RASTA

LOCAL RESEARCH
LOCAL SOLUTIONS

Volume VII

GROWTH & TAXATION

RASTA | PIDE

Pakistan Institute of Development Economics

Edited by Nadeem Ul Haque and Faheem Jehangir Khan

RASTA: LOCAL RESEARCH LOCAL SOLUTIONS

GROWTH & TAXATION
(Volume VII)

Edited by Nadeem Ul Haque and Faheem Jehangir Khan

RASTA

Research for Social Transformation & Advancement
Pakistan Institute of Development Economics, Islamabad

Published by Research for Social Transformation & Advancement (RASTA),

Competitive Grants Programme for Policy-oriented Research

at the Pakistan Institute of Development Economics (PIDE),

Islamabad, PAKISTAN.

Tel. +92 (51) 9248144, 9248026

Email: rasta@pide.org.pk

URL: rasta.pide.org.pk

March 2024

TABLE OF CONTENTS

PART I - GROWTH & TAXATION: *Research Papers*

PART II - GROWTH & TAXATION: *Policy Briefs*

PART I

GROWTH & TAXATION

Research Papers

ESTIMATING THE DISTRIBUTIONAL BURDEN OF GENERAL SALES TAX IN PAKISTAN

Iffat Ara

ABSTRACT

Pakistan's tax regime heavily relies on indirect taxes constituting 60 per cent of total tax receipts. The study assessed who bears how much burden of these taxes and examined the extent of overall incidence and the distributional burden of general sales tax (GST) levied on domestic production and sales across deciles of household expenditures for the year 2018-19. The study used an input-output model-based approach as it allows for tracing the cascading effect of indirect taxes. Hence, even if the final product is exempted from tax, it incorporates the impact of taxes levied on intermediate inputs it uses. The incidence of GST was investigated from various perspectives, i.e., for the whole country and by rural-urban areas and by commodity groups for the whole country. The results show that the overall incidence of GST was, on average, 6.7 per cent in Pakistan. The distribution of incidence of GST was found to be regressive across the board as well as in rural and urban areas. Analysis by commodity groups indicated that basic food items bore the highest magnitude of incidence and displayed the highest extent of regressivity across all deciles. This regressivity suggests that the poorer segments of society bear a relatively greater burden of GST in Pakistan.

1. INTRODUCTION [1]

The literature suggests that the burden of indirect taxes often is not evenly distributed. Who bears a higher and who bears a lower burden in proportion to their income depends on the design of the tax regime. Investigating who actually bears the burden of a tax requires a study of the incidence of taxation across different tiers of economic groups. This research is an attempt to investigate the incidence of indirect taxes in Pakistan.

Since indirect taxes are levied on goods and services that are (ultimately) consumed, they can be shifted forward (to consumers). Hence, they place an economic burden on taxpayers. Incidence analyses generally focus on economic incidence as it tells who bears the final burden of taxes that are shifted forward. A tax is progressive if the final tax burden as a percentage of income is higher on high-income individuals relative to low-income individuals, regressive if it is higher on low-income individuals relative to high-income individuals, and it is proportional if the burden is the same percentage on all individuals relative to their income.

A progressive tax is considered to be equitable because those with a greater ability to pay would pay a higher proportion of their income in the form of taxation. However, a proportional tax may also be viewed as equitable to the extent that all taxpayers would pay the same proportion of their income as tax. Consequently, higher-income taxpayers would be paying a higher absolute amount of tax than lower-income taxpayers (Jamal and Javed, 2013).

The structure of federal taxes in Pakistan heavily relies on indirect taxes which constituted 62 per cent of total tax receipts in 2018-19, whereas direct taxes constituted 38 per cent (Table 1). Of the 62 per cent of indirect taxes, general sales tax (GST) dominated with a share of 38 per cent, whereas customs duties (CD) and federal excise duty (FED) constituted 18 per cent and 6 per cent, respectively.

Table 1: Composition of Federal Taxes (% Share)

Tax Head	2000-01	2004-05	2009-10	2014-15	2018-19
A. Direct Taxes	31.8	31.1	39.6	39.9	37.8
B. Indirect Taxes	68.2	68.9	60.4	60.1	62.2
General Sales Tax (GST)	39.1	40.4	38.9	42.0	38.1
Customs Duty (CD)	16.6	19.5	12.1	11.8	17.9
Federal Excise Duty (FED)	12.5	9	9.4	6.3	6.2
Total (A+B)	100	100	100	100	100.0

Source: Federal Board of Revenue (FBR) Annual Report, various issues.

Among direct taxes, an important source is the withholding tax (WHT), which is deducted and collected at the source. This mechanism is considered effective as it is a timely source of revenue. At present, important WHT provisions having varying tax rates comprise contracts, imports, salary, bank interest, dividends, exports, telephone and electricity bills, technical fees, commission and brokerage, utilities, vehicle tax, cash withdrawal, and stock exchange-related provisions among others. Of different heads, the WHT on contracts, imports, telephone and electricity bills, technical fees, exports, cash withdrawals, and advance tax on banking transactions, together constitute over 60 per cent of the total taxes collected [2] and are indirect as these can be shifted forward

[1] The research is part of author's on-going Ph.D. research.

[2] Source: Year Book 2019-20, Federal Board of Revenue, the Government of Pakistan.

to consumers.[3] The indirect nature of these taxes creates doubts about their progressivity and makes them candidates for rigorous incidence analysis.

The literature on the incidence of taxation advocates that direct taxes (such as income tax) impose a relatively greater burden on the richer segments and, hence, are generally considered progressive. On the other hand, indirect taxes (such as taxes levied on goods and services) impose a relatively greater burden on the poorer segments of society as a large part of the income of the poor is spent on consumption, particularly food, hence, are generally considered regressive.

Among different studies done on Pakistan, Jamal and Javed (2013) estimated the incidence of GST, Wahid and Wallace (2008) estimated the incidence of all taxes, whereas Refaqat (2008) and SPDC (2004) estimated the incidence of all the federal indirect taxes. From a methodological perspective, the results of these studies (in terms of tax progressivity) were based on an average rate of progression as they compared the average tax rate across different income groups.

The major limitation of Refaqat (2008) and Jamal and Javed (2013) was that while estimating the incidence of the GST, they considered taxes levied on final consumption only and did not incorporate taxes levied on intermediate inputs used in the production of the final output, which constitutes a substantial part of total tax revenue. Furthermore, since they did not consider the taxes on intermediate inputs, they did not account for the items that are exempted from tax in their analysis. It is argued that even if the final output is exempted from tax, its price includes an implicit tax which is transferred through taxes levied on inputs that were used to produce it. Estimating the incidence of indirect taxes without capturing the impact of taxes on inputs is likely to produce misleading results.

Though Wahid and Wallace (2008) and SPDC (2004) accommodated the taxes levied on intermediate inputs, they did not estimate the incidence at a disaggregated level, i.e. by considering different consumption items such as food, utilities, etc. Analysing the distribution of incidence by item or commodity group helps understand the tax burden according to the consumption patterns of the poor and rich.

The present study estimates the incidence of GST, which is levied on domestic production and sales, in Pakistan for the year 2018-19 by taking into account the limitations of the studies discussed in the preceding paragraphs. That is, it takes into account the cascading effect of indirect taxes by using an input-output (IO) table that captures the effect of input taxes on final consumers. In addition, it looks at the distributional burden of indirect taxes by commodity groups that households consume. Since the study is limited to examining the incidence of taxes, carrying out a gap analysis of tax collection, analysing the issues of compliance and further taxation are beyond the scope of the study.

The study is organised as follows. Section 2 spells out the objective of the research. Section 3 presents a review of recent research on the subject, while Section 4 lays out the methodology used to estimate the incidence of GST and its distribution. Section 5 presents the estimation results and their explanation. Finally, Section 6 concludes the discussion and presents recommendations.

2. OBJECTIVES OF RESEARCH

This study assesses the incidence of general sales tax (GST) levied at domestic production and sales and its distributional burden in Pakistan across deciles of households for the year 2018-19.

[3] The provision of direct WHTs comprise salary, bank's interest, dividend, commission, brokerage, vehicle tax, stock exchange-related provisions, etc., having varying tax rates.

Specifically, the study:

- Examines the extent to which domestic GST in Pakistan can be considered as progressive (i.e., placing a higher tax burden on higher income groups), regressive (i.e., placing a higher tax burden on lower income groups), or proportional (i.e., placing the same tax burden on each income group).
- Analyses this by capturing implicit taxes as well, i.e., taxes paid on intermediate inputs, by using the latest available IO table for the year 2010-11.
- Estimates the incidence of domestic GST and its distribution for various commodity groups that households consume.
- Estimates the distributional burden of indirect components of the WHT. However, this estimation is possible for those heads only where data at the household level are available.

3. REVIEW OF LITERATURE

Numerous studies have examined the distribution of incidence of taxes or the distributional burden of taxation, both nationally and internationally. The majority of the studies have employed the conventional approach, i.e., an average rate of progression. This approach uses a priori assumptions from economic theory to ascertain who bears the final burden of taxes (i.e., the tax shifting assumptions) and employs household survey data to compute effective tax rates (ETR) for each household by dividing the tax liability of a household by its total income/expenditures. The ETR is then compared across households based on welfare scale (consumption or income). A tax structure is said to be progressive when the ETR rises along with the rise in the scale of individual/household income or expenditure. On the other hand, it is considered regressive when the ETR falls when the scale of individual/household income or expenditure rise. Finally, it is said to be proportional when the ETR remains constant for all individuals/households.

The work by Pechman and Okner (1974) is considered the standard analysis to compute the ETR using microdata. They assessed the burden of taxation on the US economy. They ordered taxpayers by their annual income groups, specified the tax shifting assumptions that the final burden of sales and excise taxes falls on households per their consumption patterns, and utilised taxable consumption items. Their result showed that the US tax system was nearly proportional. Later studies followed the same set of shifting assumptions to examine the distribution of tax burden for the US economy by addressing some of the methodological issues, such as Musgrave et al. (1974) and Browning (1978 and 1985).

Lovejoy (1963), McLure (1977), Wasylenko (1986), Sjoquist and Green (1992), Alleyne (1999), and Alleyne et al. (2004) assessed the incidence of direct and indirect taxes in Jamaica. They took households as the unit of analysis, employed the standard assumption of full shifting of indirect taxes to consumers, i.e., the burden of indirect taxes to be borne by consumers, and used income as a welfare indicator to compute ETRs. All these studies found indirect taxes to be proportional. Some studies found taxes to be slightly progressive for the lower-income groups and slightly regressive for the upper-income groups, while some found the opposite results. Kaplanoglou and Newbery (2003) studied the distributional impact of indirect taxation in Greece. They considered non-durable household expenditures as a welfare indicator to compute average tax rates across household deciles. Employing the similar assumption of tax shifting used in previous studies, they found that poorer households paid a higher proportion of their total expenditure in indirect taxes, while richer households paid a lower proportion.

Some studies have also assessed the incidence of various indirect taxes by comparing tax concentration curves of different types of tax categories for African countries. These include Sahn and Younger (1999), Younger et al.

(1999), Chen et al. (2001), and Rajemison et al. (2003). They all found that indirect taxes, such as commodity tax, import taxes, and excise duty on alcoholic and non-alcoholic drinks, tobacco, and automobiles were consistently progressive. Taxes on gasoline and diesel were by far the most progressive, whereas taxes on kerosene (or paraffin) were regressive, given that kerosene is widely used as a fuel for lighting and cooking by the poor and has a very low income elasticity of demand. Besides these approaches, studies have also used the general equilibrium (GE) approach pioneered by Harberger (1962) to assess the incidence of taxation. These studies include Mieszkowski (1969), McLure (1975), Bovenberg (1987), and Deverajan et al. (1980).

Some studies have also incorporated taxes on inputs while assessing the incidence of taxes on final goods. In such cases, their analyses were not based on nominal tax but on tax rates computed by using the input-output framework. These include Ahmad and Stern (1989), Malik and Saqib (1989), Bahl (1991), Rajemison et al. (2003), Alleyne et al (2004), SPDC (2004), and Wahid and Wallace (2008), among others.

Studies on Pakistan

Studies undertaken in Pakistan have employed Peckman and Okner's (1974) methodology, i.e., these studies computed effective or average tax rates to examine the incidence of indirect taxes by considering households as a unit of analysis. They assumed that indirect taxes were to be borne, i.e., full forward shifting of indirect taxes by consumers who consume taxable commodities.

One of the earliest studies on tax incidence in Pakistan was conducted by Jeetun (1978) who estimated the distribution of tax burden across different income groups by rural and urban areas for the year 1972-73. His results showed that the total incidence of all taxes exhibited slight progressivity. Rural-urban comparison indicated that higher-income groups in rural areas were greatly undertaxed compared to their urban counterparts.

Malik and Saqib (1989) also focused on the distributional aspect of federal indirect taxes by estimating their incidence across households belonging to different income groups in rural and urban areas. Their findings indicated a regressive tax system in rural areas, where all components of indirect taxes (import duties, sales taxes, and excise duties) exhibited a regressive pattern. In urban areas, import duties and excise duties were regressive, while sales tax was slightly positive. SPDC (2004) also showed that all components of the indirect tax system along with the overall system of indirect taxes portrayed regressive patterns.

Refaqat (2008) analysed indirect taxes in Pakistan from the perspective of equity and distributional considerations as a result of tax reforms initiated in the 1990s. The findings illustrated the progressivity of GST for 1990-91 (pre-reform era) with small magnitudes of incidence. However, despite exemptions for basic food items, the proportionality of GST/VAT emerged in 2001-02 (post-reform era). Commodity-wise results showed the regressivity of GST on food items, clothes, fuel and utilities, progressivity on durable items, and POL products, and proportionality on tobacco and personal care items. The ETRs computed by Wahid and Wallace (2008) indicated that the incidence of all indirect taxes combined was relatively proportional. Individually, the results suggested that the incidence of the GST and customs were proportional for the lower deciles and progressive for the upper deciles, while excise duty was regressive. Jamal and Javed (2013) indicated the proportionality of the GST structure, which was associated with progressivity for the upper end of deciles of per capita expenditure. The urban incidence of the GST was higher than the rural incidence. Of the studies discussed above, Malik and Saqib (1989), SPDC (2004), and Wahid and Wallace (2008) employed the IO table to take into account the taxes on intermediate inputs.

4. RESEARCH METHODOLOGY

The study followed an IO model-based approach to estimate the incidence of indirect taxes. This approach allows tracing the cascading effects of indirect taxes on intermediate inputs. It measures the income of a household that goes away because of both explicit taxes on taxable items and implicit taxes on exempted items.

To incorporate this feature, input-adjusted effective tax rates (ETRs) for each sector in the IO table were computed by employing the IO input coefficient matrix (see Ahmed and Stern, 1991).

In the simple IO model of production with perfect competition and constant return to scale, THE equilibrium price condition can be written as:

$$P_s = P_b A + V \quad \text{..................} \ (1)$$

Where vector Ps represents the seller's price, i.e., the price received by producers for sales, Pb represents the buyer's price, i.e., the price paid by consumers on buying goods for final consumption as well as by producers for buying intermediate inputs, A is the fixed coefficient matrix of IO, and V is a vector of payments to factors of production or value added.

In the presence of taxes, the buyer's prices become:

$$P_b = P_s + T \quad \text{.........................} \ (2)$$

Or $$P_s = P_b - T \quad \text{.........................} \ (3)$$

Substituting Equation 3 into Equation 1 gives:

$$P_b - T = P_b A + V \quad \text{......................} \ (4)$$

Or $$P_b = T(I-A)^{-1} + V(I-A)^{-1} \quad \text{....................} \ (5)$$

This indicates that the purchaser's price is the sum of two components. The following equation is the input adjusted ETR vector (product of statutory tax rates and inverse of the (I-A) matrix):

$$T_e = T(I-A)^{-1} \ \text{......................} \ (6).$$

Equation (7) below, is the per-unit resource cost vector (product of per-unit value-added and inverse of the (I-A) matrix), which is the basic price vector or prices in the absence of tax.

$$V_c = V(I-A)^{-1} \quad \text{......................} \ (7)$$

This ETR is based on the assumption of full forward shifting of indirect taxes, i.e., the burden of indirect taxes is borne by consumers in proportion to their expenditures. These input-adjusted ETRs are used to compute the tax payments of households to compute the incidence of indirect taxes.

The following methodology was followed to compute the incidence of taxes and their distribution across households.

Computation of Nominal Tax Rates

The variable T in Equation (6) is the prevailing tax rate. The question is whether to take statutory or nominal rates of taxes that are based on revenue collection. Studies have used both rates. However, the nominal rate helps overcome the issue of tax compliance and matching tax burden with revenue collection. The present study computed nominal rates for sales tax, customs duties, and federal excise duty instead of taking statutory tax rates.

1) To compute the nominal tax rate, the mapping of revenue collections of each component of indirect tax was carried out with 81 sectors in the IO table to acquire revenue collection from each sector, i.e., the mapping of:

    - the commodity-wise revenue collection of sales tax local (882 commodities); the mapping of commodity-wise revenue collection of excise duty on both local products and imports (12 commodities);

    - Pakistan Customs Tariff's chapter-wise data (99 chapters) on revenue collection of sales tax imports and custom duty.

2) The mapping of the value of imports with sectors in the IO table:

3) Pakistan Customs Tariff's chapter-wise data (99 chapters) on the value of imports.

4) Calculation of value-added

    The shares of gross value added (GVA) for each sector were obtained from the 2010-11 IO table. These shares were then applied to the total GVA (GDP at factor cost) for the year 2018-19 to obtain sector-wise GVA for 2018-19.

5) Nominal rates of each tax were computed using respective revenue collection and the GVA.

Computation of Effective Tax Rate

Nominal rates and the IO coefficient matrix, A, were then used to compute the IA-ETR for each sector as specified in equation (6).

Reference unit

The household was taken as the unit of analysis because it was assumed that household members collectively make decisions regarding work, consumption, and saving, and they often pool their resources and share them equally (see Alleyene, 2004; Refaqat, 2005, 2008; Wahid and Wallace, 2008; and Jamal and Javed, 2013).

Welfare indicator

Households' total expenditures were taken as a measure of their well-being and an indicator was constructed that ranks them by welfare level. Representing consumption as a proxy of household welfare is justified because it reflects the capacity to pay, is less volatile than current income, and is less likely to be under-reported than income (see Deaton and Grosh, 2000; Refaqat, 2005, 2008; Wahid and Wallace, 2008; Cubero and Hollar, 2010).

Tax shifting assumption

The final burden of indirect taxes was assumed to be borne by consumers based on the view that owners of factors of production have perfectly inelastic supplies and consumers have perfectly inelastic demand for commodities. The lack of reliable information on these elasticities tends to the widespread adoption of the full forward shifting of indirect taxes (Gemmell and Morrissey, 2003).

Computation of Household Tax Payment

The estimation of tax incidence requires tax payments of each household for each taxable and exempted consumption item. For this, household consumption items were mapped with the sectors in the IO table. The estimated input-adjusted ETR for each sector was then assigned to each item according to its mapping with the respective sector.

Tax payment for each item was computed by applying the respective item's ETR to its expenditure in the following manner.

$$TP_{j,h} = EXP_{j,h} \times \frac{1}{1 + ETR_j} \quad \text{.......................................} (8)$$

Where TP is tax payment, EXP is a household expenditure, *j (=1…n)* is consumption item, *h (= 1…m)* are households, and t is the type of tax (GST-Local, GST-Imports, CD, FED-Local and FED-Imports).

Estimation of Tax Incidence

Tax incidence (INC) was computed by taking a percentage share of tax payment for a particular item in the household's total expenditures.

$$INC_{j,h} = \frac{TP_{j,h}}{EXP} \times 100 \quad \text{.......................................} (9)$$

The distribution of incidence or distribution of tax burden across households was assessed by comparing the average rate of incidence across household expenditure deciles.

This allows for analysing the progressivity or regressivity of taxes. A tax is progressive when the ARP rises along with the rise in households' total expenditures, it is regressive when it falls, and it is proportional when it remains constant.

Data Sources

The following data sources were used.

- Household Integrated Economic Survey (HIES) 2018-19, Pakistan Bureau of Statistics, Government of Pakistan, for households' consumption expenditures. HIEs data were assigned survey weights provided in HIES. As a result, the analysis was based on data that is both nationally and provincially representative.
- The IO Table 2010-11, Federal Bureau of Revenue, Government of Pakistan, was used to trace the impact of taxes on intermediate inputs.
- Tax schedules of the Sales Tax Act 1990 (amended up to 11 March 2019), Federal Board of Revenue, were used to identify the taxable and exempted sectors/items.
- Income Tax Ordinance 2001 (amended up to 30th June 2019) Federal Board of Revenue, Government of Pakistan, was used to obtain WHT rates.

5. RESEARCH FINDINGS AND DISCUSSION

This section presents the estimation results. It first displays the nominal and estimated ETRs for each component of indirect taxes. It then furnishes results for the incidence of taxes and their distribution across household deciles.

Nominal and Effective Tax Rates

Computed nominal rates and estimated input-adjusted ETRs of the GST are presented only for those sectors in the IO table that are related to households' final consumption of goods (see Table A1: Annexure).

The statutory tax rate of GST is 17 per cent, but, except for a few sectors, the computed nominal tax rate for each sector, based on its revenue collection, was less than the statutory rate. This indicates the presence of leakage in tax revenue collection. A comparison of nominal rates and ETRs indicates that all sectors were affected by GST levied on intermediate inputs, which is reflected by the higher ETRs compared to nominal rates (Table A1: Annexure). In other words, it means that the burden of taxes on households was higher than the tax rate that exists due to cascading effect of taxes on inputs. In particular, nine sectors associated with crops, livestock, fisheries and milled grains are exempted from GST. However. these sectors were taxed at varying rates, in the range of one to 3 per cent, depending on the type and share of, and the nominal tax rate on intermediate inputs they used.

GST Incidence and Its Distribution Across Households

The distribution of tax incidence or distribution of tax burden of GST across household expenditure deciles is presented in Table 2. The first decile represents the households in the lowest income group or with the lowest total expenditures, while the tenth decile represents the households in the highest income group or with the highest total expenditures. This section uses these terms interchangeably while explaining research findings.

According to Table 2, the overall incidence of GST was on average 6.7 per cent in Pakistan. The distribution of incidence was found to be regressive as it declined for higher deciles. It ranged from 6.9 per cent for the lowest decile to 6.3 per cent for the highest decile. This suggests that households in the first decile, or the poorest 10 per cent households, spent, on average, Rs.7 for every Rs.100 to pay indirect taxes, while households in the tenth decile, or the richest 10 per cent, spent Rs.6, on average. The pattern of incidence of GST was also regressive in both rural and urban areas. The magnitude of incidence in rural areas was 0.2 to 0.3 percentage points higher in rural areas compared to urban areas.

Table 2: Overall Distribution of Incidence of GST (%) – 2018-19

Deciles of HH Expenditures	All Areas	Rural	Urban
1	6.90	6.93	6.71
2	6.88	6.95	6.72
3	6.90	6.93	6.62
4	6.81	6.92	6.59
5	6.74	6.85	6.57

6	6.65	6.85	6.54
7	6.61	6.71	6.44
8	6.57	6.66	6.47
9	6.47	6.59	6.32
10	6.34	6.50	6.29
Overall	**6.69**	**6.79**	**6.53**

Source: Author's estimates based on HIES 2018-19 and IO Table 2010-11.

Distribution of Incidence: Comparison with Earlier Studies

Before providing the distributional pattern of incidence of earlier studies, a few words on the structure of taxation in Pakistan which has undergone several reforms over the last three decades are in order. In 1990-91, indirect taxation was shifted away from CD and FED and moved towards GST, which is a variant of the Value Added Tax (VAT). The Sales Tax Act of 1990, introduced a GST at the rate of 12.5 per cent on imported goods and value added at each stage of production on goods manufactured and sold in Pakistan. However, goods, such as agricultural products, petroleum, electricity, pharmaceuticals, and fertilisers, were exempted from GST. By the late 1990s, the GST net was broadened to include items such as petroleum products, electricity, and natural gas. Over time, the rate of GST increased to 17 per cent and the exemptions were removed. At present, the GST net has been expanded to include food items (e.g., tea, sugar, beverages, etc.), essential consumer products, and fertiliser, among other products.

