransman & King (ed) McMillan; London.
- Law, M. T. (2000). "Productivity and Economic Performance: An Overview of the Issues". *Public Policy Sources*, No. 37.
- Lipsey, Richard, and Kenneth Carlaw. (2000). "What Does Total Factor Productivity Measure?" ResearchGate. Web. 29 Dec. 2016.
- Luong, M., and B.-P. Hebert. "Age and Earnings." *Perspectives 5 Statistics Canada*, Catalogue no. 75–001-X, 2009
- Maksimovic, Vojislav, and Gordon Phillips. (2001). "The Market for Corporate Assets: Who Engages in Mergers and Asset Sales and Are There Efficiency Gains?" *Journal of Finance*, 56(6): 2019–65.
- Marschak, J., Andrews, J. and William H. (1944). "Random simultaneous equations and the theory of production." *Econometrica* 12: 143–205.
- Mas, Alexandre. (2008). "Labour Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market." *Review of Economic Studies*, 75(1): 229–58.
- Melitz, Marc J. 2003. "The Impact of Trade on Intraindustry Reallocations and Aggregate Industry Productivity." *Econometrica*, 71(6): 1695–1725.
- Mody, Ashoka, and Fang-Yi Wang. 1997. Explaining Industrial Growth in Coastal China: Economic Reforms ... and What Else? *The World Bank Economic Review* 11 (2):293–325.
- Noman, Omar. (1989). "Small Scale Industry in Pakistan with Special Reference to the Regulatory Framework. Islamabad, Pakistan. Small Scale Industry in Pakistan with Special Reference to the Regulatory Framework." Print.
- Olley, S.G. and Pakes, A. (1996). "The dynamics of productivity in the telecommunications equipment industry." *Econometrica* 64: 1263–1297.
- Ornaghi, C. (2006). "Assessing the effects of measurement errors on the estimation of production functions." *Journal of Applied Econometrics* 21: 879–891.

- Papanek, G. (1962). "The Development of Entrepreneurship." AER: Papers and Proceedings, Vol: L11(2).
- PBIT. (2016). "Punjab at a Glance." Punjab Board of Investment & Trade. Punjab Information Technology Board. Web. 28 Dec. 2016.
- Pekka, Ilmakunnas, Mika Maliranta, and Jari Vainiomäki. (2004). "The Roles of Employer and Employee Characteristics for Plant Productivity." *Journal of Productivity Analysis*, 21(3): 249–76.
- Perkins, Dwight. (1977). "Rural Small-scale Industry in the People's Republic of China." Berkeley: U of California, Print.
- Petrin, A., Levinsohn, J. and Poi, B.P. (2003) "Production function estimation in Stata using inputs to control for unobservables". *Stata Journal* 4: 113–123.
- Petrin, Amil, and Jagadeesh Sivadasan. (2010). "Estimating Lost Output from Allocative Inefficiency, with Application to Chile and Firing Costs." Unpublished.
- Punjab Small Industries Corporations - PSIC. Census of Small & Cottage Industry in Punjab: (2011-13). Comp. Shamim Rafique and Muhammad Arif Bhatti. Vol. 1. N.p.: Government of Punjab through Annual Development Programme (ADP), 2011-13. Print.
- Rosenberg, N. (1982). "Inside the Black Box: Technology and Economics." Cambridge University Press, Cambridge.
- Schmitz, James A. (2005). "What Determines Productivity? Lessons from the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries following Their Early 1980s Crisis." *Journal of Political Economy*, 113(3): 582–625.
- Schoar, Antoinette. (2002). "Effects of Corporate Diversification on Productivity." *Journal of Finance*, 57(6): 2379–2403.
- Singh, Lakhwinder Pal, and Arvind Bhardwaj. (2011). "Respirable Suspended Particulate Matter (RSPM) and Respiratory Symptom among Casting Industry Workers: An Exploratory Study in Northern India." *International Journal of Advanced Engineering Technology* 2.1: 251-59. Technical Journals Online. Technical Journals Online, Jan. 2011. Web. 19 Jan. 2017.
- Small and Medium Enterprises Development Authority - SMEDA. (2011). Lahore. Punjab. Government of Pakistan. Web. 27 Dec. 2016. <https://www.smeda.org/index.php?option=com_content&view=article&id=69&Itemid=174>.
- Syverson, Chad. (2011). "What Determines Productivity?" *Journal of Economic Literature* 49.2: 326-65. Web.
- Syverson, Chad. (2004) "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*, 112(6): 1181–1222.

- Turab, Syed, Usman Khan, and Kashif Malik. (2012). "The Constraints to Industry in Punjab, Pakistan." *The Lahore Journal of Economics*. Web.
- Van Biesebroeck, Johannes. (2003). "Productivity Dynamics with Technology Choice: An Application to Automobile Assembly." *Review of Economic Studies*, 70(1): 167–98.
- Van Reenen, John. (1996). "The Creation and Capture of Rents: Wages and Innovation in a Panel of U.K. Companies." *Quarterly Journal of Economics*, 111(1): 195–226.
- Veracierto, Marcelo. (2008). "Firing Costs and Business Cycle Fluctuations." *International Economic Review*, 49(1): 1–39.
- Vives, Xavier. (2007). "Innovation and Competitive Pressure." IESE Business School Working Paper 634.
- Wedervang, F. (1965). "Development of a Population of Industrial Firms." Oslo: Scandinavian University Books.
- Wooldridge, Jeffrey M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- Wooldridge, J.M. (2009). "Estimating firm-level production functions." *Economics Letters* 104(3): 112–114.

Table 1. Descriptive Statistics for Small Manufacturing in Punjab					
Variable	Obs	Mean	Std. Dev.	Min	Max
Sales Revenue (log)	131198	12.806	1.701	0	29.870
Raw Materials (log)	53852	12.628	2.047	0	26.427
Fixed Capital (log)	133598	11.788	1.433	0	24.532
Total Wage (log)	46405	9.337	1.198	0	14.145
Energy (log)	117747	11.026	1.331	0	22.669
Inherited Enterprise	164654	0.073	0.260	0	1
Technical Training	136409	0.544	0.498	0	1
Proprietor's Age	137982	41.026	10.740	0	99
Mechanized	132159	0.930	0.254	0	1
Age of Firm	134293	11.102	9.679	0	213
Area	164827	0.477	0.499	0	1
Product Market	131216	1.359	1.492	1	8

Table 2. Descriptive Statistics for Cottage Industry Manufacturing in Punjab

Variable	Obs	Mean	Std. Dev.	Min	Max
Sales Revenue (log)	71462	12.4531	1.36502	0	25.30004
Raw Materials (log)	26535	12.07068	1.71442	0	23.80838
Fixed Capital (log)	71885	11.35461	1.177138	5.135798	24.12448
Total Wage (log)	16019	8.812549	0.9261124	0	13.12236
Energy (log)	63419	10.76242	1.138839	0	17.39923
Inherited Enterprise	71718	0.0842327	0.2777384	0	1
Technical Training	71276	0.5535243	0.4971304	0	1
Proprietor's Age	70893	40.54728	10.78419	0	99
Mechanized	71885	0.9225986	0.2672292	0	1
Age of Firm	71659	10.6455	9.371912	0	129
Area	71885	0.3682409	0.4823306	0	1
Product Market	70440	1.412351	1.625323	1	8

Table 3. Descriptive Statistics for Small Enterprise Manufacturing in Punjab

Variable	Obs	Mean	Std. Dev.	Min	Max
Sales Revenue (log)	59736	13.229	1.950	0	29.870
Raw Materials (log)	27317	13.170	2.194	0	26.427
Fixed Capital (log)	61713	12.293	1.535	0	24.532
Total Wage (log)	30386	9.613	1.232	0	14.145
Energy (log)	54328	11.333	1.467	0.693	22.669
Inherited Enterprise	92936	0.064	0.245	0	1
Technical Training	65133	0.534	0.499	0	1
Proprietor's Age	67089	41.533	10.669	0	98
Mechanized	60241	0.940	0.238	0	1
Age of Firm	62634	11.625	9.994	0	213
Area	92942	0.561	0.496	0	1
Product Market	60776	1.297	1.317	1	8

Table 4. Top Ten Revenue Generating Industries in Punjab

Industry Name	4-Digit ISIC Code
Manufacture of veneer sheets; manufacture of plywood, lamination board, particle board and other panels and boards	2021
Manufacture of grain mill products	1531
Manufacture of structural non-refractory clay and ceramic products	2693
Sawmilling and planning of wood	2010
Manufacture of articles of concrete, cement and plaster	2695
Manufacture of non-structural non-refractory ceramic ware	2691
Treatment and coating of metals; general mechanical engineering on a fee or contract basis	2892
Manufacture of other articles of paper and paperboard	2109
Manufacture of insulated wire and cable	3130
Manufacture of bakery products	1541

Table 5. Empirical Estimates for the Factors Affecting Productivity in Punjab

	Cottage	Small	All
Variables	TFP_Cottage	TFP_Small	TFP_Small_Manuf
Inherited Enterprise	-0.118** (0.0523)	-0.0713* (0.0397)	-0.0898*** (0.0314)
Proprietor's Age (log)	0.0587 (0.0752)	0.245*** (0.0606)	0.174*** (0.0487)
Mechanized	0.116*** 0.0398	0.0583 0.0429	0.0788*** 0.0299
Firm Age (log)	0.0316 (0.0258)	0.0537*** (0.0204)	0.0440*** (0.0164)
Urban Area	0.192*** (0.0381)	0.108*** (0.0351)	0.119*** (0.0269)
Market of Product: National	0.209 (0.151)	0.0623 (0.0607)	0.0799 (0.0552)
Market of Product: International	0.218 (0.527)	0.728*** (0.207)	0.756*** (0.191)
Education of Proprietor: Middle	0.105* (0.0592)	-0.0252 (0.0576)	0.0562*** (0.0209)
Education of Proprietor: Matric	0.0544* (0.0601)	0.0254 (0.0577)	0.0645*** (0.0210)
Education of Proprietor: Inter	0.155** (0.0689)	0.1329*** (0.0510)	0.0869*** (0.0227)
Education of Proprietor: Degree	0.0497 (0.149)	0.1859*** (0.0769)	0.0920*** (0.0338)
Organizational set-up: Partnership	-0.0478 (0.101)	0.222*** (0.0575)	0.185*** (0.0501)

Number of Products Produced	-0.0164** (0.00796)	-0.0271*** (0.00726)	-0.0256*** (0.00559)
WholeSellers_MarketChannel	-0.0102 (0.0561)	0.0742 (0.046)	0.0222 (0.0364)
Retailers_MarketChannel	-0.0207 (0.0509)	0.223*** (0.0464)	0.127*** (0.0357)
Distributors_MarketChannel	-0.203* (0.108)	0.0345 (0.101)	-0.0619 (0.0771)
DirectSales_MarketChannel	0.0889* (0.0471)	0.111** (0.0443)	0.0829** (0.0337)
Exports_MarketChannel	0.0457 (0.255)	-0.146 (0.25)	-0.086 (0.188)
OnConsignmentBasis_MarketChannel	0.066 (0.0564)	0.209*** (0.0664)	0.186*** (0.0452)
SubContract_MarketChannel	0.186** (0.0898)	0.414*** (0.0812)	0.339*** (0.0619)
Electricity_EnergySource	0.0854*** (0.0263)	0.200*** (0.0729)	0.0351 (0.0564)
Gas_EnergySource	0.0314 (0.0853)	0.0992 (0.0648)	0.0977* (0.0529)
Coal_EnergySource	0.31 (0.219)	-0.303*** (0.0842)	-0.177** (0.0751)
Generator_EnergySource	0.203*** (0.0511)	0.434* (0.242)	0.0280 (0.100)
Log_roads	-0.0919*** (0.0279)	-0.0432** (0.0187)	-0.0515*** (0.0146)
Log_crime	-0.067***	-0.0032	-0.0239**

