c d p r

Consortium for
Development
Policy Research

Mapping Economic Activity in Rural Punjab

IGC

**International
Growth Centre**

About the project

Funded by: IGC

Key Counterpart: Punjab Small Industries Corporation (PSIC)

Impact
This study was developed in response to a direct government request by PSIC to focus on cottage and small sectors in Southern Punjab. The recommendations of this project, have now been taken forward to implementation in the form of a Credit Guarantee Scheme, captured in a confidential government document called a Planning Commission (PC) 1. Furthermore, in November 2017, the State Bank announced that it was setting up an electronic register of movable assets, another recommendation from this project.

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.

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.

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 NDU's were 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 NDU's 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 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	Rest*
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
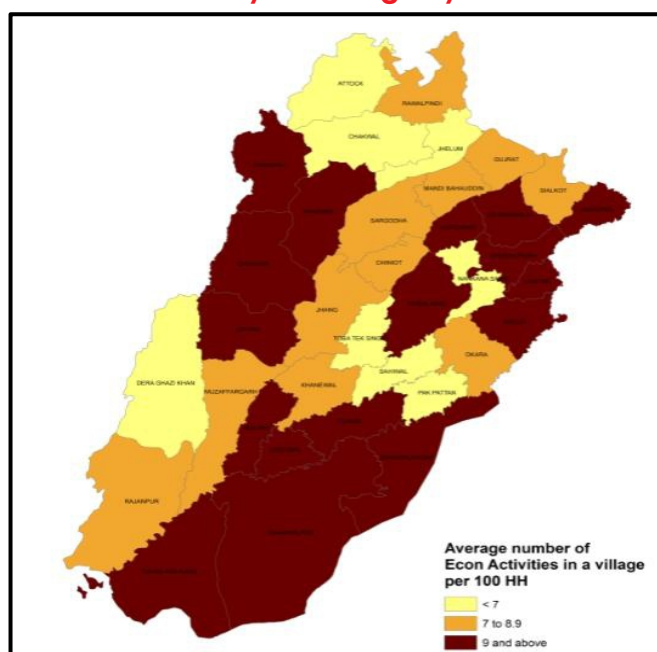
Furthermore, to see if the above patterns in IEA across regions are also reflected in the type of non-agricultural economic activities in those regions, distribution of the most important non-agricultural economic activities at the 4-digit ISIC level in each of the four regions were also studied. For 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.