As a result of these reforms, the composition of federal indirect tax receipts kept changing. In 1990-91, of the total federal tax collection, the CD constituted 55 per cent, sales tax 18 per cent, and FED 27 per cent. In 2000-01, the GST constituted 57 per cent, CD 24 per cent, and FED 18 per cent. In 2020-21, among the three components of indirect taxes, the GST dominated with a share of 66 per cent followed by CD at 25 per cent, and FED at 9 per cent.

The effect of the imposition of the GST on domestic production and sale with an expanded base and increased rate reflects on its distributional burden across different segments of society. A comparison of the results of the GST incidence of this study research with those conducted earlier is given in Table 3. [4]

While comparing the incidence in the pre- and post-reform era, Refaqat (2008) indicated that the distribution of GST changed from progressive in 1990-91 to proportional in 2001-02. Jamal and Javed (2013) also found it to exhibit a proportional pattern associated with progressivity at the upper end of income in 2010-11. It can be said that as the coverage of GST increased, incidence changed from progressivity to proportionality. However, Refaqat (2008) and Jamal and Javed (2013) considered tax levied only on final consumption and did not incorporate taxes levied on intermediate inputs, i.e., the cascading effect of tax. Due to this, they excluded the items that are exempted from GST from their analysis. This might be the factor resulting in a proportional GST burden.

On the other hand, Malik and Saqib (1989), SPDC (2004), and the present study showed the distribution of GST burden to be regressive in 1978-79, 2001-02, and 2018-19. All these studies took into account the cascading effect of the GST and, therefore, included all items even if the final consumption was exempted from the GST. And that could be one of the reasons that they found a regressive pattern of incidence. However, despite incorporating the impact of taxes on inputs, Wahid and Wallace (2008) concluded that the incidence was proportional.

[4] All studies used the Household Integrated and Economic Survey (HIES) corresponding to the year of their analyses and computed the average rate of progression to assess the incidence of the GST.

Although exempted items have zero tax, taxes paid on inputs used in producing these items are not adjusted by a refund. As a result, even in the case of exempted items, any tax on inputs is passed on to consumers based on the inputs' share in production as well as the ripple effect of these inputs in terms of the type-1 multiplier. For example, wheat flour is exempted from the GST, but taxes paid on inputs (electricity, petrol, etc.) are not adjusted. Therefore, incorporating these taxes through the IO table wheat flour is effectively taxed, which is included in the consumer price of wheat flour.

The household expenditure pattern shows that 30 per cent poorest households, on average, spent 48 per cent of their total expenditures on food, whereas 30 per cent of the richest households spent 37 per cent (HIES 2018-19). It shows that the effective tax on wheat affected poor households relatively more compared to rich households. Hence, the incidence of the effective tax rate on wheat was regressive. The literature points out that the results of the incidence analysis are different if taxes on inputs are incorporated or not. Generally, taxes are regressive if taxes on inputs are incorporated (see Rajemison et al., 2003; Younger et al., 1999).

Table 3: Distribution of GST Incidence – Comparison with Earlier Studies

Monthly income class (Rs)	Malik & Saqib (1989)	HH Deciles	Refaqat (2008)		SPDC (2004)	Wahid & Wallace (2008)	Jamal & Javed (2013)	This Research
	HIES 1978-79		1990-91	2001-02	2001-02	2004-05	2010-11	2018-19
up to 300	1.08	1	1.08	4.58	9.30	3.32	4.41	6.90
301 - 400	1.03	2	1.25	4.73	8.60	3.23	5.49	6.88
401 -500	0.95	3	1.25	4.70	8.30	3.20	4.62	6.90
501 -600	1.01	4	1.28	4.70	8.20	3.27	4.73	6.81
601 - 800	0.92	5	1.30	4.71	8.00	3.50	4.77	6.74
801 - 1000	1.00	6	1.31	4.68	7.70	3.58	4.95	6.65
1001 - 1500	0.87	7	1.34	4.69	7.40	3.30	4.97	6.61
1501 - 2000	0.83	8	1.35	4.58	7.10	3.65	5.02	6.57
2001 - 2500	0.78	9	1.39	4.70	6.70	3.39	5.26	6.47
2501 - 3000	0.76	10	1.52	4.65	5.90	3.72	5.49	6.34
3001 - 3500	0.77							
3501 & above	0.88							
Average	**0.91**		**1.31**	**4.67**	**7.72**	**3.42**	**4.97**	**6.69**

Distribution of GST Incidence by Commodity Groups

The distribution of the incidence of the GST by commodity groups across deciles of household expenditures is given in Table 4.

Basic food items had a highly regressive pattern of tax incidence across all deciles with the highest magnitude among all commodity groups.[5] For example, 10 per cent of the poorest households paid 1.6 per cent of their expenditures as GST when buying basic food items compared to 0.6 per cent paid by the 10 per cent richest households.

Other commodity groups that had regressive patterns across all deciles include transport services and tobacco products. Some groups, though, depicted an overall regressive pattern but a proportional pattern for the bottom deciles. These include personal items, household items, and pharmaceutics. The incidence of GST on transport services showed proportionality associated with regressivity for the bottom deciles.

Commodity groups that had a progressive incidence of GST include transport fuel and durable goods. The highest progressivity was in transport fuel where the poorest 10 per cent of households' 0.3 per cent expenditures were on GST, while the richest 10 per cent paid one per cent. Other commodity groups, such as utilities, non-basic food items, and books and stationery, had an overall progressive pattern accompanied by a proportional pattern for some deciles. For instance, the incidence for utilities was progressive for the bottom two deciles, proportional to the sixth decile, and progressive thereafter. The incidence of tax for communication services was proportional across all deciles.

Table 4: Distribution of Incidence of GST by Commodity Groups (%), 2018-19

Commodity Groups	Household Expenditures by Deciles									
	1	2	3	4	5	6	7	8	9	10
Regressive										
Basic Food Items	1.634	1.505	1.382	1.303	1.225	1.139	1.057	0.961	0.824	0.582
Personal Items	1.360	1.333	1.334	1.304	1.281	1.271	1.223	1.218	1.183	1.087
Household Items	0.753	0.692	0.671	0.631	0.616	0.611	0.593	0.584	0.543	0.565
Transport Services	0.156	0.135	0.136	0.139	0.137	0.135	0.133	0.133	0.128	0.114
Pharmaceutics	0.503	0.442	0.412	0.419	0.395	0.385	0.380	0.337	0.327	0.298
Tobacco & Products	0.146	0.111	0.109	0.102	0.095	0.080	0.076	0.071	0.060	0.045
Proportional										
Communication Services	0.031	0.031	0.031	0.031	0.030	0.031	0.031	0.033	0.035	0.037
Progressive										
Non-Basic Food Items	0.828	0.822	0.844	0.793	0.816	0.820	0.860	0.884	0.957	0.994
Durable Goods	0.142	0.151	0.166	0.171	0.183	0.192	0.224	0.240	0.259	0.461
Utilities	0.979	1.037	1.046	1.051	1.069	1.065	1.081	1.107	1.126	1.100
Transport Fuel	0.335	0.567	0.694	0.785	0.797	0.828	0.851	0.898	0.919	0.950
Books & Stationery	0.035	0.059	0.073	0.079	0.092	0.097	0.101	0.105	0.113	0.110

Source: Author's estimates based on HIES 2018-19 and IO Table-2010-11.

[5] Items such as wheat flour, rice, pulses, vegetables, spices, fresh dairy, ghee, sugar, tea were considered basic food items in this study. The remaining food items were included in non-basic food group.

Incidence of Withholding Tax

As mentioned earlier, the incidence of WHT was estimated only for the heads where data at the household level were available from HIES 2018-19. Table 5 shows these heads and corresponding household consumption expenditures. These heads constituted over 9 per cent of the WHT collection, which is indirect. Many of these taxes are adjustable, however, from household data, it cannot be deduced as to which households get their tax payment adjusted. Therefore, it was assumed for this analysis that tax adjustment had not occurred. Table 5 also mentions the WHT rates for each head. These rates were applied to the respective household expenditure to calculate the tax payments using Equation (8) and the incidence was computed using Equation (9) described in Section 4.

Table 5: Indirect WHT Heads, Matching HH Items, Tax Status, and Collection in 2018-19

Heads of Collection	Matching Household Items Available in HIES 2018-19	Taxation Status	Collection (Rs. million)
U/s 236 (Telephones subscribers other than Mobile Phones) @10% if bill > Rs. 1,000	Telephone, Mobile Charges (Easy load, Mobile card, etc.)	Adjustable	9,311.50
U/s 236 (Mobile Phone Subscribers -.Prepaid Cards) @10% if bill > Rs. 1,000	Expenses on electricity	Adjustable	7,875.60
U/s 235 (Electricity Bills) @ 7.5% on domestic users if monthly bill ≥ Rs. 25,000	Expenses on petrol/ diesel/ Mobil oil Expenses on generator (petrol/diesel)	Adjustable	35,558.30
U/s 235A (Advance tax on domestic electricity consumption)	Expenses on CNG	Adjustable	596.8
U/s 156 A (Petroleum Products) @ 10%	Expenses on petrol/ diesel/ Mobil oil Expenses on generator (petrol/diesel)	Final	6,948.50
U/s 234 A (On CNG Stations) @ 4%	Expenses on CNG	Minumum Tax	2,953.20
U/s 236B (Purchase of Domestic Air Ticket @ 5 %)	Expenses on travelling by air	Adjustable	687.1
Total (1)	-	-	63,931.00
Total WHT (2)	-	-	683,829.50
(1) as % of (2)	-	-	9.3%

Source: Federal Board of Revenue Yearbook 2019-20, Government of Pakistan

The results of the incidence of specific WHT heads are shown in Table 6. Considering all areas, the incidence of WHT for telephone and mobile usage portrayed a slightly regressive pattern associated with A proportional pattern for the middle deciles. However, it was roughly proportional in rural areas and regressive in urban areas. The incidence of WHT for petroleum products and electricity was progressive across all deciles in all areas as well as in both rural and urban areas. The incidence of WHT for internet usage, air travel, and CNG stations was also progressive, but its magnitude was minimal for all areas.

Table 6: Incidence of WHT (%) – Pakistan 2018-19

Deciles of HH Expenditures	Telephone & Mobile	Internet	Petroleum Products	CNG	Electricity	Air travel
All Areas						
1	0.177	0.002	0.163	0.000	0.026	0.000
2	0.173	0.005	0.273	0.001	0.057	0.002
3	0.171	0.004	0.334	0.000	0.073	0.001
4	0.170	0.006	0.377	0.001	0.105	0.002
5	0.167	0.007	0.382	0.000	0.135	0.003
6	0.166	0.009	0.397	0.001	0.163	0.006
7	0.162	0.014	0.407	0.002	0.182	0.004
8	0.166	0.020	0.432	0.002	0.233	0.008
9	0.167	0.032	0.444	0.004	0.279	0.014
10	0.157	0.051	0.458	0.017	0.329	0.033
Rural Areas						
1	0.173	0.002	0.138	0.000	0.023	0.000
2	0.168	0.002	0.267	0.000	0.041	0.000
3	0.168	0.005	0.327	0.001	0.057	0.000
4	0.168	0.004	0.367	0.000	0.063	0.001
5	0.165	0.004	0.405	0.000	0.081	0.003
6	0.165	0.005	0.409	0.000	0.111	0.002
7	0.168	0.008	0.401	0.001	0.134	0.009
8	0.168	0.010	0.417	0.002	0.129	0.005
9	0.172	0.011	0.435	0.002	0.181	0.015
10	0.162	0.029	0.472	0.009	0.223	0.033
Urban Areas						
1	0.198	0.005	0.149	0.001	0.055	0.005
2	0.185	0.009	0.275	0.001	0.148	0.000
3	0.170	0.010	0.332	0.001	0.174	0.002
4	0.164	0.012	0.380	0.000	0.219	0.003
5	0.157	0.016	0.406	0.002	0.228	0.000
6	0.161	0.027	0.411	0.002	0.260	0.003
7	0.165	0.030	0.432	0.003	0.297	0.004
8	0.160	0.040	0.424	0.004	0.323	0.010
9	0.166	0.056	0.418	0.010	0.352	0.015
10	0.150	0.062	0.485	0.025	0.405	0.037

6. CONCLUSION AND RECOMMENDATIONS

This study examined the incidence of GST-Domestic in Pakistan and its distribution across deciles of household expenditures. The findings indicate that the overall incidence of GST, on average, was 6.7 per cent. The distribution of incidence portrayed an overall regressive pattern across all deciles and in rural and urban areas.

Analysis by commodity group shows the highest rate of incidence as well as the highest extent of regressivity for basic food items. Other commodity groups that indicated regressivity include personal and household items. Commodity groups indicating a progressive pattern of incidence included non-basic food items, utilities and transport fuel.

The marked regressivity of incidence for basic food items primarily occurred on account of household spending patterns on food items. The HIES 2018-19 data revealed that 30 per cent poorest households, on average, spent 48 per cent of their total expenditures on food, whereas 30 per cent of the richest households spent 37 per cent.

Food inflation has often been a major public policy challenge for the governments in Pakistan and numerous measures are undertaken to control basic food prices to provide relief for the poor. For example, major basic food items have been exempted from indirect taxation over several years. However, indirect taxes levied on inputs used to produce these items act as implicit taxes, which are transferred to the final prices of these items and cause an increase in prices. On the other hand, to raise revenues, governments often increase taxes on necessities, which have inelastic demand, such as utilities, which put a burden on households' budgets, particularly on the poor.

Regressivity affecting the poor segment needs to be addressed, albeit without causing secondary distortions. For example, exempting selected essential items as well as their inputs from taxes would not only cause revenue losses but would also benefit the items not in the consumption basket of the poor.

An alternative way to avoid secondary distortions and support low-income groups is transfer payments, which can minimise the impact of taxes on them. Practices from other countries also demonstrate the use of transfer payments. Karageorgas (1973) pointed out a decline in inequality after the initiation of transfer payments in Greece, with the highest benefit received by the lowest income groups. Ruggeri et al. (1994) reported the progressivity of general sales tax at the lower end of the income scale due to transfer payments to these income classes in Canada. Crisan et al. (2015) also highlighted the progressive tax and transfer system in Canada, where the bottom two quintiles of the income distribution were net recipients of government transfers, while the middle and top two quintiles were net taxpayers.

REFERENCES

Ahmad, E., & Stern, N. (1986). Tax reform for Pakistan: Overview and effective taxes for 1975-76. *The Pakistan Development Review,* 43-72.

Ahmad, E., Ahmad, E., & Stern, N. (1991). *The theory and practice of tax reform in developing countries.* Cambridge University Press.

Alleyne, D., Alm, J., Bahl, R., & Wallace, S. (2004). *Tax burden in Jamaica* (No. paper0434). International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Alleyne, D. (1999). *Taxation and equity in Jamaica, 1985-1992: Who bears the burden?* (Vol. 5). New Generation Series.

Bahl, R. W. (Ed.). (1991). The Jamaican tax reform (Vol. 38). Lincoln Institute of Land Policy

Bovenberg, A. L. (1987). Indirect taxation in developing countries: A general equilibrium approach. *IMF Staff Papers,* 34(2), 333-373

Chen, D., Matovu, J., & Reinikka, R. (2001). *A quest for revenue and tax incidence*. Uganda's recovery. The role of farms, firms, and government, Washington, DC: World Bank, 271-317.

Crisan, D., McKenzie, K. J., & Mintz, J. (2015). The distribution of income and taxes/transfers in Canada: A cohort analysis. *SPP Research Paper*, (8)

Devarajan, S., Fullerton, D., & Musgrave, R. A. (1980). Estimating the distribution of tax burdens: a comparison of different approaches. *Journal of Public Economics*, 13(2), 155-182.

Gemmell, N., & Morrissey, O. (2003). Tax structure and the incidence on the poor in developing countries. *CREDIT research paper,* (03/18).

Harberger, A. C. (1962). The incidence of the corporation income tax. *Journal of Political economy,* 70(3), 215-240.

Jamal, H. (2017). *Poverty and vulnerability estimates: Pakistan*. 2016-Research Report No. 99 2017. Social Policy and Development Centre (SPDC), Karachi.

Jamal, H., & Javed, S. (2013). Incidence of General Sales Tax in Pakistan: Latest Estimates. *Pakistan Journal of Applied Economics,* 23(2), 73-95

Jeetun, A. (1978). The incidence of taxes in Pakistan. Distributional paper. *Applied Economics Research Centre, University of Karachi.*

Kaplanoglou, G., & Newbery, D. M. (2003). *The distributional impact of the proposed tax reform on Greek households*. Centre of Planning and Economic Research.

Karageorgas, D. (1973). The distribution of tax burden by income groups in Greece. *The Economic Journal,* 83(330), 436-448.

Lovejoy, R. M. (1963). The Burden of Jamaican Taxes 1958. *Social and Economic Studies*, 12(4).

Malik, M. H., & us Saqib, N. (1989). Tax incidence by income classes in Pakistan. *The Pakistan Development Review,* 13-25.

McLure Jr, C. E. (1975). General equilibrium incidence analysis: The Harberger model after ten years. *Journal of Public Economics*, 4(2), 125-161.

McLURE, C. E. (1977). *The Incidence of jamaican taxes*: *1971-1972*. Institute of Social and Economic Research, University of the West Indies.

Mieszkowski, P. (1969). Tax incidence theory: The effects of taxes on the distribution of income. *Journal of Economic Literature,* 7(4), 1103-1124.

Pechman, J. A., & Okner, B. A. (1974). *Who bears the tax burden?* Studies of Government Finance Series.

Refaqat, S. (2008). *Social incidence of indirect taxation in Pakistan* (1990-2001) (Doctoral dissertation, University of Bath).

Rajemison, H., Haggblade, S., & Younger, S. D. (2003). Indirect tax incidence in Madagascar: Updated estimates using the input-output table. *Cornell Food and Nutrition Policy Program Working Paper,* (147).

Ruggeri, G. C., Van Wart, D., & Howard, R. (1994). The redistributional impact of taxation in Canada. *Can. Tax J.*, 42, 417.

Sahn, M. D. E., & Younger, M. S. D. (1999). *Dominance testing of social sector expenditures and taxes in Africa*. International Monetary Fund.

Sjoquist, D. L., & Green, D. (1992). *Distribution of tax burdens*. Policy Research Center, College of Business Administration, Georgia State University, Jamaica Tax Review Project, Report, (3).

Social Policy and Development Centre (SPDC) (2004). *Combating Poverty: Is Growth Sufficient: Annual Review.* SPDC and Oxford University Press, Pakistan.

Wahid, U., & Wallace, S. (2008). *Incidence of taxes in Pakistan: primer and estimates* (No. paper0813). International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University.

Wasylenko, M. (1986). *The distribution of tax burden in Jamaica: Pre-1985 reform*. Jamaica Tax Structure Examination Project, Metropolitan Studies Program, Syracuse University, Board of Revenue, Government of Jamaica. Staff Paper no. 30. Syracuse, NY: Syracuse University.

Younger, S. D., Sahn, D. E., Haggblade, S., & Dorosh, P. A. (1999). Tax incidence in Madagascar: An analysis using household data. *The World Bank Economic Review*, 13(2), 303-331.

ANNEXURE

Table A1: Nominal and Estimated Effective Tax Rates of GST-Domestic 2018-19

Sectors From IOT	Nominal Rate	Effective Rate
Rice	0.016	3.375
Wheat	0.000	3.270
Sugarcane	0.000	2.609
Pulses	0.003	0.810
Potatoes	0.000	2.962
Vegetables & Condiments	0.113	2.919
Fruits	0.025	2.395
Livestock & Slaughtered Products	0.091	1.246
Fisheries	0.000	6.880
Coal	16.944	19.257
Crude Oil & Natural gas	4.936	6.691
Vegetable Oils	0.827	5.256
Milled Grains	0.003	3.019
Bakery Products	8.572	12.663
Sugar	8.098	10.123
Other Food	17.000	19.879
Beverages	17.000	21.943
Cigarettes & Tobacco	5.325	6.739
Cotton Cloth	0.055	5.160
Art Silk	0.330	6.495
Made-up Textile Goods	0.036	3.725
Knitwear	0.551	4.916
Carpets	0.434	4.674
Garments	3.827	7.966
Other Textile Products	11.932	16.907
Leather & Leather Products	2.048	5.471
Footwear	6.400	9.342
Paper & Printing	4.458	10.531
Pharmaceuticals	7.078	14.964
Chemical Consumer Products	17.000	25.384
Refined Petroleum	17.000	23.519
Rubber & Plastic Products	4.578	13.209
Bricks	0.033	5.901
Cement	8.728	15.421
Metal Products	16.994	23.151
Non-electrical Machinery	5.887	14.937
Electrical Equipment	13.120	22.596
Transport Equipment	9.714	23.671
Handicrafts	0.328	3.986
Sports Goods	0.544	6.361
Jewelry & Precious Metals	0.177	5.509
Other Manufacturing Products	16.967	22.725
Electricity, Waterworks & Supply	11.796	19.320
Gas Supply	5.644	9.826
Transport - Railway	0.000	13.007
Transport - Road	0.000	4.748
Communication	0.206	1.595

CITY DEVELOPMENT PRODUCT: DATA ARCHITECTURE FOR SUSTAINABLE ECONOMIC DEVELOPMENT IN PAKISTAN

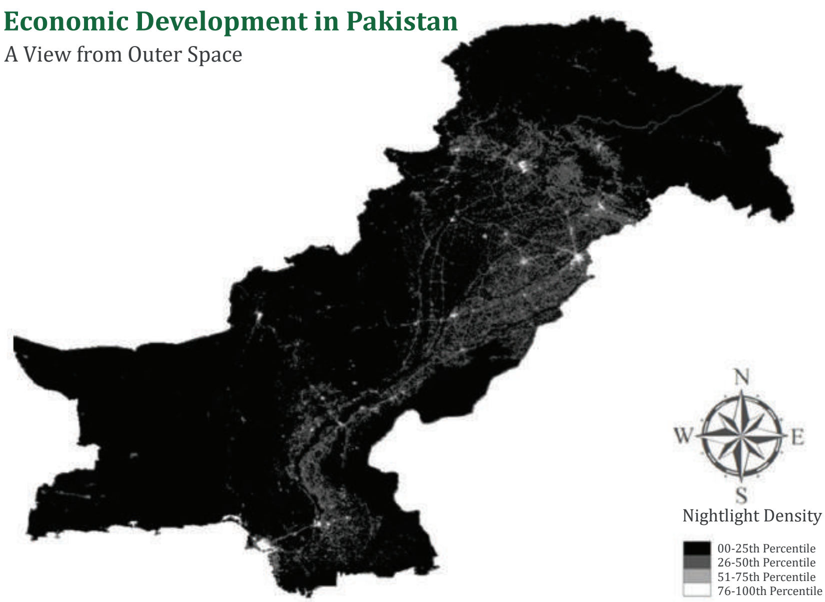
Mohammad Ahmad

ABSTRACT

This paper used multiple sources of publicly accessible remotely sensed satellite data as the primary source for estimating Pakistan's Regional Domestic Product (RDP) at the city level. In the absence of a formal System of Regional Accounts (SRA), which requires repeat economic censuses, RDP estimates can be useful for tracking economic growth and development at lower administrative levels where identification and preparation of development projects take place. The study built a unique panel dataset of city-level RDP estimates for Pakistan from 2012-2020, using satellite imagery on nighttime luminosity, human settlement patterns, and machine learning algorithms. In addition to empowering local economic planning, this dataset can serve as a baseline layer for public policy along multiple dimensions, including evidence-based impact evaluation of development spending, rural-urban migration, structural change, and urban planning.