	(0.0171)	(0.0157)	(0.0115)
EG	0.874*** (0.174)	0.680*** (0.141)	1.001*** (0.104)
HLL	-0.470*** (0.151)	-0.917*** (0.139)	-0.720** (0.101)
Professional training	0.0706 (0.0574)	0.0116** (0.0577)	0.0388 (0.0611)
Middle#Prof.training	-0.065 (0.114)	0.0659 (0.0983)	0.0106 (0.0774)
Matric#Prof.training	-0.00718 (0.117)	0.0818 (0.0981)	0.0714 (0.0777)
Inter#Prof.training	0.00655 (0.137)	0.209** (0.103)	0.195** (0.0833)
Degree#Prof.training	-0.291 (0.31)	0.389** (0.156)	0.344** (0.136)
Financial Assistance and Matric	0.658*** (0.228)	0.200 (0.210)	0.209*** (0.0768)
Cottage (Dummy)			0.0272 (0.0176)
Constant	1.755*** (0.282)	0.126 (0.251)	0.667*** (0.185)
Observations	9068	15565	25012
R-squared	0.085	0.088	0.075
Cluster Robust Standard Errors in Parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

Table 6. District HLL Index and EG Index for Top Ten Revenue Generating Industries in Punjab

District	HLL	Industry Description	ISIC	EG
Bahawalpur	0.118	Manufacture of articles of concrete, cement and plaster	2695	0.012
Bahawalnagar	0.160	Manufacture of cocoa, chocolate and sugar confectionery	1543	0.002
R.Y.Khan	0.087	Manufacture of grain mill products	1531	0.009
D.G.Khan	0.047	Manufacture of bakery products	1541	0.002
Layyah	0.277	Sawmilling and planning of wood	2010	0.023
Muzzafargarh	0.069	Manufacture of veneer sheets; manufacture etc.	2021	0.028
Rajanpur	0.050	Manufacture of veneer sheets; plywood, etc.	2021	0.028
Faisalabad	0.205	Spinning of textile fibers; weaving of textiles	1711	0.114
Chiniot	0.382	Manufacture of furniture	3610	0.010
Jhang	0.055	Manufacture of structural non-refractory clay and ceramic products	2693	0.018
T.T.Singh	0.018	Manufacture of articles of concrete, cement and plaster	2695	0.012
Gujranwala	0.071	Manufacture of insulated wire and cable	3130	0.538
Gujrat	0.079	Manufacture of articles of concrete, cement and plaster	2695	0.012
Hafizabad	0.046	Manufacture of made-up textile articles, except apparel	1721	0.054
Mandi-Bahaudin	0.180	Manufacture of structural non-refractory clay and ceramics products	2693	0.018
Narowal	0.169	Manufacture of structural non-refractory clay and ceramic products	2693	0.018
Sialkot	0.124	Manufacture of sports goods	3693	0.200
Lahore	0.058	Publishing	2219	0.309
Kasur	0.042	Manufacture of grain mill products	1531	0.009
Nankana Sahib	0.150	Manufacture of articles of concrete, cement and plaster	2695	0.012
Sheikhupura	0.038	Treatment and coating of metals; general mechanical engineering on a fee or contract	2892	0.146

		basis		
Multan	0.371	Manufacture of non-structural non-refractory ceramic ware	2691	0.040
Khanewal	0.076	Manufacture of non-structural non-refractory ceramic ware	2691	0.040
Lodhran	0.058	Manufacture of dairy products	1520	0.021
Vehari	0.242	Manufacture of grain mill products	1531	0.009
Sahiwal	0.136	Manufacture of jewelry and related articles	3691	0.034
Pakpattan	0.145	Manufacture of non-structural non-refractory ceramic ware	2691	0.040
Okara	0.027	Manufacture of other articles of paper and paperboard	2109	0.160
Rawalpindi	0.179	Manufacture of grain mill products	1531	0.009
Attock	0.063	Manufacture of articles of concrete, cement and plaster	2695	0.012
Chakwal	0.294	Manufacture of grain mill products	1531	0.009
Jehlum	0.259	Manufacture of grain mill products	1531	0.009
Sargodha	0.078	Cutting, shaping and finishing of stone	2696	0.208
Bhakkar	0.068	Spinning of textile fibers; weaving of textiles	1711	0.1147
Khushab	0.100	Manufacture of grain mill products	1531	0.009
Mianwali	0.072	Manufacture of non-structural non-refractory ceramic ware	2691	0.040

Table 7: Sample Proportion of Firms Included in Empirical Analysis.

District	Total Firms Interviewed	Firms with Complete Data	Sample Proportion
Bahawalpur	2989	767	0.256608
Bahawalnagar	4061	620	0.152672
R.Y.Khan	2979	529	0.177576
D.G.Khan	2488	624	0.250804
Layyah	4671	182	0.038964
Muzzafargarh	5429	1239	0.228219
Rajanpur	1691	259	0.153164
Faisalabad	15205	4083	0.26853
Chiniot	3096	1081	0.34916
Jhang	4138	475	0.11479
T.T.Singh	3076	613	0.199285
Gujranwala	10028	2210	0.220383
Gujrat	3774	485	0.128511
Hafizabad	2117	340	0.160605
Mandi- Bahaudin	2287	542	0.236992
Narowal	3047	242	0.079422
Sialkot	4865	604	0.124152
Lahore	4274	1160	0.271409
Kasur	4671	1001	0.214301
Nankana Sahib	1583	260	0.164245
Sheikhupura	2919	433	0.148339
Multan	5923	1938	0.327199
Khanewal	3289	560	0.170265
Lodhran	2110	450	0.21327
Vehari	2757	472	0.171201
Sahiwal	2808	767	0.273148
Pakpattan	1767	249	0.140917
Okara	3742	783	0.209246
Rawalpindi	2668	831	0.311469
Attock	1589	323	0.203273
Chakwal	1245	152	0.122088
Jehlum	645	56	0.086822
Sargodha	5376	1332	0.247768
Bhakkar	4163	614	0.14749
Khushab	2150	266	0.123721
Mianwali	2746	407	0.148216
Total	132366	26949	0.203595

Figure 1: GDP Contribution of Manufacturing Sector in Pakistan in Percentage



Figure 2: Cluster of Cottage and Small Industry in Punjab

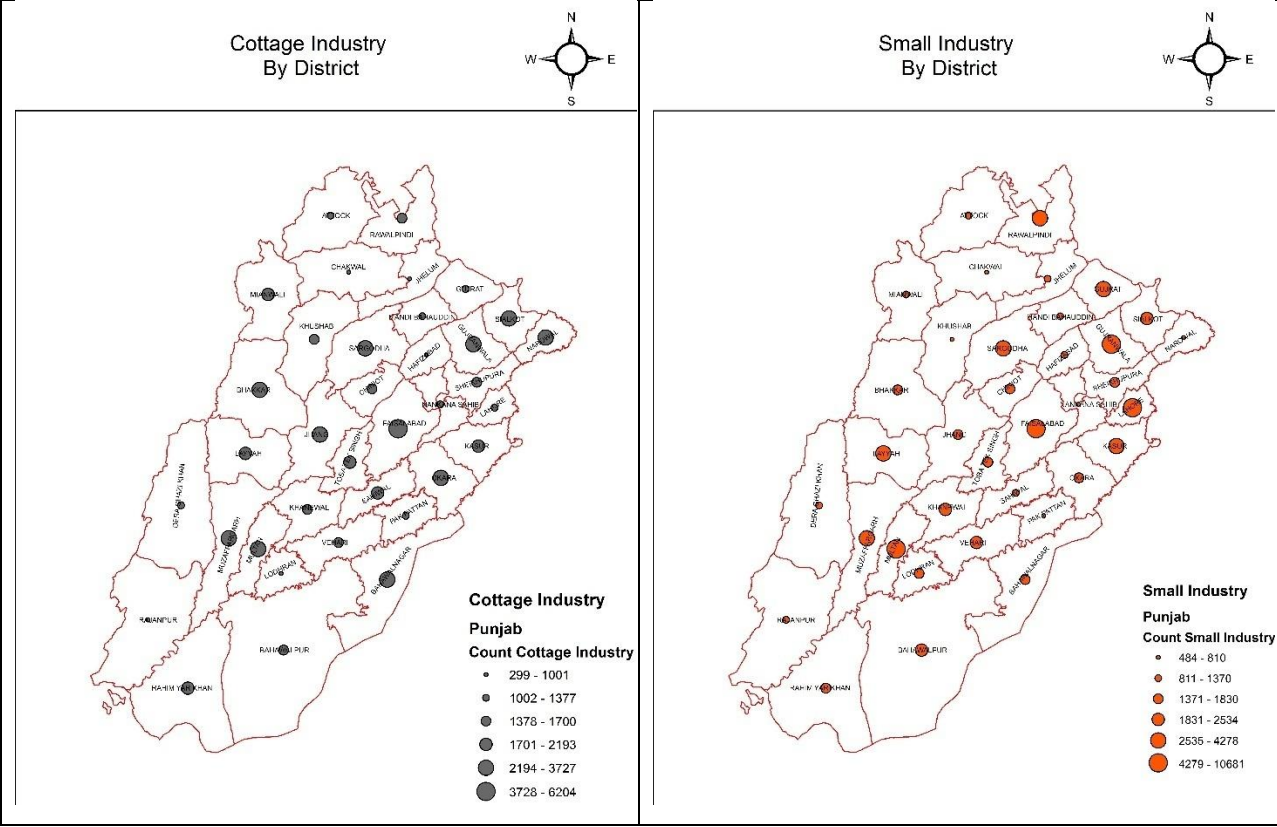


Figure 3: Log TFP Distribution for Small, Cottage Industry

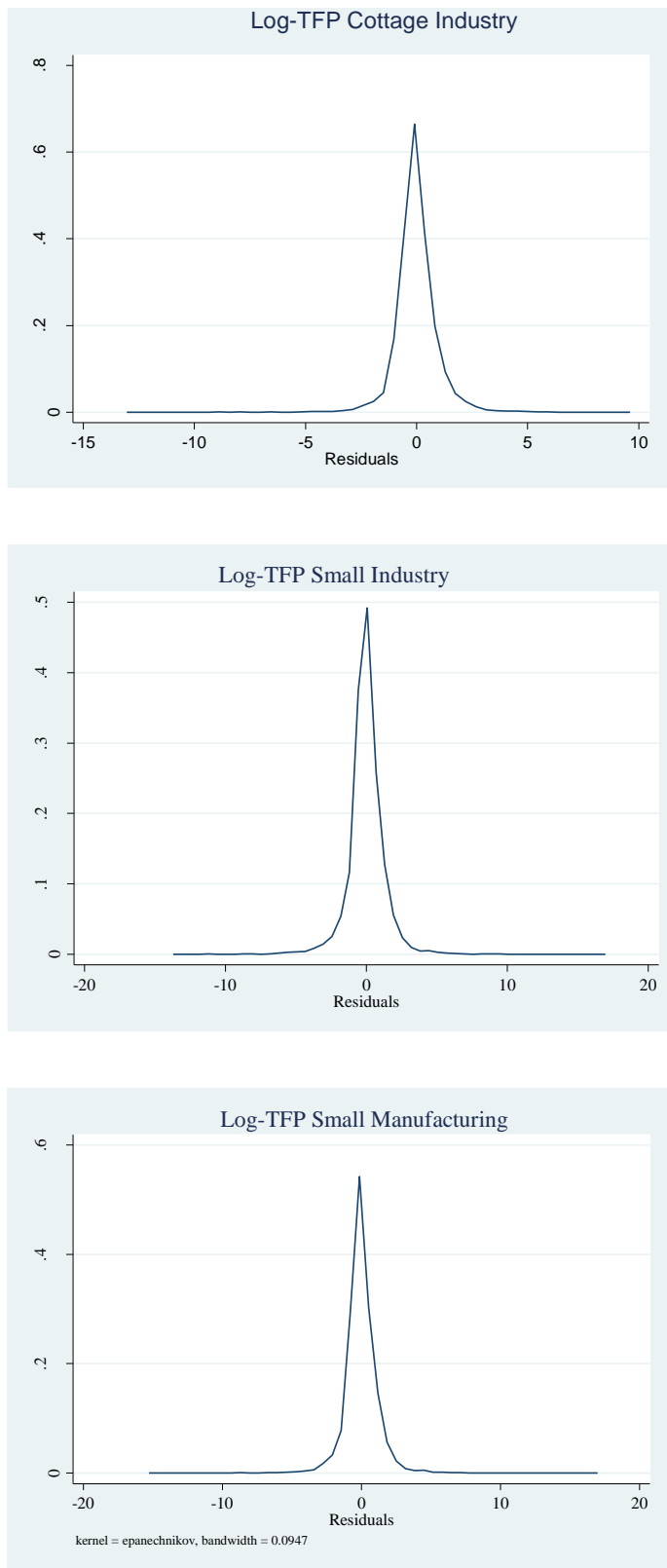


Figure 4: Top Revenue Generating Industries Across Districts in Punjab

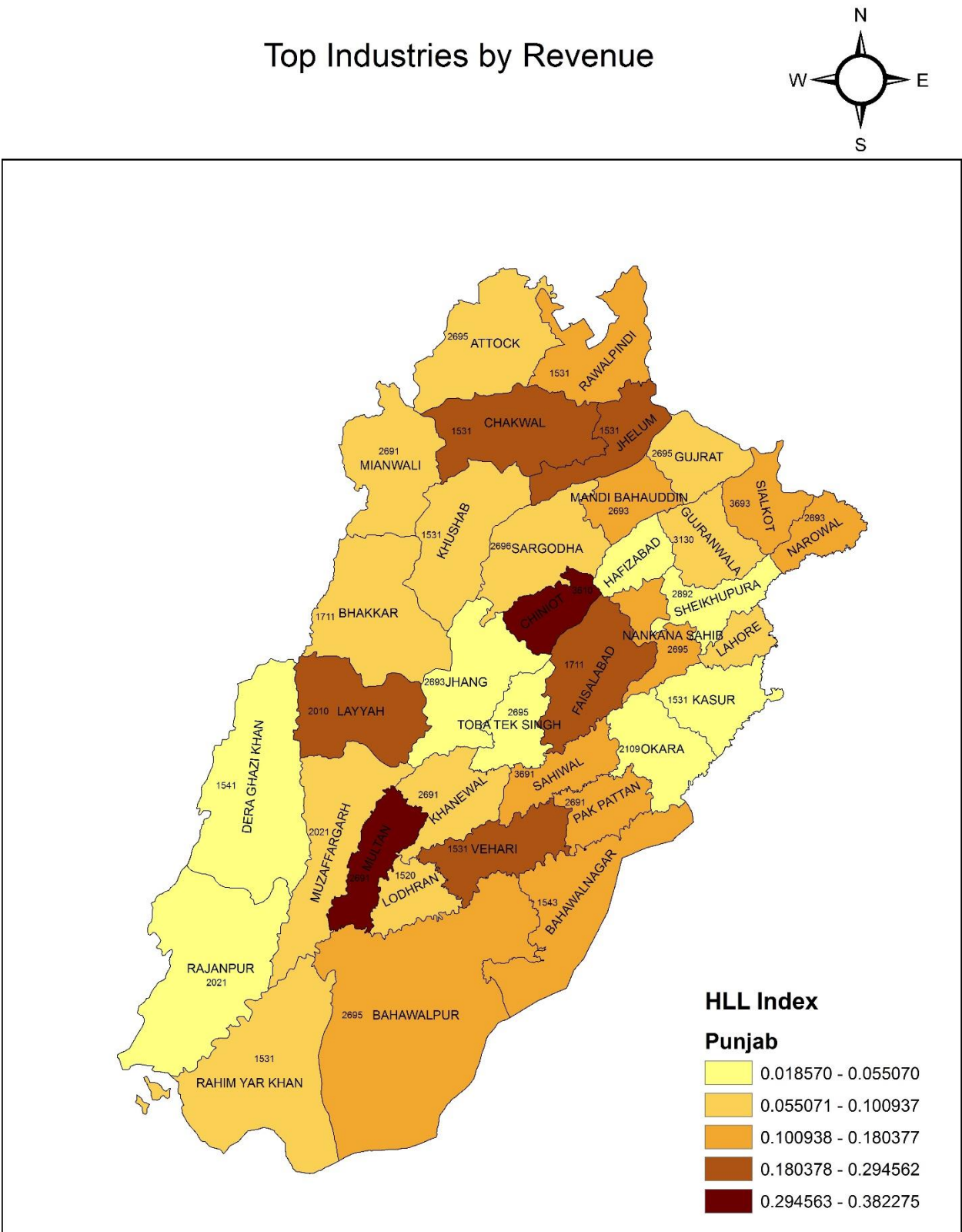


Figure 5: District-wise Maximum TFP Level and TFP Distribution Dispersion

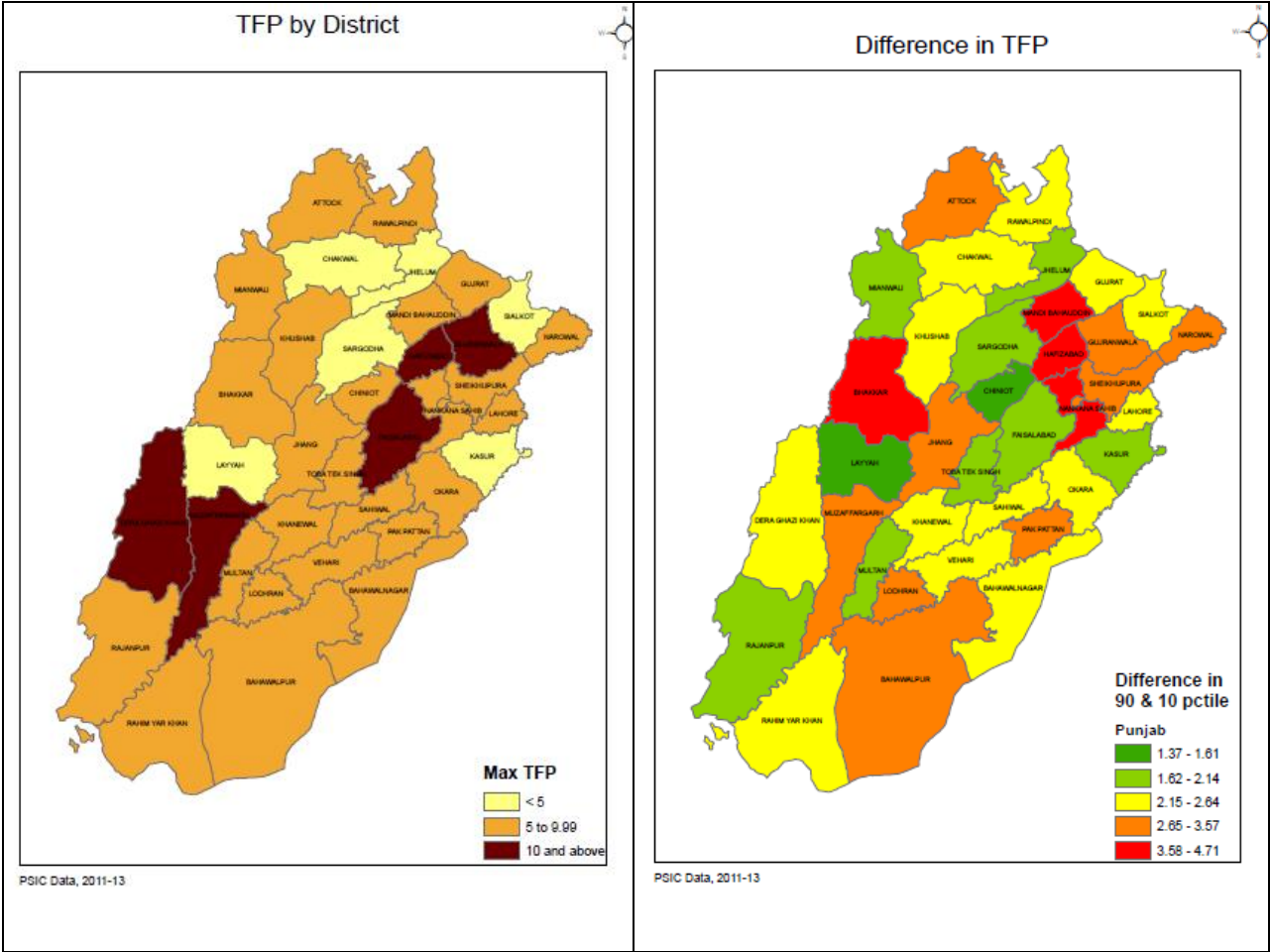


Figure 6: District-wise Maximum TFP Level and TFP Distribution Dispersion for Cottage Industry

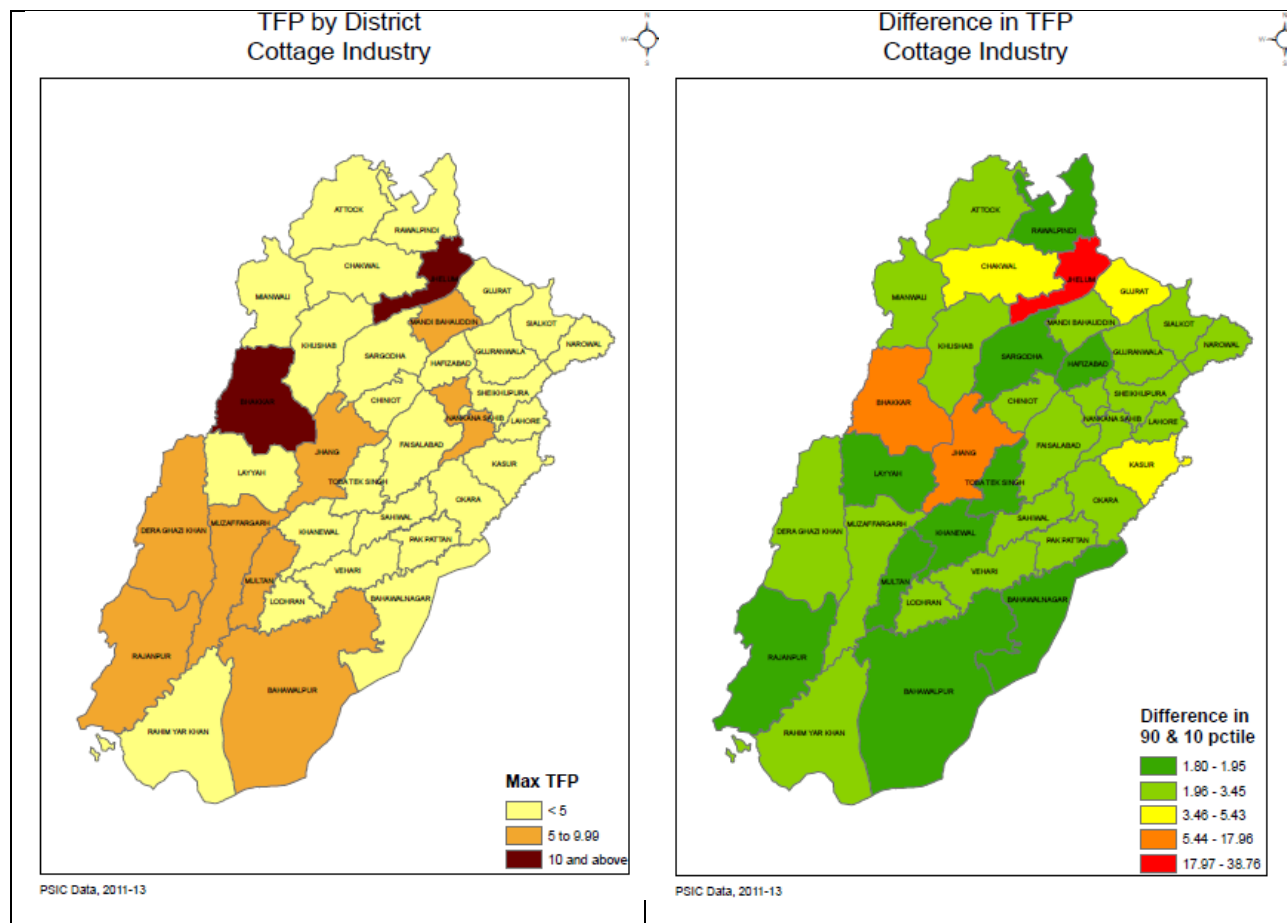


Figure 7: District-wise Maximum TFP Level and TFP Distribution Dispersion for Small Industry