1. INTRODUCTION

The Government of Pakistan currently does not have a methodology for estimating Regional Domestic Product (RDP) at the provincial, district, or city levels. The need for disaggregated estimates of economic growth has become even more important after the 18th Amendment to the constitution, under which the key areas of economic development including health, education, infrastructure, and industrial development have been devolved to the provinces. Furthermore, the 18th Amendment mandated provinces to further distribute revenues to districts, which form the next administrative tier. These reforms were widely lauded as a strong step towards fiscal federalism and improved allocative efficiency of public finances.



The constitutional steps taken towards the decentralised governance and institutionalised distribution of public resources have, however, run into capacity constraints, which include lower administrative tiers that do not have the informational architecture required for efficient planning and development. As a result, a wedge between de jure and de facto practices of fiscal federalism has emerged. As a step toward decreasing this wedge, this study identifies and addresses a key part of the problem, i.e., the lack of reliable and methodologically comparable economic data at the city level.

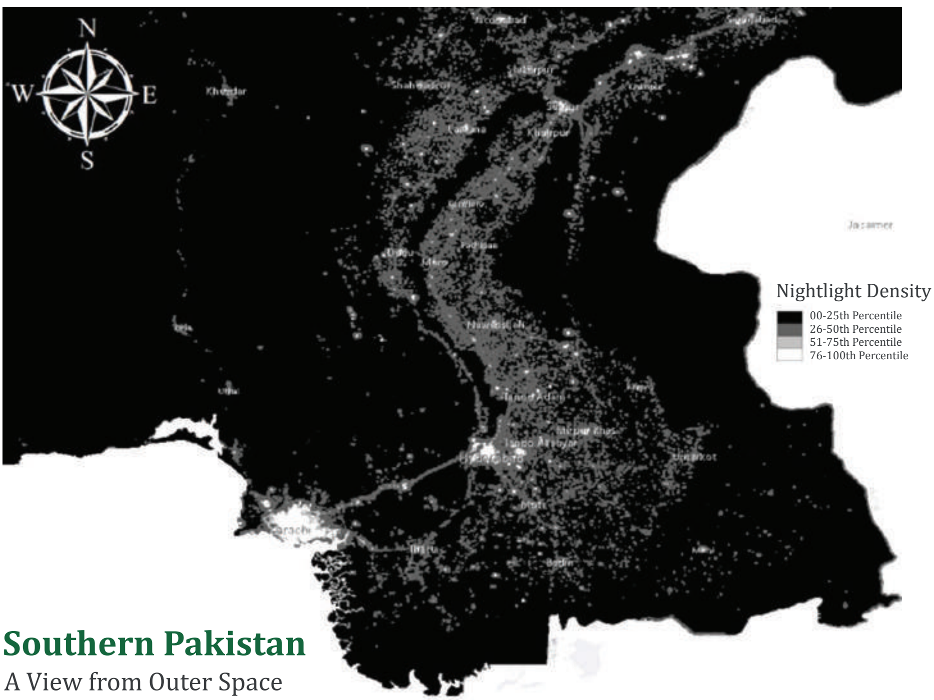
The availability of city-level data that is reliable and comparable can have multiple channels of impact on the state's planning, administrative, and public funds management capacity. Many developing countries, including Brazil, Colombia, and India have developed provincial Systems of Regional Accounts (SRA) based on extensive survey and data collection regimes that are analogous to the national income accounting exercise that takes place each year to compute aggregate GDP measures. However, with the cheap availability of high-resolution satellite imagery, a more cost-effective estimation methodology for subnational GDP estimation can be derived. Moreover, while this proposed top-down approach to RDP estimation may leave certain supranational economic activities unaccounted, which is a natural problem in subnational income accounting, it, nevertheless, has the advantage of being a consistent estimator of growth across all districts, which is an important requirement for distributive purposes under study. Economics and other research fields have incorporated this new source of data as a powerful tool for analysis, and there is little reason why public sector analysis cannot benefit from this affordable and near-real-time data source.

The estimation of regional RDP is an important exercise for growth-oriented regimes that devise part or whole of their economic policies at the subnational/provincial level. There are three broad reasons why the estimation of the district-level domestic product may not only be a valuable indicator of economic growth but also a driver of it.

1) The RDP allows provinces with economic growth as a priority to engage in data-backed comparative policy analysis. It allows for a dialogue on what "treatments" to the economy have produced policy results to be replicated, and what policies can be shelved as ineffective. A realisation of the DP can lead to an institutionalized data-driven dialogue between interlinked provincial departments, i.e., Planning & Development and the Ministry of Finance.

2) The RDP can aid intergovernmental resource distribution bodies, such as the National Finance Commission (NFC) and Provincial Finance Commissions (PFCs), in developing formulaic distribution criteria based on the RDP and provincial GDP. This also creates a performance yardstick for provincial governments to bargain with the centre for an increase in the share of the financial award.

3) The RDP provides disaggregated market signals to private investors, especially to foreign direct investment (FDI). A system of the RDP is indicative of a serious government reform effort to attract private finance. In Pakistan's case and all other middle-income, high-poverty (MIHP) countries, the same logic also applies to foreign aid.



Southern Pakistan
A View from Outer Space

Lack of economic data at lower administrative levels can potentially constrain the efficient planning and utilisation of public finances, impeding economic development. On the other hand, equitable and accountable distribution of public resources can improve democratic practices and the provision of public goods in Pakistan. The objective of this study is to provide an open-access repository of decentralised economic growth data for Pakistani cities that can be used for evidence-based research on Pakistan.

The remainder of this report proceeds as follows. A review of the literature is presented to qualify the exercise at hand, followed by a description of the data used and their sources. The next section presents a discussion of the key methodological issues in calculating the RDP for cities in Pakistan for the 2012-2020 period.

2. REVIEW OF LITERATURE

The availability of fine-grain, remotely-sensed data has become a new powerful tool for analysis within the economics discipline. Economists have found many applications for this source of data, and multiple rich streams of economic literature have developed as a result. Donaldson and Storeygard (2016) have presented a comprehensive survey of the use of satellite data in economics.

One of the most commonly used data sources is night-time luminosity, also known as nightlight density. Henderson, Storeygard, and Weil (2012) showed that nightlight data are an accurate proxy for economic activity. Lee (2016) used nightlight density to predict economic activity in North Korea, for which official data is not available. Kudamatsu, Persson, and Stromberg (2016) merged rainfall satellite data with demographic and health surveys of 28 African countries to study the effects of climate shocks on infant mortality. An important and recent validation of using nightlight data for measuring RDP comes from Perez-Sindin et al. (2021), who benchmarked the VIIRS nightlights data against municipal RDP measures prepared by Colombia's national statistics agency. The results of regressing official RDP on nightlight measures showed that nightlight can serve as a good indicator of municipal RDP.

Another source of remotely sensed data involves the mapping of human settlement and urbanisation patterns around the world. Satellite-based settlement maps are widely used for many scientific purposes. Their most common use is to distinguish urban areas from rural areas. Previously, human settlement data was available at very coarse resolutions (1km-500m). As the need for cooperation on global public goods provision, such as climate change and sustainable development, has increased, so too has the availability of micro-resolution human settlement data which can help governments around the world to plan in a sustainable, informed way. Due to rapid technological advances, the latest raster of global settlement data is now available at a resolution of only 10m with global coverage. The data and their sources are discussed in detail in the data section.

The use of satellite data has not just been restricted to academic research but has also been leveraged in many areas of public policy. Kudamatsu, Persson, and Stromberg (2016) used the normalized difference in vegetation index (NDVI) to capture a measure of agricultural productivity which can be used for agricultural policy. Holmes and Lee (2012) presented the relationship between crop choice patterns and land ownership. Maue et al. (2020) used satellite data to estimate the yields of smallholder farmers. In urban development literature, Turner, Haughwout, and van der Klaauw (2014) studied the impact of land-use regulations on urban sprawl using satellite data. Another innovative application of satellite data was done by Casaburi and Troiano (2016), who used satellite data to identify buildings in an Italian city that had not registered for tax collection. Other applications of satellite data are very diverse and include fields such as pollution monitoring, natural resource management, fire management, and many others.

Another rich stream of contemporary literature in development economics has highlighted the salience of state capacity in eradicating poverty. Page and Pande (2018) documented the new geography of poverty and showed that 49.3% of the world's extremely poor lived in eight middle-income high-poverty countries, including Pakistan. The authors argued that in contrast to low-income countries, poverty reduction in countries like Pakistan must come by enabling the state to build an invisible architecture of economic empowerment. This paper posited that a key pillar of this invisible infrastructure involved building feedback loops and data systems that make the public ask for a data-driven and accountable public policy regime.

Putting together the various pieces from this review of literature, a case can be made for an attempt to build the data architecture of Pakistan's economic policy using remotely sensed satellite data. This study is a first step in that direction.

3. DATA

The study used four sources of data to calculate the RDP estimates for all Pakistani cities. The datasets and their sources are discussed in this section, whereas the technical and methodological details are explained in the methodology section.

The first dataset measures nightlights or luminosity (henceforth NTL) captured by the Visible Infrared Imaging Radiometer Suite (VIIRS), a sensor on board Suomi NPP VIIRS/DNB satellite launched in October 2011 by the National Oceanic and Atmospheric Administration (NOAA), a scientific agency based within the United States government. This study used processed annual composites of the VIIRS NTL for 2012-2020 (Elvidge et al., 2021). A masked annual average of VNL V2 data was used to overcome NTL errors in measurement such as biomass burning and other outliers. The data has global coverage at a resolution of 15 arcseconds (450m) and is freely available for research.

The second dataset is the World Settlement Footprint (WSF 2019), a global binary raster dataset of global human settlements with unprecedented detail and precision. The WSF 2019 features data from the Copernicus Sentinel-1 and Sentinel-2 missions of the European Space Agency (ESA). The data is processed and released by the German Aerospace Center (DLR) in collaboration with the Google Earth Engine and provides raster data at a resolution of 10m for 2019. The WSF 2019 was released at the United Nations Climate Change Conference (COP26) in November 2021.

The third dataset is the World Settlement Footprint Evolution (WSF Evolution). This dataset has been processed by the German Aerospace Center (DLR) by processing seven million images from the US Landsat satellite collected between 1985 and 2015 and illustrates the worldwide growth of human settlements on a year-by-year basis. The WSF Evolution was released along with the WSF 2019 at the United Nations Climate Change Conference (COP26) in November 2021.

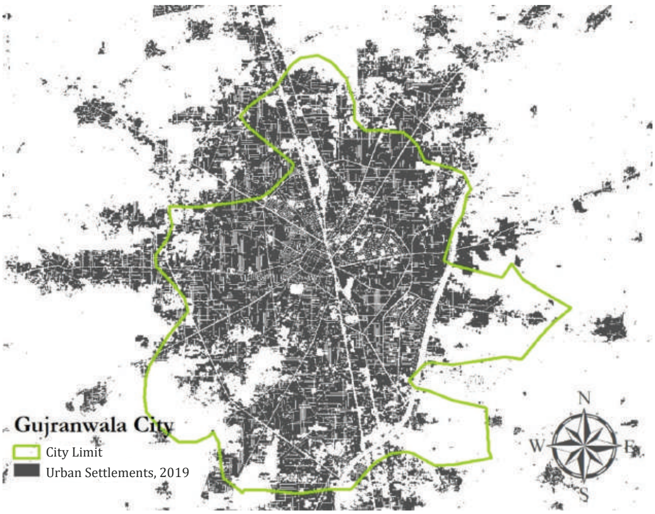
The fourth dataset is the shapefile data for Pakistani administrative divisions. Shapefiles are a digital version of maps and are used for GIS processing. The shapefiles for Pakistan's national boundaries, provincial, and lower administrative boundaries were sourced from the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA), among other sources.

4. METHODOLOGY

The methodology for calculating city-level RDP can be divided into two broad categories. The first category is concerned with defining the outer bounds of cities, whereas the second is concerned with using nightlights to arrive at growth estimates of economic activity.

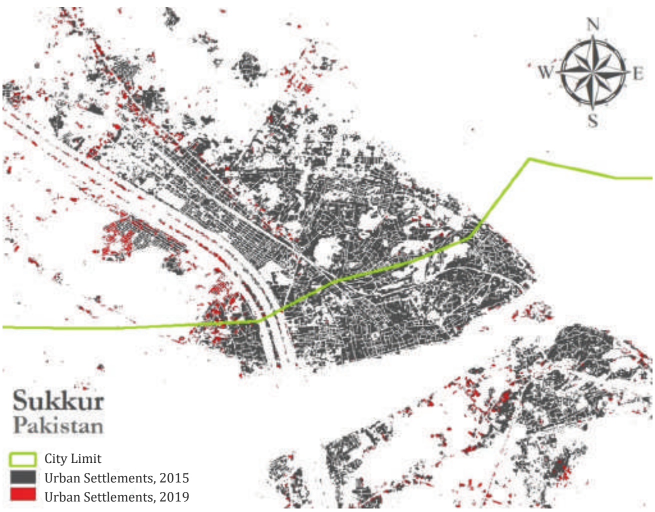
1. Defining Cities for an exercise of this nature presents itself with the typical challenge of whether to define an urban area as per its de jure boundary or to define bounds by the de facto extent of urban sprawl. From a researcher's point of view, both estimates can be valuable depending on the question at hand. For example, a researcher evaluating the impact of infrastructure development at the extensive margins of cities would consider the broader definition of a city, defined by its jurisdictional limits. This would also be a function of the fact that the source of spending would be the city government and, hence, development outside the jurisdictional bounds would bias estimates. On the other hand, a researcher studying urban-rural dynamics and structural change for an inland city would not be concerned with the de jure boundaries of the city but would simply look at the de facto spread of human settlement.

Figure 1: De jure vs de facto city limits of Gujranwala city



Source: WSF 2019.

Figure 2: Evolution of urban development between 2015 and 2019 in Sukkur, Sindh



Source: WSF 2015, 2019.

Since both definitions of city limits are valuable for research and policymaking, choosing one over the other would not be Pareto optimal. With this in mind, the present study developed growth estimates using both definitions. The first was the de jure method, which looked at growth in economic activity within legal city boundaries as defined in Dataset 4. The second definition used the richness of the WSF 2019 and the WSF Evolution to study the growth of economic activity within and at the boundaries of built-up areas as identified in WSF.

2. Measuring RDP from VIIRS NTL involves dealing with several complex tradeoffs along with quality, coverage, and time dimensions. Satellite data was sourced from two satellites, namely, the Defense Meteorological Satellite Program (DMSP) and the Suomi NPP VIIRS/DNB satellite, which carries the VIIRS sensor. The DMSP has the advantage that it provides data over a larger period and allows to look at data as far back as 1992. However, with the launch of the VIIRS satellite in 2012, the DMSP NTL data was discontinued. VIIRS provides monthly NTL data at a much finer resolution as compared to the DMSP but only starting from 2012. It also allows for significantly improved discounting of distortions caused by clouds as well as ephemeral events such as gas flares and biomass burning. In Pakistan's case, biomass burning is an important source of bias due to the large base of agricultural production.

Once the appropriate choice of NTL data has been made, the raster must be converted, cleaned, reclassified, clipped, and then averaged over the preferred geographic area for which NTL density is required. Annual composites allow for tracking the growth in NTL density over the years at very granular levels.

5. SPATIAL DISTRIBUTION OF ECONOMIC ACTIVITY IN PAKISTAN

This section presents data on the spatial or cross-sectional distribution of economic activity in Pakistan, proxied by NTL. Before doing so, it is important to investigate whether NTL data is indeed a good proxy for official use of GDP data.

Cross-Country Comparison

Since national accounts are available only at the federal level for Pakistan, a benchmarking exercise was conducted by comparing Pakistan to a set of other developing countries. Table 1 presents cross-country estimates of GDP using the NTL methodology developed in this paper for Pakistan and several comparable developing countries. To benchmark these to national accounts, the table also presents per-capita GDP (World Bank WDI) figures for all countries. The third column presents the share of NTL per million residents, which normalises NTL data by population and hence allows for a more analogous estimate of GDP per capita.

Table 1 provides key insights into the comparative validity of the CDP methodology developed in this paper. First, GDP per capita is nearly the same as NTL per million for countries that are at similar income levels. For example, India's per capita GDP relative to Pakistan is nearly the same as India's NTL per million estimates (1.37). This is also the case for Afghanistan and Colombia. Thus, within a band of per capita GDP, the NTL methodology is effective at capturing income differences. A second trend that becomes apparent from the cross-country comparisons of NTL-based estimates of GDP and national accounts data is that NTL estimates are elastic to large jumps in income levels. This can be seen in the case of Egypt and Mexico, both of which have large per capita differences relative to Pakistan. Here, the jump in NTL-based GDP measures is nearly twice that of national accounts data. This gives us a third key insight presented in Table 1, i.e. the important role of urban centres in developing agglomeration effects, which are captured both in national accounts and NTL-based measures. The large elasticity of the NTL-based measures in Mexico and Egypt may be capturing urban agglomeration effects in industrial cities like Cairo or Mexico City of upper-middle-income countries. This is confirmed by looking at the last column of Table 1, which provides the highest NTL value captured in any pixel for each country. The values for Egypt and Mexico are larger than Pakistan by a factor of 12 and 24, respectively.

Table 1: Cross-country comparisons of VIIRS VNL V2 nightlights, 2019 (Elvidge et al., 2021). Population & p.c. GDP estimates are taken from World Bank's World Development Indicators (WDI)

Country	Per Capita GDP		NTL (Aggregate)		NTL Per Million		Max	
	USD PPP	Rel.	Abs. (000s)	Rel.	Abs.	Rel.	Abs.	Rel.
Pakistan	**4,812**	**1.00**	**1,353.2**	**1**	**6126.1**	**1**	**532.5**	**1.00**
Afghanistan	2,078	0.43	89.0	0.07	2285.9	0.37	185.4	0.35
Bangladesh	5,138	1.07	264.4	0.20	1605.4	0.26	242.3	0.45
Colombia	14,931	3.10	943.4	0.70	18541.1	3.03	2842.0	5.34
Egypt	12,607	2.62	3,576.9	2.64	34953.3	5.71	6601.2	12.40
India	6,503	1.35	11,600.0	8.57	8393.3	1.37	636.2	1.19
Mexico	18,444	3.83	5,039.1	3.72	39083.2	6.38	13131.1	24.66

Other confounders may also include geographic factors, such as terrain ruggedness and forest cover. These factors can be controlled for in the cell-level analysis carried out in later sections. For this report, Table 1 provides sufficient support for the NTL-based measures of the GDP as a proxy for real GDP at lower levels of disaggregation.

Provincial RDP: Spatial Development Across Provinces

Table 2 presents the interprovincial spatial distribution of economic activity in Pakistan. The first column presents the share in the GDP of each province as estimated by the NTL-based methodology.

Table 2: Cross-province comparisons of VIIRS VNL V2 nightlights, 2019 (Elvidge et al., 2021). Population data is from PBS Census, 2017. GDP data taken from World Bank's World Development Indicators (WDI).

Province	NTL (Aggregate)		NTL Per Million		Max		Imputed RDP	
	Abs. (000s)	Share	Abs.	Rel.	Abs.	Rel.	RDP (PKR, Bil.)	P.C. RDP (PKR)
Pakistan	**1,233.9**	**1**	**6126.1**	**1**	**532.5**	**1.00**	**38,000**	**182,000**
Punjab	670.7	0.54	6098.1	0.99	319.2	0.60	20,656	188,000
Sindh	331.5	0.27	6927.2	1.13	532.5	1.00	10,209	213,000
Kh. Pakhtunkhwa	101.1	0.08	2847.6	0.46	196.5	0.37	3,113	102,000
Balochistan	61.0	0.05	4944.7	0.81	237.0	0.45	1,878	152,000
Islamabad CT	40.6	0.03	20249.7	3.31	75.1	0.14	1,249	624,000
AJK + GB	29.0	0.02	6236.6	1.02	10.6	0.02	893	190,000

Punjab

As can be seen in the first set of results, Punjab led with an estimated 54 per cent contribution to Pakistan's GDP. When normalised by population, the per-million estimate of GDP was almost identical to the national NTL-based

GDP for all of Pakistan. It is also important to note that despite being the largest contributor to economic activity, the province does not have the most productive geographic zone, i.e., Punjab's highest unit of NTL per cell was 60 per cent of the highest national NTL per cell, which is located in Karachi.

Spatial disaggregation of economic activity by province also allows to impute the per-capita RDP at the provincial level. With an estimated 54 per cent contribution to the GDP of Pakistan (PKR 38 trillion, 2019), the imputed per capita income in Punjab was PKR 188,000. This is very close to the national per capita income of PKR 182,000.

Sindh

The second largest contributor to Pakistan's GDP based on NTL-based GDP estimates was Sindh, which contributed 27 per cent to Pakistan's overall GDP. When normalised by population, Sindh's NTL per million was higher than the national level by 13 per cent. This implies that a segment of Sindh's population was relatively more productive than the national mean. This is barely a surprising result, as it is well known that Karachi is the economic capital of the country. This is reinforced by the province claiming the highest NTL per cell in all of Pakistan. As a result of Karachi's high productivity, the imputed per capita RDP for Sindh was PKR 213, 000.

Khyber-Pakhtunkhwa

As per NTL-based RDP estimates, Khyber Pakhtunkhwa province contributed 8 per cent to the GDP of Pakistan. After normalizing for population, the NTL per million turned out to be 46 per cent of the national NTL per million. This is indicative of the relatively smaller industrial base in Peshawar when compared to cities like Karachi or Faisalabad. It is also important to note that KP's estimates included the newly merged tribal districts (ex-FATA), which have one of the lowest levels of economic development in the country.

Khyber Pakhtunkhwa's imputed RDP using the NTL-based approach shows a per capita income of PKR 102, 000. This shows tremendous potential for interventions targeting productivity gains in the province. This is particularly important since KP districts have one of the fastest-growing young populations in the country. Moreover, the most low-hanging fruit in terms of income gains in the KP is the newly merged tribal districts which currently have mostly a subsistence economy.

Balochistan

NTL-based estimates show that Balochistan contributed approximately 8 per cent to the national economy of Pakistan. At 0.81 of the national mean, the NTL per million score of Balochistan was higher than that of the KP largely due to the smaller population base which was the denominator of the NTL per million calculations. Imputed er capita income for Balochistan is approximately PKR 152, 000 which again is higher than KP owing to a smaller population base. As we will see in the next section, there are stark district-level differences implicit in this provincial estimate for Balochistan.

Islamabad Capital Territory

The estimates for the Islamabad Capital Territory show the first empirical evidence so far of the importance of cities and urban centres for economic development. Despite having a very small industrial base and an economy largely centred around the seat of government, Islamabad's NTL per million estimate was more than three times the overall national mean. Moreover, at PKR 624, 000, the imputed per capita income for the city was also more than three times the national per capita income. This estimate signifies the importance of an NTL-based City Development Product (CDP) to compare how cities fare in terms of the distribution of economic activity and growth over time.

District RDP: Understanding Variation within Provinces

A similar exercise can be conducted for all districts of Pakistan to understand the regionality of economic activity within provinces. Table 3 presents descriptive statistics of the most productive districts of Pakistan. Not surprisingly, all these districts are the most urbanized districts of Pakistan.

A second significant observation is the large differences in per capita income imputed with the help of NTL-based RDP estimates. Compared to their provincial means, smaller cities like Kasur, Sheikhupura, and Attock had significantly higher per capita incomes. The same is true for cities like Lahore and Karachi, which is not a surprise. The significant income differences for smaller cities point to the important agglomeration effects induced by smaller cities. It is worth remembering that compared to their larger counterparts, local governments in these smaller cities have little to no fiscal resources that can be used towards public goods provision. This further strengthens the case for a city-level analysis of economic growth in Pakistan's cities.