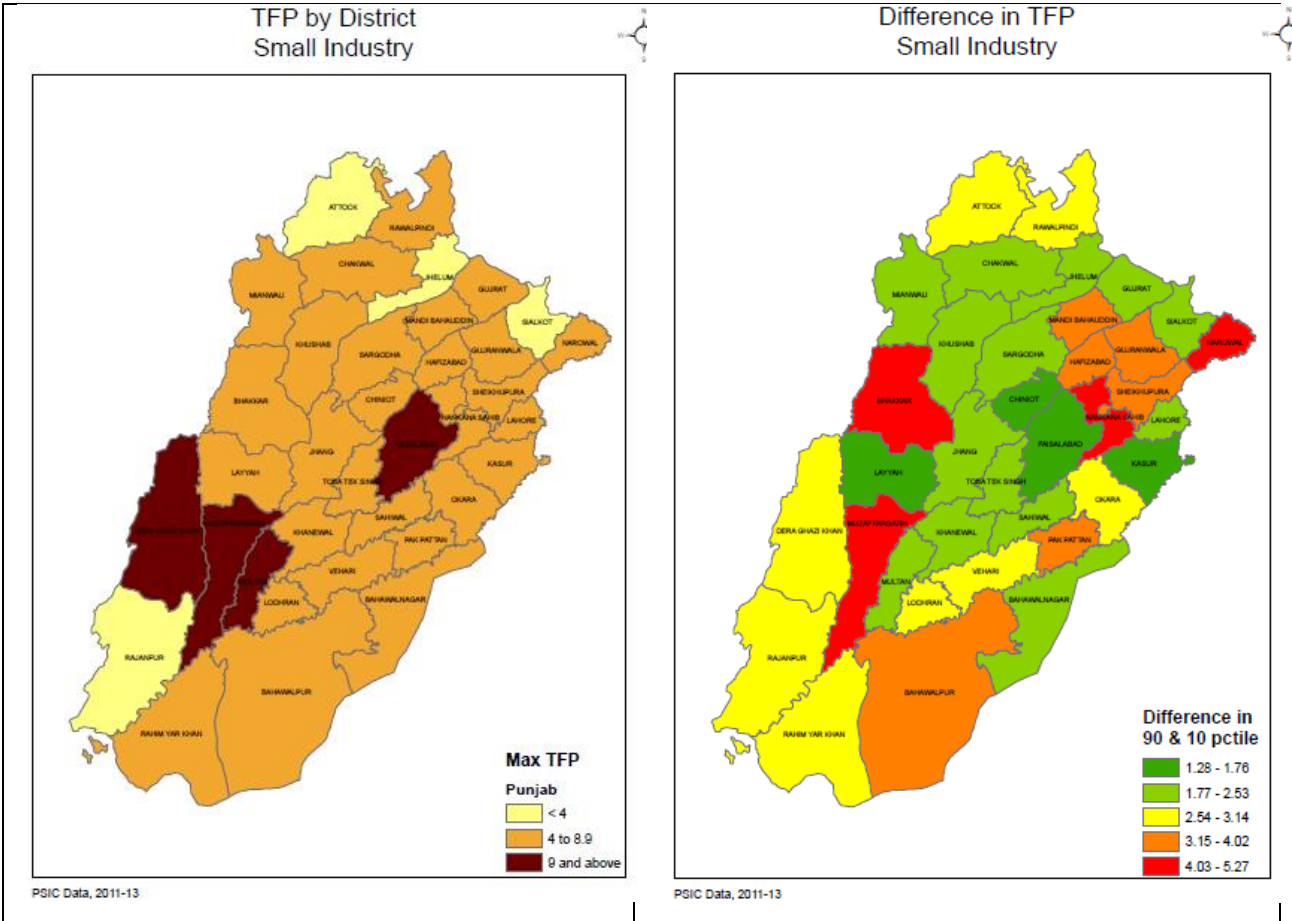
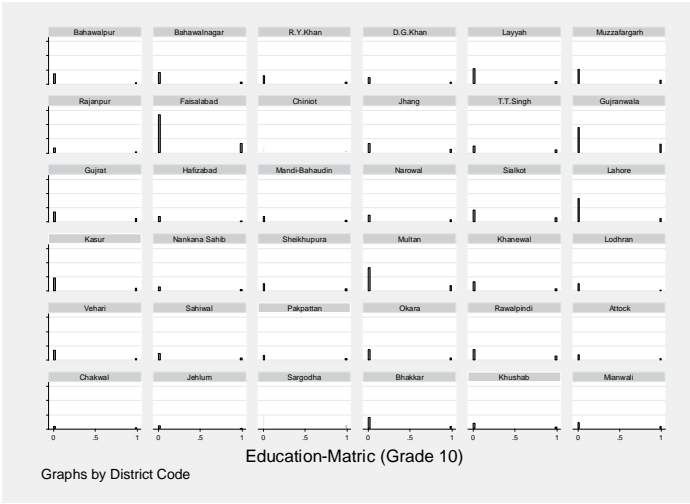
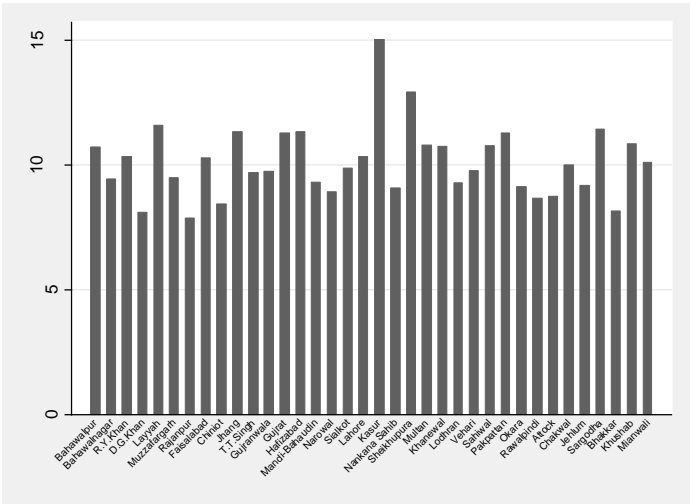


Figure 8: District -wise plots of Firm Age and Entrepreneur Education



Section B: Data Entry Component

1. Introduction

1.1 Description of Census of Small and Cottage Industry in Punjab: 2011-2013

Punjab Small Industries Corporation (PSIC), in collaboration with the Bureau of Statistics (BOS), Punjab, conducted a census of small and cottage industries in Punjab during 2011-2013 to identify industrial clusters and gather data on availability of raw material, production, manpower, technical and financial status and skills for the small and cottage industry across Punjab. The census was based on the Census Frame 1998, and covered all 40,680 clusters (including 25,914 rural clusters and 14,766 urban clusters) in Punjab. The following two paragraphs describe the two stages in which the census was conducted.

At the first stage, a cluster-level listing exercise was performed under which each structure in Punjab was categorized as either a dwelling unit (DU) or a non-dwelling unit (NDU). This exercise separated structures that were being used only for residential purposes (i.e., DU) from those where other activities (economic as well as non-economic) were being undertaken (i.e., NDU). Depending on the nature of activity that was being undertaken in each NDU, all NDUs were further categorised into four categories, namely manufacturing¹, trade, personal services and others. The information generated from this exercise was recorded on a listing form (i.e., per forma), where each listing form captured information on the location of the cluster² (i.e., cluster number, district name/ code, tehsil name/ code, etc.), the type of the structure (i.e., DU or NDU: manufacturing or trade or personal services or others) and the description of activity that was being undertaken in each NDU for all four categories of NDUs (i.e., manufacturing, trade, personal services and others). All this information, except for the description of activity, has been digitised by PSIC and is referred to as the cluster control sheet.

At the second stage³, a detailed survey was conducted just of the small and cottage manufacturing units, i.e., units not registered with the industries department. The detailed survey included firm level information on inputs, outputs, costs, profits, etc. This information was also digitised and made available by PSIC. However, for the other three categories of NDUs

¹ Apparently, the enumerator also checked if the manufacturing unit was registered with the industries department or not, and it seems that only unregistered units were shown in the manufacturing category. It is not clear whether registered manufacturing units were included in the others category or ignored. According to BOS, in the rural areas there were hardly any registered units, but if they came across any it would have been included in the manufacturing category. However, a quick check showed that there was hardly any difference between the total number of manufacturing units in the Punjab in the cluster control sheets and the total number of small and cottage industry units in the detailed industry survey.

² In the rural areas, each village, or mauza (a revenue unit in the land records), is considered a cluster in the census frame.

³ The enumerator undertaking the listing exercise, revisited all NDUs that had been categorised as manufacturing for the detailed survey – the enumerator could revisit the unit immediately after completing the assigned listing exercise or return the next day for the detailed survey.

(i.e., trade, personal services and others) no detailed survey was conducted and the only information available on them is in the description of the activity that is in the cluster control sheet.

1.2 Motivation for the Current Project

There have been a number of surveys, prior to this census, of the informal (small-scale and cottage) manufacturing sector in the Punjab, but there is almost no micro-level information on other economic activities and businesses particularly in the rural areas of the Punjab. This is an important gap in our information about the rural economy as up to 60 percent of the rural labor force may be employed in non-agricultural activities. It was felt that by digitising the information in the “description of economic activity” being undertaken in three categories of NDUs, namely trade, personal services and others, IGC could make a start in filling this gap. In addition to an analysis of the types of economic activities, and the relative importance of the various economic activities, taking place at the village level, the data set can be used to provide a framework for conducting more detailed surveys of specific business activities in rural Punjab. It can also be linked with other village level data sets, such as the Pakistan Mauza Census 2008, compiled by the Agricultural Census Organization, and the enhanced data set could be useful for researchers’ analysing the determinants of social and economic outcomes, as well as an input into the Government’s regional infrastructure investment planning process.

2. Project Implementation

2.1 Challenges Faced during Data Entry and how they were addressed

The two main challenges faced in implementing of the project were: 1) Completing the data entry in within the limited time and budget; 2) Ensuring the quality of the data entered. In the census frame there are 40,680 clusters in the Punjab of which 25,914 are rural clusters and 14,766 are urban clusters. On average, in the listing form for each cluster there were about 70 entries relating to economic activities, which meant that over 2 million records of economic activities had to be entered. As it was impossible to do this within the budget allocated, it was decided to enter data for only the rural clusters and to do it in-house instead of giving it out to a commercial data entry firm. For this purpose, a Project Coordinator (PC) was hired and a team of student RAs were recruited to enter the data under the direct supervision of PC and the PIs. It was felt that in-house data entry would also provide the project team with greater control over quality of data entry given that an element of subjectivity cannot be completely avoided in the process of interpreting the descriptions of economic activities recorded in the listing form and assigning ISIC codes to them.

2.1.1 Data Entry Management System and Quality Assurance

In order to minimise subjectivity, a macro-enabled excel-based data entry system was designed⁴ that automated much of the data entry operations⁵. In addition, three Stata programmes were written for improved data entry management and quality assurance:

1. Programme Number 1: This flagged and helped to remove RA-level duplicates and generate RA-level entry summary statistics for benchmarking.
2. Programme Number 2: It appended individual RA entry files into single master file at the end of each day and generated timeline projections for entry of remaining clusters to improve project management.
3. Programme Number 3: This helped to ensure that data entry error was consistently kept below 5 percent⁶. It calculated entry error rate for each batch entered by a RA and if the error rate exceeded 5 percent for any batch the RA concerned was required to re-enter the entire batch.

With strict adherence to the protocols listed above (and described in greater detail in the Annexure 1) data for 24,210 rural clusters in Punjab was successfully entered. The final data set includes 1,404,135 entries and took just over a year to complete.

3. Project Output

3.1 Description of the Data Set

The data set is available in its wide form, but can also be converted into its long form⁷. In the former case, each row represents a cluster and each column represents an economic activity in that cluster.

Table 1: Description (and coverage) of the Data Set

Total Observations	The data set (in its wide form) has 24,210 observations, where each observation represents one cluster.
Number of Villages (Clusters or Mauzas)	The data set covers 24,210 villages (clusters or mauzas) in rural Punjab.
Number of District	The data set covers 36 districts of the Punjab.
Number of Tehsils	The data set covers 135 tehsils of the Punjab.
Activities (economic and non-economic)	There are 352 unique activities identified in this data set of which 280 have been classified as non-agricultural economic activities and 72 as non-economic activities ⁸ .

⁴ We would like to thank Muhammad U. Ghaus for his excellent support in designing the data entry system.

⁵ Please refer to the annexure to this report for an overview of the data entry system that was designed.

⁶ Please refer to the annexure to this report for an overview of the data entry error rate protocol.

⁷ Since the size of the data set in its long form exceeds 1GB, it has been made available in its wide form. However, the researchers can reshape it, if necessary.

ISIC Code	<p>Each of the 280 economic activities have been assigned a 4-digit ISIC code. There are 92 distinct codes at the 4-digit ISIC level⁹.</p> <p>(ISIC Rev.3.1 has been used for this data set, originally accessed at http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=17 on November 5,2015.)</p>
-----------	---

4. Descriptive Data Analysis

Analysis based on this dataset is possible for Punjab as a whole and at different levels of disaggregation, from regions of the Punjab to the level of a tehsil, of which there are 135. In this section, the focus of analysis is the Punjab and its regions, with some references to the individual districts. For this purpose, Punjab is divided into four regions (North, Centre, South and West) using the classification adopted by Cheema, Khalid and Patnam (2008). There are significant differences between the four regions in terms of per capita income and poverty – North has on average the highest per capita income and the lowest poverty head count ratio in the Punjab, followed by Centre, then South and finally West, which has on average the lowest per capita income and the highest poverty. There are also important differences between the four regions in terms of the economy, agrarian structure, road and irrigation infrastructure, human development, etc. which are discussed below in brief, as they provide the context and help in understanding the variations in the observed patterns in the economic activity in the villages in the four regions.

North Punjab, has largely barani (rain-based) agriculture with a relatively small portion of its land being cultivable. In the colonial period, it was a major recruiting ground for the army and continues to be even so today. As a result, many of its people are employed outside the region and remittances, salaries and pensions are the major source of income. Possibly because of this connection, North today has the lowest level of poverty and the best human development indicators of all the regions in the Punjab. Centre Punjab was sparsely populated until the British started developing the canal irrigation system in the Punjab in the 1890s. Over the next 40 years, millions of acres of land was provided with canal irrigation and brought under cultivation by settlers – 80 percent of this land was allotted to peasant cultivators (40-50 acres), making Centre Punjab an region of predominantly small land holders. As Centre Punjab also had a large sweet water aquifer it benefited the most from the green revolution in the 1960s and private investment in tubewell irrigation, with its accompanying increase in cropping intensity. It is also the most industrialised region in the Punjab, with Faisalabad, Lahore and

⁸Please refer to the Annexure to this report for the list of 72 non-economic activities that were dropped from the total list of 352 activities for the purpose of analysis in this report.

⁹As a number of economic activities can fall in the same 4-digit ISIC code, the 280 economic activities described in the listing forms mapped to 92 economic activities at the 4-digit level. Please refer to the Annexure to this report for the list of 92 economic activities and their corresponding 4-digit ISIC codes.

districts to its north (Sheikhupura, Gujranwala, Sialkot and Gujrat) constituting the industrial heartland of Punjab.

South Punjab, which includes the old princely state of Bahawalpur (i.e., Bahawalpur, Bhawalnagar and Rahim Yar Khan Districts), also benefited from the development of canal irrigation by the British. However, as it was an old settled area with feudal and religious families (keepers of sufi shrines and *pirs*) controlling large chunks of land, the development of canal irrigation and British policies only consolidated the hold of the large land holding families. As a result, while South Punjab is a rich agricultural area with its prominent land owning families dominating the provincial and national legislatures, it has some of the worst human development indicators in the Punjab as well as very little industry. Finally, West Punjab is a combination of hilly (mostly tribal) districts and riverine (flood plains of the Indus) districts where livestock rearing and herding are major sources of income of the rural population along with agriculture, which is generally low-intensity cultivation based on rain in much of the region. Even in areas where there is canal irrigation, it is usually much more limited in nature (i.e., non-perennial or insufficient water supply) than in Centre and South Punjab and therefore not as intensive and productive as in those regions. West Punjab is probably the least developed region in the province in terms of roads and other physical infrastructure as well as human development indicators, which are generally worse than those of South Punjab.

A snapshot of the ***non-agricultural economic activities*** (hereafter called Economic Activities, or EAs) in the villages (clusters) in rural Punjab are presented at the regional and district level in Table 1. It is seen that there are over 24 thousand villages in the Punjab and on average there are 426 households (i.e., over 2,500 persons) in a village. The average size of a village, based on the number of households, varies from 377 households in the North to 462 households in the West. To see if these differences are symptomatic of the region, the average number of households per cluster in a district were mapped and there does not seem to be any pattern with regards to the regions (Figure 1). However, in the middle of the province there is a belt of districts (through West, South and Centre regions) where villages, on average, are relatively large. One can only conjecture as to why this is so, for example, the belt in the middle seems to be along the historical trade routes from Delhi and Lahore to Multan and on to Central Asia through Baluchistan and thus villages in these districts may be generally older inhabitations and therefore are today larger than in other districts. Outside this belt, there are only 3 districts (Attock in North, Mianwali in West and Mandi Bhauddin in Centre) in the Punjab where villages are on average large and, as none of these districts are part of some current major urban or industrial agglomeration, probably there are historical reasons for this.

Table 2: Non-Agricultural Economic Activity in Rural Punjab Database

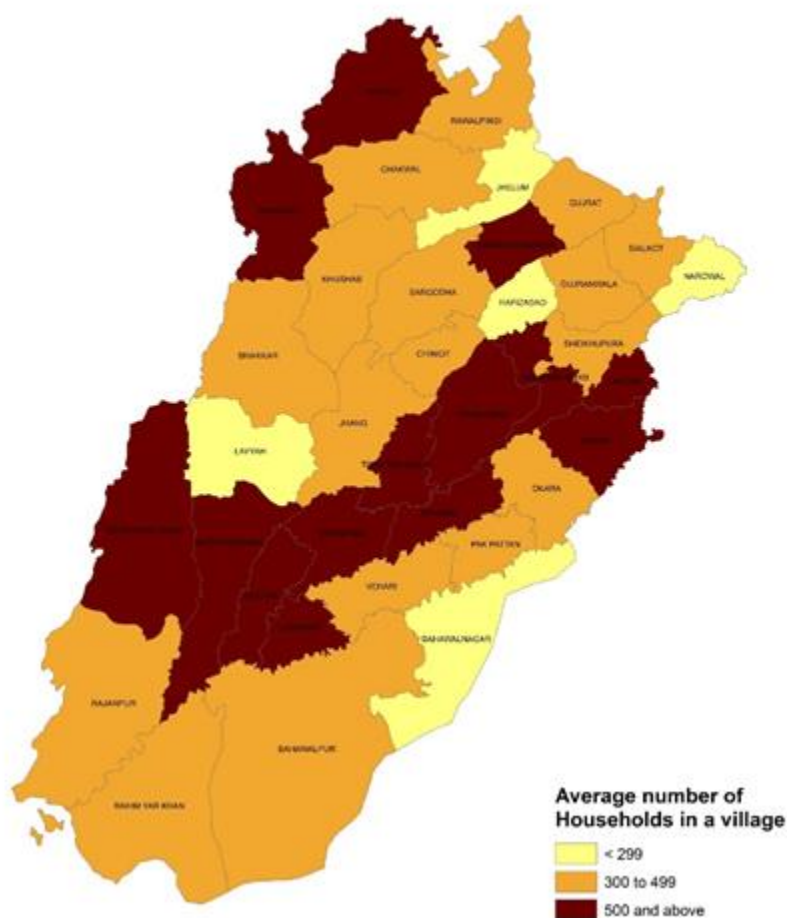
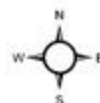
District/Region	Total Clusters	Average Per Cluster							
					NDUs: Non-Agricultural Economic Activity				NDUs
		Structures	House Holds	NDUs	Manufacturing	Trade	Personal Service	Others	Rest ¹⁰
Punjab	24049	430.84	426.43	61.89	3.83	19.95	7.54	4.26	26.31
-									
North	2309	394.36	377.34	35.96	2.28	15.06	5.90	3.54	9.18
Attock	394	526.31	518.15	46.76	3.28	17.02	9.55	4.63	12.29
Chakwal	403	554.47	495.53	40.72	2.84	18.64	5.15	4.07	10.01
Jehlum	516	267.80	246.30	25.52	1.84	8.80	2.96	2.21	9.70
Rawalpindi	996	342.95	341.71	35.17	1.87	16.09	6.28	3.59	7.34
Center	11961	435.75	428.13	70.40	3.99	19.64	6.98	4.43	35.35
Chiniot	353	361.59	428.60	55.12	4.64	15.49	6.65	3.50	24.84
Jhang	689	388.45	399.86	45.37	5.01	17.17	5.45	4.02	13.72
T.T.Singh	524	538.58	545.52	73.43	4.57	16.84	6.41	4.78	40.83
Gujranwala	731	482.15	377.99	116.16	5.78	27.15	11.17	6.71	65.36
Gujrat	991	370.13	316.57	56.35	2.00	13.25	4.18	2.68	34.24
Hafizabad	389	337.61	288.77	83.04	3.14	18.49	6.07	4.41	50.93
Faisalabad	783	790.43	779.19	133.01	9.24	37.17	15.90	8.20	62.50
Mandi-Bahaudin	394	537.82	501.02	73.78	4.79	21.86	7.64	4.40	35.08
Narowal	1147	221.95	166.32	46.09	2.46	9.63	2.91	2.37	28.72
Sialkot	1433	301.95	334.54	58.61	3.08	14.95	5.29	3.31	31.98
Lahore	223	709.43	767.41	106.00	3.50	39.06	20.30	7.79	35.35
Kasur	535	597.73	534.41	115.25	4.56	35.08	12.36	7.58	55.67
Nankana Sahib	388	427.72	611.84	64.62	2.84	18.07	4.81	3.97	34.94
Sheikhupura	592	420.91	383.68	86.20	3.26	20.97	5.98	4.90	51.08
Sahiwal	505	648.31	592.27	68.58	4.48	22.96	7.39	4.84	28.91
Sargodha	812	457.08	484.33	51.12	4.45	18.33	6.73	4.38	17.24
Pakpattan	568	428.40	455.10	50.19	3.01	16.91	4.91	3.48	21.87
Okara	904	365.93	393.49	48.27	2.91	17.12	4.93	3.74	19.58
West	4094	402.57	461.58	53.03	5.03	20.62	7.55	4.26	15.56
D.G.Khan	820	377.68	533.81	45.39	2.68	17.29	6.41	2.99	16.02
Layyah	711	303.30	298.97	42.80	6.46	13.37	5.51	3.92	13.55
Muzzafargarh	960	569.22	608.26	70.12	5.36	29.77	10.84	4.77	19.38

¹⁰ That is Agricultural and Non-Economic Activities, as well as those that were unidentified because of incomplete or illegible information.