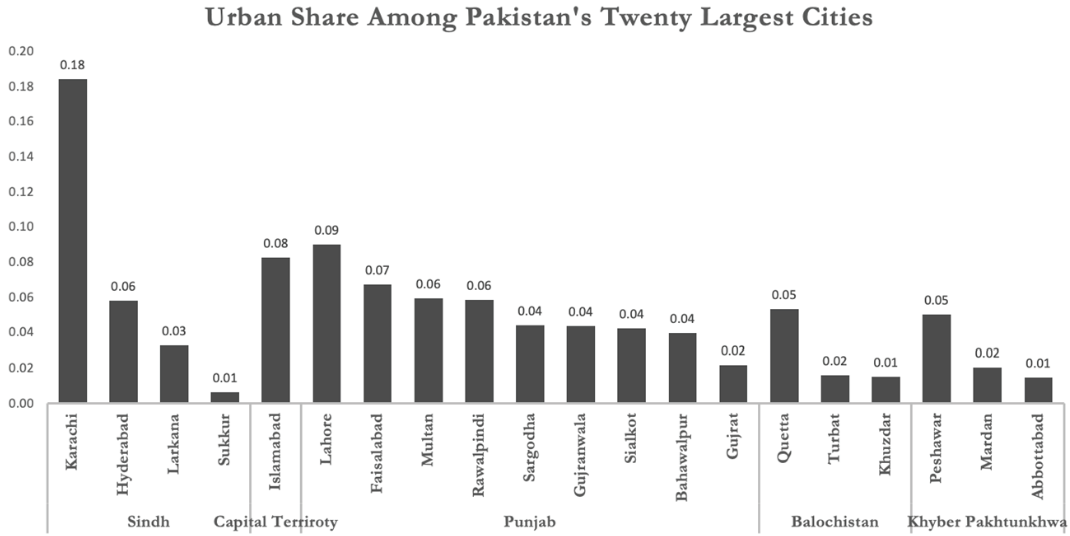
6. CITY-LEVEL DEVELOPMENT PRODUCT: THE TWENTY LARGEST CITIES

This section discusses preliminary findings of CDP estimates for Pakistan's 20 largest cities. Figure 3 below shows the urban share of economic activity in each city. Among the twenty largest cities, the city of Karachi alone contributed 18 per cent of the urban economic output. By magnitude, Karachi was by far the largest urban centre of economic activity, followed by Lahore, Faisalabad, and Islamabad with approximately 7 per cent share each. After megacities, Pakistan's medium-sized cities' contribution to Pakistan's economy is significant. Cities such as Hyderabad, Gujranwala, and Rawalpindi together contribute roughly the same economic output as Karachi. This may be due to the presence of small and medium-sized manufacturing clusters concentrated in each of these cities.

Table 3: Districts of Pakistan with highest VIIRS VNL V2 nightlights, 2019 (Elvidge et al., 2021). Population data is from PBS Census, 2017. GDP data taken from World Bank's World Development Indicators (WDI).

Rank	District	NTL (Aggregate)	NTL Per Million	Imputed RDP	
		Abs. (000s)	Abs.	RDP (PKR, Bil.)	P.C. RDP (PKR)
1	Lahore	87,185	7,840.4	2,685	242,000
2	Rawalpindi	51,504	9,533.6	1,586	294,000
3	Faisalabad	47,776	5,934.2	1,471	187,000
4	Karachi (*N=10*)	45,238	5,934.2	1,393	724,000
5	Islamabad	40,567	5,934.2	1,249	624,000
6	Kasur	29481.35	8533.245	908	263,000
7	Rahim Yar Khan	28165.33	5858.304	867	180,000
8	Sheikhupura	27772.14	8026.621	855	247,000
9	Multan	27422.66	5777.856	845	178,000
10	Attock	26207.16	13892.85	807	428,000

Figure 3: Economic activity in Pakistan's largest cities



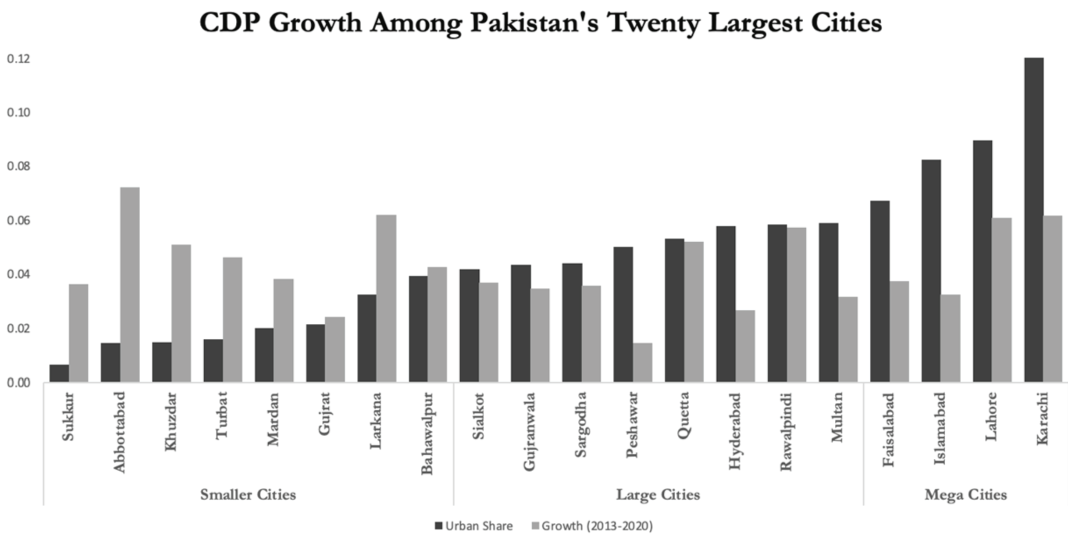
It is important to point out that the CDP estimates presented in Figure 3 are based on de jure city limits. As explained in the previous sections, many of the smaller cities in Pakistan have outgrown their defined city limits and have large de facto extensions in neighbouring administrative districts (Figure 2). Since these de facto city limits are not factored into the CDP estimates, our results may be biased downwards. A natural next step would be to account for this bias by expanding the city limits in the analysis.

CDP Growth – 2013 to 2020:

The previous sections developed an understanding of the spatial distribution of economic activity across Pakistan's provinces, districts, and cities. This section explains if these patterns result in differential rates of growth. As the largest urban population centres, these cities can safely be assumed as the most dynamic and fast-growing regions of Pakistan. In a way, the CDP represents the ceiling of economic growth in Pakistan. Figure 4 below presents the growth rates for each city along with the share of each city's CDP in the CDP of twenty cities.

Figure 4 has many important insights. First, Pakistan's cities are growing at a fast pace. Second, this growth is not restricted to megacities, such as Karachi and Lahore. In fact, the growth rate of smaller cities and medium-large cities is much higher than some of the megacities. To understand the determinants of this growth, a growth accounting exercise can be carried out by mapping this data onto socioeconomic and demographic data from the 2017 Population Census of Pakistan.

Figure 4: CDP growth rates for Pakistan's twenty largest cities



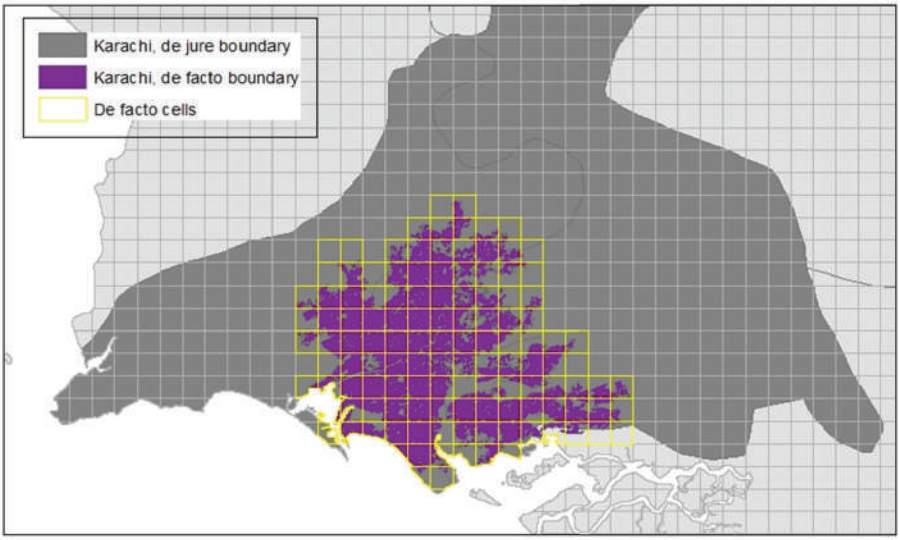
7. THE FINAL CDP DATASET

This section presents details of the final CDP dataset. This dataset is provided in the form of a spatiotemporal grid. First, an overview of the dataset is presented. Second, the possible methods of aggregation, including city, tehsil, and district-level aggregation are explained. Third, all key variables are defined and how these variables are constructed is explained. Finally, illustrations of the data at the grid-cell level are provided.

The final CDP dataset is constructed as a national spatiotemporal gridded dataset. This format divides the land area of Pakistan into square grid cells, each measuring 0.025 x 0.025 decimal degrees, or approximately 7.5 km^2. This approach results in 132,948 cell-level observations. Furthermore, the temporal coverage of the CDP data spans from 2013 to 2020, which means that the panel provides seven observations for each grid cell. The full CDP dataset, therefore, has over 1 million observations, with NTL-based RDP estimates provided for each 7.5 km2 grid cell.

The high resolution of the CDP grid provides proxy estimates of economic activity at very granular levels. Each grid cell has multiple geographic identifiers to allow researchers to collapse the dataset at higher levels of aggregation. Figures 5a and 5b present illustrations of this setup. Collapsing by de facto jurisdictions of Karachi aggregates all yellow cells in Figure 5a, whereas collapsing by de jure jurisdictions results in all yellow cells in Figure 5b being aggregated. Similarly, collapsing by tehsil aggregates NTL cells by tehsil, whereas collapsing by district aggregates all cells by district. This flexibility provides researchers with the freedom to choose the level of aggregation that is suitable for the research question under study.

Figure 5: Collapsing the CDP Dataset by de facto boundary (Figure 5a, top) and by de jure boundary (Figure 5b, bottom).



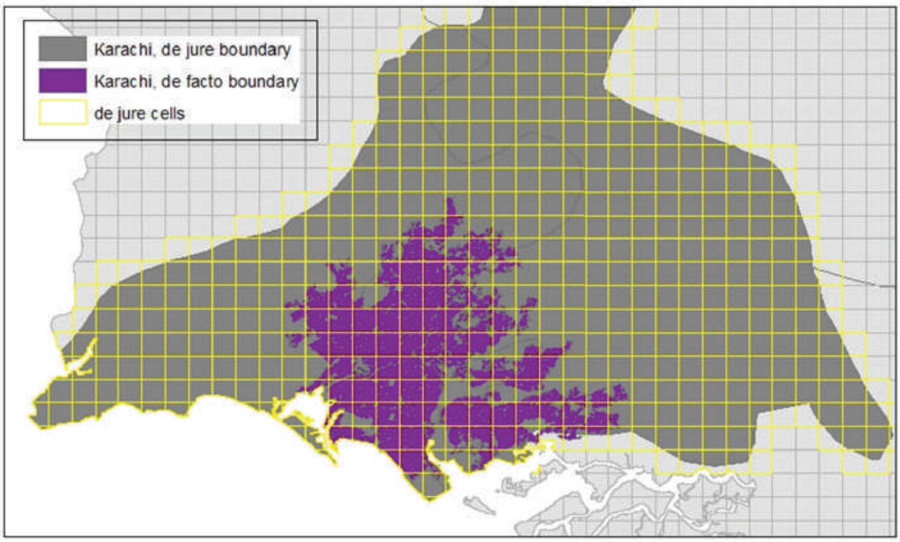


Figure 6 shows an overview of the extent of the CDP dataset. Darker shades of red and yellow show higher economic activity, captured through nightlight density, while Figure 6 only shows a macro view of the data. The true variation can only be captured by zooming in and analyzing economic activity at the level of grid cells. Figure 7, for example, zooms in to explore economic activity in Islamabad and Rawalpindi.

Figure 6: Distribution of economic activity in Pakistan as measured by NTL-based quantiles

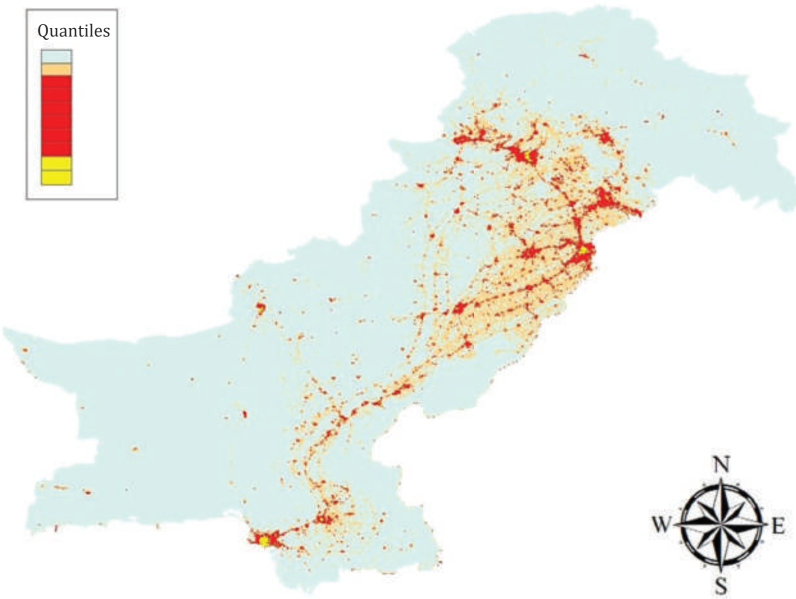
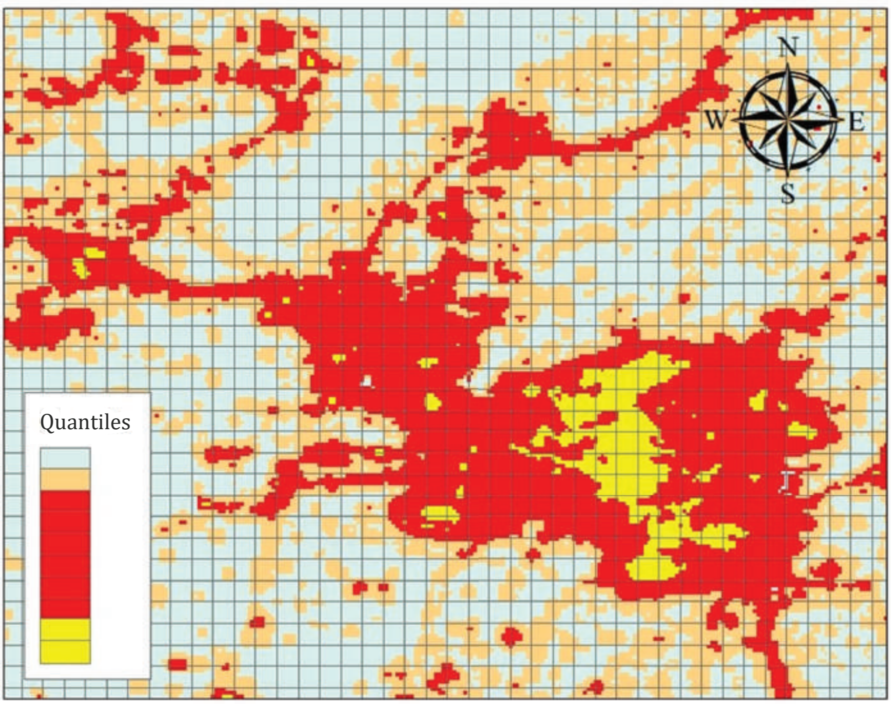


Figure 7: Economic activity in Islamabad and Rawalpindi



Overall, the CDP dataset provides high-resolution, granular proxy data of economic activity across Pakistan. The data can be spatially matched to other geospatial data sources and can also be used for RDP estimation. The CDP also provides cell-level population estimates, among other variables. All key variables are defined in the following section.

Key Variables

Cell_ID: a unique cell-level identifier for each grid cell. The full dataset has seven observations for each cell_id value, i.e., one for each year.

City_dejure: Identifies if a given grid cell is within the de jure limits of its associated city.

Cite_defacto: Identifies if a given grid cell is within the de facto limits of its associated city. De facto city limits were constructed using the World Settlement Footprint dataset. The WSF raster data was used to construct GIS polygons mapping the extent of the built-up area for each city.

ntl_lumin: mean the nightlight density at the grid cell level from 2013 to 2020. The variable was constructed using processed annual composites of the VIIRS NTL for 2012-2020 (Elvidge et al., 2021).

Wsf_area: Measures the built-up area (buildings) in m^2 for each grid cell.

Xcoord: the longitude of the grid cell centroid.

Ycoord: the latitude of the grid cell centroid

Cell_pop2017: cell-level population estimates based on the 2017 Census. Population data were matched at the level of a district and disaggregated using a consistent methodology. These estimates were constructed as follows. Each cell in a district was categorised as either urban or rural. All cells where more than 80 per cent of the land area was unbuilt, was identified as a rural cell. Likewise, all cells where 20 per cent or more of the land area was built, were identified as urban lands. District-level urban population counts were divided across urban cells, with the population weighted by the share of the built-up area in each grid cell. Similarly, district-level rural population counts were then divided based on the number of pixels per rural cell. In the aggregate, the total population of all cells in a district match the total district population. Population counts are also provided by gender.

Cell_pop1998: provides cell-level population estimates from the 1998 Census. The methodology used to construct this variable is the same as cell_pop2017.

8. CONCLUSION

The project to develop city-level RDP estimates is a promising avenue for building data architecture for economic and urban development in Pakistan. Analogous in ambition to some of the leading high-resolution geographic datasets, such as Asher et al. (2019), the CDP dataset has the potential to generate a wave of economic research centred around macro-development in Pakistan. Throughout this project, several natural extensions of this project have emerged. These include merging the dataset geographically with survey data from PSLM and from Census 2017. To complete this task, the RDP data will have to be extended beyond cities to cover the entire administrative landscape of the country. A second extension would be to use machine learning algorithms to accurately estimate agricultural production in near real-time. Given its potential, the RDP dataset for Pakistan can be a significant value-addition to Pakistan's research dataverse.

REFERENCES

Asher, S., Lunt, T., Matsuura, R., & Novosad, P. (2019). The socioeconomic high-resolution rural-urban geographic dataset on India (SHRUG). URL: https://doi. org/10.7910/DVN/DPESAK.

Casaburi, L., & Troiano, U. (2016). Ghost-house busters: The electoral response to a large anti–tax evasion program. *The Quarterly Journal of Economics,* 131(1), 273-314.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review,* 102(2), 994-1028.

Holmes, T. J., & Lee, S. (2012). Economies of density versus natural advantage: Crop choice on the back forty. *Review of Economics and Statistics,* 94(1), 1-19

Kudamatsu, M., Persson, T., & Strömberg, D. (2012). Web Appendix for Weather and Infant Mortality in Africa.

Lee, Y. S. (2018). International isolation and regional inequality: Evidence from sanctions on North Korea. *Journal of Urban Economics,* 103, 34-51.

Maue, C. C., Burke, M., & Emerick, K. J. (2020). *Productivity dispersion and persistence among the world's most numerous firms* (No. w26924). National Bureau of Economic Research.

Page, L., & Pande, R. (2018). Ending global poverty: Why money isn't enough. *Journal of Economic Perspectives,* 32(4), 173-200.

Turner, M. A., Haughwout, A., & Van Der Klaauw, W. (2014). Land use regulation and welfare. *Econometrica,* 82(4), 1341-1403.

Pérez-Sindín, X. S., Chen, T. H. K., & Prishchepov, A. V. (2021). Are night-time lights a good proxy of economic activity in rural areas in middle and low-income countries? Examining the empirical evidence from Colombia. *Remote Sensing Applications: Society and Environment,* 24, 100647.

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., ... & Strano, E. (2020). Outlining where humans live, the World Settlement Footprint 2015. *Scientific Data,* 7(1), 242.

Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F. C., & Taneja, J. (2021). Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019. *Remote Sensing,* 13(5), 922.

Marconcini, M., Metz-Marconcini, A., Esch, T., & Gorelick, N. (2021). Understanding current trends in global urbanisation-the world settlement footprint suite. *GI_Forum*, 9(1), 33-38.

AN EVALUATION OF DIFFERENT TAX REFORM PROPOSALS

Muhammad Nadeem Sarwar

ABSTRACT

Taxes are an important tool of fiscal policy for generating revenues and enabling governments to finance current and development expenditures. However, taxes leave individuals and firms with less income and, therefore, they have to compromise their consumption. Thus, the taxation system of a country affects its economic growth and the welfare of the people. A good taxation system should generate sufficient resources for the government without overburdening households and firms. Since a change in tax policy has far-reaching consequences for various interconnected economic agents, in this paper the impact of tax reforms was analysed using a general equilibrium approach, considering the interrelationships between all the sectors of the economy. The study used the computable general equilibrium model to quantify the impact of changes in direct and indirect tax rate policies on various economic indicators including economic growth, consumption, investment, exports, sectoral shifts, and income. For this, first a social accounting matrix based on 2017 data was developed and simulations were run subsequently. The results show that in the long run under the unbalanced budget condition, reducing personal income tax rates results in increased consumption expenditures, government expenditures, and incomes of various types of labour, but decreased economic growth and exports. However, introducing a flat and low-income tax rate along with decreasing corporate tax, sales tax, and customs duties results in higher economic growth, exports, consumption expenditures, and household income. On the other hand, a balanced budget condition produces better economic results. Therefore, it is recommended to simplify the tax regime by abolishing inefficient and distortionary taxes and reducing rates of all types of taxes and controlling government expenditures to keep the budget in balance even in the short run as a suitable policy for economic growth and household welfare.

1. INTRODUCTION

To provide people with public goods, infrastructure, and foster economic activities, governments need funds which are collected through various means including taxation, foreign aid and borrowing. However, after the Global Financial Crisis of 2008, it was realised that domestic resource mobilisation is the only sustainable and reliable way to finance such public expenditures (Fossat & Bua, 2013; Gordon, 2010; Keen, 2012). In this context, taxes of various kinds become important fiscal policy tools that are also used for stabilising the economy and income redistribution (Wawire, 2017).

A good taxation system should be efficient and equitable. However, there seems a tradeoff between these two. This has generated an ongoing debate about optimal taxation theories (see Feldstein, 1973; Martimort, 2001). There is also rich literature on the relationship between taxation and economic growth (see Engen & Skinner, 1996; Gemmell, 1988; Goulder & Summers, 1989; Lee & Gordon, 2005). These studies reach different conclusions while investigating the relationship. According to Auerbach (1996) and Eicher et al. (2003), these contradictory results are because of different socioeconomic and political systems prevailing in different countries. Therefore, while developing a comprehensive, efficient, and equitable taxation system, governments must take a proper account of the system's macroeconomic and distributional impacts (Sahn & Younger, 2000).

In the literature, the taxation–economic growth nexus and the impact of tax reforms is usually analysed using the general equilibrium approach by considering the interrelationships between all the sectors of the economy. Such an analysis shows the complete picture of the economy and gauges the effects of any tax policy change on all the sectors of the economy. Unfortunately, there is no such study on Pakistan that discusses the relationship between various kinds of taxes and macroeconomic indicators to evaluate various tax reform proposals by studying their impact on economic growth, fiscal deficit, exports, and income. This study aims to fill this gap.

Tax Structure in Pakistan

The structure of taxes in Pakistan is quite complex. There are multiple taxes in two broader categories of direct and indirect taxes. Direct tax includes income tax, which is further divided into income and corporate tax, wealth tax, corporate value tax (CVT), workers' welfare fund (WWF), and Workers' Profit Participation Fund (WPPF). Similarly, the broad categories of indirect taxes include customs duties, federal excise duty (FED), and sales tax (ST) from domestic production and imports. The broader categories are further subdivided into many sub-categories and the frequency of these taxes is also different. The sales tax on services falls in the domain of provinces and, therefore, its rates are decided by provinces and the revenue is collected by provincial tax authorities. Differences in sales tax rates in different provinces create additional issues for firms that operate in multiple provinces. Moreover, the share of indirect taxes is higher (60 per cent or above) in the total revenue collection and out of it, most of the taxes are collected from the international trade of goods and services. Therefore, this has additional effects on productivity, resource utilisation, balance of payments, and economic growth (Jamal & Javed, 2013; Pasha & Ghaus-Pasha, 2015). Moreover, a part of taxes is collected through withholding tax, which is by nature an indirect tax and has additional compliance costs. The revenue collected through different kinds of taxes is given in Table 1.