Rajanpur	520	277.59	437.00	41.43	2.75	19.33	6.28	2.99	10.09
Bhakkar	539	374.89	338.05	57.28	6.69	18.57	7.47	4.97	19.57
Khushab	301	370.66	358.14	47.50	5.73	19.25	6.15	4.61	11.76
Mianwali	243	486.96	568.92	63.40	7.81	25.95	9.01	8.26	12.36

South	5685	455.69	417.46	60.92	3.26	22.11	9.39	4.19	21.97
Bahawalpur	891	458.78	402.13	60.01	2.75	21.04	8.73	4.08	23.42
Bahawalnagar	1081	330.15	282.32	47.44	3.22	16.11	6.01	3.55	18.55
R.Y.Khan	1337	371.23	330.36	43.35	1.83	18.03	6.30	3.15	14.04
Khanewal	662	607.75	524.04	63.27	3.86	25.27	9.61	5.10	19.43
Lodhran	446	517.41	543.58	78.39	3.93	28.28	12.83	4.41	28.94
Vehari	768	454.26	449.29	75.04	3.81	21.93	11.39	4.24	33.68
Multan	500	693.31	667.39	98.27	5.81	38.51	19.71	7.09	27.16

Figure 1:
Average Number of Households
per Village by District



A village in the Punjab, on average, has 430 structures, of which about 85 percent are dwellings and the rest have been labeled as NDUs in the PSIC survey. About 58 percent of the NDUs are being used for some economic activity. Based on the PSIC classification, trade accounts for about 56 percent, personal services about 21 percent and manufacturing and others account for 11 and 12 percent, respectively, of the economic activities in the village. As manufacturing has been discussed in detail in Section A of this report, analysis in this section is restricted to economic activities other than manufacturing. A breakdown of EAs in rural Punjab at the ISIC 4-digit level is presented in Table 2, and it shows both the diverse nature of economic activities being undertaken at the village level and the dominant position of a few activities. For example, retail sale of food and beverages (ISIC codes 5211 and 5220) is the most important activity, accounting for about 37% of EAs, followed by repair and maintenance services (ISIC codes 5260 and 5020), education (ISIC codes 8000 and 8010) and tailoring and related activities (ISIC code 1810) with about 8%, 7% and 5% of EAs, respectively. Thus out of the over 90 activities at the ISIC 4-digit level recorded in the survey, just 8 account for about 57% of all EAs in the village.

The data can also be used to see if there is any distinct pattern in the nature and extent of economic activities across districts and regions. The intensity of economic activity (IEA) in villages, defined as the number of EAs in a village per 100 households, is mapped at the district level in Figure 2. It is interesting to note that IEA in the North region, which is a relatively better-off region of the Punjab, is generally low. The reason probably is that, since remittances, salaries and pensions are the primary source of income and a large proportion of working age males are employed outside the village (either in the army, or on government and private sector jobs in the cities) only those activities that cater to the immediate needs of the inhabitants are undertaken in the village. Thus, one might say that villages in this region are well integrated in the national labor market and economic activities in the village, besides agriculture, are limited to retail and essential services.

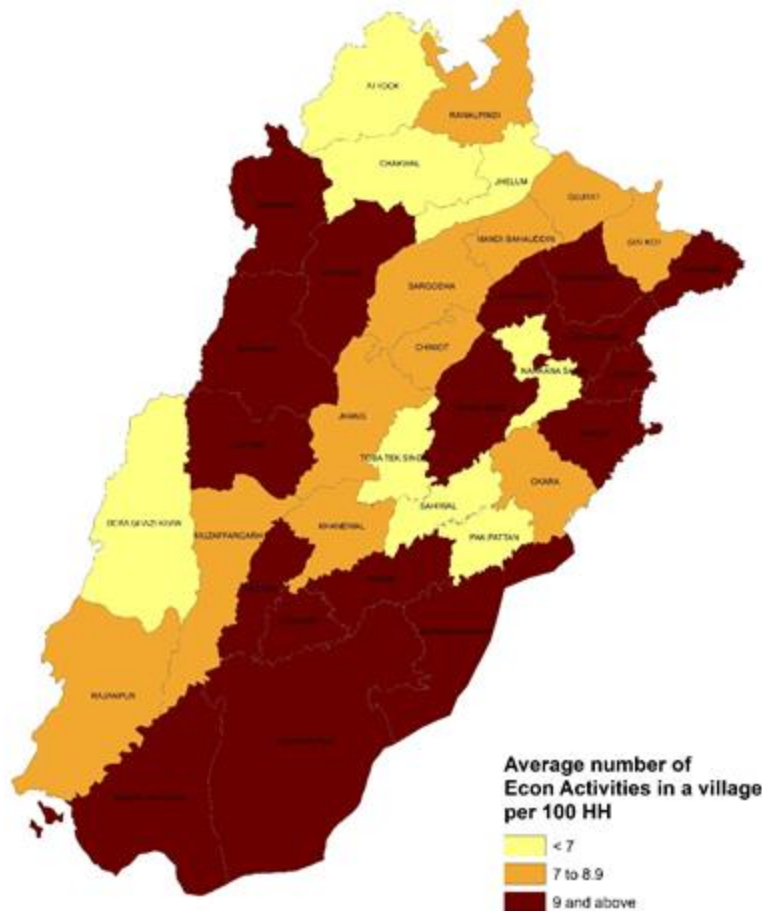
In contrast, IEA in villages is high in four out of the seven districts in the West region and in all but one district in the South region. As mentioned earlier, both regions have relatively weak road infrastructure and high levels of poverty, and it is likely that IEA in the rural areas is generally high in these two regions because of the greater need for self-sufficiency at the village level. The only district with relatively low IEA in these two regions is D.G. Khan, which may be because unlike the rest of the region it is dominated by Baloch tribes with possibly different attitudes towards engagement in non-agricultural activities. In the Centre region road infrastructure in rural areas is generally good and as result villages in region on average are much more integrated with the market than villages in South and West regions. There are three large cities in the Centre region, i.e., Lahore, Faisalabad and Gujranwala, and high IEA is observed in the rural areas in these and adjoining districts. It is likely, that the high IEA in these

Table 3: Most Important Non-Agricultural Economic Activities in Rural Punjab

ISIC Code	Description of ISIC Code	Percentage Share*	Rank
5211	Retail sale in non-specialized stores with food, beverages and tobacco predominating	32.5	1
1810	Manufacture of wearing apparel, except fur apparel	5.3	2
5220	Retail sale of food, beverages and tobacco in specialized stores	4.5	3
5260	Repair of personal and household goods	4.2	4
8000	Education (secondary and unidentified)	4.0	5
5020	Maintenance and repair of motor vehicles	3.9	6
8511	Hospital activities	3.4	7
5233	Retail sale of household appliances, articles and equipment	3.4	8
5520	Restaurants, bars and canteens	3.3	9
5219	Other retail sale in non-specialized stores	2.9	10
8010	Primary education	2.7	11
9302	Hairdressing and other beauty treatment	2.6	12
5121	Wholesale of agricultural raw materials and live animals	1.7	13
2892	Treatment and coating of metals and general mechanical engineering on a fee or contract	1.6	14
5050	Retail sale of automotive fuel	1.5	15
5200	Retail trade, except of motor vehicles and motorcycles	1.5	16
5131	Wholesale of textiles, clothing and footwear	1.4	17
5231	Retail sale of pharmaceutical and medical goods, cosmetic and toilet articles	1.4	18
5232	Retail sale of textiles, clothing, footwear and leather goods	1.3	19
5234	Retail sale of hardware, paints and glass	1.3	20
6302	Storage and warehousing	1.0	21

* Only those non-agricultural economic activities that contribute 1 percent or more of the total have been listed here – these collectively account for about 83 percent of the total.

Figure 2:
Average Intensity of Economic Activity
in a Village by District



districts is because a lot of economic activity in the villages here maybe catering to the market outside the village. In the rest of the districts in Centre, except for four (i.e., Nankana Sahib, Tob Tek Singh, Sahiwal and Pakpattan), middle level of IEA is observed. While there are certain patterns in IEA across regions, there is also ambiguity as high IEA can be the result of either the need for self-sufficiency at the village level, generally in a less developed region, or the result of production in the village for the larger market in a more developed and urbanised region.

It would be interesting to see if these patterns in IEA across regions are reflected in the type of EAs in those regions, and for that we look at the distribution of the most important EAs at the ISIC 4-digit level in each of the four regions. For purposes of discussion, these EAs have been grouped into 7 economic categories, which do not follow the PSIC classification of trade, personal services and others. It is seen that retail of food and beverages (including restaurants) is the most important, ranging from 36 percent of the EAs in the North to 42 percent in the Centre (Table 3). Next in terms of importance are other retail activities (which include, besides

general stores, specialised shops selling clothing, footwear, hardware, pharmaceutical and medical goods, etc.), which account for 15 percent of all EAs in the North and less than 8 percent in the West and South regions¹¹. This large difference between the two areas provides support for the argument above that the high IEA in West and South regions is largely because of the self-sufficiency of villages in these regions. Similar reasons may be behind the fact that in West and South regions “motor vehicles related activities” account for almost 50% more of the EAs than in North and Centre regions, i.e., just under 9% in the former against around 6% in the latter. In the more developed North and Centre regions there is a greater availability of public transport in rural areas as well as of automotive fuel, vehicle parts and repair services in a town nearby (relative to West and South regions) and therefore there is less of a need for undertaking these EAs within the village. It may also be noted that EAs related education have a larger share in North (8%) compared to West and South regions (6%).

The purpose of this descriptive data analysis section has been to illustrate the richness of the data available and to show that even at a very broad level it can yield interesting insights into the development of the rural economy of the Punjab. As discussed below, this data is being made available to researchers, who can use it separately or together with other available data sets for more in-depth analysis of the Punjab economy.

Table 4: Main Non-Agricultural Economic Activities by Region

	North	Center	West	South
Retail Food /Beverages	35.93	41.66	38.60	39.89
Retail sale in non-specialized stores with food, beverages or tobacco predominating	26.92	34.49	29.89	31.93
Retail sale of food, beverages and tobacco in specialized stores	5.94	4.52	4.37	4.27
Restaurants, bars and canteens	3.07	2.65	4.35	3.69
Retail Others	15.05	11.97	7.95	7.60
Retail sale in non-specialized stores	4.79	3.19	2.26	2.09
Retail sale of textiles, clothing, footwear and leather goods	2.28	1.39	0.96	1.10
Retail sale of hardware, paints and glass	1.98	1.39	0.97	1.08
Retail sale of pharmaceutical and medical goods, cosmetic and toilet articles	1.75	1.30	1.34	1.44
Other retail sale in specialized stores	1.58	3.19	0.89	0.80
Retail Unidentified	2.66	1.52	1.53	1.08
Household Goods and Home Appliances	6.65	7.27	7.83	8.29
Repair of personal and household goods	2.82	4.19	4.23	4.59
Retail sale of household appliances, articles and equipment	3.83	3.08	3.60	3.70

¹¹ Some retail activities, such as sale of household appliances, motor vehicle parts and accessories, automotive fuel, etc., which account for 5 to 6 percent of all EAs, are included in other categories in table 3.