Table 1.1: Breakdown of Federal Tax Revenues (Rs. in Billion)

	FY13	FY14	FY15	FY16	FY17	FY18	FY19	FY20	FY21
Fed. Tax Revenue (a + b)	**1,946**	**2,255**	**2,590**	**3,112**	**3,368**	**3,844**	**3,828**	**3,997**	**4,745**
a. Direct Tax	743	877	1034	1217	1344	1537	1446	1523	1731
I. Income Tax	723	855	1,007	1,192	1,324	1,515	1,426	1,502	1,711
II. Wealth Tax	0	0	0	0	0	0	0	0	0
III. CVT	1	1	1	2	2	5	5	2	0
IV. WWF/WPPF	20	21	26	23	18	16	14	19	20
b. Indirect Tax	1,203	1,377	1,556	1,895	2,024	2,307	2,383	2,474	3,014
I. Custom Duty	239	243	306	405	497	608	686	627	748
II. FED	121	138	162	188	198	213	238	250	277
III. ST (Import)	430	495	553	678	703	824	810	876	1,116
IV. ST (Domestic)	413	501	535	624	626	661	649	721	872

Source: FBR Revenue Division Year Book 2020-21

An overview of the fiscal indicators shows that Pakistan's fiscal issues are severe. High fiscal deficit, low tax-to-GDP ratio and greater reliance on indirect taxes are making it difficult for the government to finance public expenditures. As a result, expenditures on human resources, law and order, and important infrastructure projects are low. This low spending may compromise future GDP growth as well. Therefore, there is a need to reform the tax structure to increase tax collection. In the recent past, the government attempted to experiment with decreasing the number of slabs and personal income tax rates. However, most of the changes introduced were undone after a few months only. Therefore, we do not have actual outcomes to study the impact of the changes. Moreover, a group of tax experts proposed to limit the number of taxes to four only, a flat income tax rate, a low corporate income tax rate, and flat and low sales tax and customs duties. According to the experts, the simplified and low rate-based system will help to boost economic activity and, ultimately, result in higher tax collection and increase national wealth.

Scope of Research

The study aims to identify and quantify the direction and magnitude of impacts of reducing the marginal income tax rate, decreasing the number of slabs, and introducing flat income and corporate tax rates with a reduction in sales tax and customs duties on the economy at both macro and micro levels. This includes the effects of such changes on economic growth, private consumption, investment, government budget, sectoral impacts, and labour income.

This is the first study in Pakistan that uses the computable general equilibrium (CGE) model to analyse the proposed tax reforms, especially in the income tax system. We utilised the latest input-output (IO) table, an updated social accounting matrix (SAM) based on 2017 data from the Labour Force Survey (LFS) and the Household Integrated Economic Survey (HIES). This study hopes to add to the debate on income tax issues in developing economies and reforming taxation systems in developing countries.

The plan of the study is as follows. Section 2 presents a brief review of the literature. Section 3 is on the social accounting matrix followed by a section on the CGE model. Section 5 presents the results and discusses the findings, followed by the concluding session.

2. REVIEW OF LITERATURE

The recent literature has concentrated on studying the effects of fiscal stimulus through tax cuts and increases in government expenditures on economic and social indicators. Hamilton & Whalley (1989) evaluated the outcomes of various changes in the Canadian indirect taxation system using a general equilibrium tax model. The results showed an improvement in both welfare and revenue collection by adopting a broad-based sales tax instead of federal or provincial sales taxes. Fortin et al. (1997) examined the impact of taxation and wage-setting in a developing economy with an informal sector. Analysis using the CGE model showed that an increase in corporate taxes, payroll taxes, and minimum wage rate led to growth in the informal sector, an increase in unemployment, and efficiency costs. Diao et al. (1998) used a dynamic general equilibrium model to study various debt management policies in the Turkish economy and concluded that although reliance on indirect taxes had distortionary effects and resulted in the loss of welfare, fiscal targets were achieved.

Knudsen et al. (1998) studied the Danish tax reforms of 1993 using a dynamic CGE model. The simulations showed that reducing taxes, the progressivity of labour income taxation, and a restructuring of capital income taxation resulted in the accumulation of wealth and increased consumption. The reforms brought Pareto improvement. Damuri & Perdana (2003) studied the effect of a 20 per cent increase in government spending under different financing conditions on income distribution and poverty in Indonesia using a comparative static CGE model. They found that an increase in spending had a significant and large positive impact on the GDP if it was not followed by an increase in taxes and financed through an increase in loans. However, Begg et al. (2003) found the opposite results as an increase in spending financed by an increase in income taxes showed an improvement in GDP through the balanced budget multiplier effect. On the same lines, Mabugu et al. (2013) studied the impact of a 6 per cent increase in government spending on South Africa's economy using the dynamic CGE model. They concluded that an increase in government spending resulted in higher GDP no matter if it was financed through a higher income or output tax, or all the taxes.

Mountford & Uhlig (2009) analysed the impact of changes in tax on the economy and concluded that unanticipated deficit-financed tax cuts stimulated the economy in the short term. However, the growing deficit might have consequences in the long run which overweigh the short-term gains. Cororaton & Orden (2010) evaluated the effects on poverty reduction of trade liberalization when tariff revenue was replaced with either direct or indirect taxes to keep the government budget balanced, with a greater reduction in poverty when a direct tax was imposed. Romer & Romer (2010) found that tax changes had very large effects on output and investment. Particularly, they showed that an exogenous tax increase of one per cent of GDP lowered real GDP by approximately three per cent. Amir et al. (2013) identified and quantified the impacts of income tax reforms on the Indonesian economy using key macroeconomic and socioeconomic indicators. The results of the CGE model showed that reducing income tax and introducing a low and flat tax rate for corporate tax led to higher economic growth and poverty reduction.

Gale & Samwick (2014) suggested that though the tax cuts may encourage individuals to work, save, and invest more, such policy must be backed by spending cuts to avoid large deficits. Otherwise, it may result in reducing national savings, increasing interest rates and, thus, a drop-in investment in the long run. Hasudungan & Sabaruddin (2016) investigated the impact of choosing between increasing borrowing to support increased government expenditures or simultaneous increase in both borrowing and exogenous output tax rates or a reduction in subsidies on the Indonesian economy using the CGE model. The simulations showed that the first proposal improved GDP but also increased the fiscal deficit, whereas the other two alternatives resulted in lowering the GDP because both resulted in increasing the cost of production and thereby increaseing inflation and decreasing consumption.

Huang & Rios (2016) derive the framework for optimal taxation when households are involved in tax evasion. The paper derives the mix of linear optimal consumption and non-linear optimal income tax for redistribution purposes. It is assumed that consumption taxes are enforceable, while income taxes can be evaded. To achieve the goal of income redistribution in economies with low compliance, the two tax instruments are complementary. As the social planner puts more weight on the lower-ability households, the income tax becomes more progressive, but the optimal consumption tax rate also increases because of higher evasion at higher marginal tax rates.

Hussain & Malik (2016) investigated the asymmetric response of output to changes in average marginal tax rates using Romer & Romer's (2010) data and found that only a tax decrease resulted in a significant and permanent increase in output whereas the tax increase had no significant impact. Using a simple model, it was shown that this asymmetry was derived from the asymmetric response of individual consumption to change in taxes as households face asymmetric consumption adjustment costs. Bhattarai & Trzeciakiewicz (2017) developed a DSGE model and analysed the fiscal policy in the UK. The findings showed that public consumption and capital income tax were the most effective fiscal tools in the short and long runs, respectively, whereas public investment was effective in both short and long runs and transfer payments were the least effective tool. On the other hand, when the interest rate fell to a zero lower bound, the effectiveness of consumption taxes and public expenditures increased, and the income taxes became the least effective. The analysis also showed that non-Ricardian households make fiscal policy more effective and nominal rigidities enhance the effectiveness of public spending and consumption taxes and decrease the effectiveness of income taxes.

Giraldo & García (2018) examined the effects of changes in the tax system on economic growth, welfare, and income distribution in the Colombian economy using a CGE model. Considering three alternatives of increasing the VAT, extending the VAT to all products, or decreasing the corporate income tax by 20 per cent and a progressive income of the tax rate on wealthy people, they found that an increase in indirect taxes did not have a large significant impact on the welfare of low-income households and taxing production. Mertens & Montiel Olea (2018) provided empirical evidence that a cut in marginal tax rates increased output and decreased unemployment. Belayneh (2018) examined the impacts of a cut in direct taxes on macroeconomic variables, fiscal balance, income distribution, and the welfare of households using the dynamic CGE model. The simulations showed that such a reform would result in increasing the income of the households. However, non-poor urban households would enjoy more benefits. The manufacturing sector would receive more benefits from such reform than any other sector of the Ethiopian economy.

Abdisa (2018) studied the effect of tax reforms on major macroeconomic indicators in the Ethiopian economy of tax reforms using the dynamic CGE model. The results showed that reducing direct tax or increasing the sales tax would boost overall economic activity, whereas reducing tariffs would have negative consequences. Lin & Jia (2019) analysed the impact of taxes on energy production sectors energy, $CO2$, and the Chinese economy using a dynamic recursive CGE model. They found that the tax rate in the ad valorem tax system affected the GDP negatively, while the tax rate in a specific and fixed tax regime had a limited positive relationship with the GDP. Switching to a fixed tax system would also result in decreasing inflation. Nandi (2020) proposed and calibrated a DSGE model for the Indian economy to study the impact of fiscal policy shocks. The results showed that the GDP and employment were positively related to government spending, negative consumption tax reduced inflation and induced consumption, while negative labour income tax had an asymmetric effect on the economy. Results also showed that an increase in public investment did not crowd out private investment.

The US Senate approved a new tax plan that reduced almost all kinds of taxes. The supporters of this move argued the workers would enjoy higher wages, while the opponents argued that a reduction in government expenditure because of this would be costly for workers. Using Romer & Romer's (2010) average marginal tax rate data, Berisha (2020) studied the effect on middle-class workers' earnings of these changes. The results suggested that a one percentage point increase in tax liabilities (relative to the GDP) led to about a 1.5 per cent decrease in real GDP growth and a 0.5 per cent decrease in median weekly earnings. However, the direct effect of decreasing taxes on median weekly earnings was not statistically significant. The outcomes also suggested that deficit-driven tax increases contributed to lower median weekly earnings.

This review of selected literature shows that most economists view that a fiscal stimulus results in higher GDP and poverty. However, the choice of mechanism is critical, and the optimal choice depends on a particular economy's conditions. Moreover, we find only a few studies on Pakistan and even those are very limited in scope. For example, the study of Iqbal et al. (2019) looked at the impact of the GST only on household consumption patterns. Similarly, the focus of Ahmed et al. (2011) was on the GST only and it is conducted by using SAM for 2004, which is quite old now. Naqvi et al. (2011) covered agricultural income tax by using SAM 2001-02. A comparative study of different income tax rate proposals that examines the impact on key economic variables of Pakistan's economy is missing and the current study aims to fill this gap.

3. SOCIAL ACCOUNTING MATRIX 2017

A social accounting matrix (SAM) is one way to represent the economy. It is based on a single-entry accounting system, which assigns values to incomes and expenditures in a circular flow and records all the transactions in an economy (Breisinger et al., 2009; Dorosh et al., 2004). It is an extended form of national accounts in which different activities employ factors of production to produce various commodities, earn revenues or income by selling these commodities, and undertake expenditures to pay the factors, the government, and related industries. Similarly, households, who own the factors of production, spend their income on buying commodities and government earns taxes which are spent back on households and other sectors.

Mathematically, a SAM is a square matrix each row and column of which represents an account and each cell shows an expenditure made by the sector/agent (column) to purchase the goods or services of the sector/agents (row). The income-expenditure equality is maintained in the SAM. Thus, on one hand, macroeconomic consistency is maintained and, on the other, details of the income of the factors, expenditures of the households, and production of various goods and services are also recorded. Rich multisectoral data helps policymakers to quantify the impact of change in a policy on various sectors of the economy (Robinson et al., 2001).

Building a SAM requires collecting data from various sources such as input-output (IO) tables, national accounts, the desegrated balance of payment, fiscal account, household income and expenditures surveys, and labour force surveys. The rich information gathered from all these resources captures the heterogeneity of production activities, incomes and expenditures. This strongly interconnected information helps policymakers to perform structural analysis, and allows the study of the distributional impact of a change in a policy parameter.

The section describes the construction of an updated and highly desegregated SAM that can be used for various policy analyses by policymakers and researchers. First, a review of the literature is presented and previously developed SAMs for Pakistan are discussed. Second, the step-by-step methodology adopted to construct the SAM based on data for the financial year 2017 is elaborated. In the last part, some analysis based on the 2017 SAM is presented.

An Overview of SAMs of Pakistan's Economy [1]

The first social accounting matrix for Pakistan's economy was constructed in 1985 by the Pakistan Institute of Development Economics (PIDE). The base year for this SAM was 1979. The Federal Bureau of Statistics, in collaboration with the Dutch government, produced the second SAM for Pakistan in 1993, which was constructed based on 1984-85 data. It was one of the outputs of a research project aimed at improving the national accounting system of Pakistan. This SAM had one household only and, therefore, it was not suitable for a distributional analysis.

Siddiqui & Iqbal (1999) developed a comparatively large-sized SAM in which the factor account consisted of labour and capital. Total production of the economy was desegregated into agriculture, industry, health, education, and other sectors with each sector which were assumed to be participating in local and global markets. Two factors of production were owned by eight various types of households. The record of international trade was recorded in the capital account. This SAM was based on 1989-90 data. This SAM presented some disaggregation of the economy, which was helpful for distributional analysis. However, the detailed breakdown of the firms based on various production activities and the factors of production based on different skill levels of workers were missing.

Dorosh et al. (2004) developed the SAM for 2001–02. The SAM was specially developed to analyse the rural economy. It was composed of 34 activities, 33 commodities, and 27 classifications of factors of production, including land, labour and other factors. It consisted of 17 rural and 2 urban households, firms, government, and the rest of the world. However, this SAM used the same IO table used by Siddiqui & Iqbal (1999). Since it was developed for the analysis of the rural economy, its main segregation was based on that objective and, therefore, it provided very detailed information about the rural economy, but the decomposition of various industrial productive activities and labours based on their skill set were missing.

Waheed & Ezaki (2008) constructed another SAM using 1999-00 as the base year and used the same IO table used by Siddiqui & Iqbal (1999) and Dorosh et al. (2004). They expanded sectors and brought in financial institutions as well. The SAM by them also had two factors of production i.e., labour and capital, and six production sectors. The agents consisted of households, firms, the government, commercial banks, the central bank, and the rest of the world. This SAM integrated the financial sector with the real sector. However, the segregation of households and production sector was missing, therefore, it was not very useful for a distributional analysis.

Debowicz et al., (2012) worked on SAM for 2008-09 following the SAM 2001-02. This SAM consisted of 12 agriculture activities (as in SAM 2001-02) and segregated textile and chemical industries to make a total of 22 industrial activities. Similarly, trade, the transportation sector, household and private sector services were also further classified. The division of commodities and rural households was the same as in SAM 2001-02, but urban households were divided into three groups based on their income. Similarly, in institutional accounts, separate subaccounts were introduced for import taxes, direct taxes, and sales taxes.

Zeshan (2020) developed a SAM for Pakistan based on the 2014 data. This SAM included 65 activities, producing the same number of commodities. However, the desegregation was quite detailed for the agriculture sector but not as detailed for the industrial or services sector. There were 3 factors of production, namely, unskilled labour, skilled labour, and capital. Similarly, tax desegregation was also limited to production taxes, trade taxes, and direct taxes only. However, a quite detailed breakdown of the rest of the world account makes it ideal to use for multi-country analysis.

[1] This section only gives a brief review of the previously developed SAMs for Pakistan's economy so that the differences between the SAM developed for this study and those developed previously can be highlighted. The literature review that discusses tax reforms is presented in the previous section.

Parts of a Social Accounting Matrix

The main parts of a SAM are activities and commodities, domestic institutions, saving and investment, and the rest of the world. Each of these parts is composed of many parts which together complete the SAM. These parts are elaborated below.

By activities, we mean the producers of various goods and services and commodities mean the goods and services produced. These commodities are distinguished as an activity because a sector may produce more than one good. For example, the agriculture sector produces wheat, rice, cotton, pulses, etc. Similarly, a commodity can also be produced by more than one agent, like shoes can be produced by a small shoemaker and a large shoe-making firm. Therefore, the separation of the two provides better information.

An activity uses intermediate inputs produced by other activities and produces the final commodity with the help of the factors of production. These factors receive payment in the form of wages and rent, etc. as per their contribution to producing the commodities. Similarly, payments are also made to commodities for the use of intermediate inputs. Commodities, on the other hand, are supplied either by domestic activities or are imported. Indirect taxes, such as sales tax and taxes on imports (e.g., import tariffs) are paid on these commodities to the government. Therefore, the values in commodity accounts are at market prices.

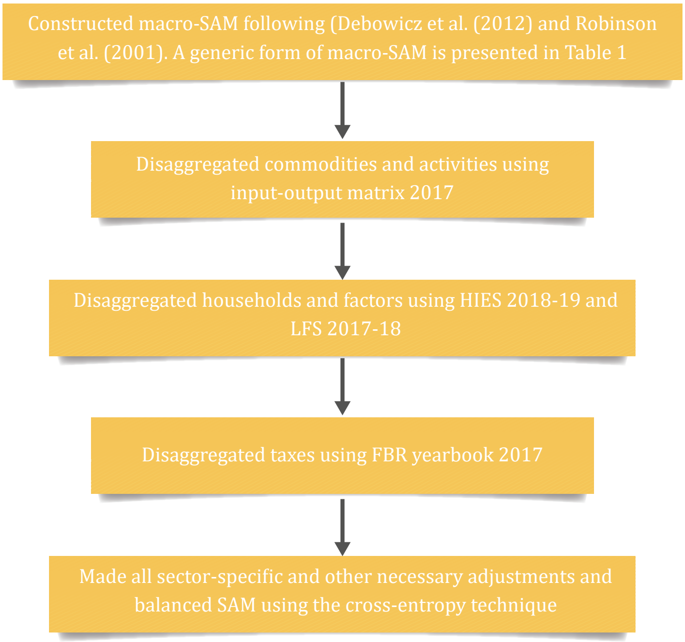
The information needed to construct detailed activities and commodities accounts usually come from the IO table, national accounts data, and some research on the share of various factors or value added by the factors.

One aspect that differentiates a SAM from an IO table is that a social accounting matrix, besides recording the flow of income and expenditures between activities and commodities, also gives information about the income and expenditures of various institutional accounts, such as households and government. Households usually own the factors of production and earn income by supplying these factors. Households may also receive some income from the government in the form of transfer payments and from the rest of the world in the form of remittances. On the expense side, households pay direct taxes to the government and purchase commodities to consume. The leftover income is saved. Different direct and indirect taxes are the income of the government. Other than this, the government may hold some capital and may earn income from engaging in some production activities. A part of the government income also comes from the rest of the work in the form of aid, grants, and development assistance. The government spends this income to make payments to factors it hires, transfer payments or subsidies to households or other activities, and the remaining amount, usually negative, is recorded in savings and the investment account. The information on the household account is usually found in national income accounts and household surveys, which are normally held regularly.

The investment comes from national savings, which is the sum of private and public savings. However, in open economies, like the one we have, the investment also comes from foreigners' savings. The information on capital inflow from the rest of the world comes from the balance of payment account, usually published by the central bank of the country.

Steps to Build a SAM

Figure 3.1 Building SAM



The macro-SAM is based on information from the National Income Accounts Handbook of Statistics on Pakistan Economy 2020 and FBR Yearbooks. These sources are published by the Finance Ministry, the State Bank of Pakistan, and the Federal Board of Revenue Pakistan, respectively. In line with Debowicz et al. (2012), Golan et al. (1994; 1997), and Robinson et al. (2001), throughout the analysis, the Bayesian view of efficient use of information is followed.

Current SAM

In developing SAM based on the Financial Year 2017, we used information on accounts from Debowicz et al. (2012), Dorosh et al. (2004), and Zeshan (2020). Debowicz et al. (2012) disaggregated the agriculture sector into various sub-categories with respect to the crop and farm size and location. Similarly, the activities in allied industries, such as textiles, trade and transport and services, were also disaggregated. In the current SAM, the agriculture sector was not split based on crops, however, mining and food, beverages, and tobacco were introduced separately because they are treated differently for tax purposes. Similarly, the textile and clothing sector was not split into various sub-categories. Instead, other manufacturing was split into various categories, such as electrical and optic equipment, rubber and plastic, chemical and chemical products, paper, printing and publishing, etc. Similarly, instead of rail and road transport, the term inland transport was used and other means of transport were defined as water transport and air transport. Similarly, transport-supporting activities, such as the services of travel and transport agencies, were introduced separately. Besides common public and private services, such as education, healthcare, and public administration, hotel and restaurant services were also introduced as these represent a growing tourism and hospitality industry. For most of the disaggregations described above, the IO 2017 Table was used and the main objective of the disaggregation was to facilitate industry-related analysis, such as changes in taxes, etc. the final 2017 SAM included 34 commodities produced by 34 activities with detailed disaggregation of industries and services sectors, but limited disaggregation of the agriculture sector. Detailed interconnections between various industries help gauge the impact of change in any such policy on various sectors and, thus, on the overall economy.

Next, we introduced 24 factors. Two basic economic factors of production, labour and capital, were divided into three categories, namely, low-skilled labour, high-skilled labour and capital. These three categories were further split into rural and urban geographies of all four provinces. The information was collected from Zeshan (2020), HIES 2017-18, and LFS 2017. The wage differences between skilled and skilled labour in rural and urban areas of four provinces were calculated from the LFS. The non-labour income was used as capital income.

We introduced 8 categories of households based on the rural-urban divide in each province. These eight types of households owned 24 factors of production, i.e., each household owned one unit of skilled labour, one unit of unskilled labour, and one unit of capital. However, some of the capital being used in the production process was also supplied by the rest of the world or by the government. The households earned an income equal to the value-added of these factors of production they own. Remittances from foreign and transfer payments from the government were the other sources of income for these households. Out of their income, they paid direct tax to the government, paid firms for consuming their goods and services and the leftover income s saved.

The government earns income by collecting tax revenue. Taxes are basically of two types – direct tax and indirect tax. While developing the current SAM, we considered various direct and indirect taxes such as income tax on individuals, firms and associations of persons etc. Besides these taxes, the government also receives income against the capital it owns and also receives loans, aids and grants from domestic and foreign financial institutions and agencies. On the expenditure side, the government provides public goods to the general public which needs various commodities as inputs. Similarly, the government needs services of various factors of production to enable itself to produce and supply public administration. It makes transfer payments and gives subsidies to households and firms. Along with all these, some of the government expenditures are on debt servicing. These expenditures include interest and principal payments to local and foreign financial institutions. The difference between government income and expenditures is called budget surplus (if it is positive) or budget deficit (if it is negative), which is recorded as public savings. The information on all such incomes and expenditures was obtained from the FBR Yearbook 2017, National Income Accounts, and Handbook of Statistics on Pakistan Economy 2020.

The rest of the world account records the flow of funds from and to foreign countries. These include payments made against imports, payments received against exports, the flow of remittances, capital payments and the flow of savings and loans, grants, and aid. The information on all these was obtained from the Balance of Payment (BOP) account published by the State Bank of Pakistan (SBP), National Income Accounts published by the Finance Ministry, and trade statistics published by SBP and Pakistan Bureau of Statistics (PBS). While developing the SAM, this account was not disaggregated, but it can be done using the IO table and information from the sources cited above.

After cross-checking each value from multiple sources and minimizing row-column sum differences, we used the cross entropy approach following Golan et al. (1994, 1997), Judge & Mittelhammer (2011), and Robinson et al. (2001).