<u>Miscellaneous Services</u>	13.27	10.15	10.89	12.58
Tailoring related activities	6.11	4.70	5.50	5.91
Hairdressing and other beauty treatment	4.69	2.57	2.73	2.00
Treatment and coating of metals; general mechanical engineering on a fee or contract basis	1.06	1.48	1.65	1.77
Embroidery related activities	0.33	0.56	0.28	2.05
Other business activities n.e.c.	1.07	0.84	0.73	0.84

<u>Social Services</u>	9.59	11.21	9.44	8.76
Education (secondary and unidentified)	4.85	4.52	3.53	3.09
Primary education	2.94	2.64	2.90	2.67
Hospital activities	1.80	4.05	3.01	3.00

<u>Wholesale</u>	4.57	4.23	7.49	5.15
Wholesale of textiles, clothing and footwear	1.78	1.34	1.56	1.37
Wholesale of agricultural raw materials and live animals	0.74	1.49	2.25	2.13
Storage and warehousing	1.93	0.79	1.55	0.93
Wholesale on a fee or contract basis	0.11	0.61	2.14	0.71

<u>Motor Vehicles Related Activities</u>	5.87	6.37	8.82	8.54
Maintenance and repair of motor vehicles	3.60	3.63	3.90	4.61
Retail sale of automotive fuel	0.87	1.19	2.20	1.79
Sale of motor vehicle parts and accessories	0.77	0.75	1.48	1.01
Sale, maintenance and repair of motorcycles and related parts and accessories	0.64	0.79	1.24	1.13

Activities not included above	9.07	7.13	8.97	9.20
--------------------------------------	-------------	-------------	-------------	-------------

5. Future Plans

5.1 How to Access the Data Set?

To the best of our knowledge, this is the first data set that gives a detailed breakdown of the economic activities that are being undertaken at the village level in rural Punjab. The data set is available on request at jgc.pakistan@theigc.org. Anyone wanting access to the data set should send in a request with his/ her name, title, organisation and a brief write-up on the proposed research, to the above address. Alternatively, they can also write directly to members of the research team on this project, at syed.hasan@lums.edu.pk, ijaz.nabi@gmail.com, navedhamid@gmail.com and usman.naeem@theigc.org, while copying jgc.pakistan@theigc.org on the email.

5.2 Merging the Data Set with Other Data Sets

This data set can be merged with other mauza level data sets, such as the Pakistan Mouza Census 2008, Population Census 1998, Agriculture Census 2010, etc. This will give researchers access to a much larger set of variables at the village (mauza) level and that can facilitate more rigorous research on, for example, determinants of rural labor markets and human development outcomes, than has been possible so far. Most of these data sets are available from Pakistan Bureau of Statistics. However, researchers can also write to the research team members at addresses listed above and they will try to facilitate access to these data sets.

Researchers can also assign cluster level Geographic Information System (GIS) coordinates to this data set for any spatial analysis. However, for this they will have to contact organisations such as The Urban Unit (<http://uu.urbanunit.gov.pk/index.aspx>), Technology for People Initiative (TPI) (<http://www.tpilums.org/>), etc., who have access to this information – the information might not be available free of charge.

Finally, it is also possible for researchers to extend this project by entering data on economic activities at the cluster level for small towns or cities in the Punjab. During data entry operations, listing forms for urban census clusters in small towns were packed in separate boxes from those for the rural clusters in the district. In case any researcher is interested in entering data on small towns, the IGC Pakistan office can facilitate contact with the relevant person at PSIC.

Annexure I

Designing the Data Entry System

First a selection of listing forms for several districts across Punjab was taken to prepare a comprehensive excel-based list of the description of economic activities that were being undertaken in these districts and to each of these activities assigned a 4-digit ISIC code. This list was then embedded as a drop-down menu into the data entry system so that once the data entry operations began an RA could simply select the description of an economic activity from the drop-down menu as opposed to manually entering it each time. This reduced data entry error rate significantly. In cases where the description of economic activity was not listed in the drop-down menu, the RA had the option to manually enter it into the system.

Maximum Data Entry Error-Rate Protocol

Each RA was provided with a basket in which he/ she was required to place his/ her entered forms at the end of the day. After the basket was full, 1-2 forms (about 200-300 entries) were randomly picked from the basket by the project manager and re-entered for double-checking. If the data entry error rate for these forms exceeded 5 percent, the RA had to re-enter all the forms in the basket. It may be noted that there were only four instances, throughout the data entry operations, when the entry error rate exceeded 5 percent and the basket had to be re-entered. Otherwise the entry error rate was consistently below 5 percent throughout.

List of ISIC Codes Assigned to Economic Activities

As stated above, 356 unique activities were identified in this data set, of which 280 were classified as economic activities. Each of the 280 economic activities were assigned an ISIC code at the 4-digit level. Economic activities in the data set that could not mapped to the 4-digit ISIC level, but could be mapped to the 2-digit ISIC were assigned a 4-digit number by adding two zeros at the end of the 2-digit ISIC code. For example, an activity which obviously belonged in the area of 'Construction' (i.e., 2-digit ISIC Code 45) was assigned a 4-digit ISIC Code 4500 for the purposes of aggregation and analysis. As a number of the economic activities described in the listing form could be mapped to the same 4-digit ISIC Code, this process resulted in 92 distinct economic activities at the 4-digit ISIC level. These are listed below:

ISIC Code	Description of ISIC Code
1422	Extraction of salt
1511	Production, processing and preserving of meat and meat products
1531	Manufacture of grain mill products
1533	Manufacture of prepared animal feeds
1541	Manufacture of bakery products

1542	Manufacture of sugar
1549	Manufacture of other food products n.e.c.
1712	Finishing of textiles
1729	Manufacture of other textiles n.e.c.
1810	Manufacture of wearing apparel, except fur apparel
2010	Sawmilling and planing of wood
2022	Manufacture of builders' carpentry and joinery
2212	Publishing of newspapers, journals and periodicals
2221	Printing
2222	Service activities related to printing
2520	Manufacture of plastics products
2892	Treatment and coating of metals; general mechanical engineering on a fee or contract basis
2893	Manufacture of cutlery, hand tools and general hardware
2927	Manufacture of weapons and ammunition
3110	Manufacture of electric motors, generators and transformers
3610	Manufacture of furniture
3691	Manufacture of jewellery and related articles
4020	Manufacture of gas; distribution of gaseous fuels through mains
4030	Steam and hot water supply
4500* ¹²	Construction
5010	Sale of motor vehicles
5020	Maintenance and repair of motor vehicles
5030	Sale of motor vehicle parts and accessories
5040	Sale, maintenance and repair of motorcycles and related parts and accessories
5050	Retail sale of automotive fuel
5100*	Wholesale trade and commission trade, except of motor vehicles and motorcycles
5110	Wholesale on a fee or contract basis
5121	Wholesale of agricultural raw materials and live animals
5122	Wholesale of food, beverages and tobacco
5131	Wholesale of textiles, clothing and footwear
5141	Wholesale of solid, liquid and gaseous fuels and related products
5142	Wholesale of metals and metal ores
5143	Wholesale of construction materials, hardware, plumbing and heating equipment and supplies
5200*	Retail trade, except of motor vehicles and motorcycles; repair of personal and household goods
5211	Retail sale in non-specialized stores with food, beverages or tobacco predominating
5219	Other retail sale in non-specialized stores
5220	Retail sale of food, beverages and tobacco in specialized stores
5231	Retail sale of pharmaceutical and medical goods, cosmetic and toilet articles

¹² *' Activities that fall under this 2-digit code, but could not be identified at the 4-digit ISIC level are included here - two zeros are added at the end of the 2-digit code to convert it into a 4-digit number for the purposes of consistency.

5232	Retail sale of textiles, clothing, footwear and leather goods
5233	Retail sale of household appliances, articles and equipment
5234	Retail sale of hardware, paints and glass
5239	Other retail sale in specialized stores
5252	Retail sale via stalls and markets
5260	Repair of personal and household goods
5510	Hotels; camping sites and other provision of short-stay accommodation
5520	Restaurants, bars and canteens
6000*	Land transport; transport via pipelines
6010	Transport via railways
6021	Other scheduled passenger land transport
6302	Storage and warehousing
6303	Other supporting transport activities
6304	Activities of travel agencies and tour operators; tourist assistance activities n.e.c.
6309	Activities of other transport agencies
6400*	Post and courier activities
6420	Telecommunications
6500*	Monetary intermediation
6519	Other monetary intermediation
6592	Other credit granting
6601	Life insurance
7010	Real estate activities with own or leased property
7100*	Renting of machinery and equipment without operator and of personal and household goods
7111	Renting of land transport equipment
7411	Legal activities
7421	Architectural and engineering activities and related technical consultancy
7494	Photographic activities
7499	Other business activities n.e.c.
8000*	Education
8010	Primary education
8021	General secondary education
8022	Technical and vocational secondary education
8030	Higher education
8511	Hospital activities
8512	Medical and dental practice activities
8519	Other human health activities
8520	Veterinary activities
9000	Sewage and refuse disposal, sanitation and similar activities
9199	Activities of other membership organizations n.e.c.
9211	Motion picture and video production and distribution
9212	Motion picture projection

9214	Dramatic arts, music and other arts activities
9220	News agency activities
9232	Museums activities and preservation of historic sites and buildings
9233	Botanical and zoological gardens and nature reserves activities
9241	Sporting activities
9249	Other recreational activities
9301	Washing and (dry-)cleaning of textile and fur products
9302	Hairdressing and other beauty treatment

List of Agriculture & Non-Economic Activities Not Included in the Analysis¹³

As stated above, out of the 352 unique activities identified in this data set only 280 were classified as non-agricultural economic activities. The remaining 72 activities described in the listing forms, and entered in the data set, were classified as non-economic activities. The 72 activities classified as agricultural or non-economic are listed below:

List of Activities not Included in the Analysis		
AafiyatKaNishan	Garden	Patwar Khana
Ablution Center	Government Office	PMLN office
Agri Farm	Grass Shop	Police Choki
Agri Service	Grave Yard	Population Welfare Office
Agri Sprayer	Hamam	Post Office
Army Barracks	Haweli	Poultry Farm
AwamiMarkaz	Imam Bargah	Public Admin
Bethak	Law Associate	Qadiyan iMasjid
Bhanah	Law Chamber	Railway Quarter
BISP Office	Library	Rescue1122
Cattle Field	Loam and Seed	Sewage GandaTallab
Chobara	Loam Spray	Sewerage Water
Church	Madrasa	Showroom
Community Service NGO	Markaz-e-AhleSunat	Stable
Cotton Factory	Market Plaza	Sugar Cane
Dairy Farm	Masjid	Sui Gas
Dara	Misc Others	Tenki House Water Tank
Darbar	Misc PS	Toka
Dars	Municipal Committee	Toll Plaza Chungi
Dera	Nursery	Tower

¹³ It may be noted that these activities have been only excluded for the purposes of analysis in this report. As far as the data set is concerned all 352 activities are included and properly labeled.

DSP office	Other Mills Factory	Tube well
Electricity House	Other Worship Place	Union Council
Fish Farm	Park	Vacant Shop
Forestry Department	Party secretariat	Water Treatment Plant

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about
our work on our website
www.theigc.org

For media or communications
enquiries, please contact
mail@theigc.org

Subscribe to our newsletter
and topic updates
www.theigc.org/newsletter

Follow us on Twitter
[@the_igc](https://twitter.com/the_igc)

Contact us
International Growth Centre,
London School of Economic
and Political Science,
Houghton Street,
London WC2A 2AE

IGC
International
Growth Centre

DIRECTED BY



FUNDED BY



Designed by soapbox.co.uk

Mapping Economic Activity in Rural Punjab

This policy brief is written by Attique ur Rehman (CDPR Researcher) and Usman Naeem (IGC Country Economist) and is based on the report "Development of an electronic database of Industrial and Commercial Activity and a Spatial Analysis of Small and Cottage Industries in Punjab, Pakistan" authored by Attique Ur Rehman, Dr. Ijaz Nabi (IGC), Dr. Naved Hamid (Lahore School of Economics), Dr. Syed M. Hasan (LUMS) and Usman Naeem. The research is funded by the International Growth Centre.