Table 3.2 A Generic Schematic SAM

	Activities	Commodities	Labour	Capital	Households	Government	Changes in Stock	Savings	Rest of the World
Activities		Supply matrix							
Commodities	Intermediate consumption				Final private consumption	Final public consumption	Change in stocks	Fixed Investment	Exports
Labour	Value added by labour								
Capital	Value added by capital								
Households			Payment from labour	Payment from the capital to households		Transfer from the government to households			Remittances to households
Government		Indirect taxes		Payment from the capital to the government	Direct tax				Transfer from non-residents to Govt
Changes in Stock								Changes in stocks	
Savings				Household savings		Public savings			Foreign savings
Rest of the World		Imports		Repatriation of dividends		Payment to non-residents			

Based on Robinson et al. (2001) and D. Debowicz et al. (2012)

Structure of the Economy

A SAM gives very useful information which helps understand the structure of the economy. In this section, some economic insights from the social accounting matrix developed for this study are shared [2].

Shares of Factors of Production in Value-Added

Table 2.2 presents the value-added share of various factors of production. We combined 32 sectors into 8 sectors in the SAM, the detail of which is given in the appendix. This value-added breakup shows that Pakistan is a labour-intensive economy. The mining, metals, and chemical sector was the most capital-intensive sector. In this sector, 59.94 per cent share of the value-added went to the capital, 13.07 per cent to skilled labour, and 26.99 per cent to unskilled labour. Agriculture and food is the most labour-intensive sector in which 66.69 per cent of the value-added went to labour, 13.21 per cent to rural capital, and 20.10 per cent to urban capital.

The share of skilled labour was the highest in the services sector (38.41 per cent) and the lowest in the agriculture and food sector (8.21 per cent), whereas the share of unskilled labour was the highest in the agriculture sector (58.48 per cent), followed by textile, leather, and rubber sectors (41.38 per cent).

These findings clearly show that urban labour and capital received a higher income than rural labour and capital. This is because most of the industries, high-paying market activities, and skilled jobs are available in urban areas, whereas the economic activity in rural areas is mostly centred around agriculture activities. Similarly, most of the labour in Pakistan is low-skilled, therefore, collectively they earn higher than skilled labour. However, this also shows the limited ability of the economy to produce higher value-added products that require greater use of technology and skills. This drawback is one of the reasons for the low earnings of the workforce and poor economic growth.

[2] The full SAM is available upon request.

Table 3.2: Shares in Value Added

	Capital (Urban)	Capital (Rural)	Skilled Labour (Urban)	Skilled Labour (Rural)	Unskilled Labour (Urban)	Unskilled Labour (Rural)	Total
Agriculture & Food	20.102	13.205	5.458	2.753	39.343	19.138	100
Textile, Leather & Rubber	23.414	15.381	12.291	7.533	25.838	15.543	100
Mining, Metals & Chemical	36.173	23.763	7.720	5.353	15.670	11.321	100
Construction & Energy	23.105	15.178	12.918	9.213	23.181	16.404	100
Other Manufacturing	23.776	15.619	14.533	9.766	21.736	14.571	100
Trade	29.542	31.628	11.432	7.804	11.471	8.122	100
Transport & Communication	29.581	19.432	11.677	9.255	16.750	13.305	100
Other Services	27.200	17.868	21.523	16.888	9.479	7.042	100

Sectoral Trade Shares

Pakistan is an open economy. As shown in Table 3.3 below, textile, garments, leather and rubber was the largest exporting sector of Pakistan's economy with a 48.73 per cent share in total exports, whereas the construction and energy sectors were the most import-dependent sectors with a share of 30.58 per cent in imports. This is because Pakistan relies on imports for its energy needs, especially oil and gas. The other prominent sectors that strengthen Pakistan's link with the world through trade are agriculture and food, mining, metals and chemical, trade services, and transport and communication. The agriculture and food sector had a 19.85 per cent share in total exports and a 14.66 per cent share in imports. Pakistan exports fruits, vegetables, and rice, among other products, and imports edible oil, packed juices, dry milk etc. Similarly, the transport and communication sector had an import share of 16.01 per cent and an export share of 6.19 per cent

Table 3.3: Trade Share of Each Sector

	Agriculture & Food	Textile, Leader & Rubber	Mining, Metals & Chemical	Construction & Energy	Other Manufacturing	Trade	Transportation & Communication	Services
Import	14.66	7.46	18.21	30.58	4.29	3.31	16.01	5.49
Export	19.85	48.73	6.21	0.58	2.47	10.39	6.19	5.59

Note: Calculated by the author based on the SAM developed for this study.

Intermediate and Final Demand

Table 3.4, given below, shows the share of intermediate and final demand for each sector. Intermediate demand is the demand for goods and services that are used in the production process. The table shows that out of the total output of the mining, metals and chemicals sector, 94.71 per cent was used up as intermediate demand by other industries and the remaining was sold to final consumers. The shares of intermediate demand and final demand were 93.02 per cent and 6.98 per cent out of the output produced by the construction and energy sectors.

Table 3.4: Demand Breakup (%)

	Agriculture & Food	Textile, Leader & Rubber	Mining, Metals & Chemical	Construction & Energy	Other Manufacturing	Trade	Transportation & Communication	Services
Intermediate Demand	85.906	84.556	94.708	93.022	79.432	91.460	87.604	81.082
Final Demand	14.094	15.444	5.292	6.978	20.568	8.540	12.396	18.918

Note: Calculated by the author based on the SAM developed for this study.

4. COMPUTABLE GENERAL EQUILIBRIUM MODEL

To study the impact of various policy interventions on Pakistan's economy, researchers have utilised different CGE models. Siddiqui & Iqbal (2001) developed the CGE model for Pakistan and used it to analyse the impact of tariff reduction. The same model was used by Siddiqui et al. (2008) for studying the impact of fiscal and trade policy changes on poverty. Ahmed et al. (2011) used the CGE model developed by Poverty and Economic Policy (PEP) Research Network to examine the impact of changes in indirect taxes in Pakistan. Khan et al. (2018), Shaikh (2009), and Shaikh & Rahpoto, (2009) used the Global Trade Analysis Project (GTAP) model to investigate the effects of various trade-related policies on Pakistan's economy. Robinson & Gueneau (2013) used the basic CGE model developed by International Food Policy Research Institute (IFPRI) and extended it for exploring the impact of changes in water resources in the Indus River, especially focusing on the impact of water shocks on Pakistan's economy.

The main inspiration for developing a CGE model for this study was based on ORANI-G (J. Horridge, 2000; M. Horridge, 2003), Applied General Equilibrium Model for Fiscal Policy Analysis (AGEFIS) by Yusuf et al. (2007), Amir et al., (2013), Siddiqui & Iqbal (2001) and Siddiqui et al. (2008). ORANI-G is a generic CGE model developed for various kinds of policy analysis and it is being used in many countries with slight modifications whereas AGEFIS is a CGE model based on SAM unlike ORANI-G, which is based on IO table, specially developed for conducting fiscal policy analysis, which is aligned with the objectives of the current study. The last two studies clarify relevant modifications needed for specifying Pakistan's economy. The main differences between the CGE model developed for this study and one earlier developed by Siddiqui & Iqbal (2001) is that in the previously developed model, domestic production is divided into five sectors, whereas in the current model, we divide it into 34 sectors, labour is assumed to be homogenous in the model of Siddiqui & Iqbal (2001), whereas in our model we introduce 16 different types of labour based on geographical local and skill level and 8 categories of capital. Similarly, we also introduce eight different types of households based on rural-urban localities of each province whereas the older study included only one household. Other changes are discussed at relevant places. Because of these additions, we believe that the current model is more flexible as it can show mobility of labour and capital between different areas and sectors, the kind of labour, i.e., low skilled or high skilled, being chosen by different industries, the labour-capital intensity of various sectors, rate of unemployment, and wage rigidities. Since labour income is a major share of household earnings, the ability to study these labour market adjustments is an important addition to the model. Moreover, the model may also be used for analysing the impact of various policies on poverty and income inequality.

Following other CGE models, such as Dixon (2006), Dixon et al. (1982; 1992), and Dixon & Rimmer (2002), the equations of the model are linearised using percentage changes based on the Johansen approach instead of the levels of variables. Moreover, for each component of demand, the price formation process is described in various factors such as basic value, margin, taxes, etc. The main features of the model, such as the dimensions of the model and equation system and model closure are discussed in the following subsections.

Dimensions of the Model

This model consists of 34 commodities produced by a similar number of activities or industries and are also imported from the rest of the world. The commodity set was further divided into margins and non-margins. Margins are the commodities such as wholesale and retail trade, transportation, and supported services whereas non-margins are the rest of the commodities. Details on the sets of the model are summarised in Table 4.1.

Table 4.1: Details of Sets

Commodities (COM) and Activities (IND)		**Margins (MAR)**	
1	Agriculture, Hunting, Forestry, and Fishing	1	Sale, maintenance, and repair of motor vehicles and retail sale of fuel
2	Mining and Quarrying	2	Wholesale trade and commission trade, except for motor vehicles and motorcycles
3	Food, Beverages, and Tobacco	3	Retail trade, except for motor vehicles and motorcycles; repair of household goods
4	Textiles and Textile Products	4	Inland transport
5	Leather, Leather Products, and Footwear	5	Water transport
6	Wood and Wood Products	6	Air transport
7	Pulp, Paper, Paper Products, Printing, and Publishing	7	Other supporting and auxiliary transport activities
8	Coke, Refined Petroleum, and Nuclear Fuel	**Source (SRC)**	
9	Chemicals and Chemical Products	1	Domestic
10	Rubber and Plastics	2	Imported
11	Other Non-metallic Minerals	**Occupations (OCC)**	
12	Basic Metals and Fabricated Metal	1	Pun-r-low skilled
13	Machinery	2	Pun-r-high skilled
14	Electrical and Optical Equipment	3	Pun-u-low skilled
15	Transport Equipment	4	Pun-u-high skilled
16	Manufacturing, and Recycling	5	Sin-r-low skilled
17	Electricity, Gas, and Water Supply	6	Sin-r-high skilled
18	Construction	7	Sin-u-low skilled
19	Sale, Maintenance, and Repair of Motor Vehicles and Retail Sale of Fuel	8	Sin-u-high skilled
20	Wholesale Trade and Commission Trade, Except Motor Vehicles and Motorcycles	9	KP-r-low skilled
21	Retail Trade, except Motor Vehicles and Motorcycles; Repair of Household Goods	10	KP-r-high skilled

22	Hotels and Restaurants	11	KP-u-low skilled
23	Inland Transport	12	KP-u-high skilled
24	Water Transport	13	Bal-r-low skilled
25	Air Transport	14	Bal-r-high skilled
26	Other Supporting and Auxiliary Transport Activities	15	Bal-u-low skilled
27	Post and Telecommunications	16	Bal-u-high skilled
28	Financial Intermediation	**Capital (CAP)**	
29	Real Estate Activities	1	Pun-r
30	Renting and Other Business Activities	2	Pun-u
31	Public Administration and Defence	3	Sin-r
32	Education	4	Sin-u
33	Health and Social Work	5	KP-r
34	Other Community, Social, and Personal Services	6	KP-u
Households (HOU)		7	Bal-r
1	Pun-r	8	Bal-u
2	Pun-u		
3	Sin-r		
4	Sin-u		
5	KP-r		
6	KP-u		
7	Bal-r		
8	Bal-u		

Naming System

Variables, parameters, and coefficients of the CGE model are based on the following conventions:

Table 4.2: Name System

Symbol	Full Name	Symbol	Full Name
A	Change of Technology	bas	Basic – not including margins
del	Ordinary Change	cap	Capital
F	Shifting Variable	cif	Imports at border prices
P	Price in Local Currency Unit (LCU)	Imp	Imports (duty paid)
pf	Price in Foreign Currency	lab	Labour
S	Share of Input	lux	LES (supernumerary part)
SIGMA	Elasticity of Substitution	mar	Margins
T	Tax	**oct**	**Other Cost Tickets**
V	Value in LCU	prim	primary factors of production
W	Percentage-Change Value in LCU	pur	At purchasers' prices
X	Input Quantity	**sub**	**LES (Subsistence part)**
_c	Over COM	tar	Tariffs
_s	Over SRC (local + Imp)	tax	Taxes (indirect)
_i	Over IND	tot	Total (average) inputs for some users
_io	Over IND and OCC		
_o	Over OCC		
_gi	Over CAP		

These conventions adopted in the present study match with the GEMPACK system.

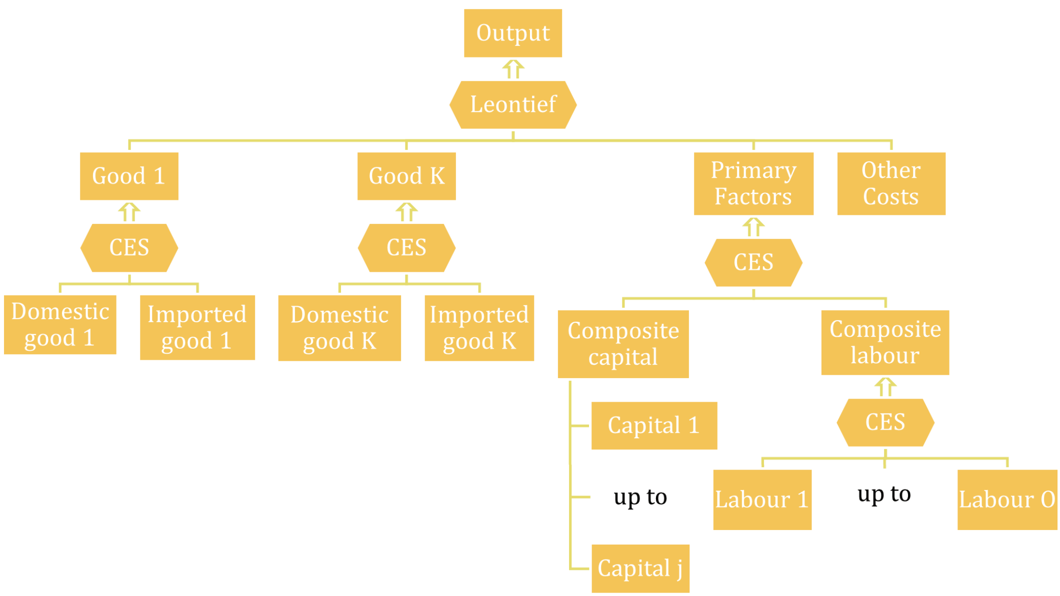
Production Block

In this model, 34 commodities were produced by 34 activities or industries, which means that all the industries were single output-producing industries. The input required to produce output consists of local or imported commodities, factors of production, and other inputs. Factors of production included eight types of capital and sixteen types of labour. The nesting structure of the model is presented in Figure 1 below.

The process represented in Figure 1 shows that production is a multistage process. The top nest of the output is produced using intermediate commodities as inputs, factors of production, and other costs, such as taxes, subsidies, etc. At this stage, the production process can be expressed using the Leontief technology according to which, all the inputs are combined in a fixed proportion to produce the output. Therefore, an excess supply of one or a few inputs will not guarantee higher output.

At the lowest stage, a nest represents the composition of intermediate inputs. The intermediate inputs may be domestic or imported. This choice of imported or domestic intermediate inputs depends on various factors such as prices in local and import markets and the elasticity of substitution between locally produced and imported inputs, which follows the constant elasticity of substitution (CES) parameter. All these decisions follow the cost minimisation principle while choosing local and domestic markets using Armington assumptions (Armington, 1969). In the second nest, the cost of primary factors is minimised using the CES function. Here, it is also assumed that a costly input will be substituted by a cheaper input. Both of the primary factors, composite labour and composite capital, are also formed using cost minimisation from various types of labour and capital, which, for this study, are 16 for labour and 8 for capital.

Figure 4.1: Production Structure



The equation that represents this production structure can be written as:

$$\mathrm{X1TOT(i)} = \frac{1}{\mathrm{A1TOT(i)}} \mathrm{MIN}\left(\mathrm{All, c, COM}\ \frac{\mathrm{X1\_S(c,i)}}{\mathrm{A1\_S(c,i)}}, \frac{\mathrm{X1PRIM(c,i)}}{\mathrm{A1PRIM(c,i)}}, \frac{\mathrm{X1OCT(i)}}{\mathrm{A1OCT(i)}}\right) \quad (4.1)$$

The A1TOT(i) represents the Hicks-neutral technical-change term.

The choice of using imported or local commodities as intermediate inputs can be represented as

$$X1\_S(c,i) = \text{CES}\left(\text{All}, s, \text{SRC}: \frac{X1(c,s,i)}{A1(c,s,i)}\right) \qquad (4.2)$$

The demand for primary input factors which follows the cost minimisation principle subject to production function is:

$$X1PRIM(i) = \text{CES}\left(\frac{X1LAB\_O(i)}{A1LAB\_O(i)}, \frac{X1CAP\_CA(i)}{A1CAP\_CA(i)}\right) \qquad (4.3)$$

The demand for labour and capital from various kinds of labour and capital can be represented by:

$$X1LAB\_O(i) = \text{CES}(All, o, OCC: X1LAB(i,o)) \qquad (4.4)$$

And

$$X1CAPT\_CA(i) = \text{CES}(All, o, CAPT: X1CAP(i,o)) \qquad (4.5)$$

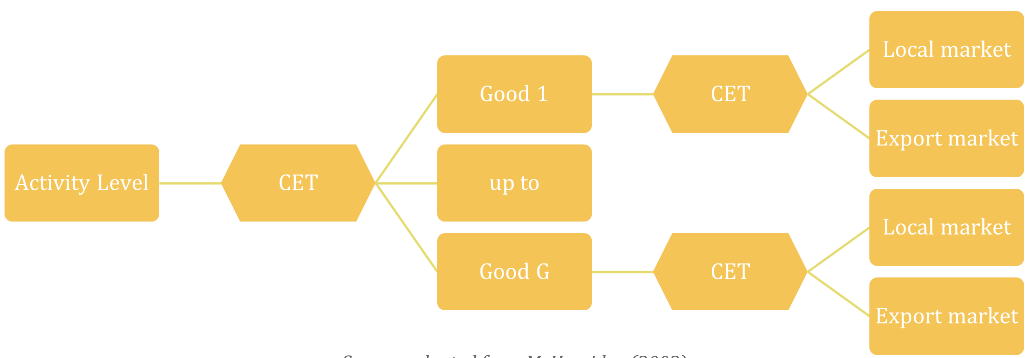
Commodity Output Block

In SAM and, subsequently, in CGE, a commodity can be produced by more than one industry and an industry can also produce more than one commodity. In such a case, an increase in the relative price of a commodity stimulates firms to transform their production in a way to produce more output of that commodity but because of the competitive market, the price of the final good produced by any firm is almost the same. Moreover, for each industry, the input mix may not be the same. Therefore, the model should be flexible enough to incorporate this to allow to analyse a policy change in such a situation. An industry seeks to maximise total revenue given the production function. This can be represented as:

$$X1TOT(i) = \text{CES}(All, c, COM: Q1(c,i)) \qquad (4.6)$$

Constant elasticity of transformation (CET) and CES functions are identical except for the sign of substitution parameter, which is opposite in the case of the CET aggregation function. The composition of output is also represented by the following diagram:

Figure 4.2: Composition of Outputs



Source: adopted from M. Horridge (2003).

In the present study, each commodity is produced by one industry only, therefore, all off-diagonal elements of the multiproduction matrix (MAKE) are zero.

Firms can sell their output in local or export markets. However, it is also possible that the goods produced for the two markets are differentiated. This possibility is taken care of by using the CET function and setting TAU, which is the reciprocal of elasticity of transformation between export and local market, equal to zero. The following equation presents the total revenue or total sale:

$$\mathrm{V1}TOT(i) = \mathrm{SALES(c)} = \sum_i MAKE(c,i) = \mathrm{DOMSALES(c)} + \mathrm{V4BAS\ (c)} \qquad (4.7)$$

Final Demand Block

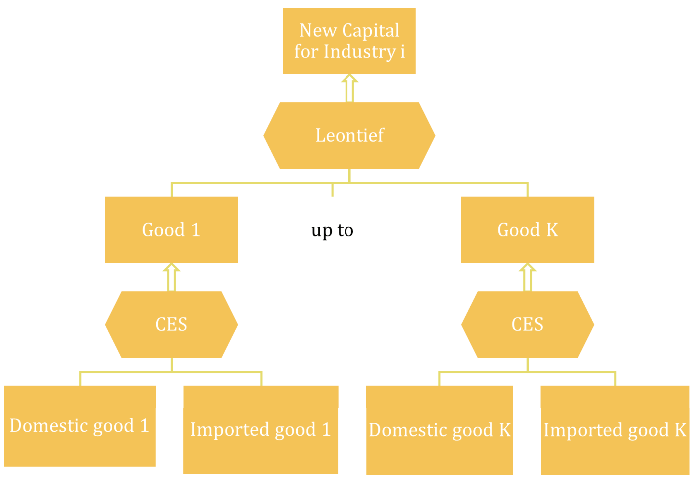
This block consists of the demand for investment by firms, demand for goods by households, demand by the rest of the world, which represents our exports, and demand by the government and inventories. Inputs from domestic and imported commodities are used in producing capital. Like the production function, the investment demand is also nest structured in which in the lowest nest, the cost of local and imported commodities is minimised subject to the production function and at the top level, the total cost of commodity composite is optimised against the given constraint of Leontief production function. This is represented in the following two equations and Figure 4.3.

$$\mathrm{X2\_}S(c,i) = \mathrm{CES}\left(All, s, SRC\colon \frac{\mathrm{X2(c,s,i)}}{\mathrm{A2(c,s,i)}}\right) \qquad (4.8)$$

And

$$\mathrm{X2}TOT(i) = \frac{1}{\mathrm{A2TOT(i)}}\mathrm{MIN}\left(All, s, COM\colon \frac{\mathrm{X2\_S(c,i)}}{\mathrm{A2\_S(c,i)}}\right) \quad (4.9)$$

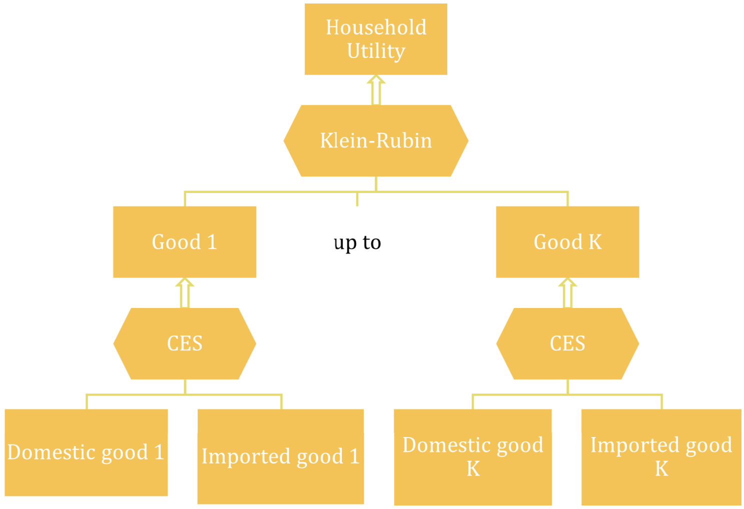
Figure 4.3: Investment Demand



Source: (Amir et al., 2013; M. Horridge, 2003).

Households also demand goods and services for their consumption. Each of the households in the model aims to maximise utility subject to a budget constraint. The nested structure of household demand is similar to the investment demand function but instead of Leontief technology, we assume the Klein-Rubin utility function, which further leads to a linear expenditure function (LES). This is represented in Figure 4.4 below.