Background of the project

This project is based on the census of small and cottage industry in Punjab that was conducted by the Punjab Small Industries Corporation (PSIC) in 2011-2013. The purpose of the census was to identify industrial clusters and gather data on raw material, manpower, production, technical and financial status and skills for the small and cottage industry. The census was carried out across all 36 districts of Punjab covering both rural and urban areas, and was based on the Population Census 1998 frame. The urban frame consists of small, dense areas with well-defined boundaries referred to as Enumeration Blocks (EB's), where each EB on average holds 250 households. There are 14,766 EB's or urban clusters in Punjab. The rural frame consists of Mouza's (i.e., villages), which are the smallest revenue units recognised by the unique Hadbast number, within a Tehsil (i.e., an administrative sub-division of a district). There are a total of 25,914 Mouza's or rural clusters in Punjab.

The census was conducted in two stages:

At the first stage, a cluster-level listing exercise was performed under which each structure in Punjab was categorised as either a Dwelling Unit (DU) or a Non-Dwelling Unit (NDU). This exercise separated structures that were being used only for residential purposes (i.e., DU's) from those where other activities (economic as well as non-economic) were being undertaken (i.e., NDU). Depending on the nature of activity that was being undertaken in each NDU, all NDUs were

In brief

- PSIC conducted a census to identify Punjab's industrial clusters of small and cottage firms across 36 districts
- There is a data gap on non-manufacturing activity in categories such as 'trade', 'personal services', and 'others'.
- IGC digitized this data of non-manufacturing units in Punjab and conducted a spatial analysis.

further categorised into four categories, namely manufacturing, trade, personal services and others. The information generated from this exercise was recorded on a physical listing form (i.e., performa), where each form captured information on the location of the cluster (i.e., cluster number, district name/ code, tehsil name/ code, etc.), the type of the structure (i.e., DU or NDU: manufacturing or trade or personal services or others) and the description of activity that was being undertaken in each NDU. All this information, except for the description of activity for three categories of NDUs including trade, personal services and others, was later digitised by PSIC and made publicly available.

At the second stage, a detailed survey was conducted just of the small and cottage manufacturing units. The detailed survey included information on inputs, outputs, costs, profits, etc. This information was also digitised and made available by PSIC. However, for the

¹ PSIC is an affiliated organisation of the Industries, Commerce and Investment Department (IC&ID) with a mandate to "promote sustained industrial development through provision of market driven credit, infrastructure and technological support to contribute towards poverty alleviation through job creation and socioeconomic uplift of the province".

remaining three categories of NDU's no detailed survey was conducted either.

Objectives of the IGC for the project

Prior to this census, there have been several surveys of the small and cottage manufacturing industry in Punjab, but there is almost no micro-level information on other types of economic activities and businesses in the rural areas of Punjab. This is an important gap in the information about the rural economy of Punjab, especially when up to 60 percent of its rural labour force maybe employed in non-agricultural activities. Therefore, IGC felt that by digitising the information on the description of the type of economic activity being undertaken in non-manufacturing NDU's, namely trade, personal services and others, it would start to fill this gap. This data can be used for the following purposes. 1) Analysis of the types of economic activities and their relative importance at the village-level. 2) Provide a framework for conducting more detailed surveys of specific business activities in rural Punjab. 3) Link with other village-level data sets, such as the Pakistan Mouza Census 2008 (compiled by the Agricultural Census Organisation), where the enhanced data set with greater number of variables for analysis could be useful for researchers and policymakers.

Implementation of the project

In the census frame of 40,680 clusters in Punjab (14,766 urban and 25,914 rural), there were, on average, 70 economic activity-related entries in the listing form for each cluster, which meant that over 2 million records had to be entered. As this was impossible to do within the available project budget, it was decided to enter data only for the rural clusters and to do it in-house instead of giving it out to a commercial data entry firm. For this purpose, a team of student Research Assistants/ Associates (RA's) was recruited to enter the data under the direct

supervision of a Project Coordinator (PI) and the Principal Investigators (PI's) on the project. It was felt that in-house data entry would also give the project team greater control over quality of data given that an element of subjectivity cannot be completely removed in the process of interpreting the description of economic activities recorded on listing forms and assigning International Standard Industrial Classification of All Economic Activities (ISIC) codes to them.

In order to minimise subjectivity, a macro-enabled excel-based data entry system was designed that automated much of the data entry operations. In addition, three Stata programmes were written for improved data entry management and quality assurance:

- 1) Programme number 1: This flagged and helped to remove RA-level duplicates and generate RA-level entry summary statistics for benchmarking.
- 2) Programme number 2: It appended individual RA entry files into single master file at the end of each day and generated timeline projections for entry of remaining clusters to improve project management.
- 3) Programme number 3: This helped to ensure that data entry error rate was consistently kept below 5 percent. It calculated error rate for each batch entered by an RA and if the error rate exceeded 5 percent for any batch the RA concerned was required to re-enter the entire batch.

With strict adherence to the protocols listed above, data for 24,210 rural clusters in Punjab was successfully entered. The final data set includes 1,404,135 entries and took just over a year to complete.

The data set is available in its wide form, but can also be converted into its long form. In the former case, each row represents a cluster and each column represents an economic activity in that cluster.

Table 1: Description (and coverage) of the data set.

Total Observations	The data set (in its wide form) has 24,210 observations, where each observation represents one cluster.
Number of Villages (mouzas or clusters)	The data set covers 24,210 villages in rural Punjab.
Number of Districts	The data set covers 36 districts of Punjab.
Number of Tehsils	The data set covers 135 tehsils of Punjab.
Activities (economic and non-economic)	There are 352 unique activities identified in this data set of which 280 have been classified as non-agricultural economic activities and 72 as non-economic activities.
ISIC Code	Each of the 280 non-agricultural economic activities have been assigned a 4-digit ISIC code. There are 92 distinct codes at the 4-digit ISIC level. (ISIC Rev.3.1 has been used for this data set, originally accessed at http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=17 on November 5, 2015.)

The data set is available upon request at igc.pakistan@theigc.org. Anyone wanting access to the data set should send in a request with his/ her name, title, organisation and a brief write-up on the proposed research, to the above address.

Descriptive data analysis

Analysis based on this data set is possible at the level of Punjab and at different levels of disaggregation, from regions of Punjab to the

level of a tehsil. This project divided Punjab into four regions (North, Centre, South and West) using the classification adopted by Cheema, Khalid and Patnam (2008) to perform analysis that revealed significant differences between the four regions in terms of the type of economic activity being undertaken in their villages. These differences can be explained by historical differences in the economy, agrarian structure, road and irrigation infrastructure, human development, etc. between each of these regions, respectively.

Table 2: Snapshot of the non-agricultural economic activities at Punjab and regional level

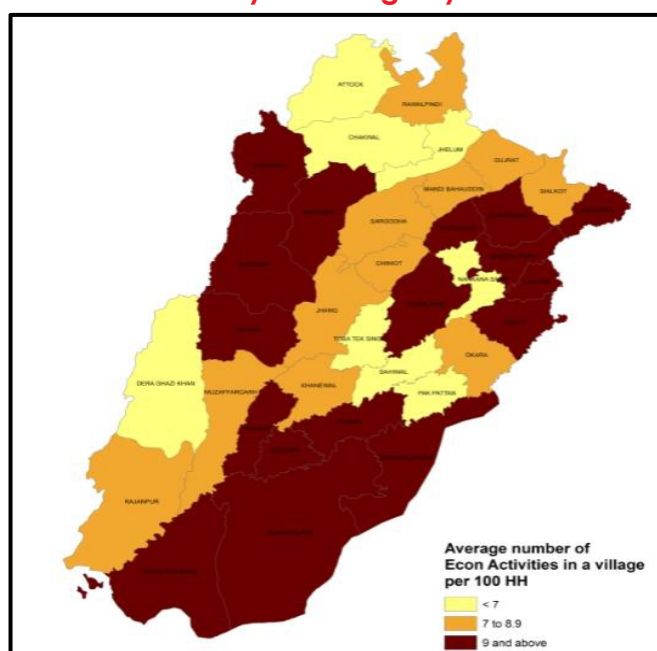
Region	Total clusters	Average per cluster							
		Structures	Households	NDU's	NDU's: non-agricultural economic activity				NDU's
					Manufacturing	Trade	Personal services	Others	
Punjab	24049	430.84	426.43	61.89	3.83	19.95	7.54	4.26	26.31
North	2309	394.36	377.34	35.96	2.28	15.06	5.9	3.54	9.18
Center	11961	435.75	428.13	70.4	3.99	19.64	6.98	4.43	35.35
West	4094	402.57	461.58	53.03	5.03	20.62	7.55	4.26	15.56
South	5685	455.69	417.46	60.92	3.26	22.11	9.39	4.19	21.97

* This includes agricultural and non-economic activities and those that were unidentified because of incomplete or illegible information

The data was also used to see if there is any distinct pattern in the nature and extent of non-agricultural economic activities across districts and regions. The intensity of economic activity (IEA) in villages, defined as the number of non-agricultural economic activities in a village per

100 households, mapped at the district level in Figure 1 shows that IEA is low in the North region, high in four out of the seven districts in the West region and in all but one district in the South region (with the exception of D. G. Khan) and it varies from high to medium in the Centre region.

Figure 1: Average intensity of non-agricultural economic activity in a village by district



Furthermore, to see if the above patterns in IEA across regions are also reflected in the type of non-agricultural economic activities in those regions, distribution of the most important non-agricultural economic activities at the 4-digit ISIC level in each of the four regions were also studied. For this, these non-agricultural economic activities were grouped into 7 categories, which do not follow the PSIC classification of trade, personal services and others. As table 3 shows, retail of food and beverages (including restaurants) is the most important, ranging from 36 percent of the economic activities in the North to 42 percent in the Centre. Next in terms of importance are other retail activities (which include, besides general stores, specialised shops selling clothing, footwear, hardware, pharmaceutical and medical goods, etc.), which ranges from being above 15 percent in the North region to below 8 percent in the West and South regions. Similar patterns are seen for the remaining categories.

Table 3: Main non-agricultural economic activities by region

Non-agricultural economic activities' categories	Region			
	North	Centre	West	South
Retail Food / Beverages	35.93	41.66	38.6	39.89
Retail Others	15.05	11.97	7.95	7.6
Household Goods and Home Appliances	6.65	7.27	7.83	8.29
Miscellaneous Services	13.27	10.15	10.89	12.58
Social Services	9.59	11.21	9.44	8.76
Wholesale	4.57	4.23	7.49	5.15
Motor Vehicles Related Activities	5.87	6.37	8.82	8.54
Activities not included above	9.07	7.13	8.97	9.2

Possible way forward

The data set created on this project can be used for multiple purposes:

- 1) This data set can be merged with other mouza level data sets, such as the Pakistan Mouza Census 2008, Population Census 1998, Agriculture Census 2010, etc. This will give researchers access to a much larger set of variables at the village (i.e., mouza) level and facilitate more rigorous research. Most of these other data sets are available from the Pakistan Bureau of Statistics, however, researchers can also request the project team at igc.pakistan@theigc.org who will try to facilitate access to these data sets.
- 2) Researchers can also assign Geographic Information System (GIS) coordinates to this

data set to perform any spatial analysis. For this, they will have to contact organisations such as The Urban Unit

(<http://uu.urbanunit.gov.pk/index.aspx>), Technology for People Initiative (TPI) (<http://www.tpilums.org/>), etc., who have access to this information – this information might not be available free of charge.

- 3) Finally, it is also possible for researchers to extend the data entry component of this project to small towns or cities in Punjab. At the time of data entry operations, listing forms for urban clusters in small towns were packed in separate boxes from those for rural clusters in districts. So, if a researcher is interested in entering this data, to make the data set more comprehensive, the IGC office can facilitate contact with the relevant person at PSIC.