Figure 4.4: Household Demand for Commodities



As per the Klein-Rubin utility function, a household first consumes the subsistence quantity of each good, which does not depend on the price (X3SUB(c) in Equation 4.10) and then allocates the remaining budget (S3LUX(c) to the other goods. This is shown as:

$$Utility\ per\ household = \frac{1}{Q}\prod[X3\_S(c) - X3SUB(c)]^{S3LUX(c)} \quad (4.10)$$

The household's total demand for the composite commodity represented by X3_*S(c)* is:

$$X3\_S(c) = X3SUB(c) + S3LUX(c).\frac{V3LUX\_C}{P3\_S(c)} \quad (4.11)$$

and

$$V3LUX\_C = V3TOT - \sum_i [X3SUB(c).P3\_S(c)] \quad (4.12)$$

Export Demand

Exports demand is categorised into individual export demand, which is negatively related to the price of the commodity, and it includes all main export commodities. The other category is collective demand, which is inversely related to the average price of all the export commodities. The equation specifying negatively slopped demand for the commodities categorised as individual export demand is:

$$X4(c) = F4Q(c)\left[\frac{P4(c)}{PHI.F4P(c)}\right]^{EXP\_ELAST(c)} \quad (4.13)$$

In Equation 4.13, EXP_ELAST(c) is a negative parameter of the constant elasticity of demand, X4(c) is export volume, and the term in brackets is the prices in foreign currency, which is converted to local currency unit by multiplying with the exchange rate, PHI. F4Q(c) and F4P(c) are responsible for quantity and price shifts.

The collective export group, represented by the set NTRADEXP, consists of services and other commodities for which the export quantity does not react significantly to the price. These collective exports are treated as Leontief aggregate and their quantity is related to the average price.

Other Demands

All other demands, such as government demand, inventory demand, and margin demand are treated differently. We assume that government demand is determined exogenously, whereas inventory demand is determined such that a percentage change in the volume of each commodity being added to the inventory is the same as the percentage change in domestic production of that commodity. Lastly, the demand for margins is assumed to be proportional to the flow of commodities with which margins are associated.

Market clearing conditions are assumed, which ensure that for all commodities, the total supply is equal to the total demand. The total supply is composed of domestic production and imports and the total demand is the aggregate level of supply for domestic use and exports.

Price System

The CGE model is based on the competitive market assumption. Therefore, for each commodity, there is a single price charged to the consumers and there is no pure profit in any commodity for any producer or distributor (P. B. Dixon et al., 1982). However, there are several sets of prices, like the purchaser's price, which are paid by the consumers and, thus, include basic price, margins, indirect taxes and/or subsidies, basic value, the price of a capital unit, f.o.b. foreign currency export price, and c.i.f. foreign currency import prices. The price of imported goods is also affected by the exchange rate and tariffs and other such duties. The transportation cost and profit margins of wholesalers and retailers, along with taxes imposed at different stages of the value chain, are sources of the difference between the prices paid by the consumers and the prices that producers receive, which are the sources are distortion in the model.

Trade Balance and Other Aggregates

The ordinary change in the trade balance is modelled as a fraction of the ordinary change in the GDP, which is calculated by taking the ratio of the change in the nominal balance of trade to the nominal GDP.

Factor Markets

The capital is created through investment and the investment in each industry is governed by one of the following three rules:

a) More profitable an industry is, the more investment it will attract.

b) Industries in which government policy determines the level of investment, get the investment as per national trends.

c) The investment follows the capital stock in an industry.

The last one is intended for long-term simulations. We may assume fixed capital at the aggregate level and let it be mobile within various sectors of the economy. Moreover, the capital supply of all categories is assumed to be elastic.

For the labour market, following ORANI-G, the current model has two options; either to employ exogenously and let the market determine the market clearing wages rate or set the wage rate exogenously and let the market determine the level of employment in the economy. In the short run, we take wages as exogenously determined. Moreover, labour is assumed to be completely mobile between all industries and labour supply is elastic for all skill types.

Market clearing equations are used to equate the demand for and the supply of each factor. The income of the factors is determined by the payments made to them for their supply in the production activities. The income is defined by:

$$XLABSUP(o).ylab(o) = \sum_i X1LAB(o,i).[p1lab(o,i) + x1lab(o,i)] + XLABRO(o)[xlabro(o) + p1cap\_i(o)] \qquad (4.14)$$

And

$$XCAPSUP(ge).ylab(ge) = \sum_i X1CAP(ge,i).[p1cap(ge,i) + x1cap(ge,i)] + XCAPRO[xcapro + p1cap_i] \qquad (4.15)$$

Institutions

In our model, there are four institutions, namely, households, firms or corporations, government, and the rest of the world. The details are given below.

Households

Households own factors of production, thus, their income comes mainly from the supply of labour and capital. Other sources of income include transfers from the government, the rest of the world, and other households. From their total income, households pay taxes to the government, spend on buying goods and services, and the rest is saved. The taxes paid by the household are based on the marginal income tax rate structure. This can be represented as

$$yh(h) = \sum_o SLABSH(o,h).ylab(o,h) + \sum_{ge} SCAPSH(ge,h).ycap(ge,h) + TRHOGO(h) + TRHOCO(h) + TRHORO(h) + \sum_g TRHOHO(g,h) \qquad (4.16)$$

And

$$eh(b) = MPCH(1 - TAXH).yh(h) = (1 - MPSH)(1 - TAXH).hy(h) \qquad (4.17)$$

Where MPCH is a household's marginal propensity to consumer and MPSH is a household's marginal propensity to save.

Government

Taxes are a major part of the government's income. There are various kinds of direct and indirect taxes that households and firms pay. Other sources of government revenue include transfers from the rest of the world and revenues earned from state-owned production factors. This can be summed up as:

$$VYGC = V1TAX\_CSI + V2TAX\_CSI + V3TAX\_CS + V4TAX\_C + V5TAX\_CS + V0TAR\_C$$
$$+ \sum_h TAXH.YH(h) + VCORTAX + VTRGORO$$
$$+ \sum_o SXLG(o)YLAB(o) + \sum_{gi} SXCG(gi).YCAP(gi) + VTRGOGO \quad (4.18)$$

Government expenditures constitute expenditures on the purchase of goods and services, subsidies to households and firms, and transfers to foreign and local parties. This is summarised as:

$$VEGC = \sum_c V5PUR\_S(c) + \sum_h VTRHOGO(h) + VTRROGO + VTRGOGO + \sum_c VSR(c)$$
$$+ \sum_c V1OCT(i) \qquad (4.19)$$

The difference between government income (VYGC) and expenditures (VEGC) is called budget balance.

Firms

Firms earn income from the ownership of production factors, which is capital in our model, net of the taxes paid, and transfers received from other institutes. Their spending is composed of payments to factors and transfers to other institutions, that is:

$$VYCO = \sum_{gi} SCCO(gi).YCAP(gi) - VCORTAX + VTRCOGO + \sum_h VTRCOHO(h) + VTRCORO$$
$$+ VTRCOCO \qquad (4.20)$$

And

$$VECO = \sum_h TRHOCO(h) + TRROCO + TRCOCO \qquad (4.21)$$

The difference between income and expenditures are savings of the firms.

Rest of the World

Foreign income is the revenue of the rest of the world earned against owned factors of production, import of commodities, and transfer from other institutions, while foreign expenditure is the spending on exports, payments to factors of production, and transfers to other institutions. This is represented as:

$$VYRO = \sum_o SLRO(o).YLAB(o) + SCRO.YCAP + \sum_c XCIF(c) + VTRROGO + \sum_h VTRROHO(h)$$
$$+ VTRROCO + VTRRORO \qquad (4.22)$$

And

$$VERO = \sum_c V4PUR(c) + \sum_o XLABRO(o) + XCAPRO + \sum_h VTRHORO(h) + VTRGORO$$
$$+ VTRCORO + VTRRORO \qquad (4.23)$$

Here the difference between income and expenditures is foreign savings.

Model Closure

For any model to reach a stable solution, the number of equations and endogenous variables must be the same. This is only possible by assuming some of the variables as exogenous, that is, determined outside the model. In the present mode, the short-run closure is achieved by assuming that capital is fixed capital, which implies no new investment. The rate of return on capital adjusts to equate the demand for and the supply of capital. Similarly, short-run closure also assumes that the real wage rate is predetermined. These are all assumed to be fully flexible in the long run. However, the tax rates, technological change, and transfer between institutions are assumed to be exogenous in both the short and the long run. The exchange rate is assumed to be numéraire.

It is also worth mentioning that the model does not generate a recursive dynamic path of the results from the short run to the long run; rather the two results are generated because of the difference coming from the model closure.

Policy Scenarios and Impacts

To widen the scope of the model, corporate income tax and indirect tax were also added. The following are the alternative scenarios that were tested against the baseline scenario.

Simulation 1

Income Tax rate brackets

Table 4.3 Proposed Personal Income Tax Rate

Income	Tax Rate
≤ 400,000	0
400,001 – 800,000	Rs. 1,000
800,001 – 1,200,000	Rs. 2,000
1,200,001 – 2,400,000	5%
2,400,001 – 4,800,000	Rs. 60,000 + 10%
4,800,001 and above	Rs. 300,000 + 15%

Simulation 2

In this scenario, a flat income tax rate of 10 per cent for households having a taxable income of Rs. 400,000, a corporate tax rate of 20 per cent, a sales tax rate of 5 per cent, and a custom duty of 5 per cent across all commodities and no other tax as proposed by Bukhari & Haq (2016; 2020) was assumed.

Both scenarios were simulated for two conditions, i.e., the unbalanced budget condition and the balanced budget condition. The unbalanced budget condition allows the government to spend more than tax revenue in the short run through borrowing, whereas the balanced budget condition does not allow the government this luxury. In the long run, however, the fiscal budget must be balanced under any of the two conditions. These conditions were modelled by adding an additional closure condition of keeping the government expenditures equal to government revenues in the short run for the balanced budget condition.

Table 4.4 and Table 4.5 give the personal tax rates applicable to non-salaried and salaried individuals, respectively, in 2016-17.

Table 4.4 Personal Income Tax Rates for Non-Salaried Persons

Taxable Income (in Rs.)	Tax Rate
0 to 400,000	0%
400,001 – 500,000	7%
500,001 – 750,000	Rs. 7000 + 10%
750,001 – 1,500,000	Rs. 32,000 + 15%
1,500,001 – 2,500,000	Rs. 144,500 + 20%
2,500.001 – 4,000,000	Rs. 344,500 + 25%
4,000,001 – 6,000,000	Rs. 719,500 + 30%
6,000,001 and above	Rs. 1,319,500 + 35%

Source: Federal Bureau of Revenue.

Table 4.5 Personal Income Tax Rates for Salaried Persons

Taxable Income (in Rs.)	Tax Rate
0 to 400,000	0%
400,001 – 500,000	2%
500,001 – 750,000	Rs. 2000 + 5%
750,001 – 1,400,000	Rs. 14,500 + 10%
1,400,001 – 1,500,000	Rs. 75,500 + 12.50%
1,500.001 – 1,800,000	Rs. 92,000 + 15%
1,800,001 – 2,500,000	Rs. 137,000 + 17.5%
2,500,001 – 3,000,000	Rs. 259,500 + 20%
3,000,001 – 3,500,000	Rs. 359,500 + 22.5%
3,500,001 – 4,000,000	Rs. 472,000 + 25%
4,000,001 – 7,000,000	Rs. 597,000 + 27.5%
7,000,001 and above	Rs. 1,422,000 + 30%

Source: Federal Bureau of Revenue.

The income tax rates given in the above tables were applied to the amount exceeding the minimum amount. The tables show that the personal income tax structure is quite complex. Tax rates for salaries and non-salaried persons are separate and the number of slabs is also high. The kinks created because of the slabs not only add complexity to the tax structure but also put an additional burden on the person earning a high income. This naturally creates an incentive for tax evasion.

Corporations had to pay corporate tax at 31 per cent (effective) which was quite high. The general sales tax rate was as high as 17 per cent. However, the customs duties were charged at various rates – 0%, 3%, 11%, 16%, 20%, 35%, 90%, and even higher – based on the categories of goods.

5. RESULTS AND INTERPRETATION

Modelling the policy changes requires changing the marginal income tax rate. However, the equations of the model are not based on marginal income tax rates, but rather on average tax rates. Therefore, average tax rates were calculated based on new marginal tax rates and these values were used as new tax rates.

In this section, first, the simulation results of changes in marginal tax rates on key macroeconomic indicators are presented followed by sectoral impacts and impacts on labour income. The two simulation scenarios were decreasing the personal income tax rate along with a smaller number of slabs (SIM 1) and introducing fixed personal income, and a reduction in corporate tax, sales tax, and customs duties and abolishing all other taxes (SIM 2).

Key Macroeconomic Indicators

Simulation results on key macroeconomic indicators such as real gross domestic product (GDP), private consumption expenditures, investment expenditures, government consumption expenditures, exports, imports, and consumer price index (CPI) are presented in the table below. The results show that reducing personal income tax rates left households with higher disposable income. As a result of which, in the long run, the consumption expenditures of the households would increase by 0.4% and investment by 0.006%. This increase in household disposable income would lead to more demand, which is reflected by an increase in imports by 0.069% and a reduction in exports by 0.389%. Government expenditures would also rise by 0.032% and the consumer price index would rise by 0.119%. An increase in government expenditures would result in increasing the budget deficit and, hence, future interest and capital payments by the government. Together, all these components of demand would result in reducing the real GDP by 0.102%.

Table 5.1. Simulation Results for Key Macroeconomic Indicators Under Unbalanced Budget

Indicators	Long Run Impact		Short Run Impact		Average Impact
	SIM 1	SIM 2	SIM 1	SIM 2	
GDP	- 0.102	0.158	0.024	0.031	0.028
Private Consumption	0.400	0.417	0.422	0.455	0.424
Investment	0.006	0.019	0.001	0.002	0.007
Government Consumption	0.032	0.029	0.041	0.059	0.040

Exports	- 0.389	0.162	- 0.189	0.015	-0.100
Imports	0.069	0.098	0.130	0.131	0.107
Consumer Price Index	0.119	- 0.079	0.298	0.137	0.119

Note: Simulation Results.

The reduction in the personal income tax rate, adds more income to the economy. However, most of this income is used to finance increased consumption expenditures. As savings grow slowly, which was reflected by smaller growth in investment, domestic production fails to match the higher domestic demand. This is also fuelled by higher government expenditures and in the case when balancing the budget is not binding, leads governments to accumulate more debt, leaving little for the private sector. As the model was built to account for the sale of goods produced both in domestic and foreign markets based on the prices producers receive, exports reduce and demand for imported goods increases. This would result in decreasing the GDP. This suggests that along with decreasing income tax, the government should also cut down its expenditures so that government would borrow less to make more funds available to the private sector to increase production. This would also moderate increased aggregate demand, which would reduce the demand for imports and increase exports, resulting in a lower trade gap, which could result from the reduced personal income tax.

The results of Simulation 2 can also be interpreted along the same lines. In this case, real GDP would increase because of positive growth in private consumption, investment, government consumption, and higher trade. A significant difference can be noted in exports which showed an increase of 0.162 per cent compared to a decline of 0.389 per cent in the case of lower PIT only. This could be because of the low financial cost under the simplified tax system with lower taxes, which encourages more investment and also leads to improving competitiveness as noted by Cororaton & Orden (2010).

Short-run results are also reported which can be interpreted along the same lines. In the short run, GDP growth was positive even in Scenario 1 when there was a decrease in income tax only. The other difference is that there was a price increase even in the case when all the taxes were lower. This shows that a decrease in the cost of production due to lower taxes is not passed through to the consumers in the short run. This is possible, according to economic theory, because of some of the frictions in the economy, which may lead to some kind of market power that results in delaying passing the benefit of the decrease in cost to the consumers. The model also incorporates these frictions. As mentioned above, the sources of friction in the model are margins and transportation costs.

The reduction in the personal income tax rate would add more income to the economy. However, most of this income would go into financing the increased consumption expenditures. As savings grow slowly, which was reflected by smaller growth in investment, domestic production would fail to match the higher domestic demand. This is also fuelled by higher government expenditures and, therefore, in the case when balancing the budget is not binding, this would lead the government to accumulate more debt leaving less for the private sector. As the model was constructed in a way that goods produced could be sold in domestic as well as in foreign markets based on the prices producers receive, exports would reduce and demand for goods produced in foreign countries would increase, which would lead to a decrease in the GDP. This suggests that, along with decreasing income tax, the government should also cut down its expenditures so that the government has to borrow less and more funds are available to the private sector for increasing production. Moreover, this might also moderate the increased aggregate demand leading to reduced import demand and increased exports. This would improve the trade gap which resulted from the reduced personal income tax.

The results of Simulation 2 can also be interpreted along the same lines. In this case, real GDP would increase because of positive growth in private consumption, investment, government consumption, and higher trade. A significant difference can be noted in exports which showed an increase of 0.162 per cent compared to a decline of 0.389 per cent in the case of lower PIT only. This is primarily because lower taxes would reduce financial costs resulting in higher profit and, thus, encouraging more investment.

The short-run results show that in the short run, GDP growth would be positive even in Scenario 1 when there was a decrease in income tax only. The other difference is that there was a price increase even in the case when all the taxes were lower. This shows that a decrease in the cost of production due to lower taxes would not be passed through to the consumers in the short run which is an indication of some kind of friction in the system.

The results of simulations for both scenarios under balanced budget conditions are presented in Table 5.2. The results show that in the long run, the GDP would grow at a higher rate under both kinds of tax reforms when the balanced budget condition is binding. However, in the short run, GDP growth is negative in both scenarios. This shows that under the balanced budget condition, the government would have to cut its expenditures, which would negatively affect economic growth in the short run. However, in this condition, the financial needs of the government would not create more debt leaving more liquidity for households and firms, which may be the key to economic growth in the long run.

Table 5.2: Key Macroeconomic Indicators in Balanced Budget Condition

Indicators	Long-Run Impact		Short-Run Impact		Average Impact
	SIM 1	SIM 2	SIM 1	SIM 2	
GDP	0.014	0.213	- 0.008	- 0.019	0.050
Private Consumption	0.391	0.402	0.121	0.119	0.258
Investment	0.037	0.128	0.000	0.000	0.041
Govt Consumption	- 0.229	- 0.278	- 0.311	- 0.513	-0.333
Exports	0.0412	0.197	0.015	0.102	0.089
Imports	0.061	0.058	- 0.018	0.006	0.027
CPI	- 0.029	- 0.126	- 0.020	- 0.123	-0.075

Note: Simulation Results.

Sectoral Impacts

Long-run sectoral impacts in terms of per cent changes in output and prices are reported below. These impacts suggest that decreasing income tax rates and slabs only, as for simulation 1 (Sim 1), would result in decreasing the output of mining and related activities, textile, machinery, manufacturing, and construction sectors, whereas it would increase the output of electricity, trade at various levels, hotelling, rent, financial services, education, and health. The prices of almost all the items would increase because of higher demand driven by an increase in the take-home income of the households. However, there would be a prominent increase in the prices of mining, textile, leather, agricultural goods, machinery, transportation services, and real estate services.

Analysing the impacts of cuts in both direct and indirect taxes across the board, we can observe that the output would increase and the price of the output of most of the sectors would decrease. This shows that with a decrease in income tax, households would increase their consumption but most of the additional supply would come from the increase in imports rather than from the increase in local production. This may be because only households were given tax relief which resulted in increasing the demand but firms were not given any incentive or additional benefit that could have resulted in decreasing their cost of production. Therefore, domestic firms had little margin to increase their supply and, hence, the additional demand was fulfilled largely from the imported goods. Therefore, significant growth in the output of the firms was not observed. On the other hand, if we look at the second scenario where a flat personal income tax rate was combined with a decrease in corporate income tax, sales tax, customs duty and abolishing all other taxes, it would result in decreasing the financial cost of the firms. Therefore, the firms could earn higher profits and look forward to expanding their production capacity. This is observed in increasing the output level as well as a decrease in the price of several commodities which may be the result of decreasing the indirect taxes which are passed on to consumers.

Short-run sectoral impacts are reported for both simulation conditions in the last two columns of Table 5.3. Overall, short-run impacts are quite similar to long-run outcomes, but there are slight differences between the two cases, such as wood, paper making, chemicals, and construction sector in terms of output and textile, coke and public administration in terms of prices.

Table 5.3: Sectoral Impacts of Tax Reforms Under Unbalanced Budget Condition

Commodities/Industries	Long-Run Impact				Short-Run Impact			
	SIM 1		SIM 2		SIM 1		SIM 2	
	Output	Price	Output	Price	Output	Price	Output	Price
Agriculture	0.096	0.205	0.107	0.012	0.101	0.199	0.103	0.013
Mining	- 1.023	0.283	0.210	0.016	- 0.233	0.263	0.119	0.019
Food	0.062	0.124	0.114	- .103	0.132	0.167	0.122	- 0.094
Textile	- 0.413	0.249	0.179	- .002	- 0.019	0.255	0.154	0.001
Leather	- 0.104	0.201	0.246	- .011	0.043	0.198	0.260	- 0.019
Wood	- 0.219	0.103	- 0.097	0.037	- 0.037	0.110	0.008	0.042
Paper	0.023	0.021	- 0.107	0.011	0.040	0.073	0.067	0.013
Coke	- 0.017	0.107	0.109	- .005	0.001	0.113	- 0.013	0.001
Chemicals	0.132	0.128	0.140	- 0.01	0.122	0.129	0.144	- 0.007
Rubber	0.097	0.094	0.107	0.004	0.101	0.100	0.121	0.010
Nonmetallic Minerals	- 0.521	0.066	- 0.877	- .016	- 0.239	0.072	- 0.767	- 0.008
Metals	0.012	0.100	0.093	0.009	0.107	0.106	0.104	0.012
Machinery	- 0.059	0.223	0.108	- .031	0.011	0.230	0.112	- 0.024

Electric Equipment	0.394	0.195	0.455	0.009	0.104	0.202	0.461	0.011
Transport Equipment	- 0.021	0.197	- 0.122	0.003	- 0.009	0.214	- 0.013	0.004
Manufacturing	- 0.031	0.182	0.140	- .011	0.016	0.186	0.140	- 0.017
Utility Supply	0.173	0.132	- 0.061	0.004	0.214	0.129	- 0.003	0.005
Construction	- 0.109	0.114	- 0.002	0.001	- 0.021	0.130	0.010	0.004
S&M of Vehicles	0.104	0.092	0.113	0.003	0.022	0.099	0.142	0.090
Wholesale Trade	0.098	0.057	0.102	- .017	0.100	0.070	0.079	- 0.009
Retail Trade	0.084	0.103	0.084	0.008	0.069	0.111	0.103	0.010
Hotels	0.102	0.034	0.214	0.011	0.092	0.053	0.200	0.012
Inland Transport	- 0.034	0.192	0.098	0.009	- 0.043	0.199	0.106	0.008
Water Transport	0.117	0.279	0.216	0.010	0.124	0.286	0.223	0.009
Air Transport	0.097	0.226	0.100	- .017	0.103	0.233	0.099	- 0.012
Transport Services	0.037	0.198	0.049	0.007	0.078	0.201	0.063	0.014
Telecom	0.010	0.154	0.021	- .006	0.031	0.193	0.101	- 0.001
Financial Institutions	0.242	0.245	0.249	- .011	0.098	0.267	0.216	- 0.003
Real Estate	0.131	0.271	0.102	0.018	0.129	0.290	0.113	0.012
Renting Business	0.034	0.109	0.021	- .007	0.029	0.111	0.016	- 0.003
Public Administration	- 0.140	0.112	0.138	- .010	- 0.024	0.109	0.171	0.002
Education	0.152	0.158	0.168	0.001	0.155	0.169	0.201	0.009
Health	0.126	0.151	0.159	- .005	0.121	0.162	0.189	- 0.001
Communication Services	- 0.042	0.023	0.003	0.008	- 0.019	0.030	0.021	0.012
Average Impact	-0.012	0.151	0.076	-0.002	0.044	0.161	0.093	0.003

Note: Simulation Results.

As different sectors of an economy have strong forward and backward linkages, the effects of changes in the cost of production through prices transmit from one firm to another and the transmission mechanism is stronger for input-producing industries. According to Carvalho et al. (2021), the effects of change in the price of a good, produced by an industry impact all industries that use this good as input especially when the elasticities of substitution between various intermediate inputs or between intermediate goods and factors of production are not equal to one. Blöchl et al. (2011), Fadinger et al. (2016), and McNerney et al. (2013) document that the distribution of sectoral impacts is highly heterogeneous. The magnitude of the impact on other industries also depends on the size of the industry. Carvalho et al. (2021) and Bernard et al. (2019) report that large firms in terms of sales and employment also have a large number of buyers and suppliers and, therefore, have deeper effects on the input suppliers and output buyers. According to Barrot & Sauvagnat (2016) and Boehm et al. (2019), these effects may have a significant impact on the overall economy.

Both alternatives that this study tested, focussed on decreasing the tax burden. In Scenario 2, only the tax burden on individuals was decreased, whereas in Scenario 2 the tax burden on both the individuals and the firms was decreased. An increase in disposable income of the households following the decrease in income taxes would lead to an increase in consumption demand and savings. The increased savings then would lead to higher investment and, therefore, higher production. As a result, firms would hire more factors of production, which would decrease unemployment and increase labour income and the GDP. Similarly, a decrease in corporate income tax and customs duties led to lowering the cost of production and increasing the output produced. Moreover, since, at present, the different sectors are treated differently as a part of protection policies through various kinds of indirect taxes, such as tariffs, customs duties, and regulatory duties, opting for similar tax treatment for all the sectors would result in impacting different sectors differently. For example, in our case, we observed a resource shift from the textile sector to other sectors, like the manufacturing of electric equipment and financial institutions as a result of the change in the tax treatment. However, lowering taxes would also decrease government revenue collection, at least in the short run, which might affect the provision of public goods or lead to debt accumulation.

Sectoral impacts of tax reforms under balanced budget conditions are presented in the table below.

Table 5.4: Sectoral Impacts of Tax Reforms under Balanced Budget Condition

Commodities/Industries	**Long-Run Impact**				**Short-Run Impact**			
	SIM 1		**SIM 2**		**SIM 1**		**SIM 2**	
	Output	**Price**	**Output**	**Price**	**Output**	**Price**	**Output**	**Price**
Agriculture	0.090	0.113	0.101	0.101	0.087	0.028	0.080	0.031
Mining	- 2.287	0.213	0.011	0.022	- 0.013	0.067	0.011	- 0.009
Food	0.071	0.083	0.193	0.009	0.117	0.122	0.113	0.003
Textile	0.132	0.034	- 0.149	- 0.010	- 0.012	- 0.03	- 0.045	0.010
Leather	- 0.122	0.029	- 0.021	0.102	0.003	0.022	- 0.002	0.013
Wood	0.109	0.017	- 0.013	0.021	- 0.021	- 0.11	- 0.101	- 0.002
Paper	- 0.031	- 0.011	0.102	- 0.002	0.029	- 0.04	0.031	- 0.022
Coke	- 0.013	0.011	0.212	- 0.006	0.000	0.026	0.004	- 0.003
Chemicals	0.210	- 0.025	0.140	- 0.010	- 0.013	0.014	- 0.013	0.002
Rubber	- 0.013	0.008	0.197	0.004	0.032	- 0.10	- 0.021	0.010
Nonmetallic Minerals	- 0.112	0.056	- 0.110	- 0.012	0.002	0.025	0.015	- 0.002

Metals	0.013	- 0.070	0.032	0.002	- 0.091	0.016	- 0.011	0.012
Machinery	- 0.031	- 0.002	- 0.013	- 0.017	0.022	- 0.03	0.101	- 0.04
Electric Equipment	0.344	- 0.079	0.155	0.009	- 0.033	- 0.17	0.076	- 0.03
Transport Equipment	- 0.002	- 0.011	- 0.122	0.003	- 0.017	0.130	- 0.011	0.002
Manufacturing	0.010	0.151	0.224	- 0.009	0.012	0.092	0.121	- 0.02
Utility Supply	0.105	- 0.012	0.013	- 0.001	0.101	0.091	0.004	0.004
Construction	- 0.011	0.101	0.230	- 0.011	- 0.043	- 0.01	- 0.009	0.004
S&M of Vehicles	0.101	0.071	0.111	0.006	- 0.006	0.023	- 0.111	0.079
Wholesale Trade	0.066	0.023	0.153	- 0.002	0.120	- 0.05	0.021	- 0.009
Retail Trade	0.079	0.107	0.082	- 0.003	0.051	0.111	- 0.001	- 0.003
Hotels	0.098	0.029	0.323	0.020	0.065	- 0.05	0.018	0.011
Inland Transport	- 0.003	- 0.010	0.082	0.001	- 0.031	0.170	0.009	- 0.002
Water Transport	0.124	0.283	0.229	0.011	0.122	0.201	0.021	0.010
Air Transport	0.092	0.199	0.137	- 0.012	- 0.009	0.023	0.069	- 0.002
Transport Services	0.040	0.168	0.051	- 0.002	- 0.101	0.102	0.041	0.021
Telecom	0.018	0.133	0.043	0.013	0.009	0.112	0.100	- 0.013
Financial Institutions	0.199	0.234	0.244	0.009	0.031	0.037	0.127	- 0.010
Real Estate	0.170	0.310	0.112	0.019	0.132	- 0.02	- 0.101	0.011
Renting Business	0.043	0.009	0.041	- 0.011	0.030	- 0.01	- 0.009	- 0.001
Public Administration	- 0.009	0.110	- 0.013	0.107	- 0.106	0.012	- 0.054	0.021
Education	0.133	0.065	0.209	0.021	0.103	0.113	0.131	0.0010
Health	0.134	0.130	0.189	0.002	0.079	- 0.03	0.138	- 0.003
Communication Services	- 0.014	- 0.006	0.019	0.010	- 0.008	0.020	- 0.018	0.008
Average Impact	-0.008	0.072	0.094	0.011	0.019	0.027	0.021	0.002

Note: Simulation Results.

Effects on Labour Income

Lastly, the effect of changes in tax rates on the income of different kinds of labour used in the model is discussed. The long-run and short-run results reported in the table below show that all the various categories of labour would experience an increase in income under both scenarios of a tax rate decrease. However, the increase in labour income would be higher in the case of Scenario 2 in which there was a decrease in the rate of all kinds of taxes which would benefit not only households and result in increasing their demand for the product but would also reduce the cost of production for the firms making it more profitable for corporations to increase their production.

Table 5.5 Impact on Labour Income

Labour Classification	Long-Run Income Effect		Short-Run Income Effect		Average Impact
	SIM 1	SIM 2	SIM 1	SIM 2	
Punjab Rural Low-Skilled	0.245	0.297	0.099	0.313	0.239
Punjab Rural High-Skilled	0.341	0.439	0.162	0.492	0.359
Punjab Urban Low-Skilled	0.279	0.301	0.103	0.381	0.266
Punjab Urban High-Skilled	0.358	0.513	0.217	0.599	0.422
Sindh Rural Low-Skilled	0.242	0.289	0.064	0.294	0.222
Sindh Rural High-Skilled	0.281	0.357	0.103	0.401	0.286
Sindh Urban Low-Skilled	0.299	0.348	0.199	0.481	0.332
Sindh Urban High-Skilled	0.446	0.792	0.342	0.829	0.602
KP Rural Low-Skilled	0.241	0.331	0.197	0.367	0.284
KP Rural High-Skilled	0.34	0.392	0.223	0.396	0.338
KP Urban Low-Skilled	0.282	0.310	0.203	0.344	0.285
KP Urban High-Skilled	0.353	0.412	0.299	0.396	0.365
Balochistan Rural Low-Skilled	0.221	0.299	0.193	0.334	0.262
Balochistan Rural High-Skilled	0.253	0.398	0.210	0.402	0.316
Balochistan Urban Low-Skilled	0.25	0.351	0.144	0.377	0.281
Balochistan Urban High-Skilled	0.316	0.443	0.231	0.476	0.367

Note: Simulation Results.

Table 5.6 Impact on Labour Income Under Balanced Budget Conditions

Labour Classification	Long-Run Income Effect		Short-Run Income Effect		Average Impact
	SIM 1	SIM 2	SIM 1	SIM 2	
Punjab Rural Low-Skilled	0.251	0.304	0.101	0.340	0.249
Punjab Rural High-Skilled	0.356	0.453	0.170	0.499	0.370
Punjab Urban Low-Skilled	0.291	0.322	0.121	0.393	0.282
Punjab Urban High-Skilled	0.339	0.499	0.209	0.624	0.418
Sindh Rural Low-Skilled	0.239	0.297	0.099	0.323	0.240
Sindh Rural High-Skilled	0.292	0.361	0.100	0.370	0.281
Sindh Urban Low-Skilled	0.297	0.386	0.210	0.253	0.287
Sindh Urban High-Skilled	0.460	0.799	0.4012	0.843	0.626
KP Rural Low-Skilled	0.270	0.437	0.210	0.388	0.326
KP Rural High-Skilled	0.279	0.282	0.283	0.312	0.289
KP Urban Low-Skilled	0.268	0.299	0.229	0.371	0.292
KP Urban High-Skilled	0.365	0.423	0.308	0.399	0.374
Balochistan Rural Low-Skilled	0.200	0.264	0.198	0.254	0.229
Balochistan Rural High-Skilled	0.231	0.299	0.252	0.456	0.310
Balochistan Urban Low-Skilled	0.248	0.362	0.160	0.401	0.293
Balochistan Urban High-Skilled	0.280	0.490	0.245	0.489	0.376

Note: Simulation Results

6. CONCLUSION AND POLICY RECOMMENDATIONS

This study was conducted to quantify the impact of changes in tax rates on the overall economy of Pakistan. For changing the tax rates, we tested two scenarios. In the first scenario, the marginal tax rate and the number of slabs for the individuals paying personal income tax were decreased but kept the taxes progressive. In the second scenario, a flat personal income tax rate was introduced, corporate income tax, sales tax, and customs duties were decreased, and all other direct and indirect taxes were abolished. Both of these scenarios simplified the tax structure and reduced the tax burden, leaving the agents with higher after-tax income. We used the CGE model to study the sectoral and macroeconomic impacts of the said changes. However, we first developed an updated SAM based on the 2017 data taken from the 2017 IO table, national accounts data, HIES and LFS for the corresponding year. The SAM developed for this study consisted of 34 industries, all producing one commodity, multiple types of labour, capital, and households and incorporated direct and indirect taxes paid by the households and firms to the government. It presented a useful picture of the economy using the double-entry system in which each entry in a cell represents the flow of income from one agent to another. After that, ORANI-G modifications of the CGE were made to make it better applicable to Pakistan's economy and the objectives of the study.

Our analysis shows that decreasing the personal income tax rate applied to individuals would only result in increasing the disposable income of the households, which, in turn, would result in increasing household consumption expenditures and decreasing government income, consequently increasing the fiscal deficit. The increased demand would be mostly fulfilled by imports, which would also widen the trade deficit. On the other hand, reducing rates of all the taxes, as modelled in Scenario 2, would enable firms to reap higher profits increasing the demand due to higher after-tax income, which would be matched by higher supply resulting from higher production motivated by lower financial and psychic costs of production and higher profits. However, the rate of growth in output and prices would be different for different sectors. Scenario 2 especially suits the export industry as it would reduce its cost making the exports more competitive. This was noted by an increase in the exports reflected in the analysis. Both these scenarios would result in increasing the take-home income of various categories of labour and the income of the households. Higher consumption due to higher income would increase the welfare of the households and improve their living standards. The expenditures on health and education would also increase. The results also show that the overall positive impact of tax reforms on the economy would be more pronounced when the balanced budget condition is binding.

This analysis leads to some simple but important policy recommendations. One of the policies that can be recommended based on the analysis is that simplifying the tax regime and lowering taxes will result in higher income of the citizens and corporations, a sectoral shift in favour of competitive and efficient sectors and, resultantly, higher economic growth. This higher growth will result in increased tax revenue without overburdening the citizens and businesses. Therefore, if the government wants to raise the living standard of the people, it should introduce a simplified tax system which is broad-based with a low tax burden. Secondly, reducing rates of only one or few taxes will not work as effectively as lowering all the tax rates, reducing the total number of taxes to be paid by firms and individuals, and letting various sectors compete based on productivity and efficiency rather than using tax as a tool for creating favourable grounds for a few sectors. The results of the study also show that reducing tax rates will result in increasing the fiscal deficit when the balanced budget condition is not binding. However, if the government is restricted to keeping the budget balanced or the deficit under control, it will compel the government to cut down or abolish unnecessary expenditures and reduce its footprint on the economy, which will result in lowering labour demand in the public sector and release it for private firms, which will result in reducing market distortions. Therefore, we recommend that the government should be restricted to keep the fiscal deficit within the target. Although this study did not extend to that area, the literature suggests that combining a simplified tax regime based on low tax rates benefits higher-income groups more than lower-income groups. Such a situation, on the one hand, encourages wealth creation but, on the other hand, it increases inequalities which need to be taken care of using suitable policies.

REFERENCES

Abdisa, T. (2018). Macroeconomic Impact of Tax Reform in Ethiopia. *Journal of Economics and Sustainable Development,* 9.

Ahmed, S., Ahmed, V., & O'Donoghue, C. (2011). Reforming indirect taxation in Pakistan: A macro-micro analysis. *EJTR,* 9, 153.

Amir, H., Asafu-Adjaye, J., & Ducpham, T. (2013). The impact of the Indonesian income tax reform: A CGE analysis. *Economic Modelling,* 31, 492–501.

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *Staff Papers,* 16(1), 159–178.

Auerbach, A. J. (1996). Measuring the impact of tax reform. *National Tax Journal*, 665–673.

Barrot, J.-N., & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *The Quarterly Journal of Economics,* 131(3), 1543–1592.

Begg, D., Fischer, S., & Dornbusch, R. (2003). *Economics* (7th ed.). McGraw-Hill Education.

Belayneh, A. (2018). *Economy-wide impact of direct tax reform in Ethiopia: A recursive dynamic computable general equilibrium analysis* [PhD Thesis]. AAU.

Berisha, E. (2020). Tax cuts and" middle-class" workers. *Economic Analysis and Policy,* 65, 276–281.

Bernard, A. B., Moxnes, A., & Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy,* 127(2), 639–688.

Bhattarai, K., & Trzeciakiewicz, D. (2017). Macroeconomic impacts of fiscal policy shocks in the UK: A DSGE analysis. *Economic Modelling*, 61, 321–338.

Blöchl, F., Theis, F. J., Vega-Redondo, F., & Fisher, E. O. (2011). Vertex centralities in input-output networks reveal the structure of modern economies. *Physical Review E,* 83(4), 046127.

Boehm, C. E., Flaaen, A., & Pandalai-Nayar, N. (2019). Input linkages and the transmission of shocks: Firm-level evidence from the 2011 Tōhoku earthquake. *Review of Economics and Statistics,* 101(1), 60–75.

Breisinger, C., Thomas, M., & Thurlow, J. (2009). *Social accounting matrices and multiplier analysis: An introduction with exercises* (Vol. 5). International Food Policy Research Institute.

Carvalho, V. M., Nirei, M., Saito, Y. U., & Tahbaz-Salehi, A. (2021). Supply chain disruptions: Evidence from the great east japan earthquake. *The Quarterly Journal of Economics,* 136(2), 1255–1321.

Cororaton, Caesar B., and David Orden. 2010. "Pakistan." in *Agricultural Price Distortions, Inequality, and Poverty.* Washington DC: The World Bank.

Damuri, Y. R., & Perdana, A. A. (2003). The impact of fiscal policy on income distribution and poverty: A computable general equilibrium approach for Indonesia. Economics Working Paper No. WPE068, Center for Strategic and International Studies, Jakarta, Indonesia.

Debowicz, D., Dorosh, P., Haider, H., & Robinson, S. (2012). A 2007-08 social accounting matrix for Pakistan. Pakistan Strategy Support Program (PSSP) Working Paper, 001.

Diao, X., Roe, T. L., & Yeldan, A. E. (1998). Fiscal debt management, accumulation and transitional dynamics in a CGE model for Turkey. *Canadian Journal of Development Studies/Revue Canadienne d'études Du Développement,* 19(2), 343–375.

Dixon, P. (2006). *Evidence-based trade policy decision making in Australia and the development of computable general equilibrium modelling.* Centre of Policy Studies (CoPS).

Dixon, P. B., Parmenter, B. R., Powell, A. A., Wilcoxen, P. J., & Pearson, K. R. (1992). *Notes and problems in applied general equilibrium economics.*

Dixon, P. B., Parmenter, B. R., Sutton, J., & Vincent, D. P. (1982)*. Orani, a multisectoral model of the Australian economy* (Vol. 142). North-Holland.

Dixon, P., & Rimmer, M. T. (2002). *Dynamic general and equilibrium modelling for forecasting and policy: A practical guide and documentation of MONASH* (Vol. 256). Elsevier.

Dorosh, P., Niazi, M. K., & Nazli, H. (2004). A social accounting matrix for Pakistan, 2001-02: methodology and results. A Background Research Paper for the Pakistan Rural Factor Markets Study, World Bank. South Asia Rural Development Unit of the World Bank, Washington DC.

Eicher, T. S., Turnovsky, S. J., & Prunera, M. C. R. (2003). The impact of tax policy on inequality and growth: An empirical and theoretical investigation. In *Inequality and growth: Theory and policy implications.* MIT Press, Cambridge, MA.

Engen, E. M., & Skinner, J. (1996). *Taxation and economic growth*. National Bureau of Economic Research.

Fadinger, H., Ghiglino, C., & Teteryatnikova, M. (2016). *Income differences and input-output structure.*

Feldstein, M. (1973). On the optimal progressivity of the income tax*. Journal of Public Economics,* 2(4), 357–376.

Fortin, B., Marceau, N., & Savard, L. (1997). Taxation, wage controls and the informal sector. *Journal of Public Economics,* 66(2), 293–312.

Fossat, P., & Bua, M. (2013). *Tax administration reform in the Francophone countries of Sub-Saharan Africa.*

Gale, W. G., & Samwick, A. A. (2014). Effects of income tax changes on economic growth. *Economic Studies, Https://www. Brookings. Edu/Wpcontent/Uploads/2016/06/09_Effects_Income_Tax_Changes_Economic_Growth_Gale_Sa Mwick. Pdf.*

Gemmell, N. (1988). Tax systems, tax revenue and growth in LDCs: A review of empirical evidence. *Intereconomics,* 23(2), 84–90.

Giraldo, M. C., & García, J. A. B. (2018). *Fiscal Policy and Inequality in a CGE Model for Colombia.*

Golan, A., Judge, G., & Miller, D. (1997). *Maximum entropy econometrics: Robust estimation with limited data.*

Golan, A., Judge, G., & Robinson, S. (1994). Recovering information from incomplete or partial multisectoral economic data. *The Review of Economics and Statistics,* 541–549.

Gordon, R. (2010). T*axation in developing countries: Six case studies and policy implications*. Columbia University Press.

Goulder, L. H., & Summers, L. H. (1989). Tax policy, asset prices, and growth: A general equilibrium analysis. *Journal of Public Economics,* 38(3), 265–296.

Hamilton, B., & Whalley, J. (1989). Reforming indirect taxes in Canada: Some general equilibrium estimates. *Canadian Journal of Economics,* 561–575.

Hasudungan, H. W., & Sabaruddin, S. S. (2016). The Impact of Fiscal Reform on Indonesian Macroeconomy: A CGE Framework. *Central European Journal of Economic Modelling and Econometrics,* 8(3), 181–202.

Horridge, J. (2000). ORANI-G: *A general equilibrium model of the Australian economy*. Centre of Policy Studies (CoPS).

Horridge, M. (2003). *ORANI-G: A generic single-country computable general equilibrium model*. Centre of Policy Studies and Impact Project, Monash University, Australia.

Huang, J., & Rios, J. (2016). Optimal tax mix with income tax non-compliance. *Journal of Public Economics*, 144, 52–63.

Hussain, S. M., & Malik, S. (2016). Asymmetric effects of exogenous tax changes. *Journal of Economic Dynamics and Control*, 69, 268–300.

Iqbal, Z., Ayyubi, M. S., Farooq, A., & Lodhi, S. (2019). Microeconomic Impact of GST on Household Consumption Patterns in Pakistan. *Forman Journal of Economic Studies*, 15.

Jamal, H., & Javed, S. (2013). Incidence of general sales tax in Pakistan: Latest estimate. *Pakistan Journal of Applied Economics,* 23(2), 73–95.

Judge, G. G., & Mittelhammer, R. C. (2011). *An information theoretic approach to econometrics.* Cambridge University Press.

Keen, M. (2012). Taxation and development-Again, IMF Working Paper, WP/12/220. International Monetary Fund Washington^ eDC DC.

Khan, M. A., Zada, N., & Mukhopadhyay, K. (2018). Economic implications of the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP) on Pakistan: A CGE approach. *Journal of Economic Structures,* 7(1), 1–20.

Knudsen, M. B., Pedersen, L. H., Pedersen, T. W., Stephensen, P., & Trier, P. (1998). *A CGE analysis of the Danish 1993 tax reform.* Danish Rational Economic Agents Model, DREAM.

Lee, Y., & Gordon, R. H. (2005). Tax structure and economic growth. *Journal of Public Economics,* 89(5–6), 1027–1043.

Lin, B., & Jia, Z. (2019). How does the tax system on energy industries affect energy demand, CO2 emissions, and the economy in China? *Energy Economics,* 104496.

Mabugu, R., Robichaud, V., Maisonnave, H., & Chitiga, M. (2013). Impact of fiscal policy in an intertemporal CGE model for South Africa. *Economic Modelling,* 31, 775–782.

Martimort, D. (2001). Optimal taxation and strategic budget deficit under political regime switching. *The Review of Economic Studies,* 68(3), 573–592.

McNerney, J., Fath, B. D., & Silverberg, G. (2013). Network structure of inter-industry flows. *Physica A: Statistical Mechanics and Its Applications,* 392(24), 6427–6441.

Mendoza, E. G., Razin, A., & Tesar, L. L. (1994). Effective tax rates in macroeconomics: Cross-country estimates of tax rates on factor incomes and consumption. *Journal of Monetary Economics,* 34(3), 297–323.

Mertens, K., & Montiel Olea, J. L. (2018). Marginal tax rates and income: New time series evidence. *The Quarterly Journal of Economics,* 133(4), 1803–1884.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The Review of Economic Studies,* 38(2), 175–208.

Mountford, A., & Uhlig, H. (2009). What are the effects of fiscal policy shocks? *Journal of Applied Econometrics,* 24(6), 960–992.

Nandi, A. (2020). Indian Fiscal Policy: A DSGE Primer. *The Journal of Developing Areas,* 54(2).

Naqvi, H. A., Hakeem, M. M., & Naeem, R. A. (2011). Impact of Agricultural Income Tax on Household Welfare and Inequality: Pakistan A Case-in-Point. *International Journal of Business and Social Science,* 2(6), 103–118.

Pasha, H. A., & Ghaus-Pasha, A. (2015). The Future Path of Tax Reforms in Pakistan. *Moving the Economy Forward,* 171.

Robinson, S., Cattaneo, A., & El-Said, M. (2001). Updating and estimating a social accounting matrix using cross-entropy methods. *Economic Systems Research,* 13(1), 47–64.

Robinson, S., & Gueneau, A. (2013). *CGE-W: An integrated modeling framework for analyzing water-economy links applied to Pakistan.*

Romer, C. D., & Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review,* 100(3), 763–801.

Sahn, D. E., & Younger, S. D. (2000). Expenditure incidence in Africa: Microeconomic evidence. *Fiscal Studies,* 21(3), 329–347.

Samuelson, P. A. (1951). *Theory of optimal taxation.*

Shaikh, F. M. (2009). Analysis of bilateral trade liberalization and South Asian Free Trade Agreement (SAFTA) on Pakistan's economy by using a CGE model. *Journal of International Trade Law and Policy.*

Shaikh, F. M., & Rahpoto, M. S. (2009). Impact of trade liberalization and SAFTA on Pakistan's Economy by Using CGE Model. *International Journal of Business and Management,* 4(4), 192.

Siddiqui, R., & Iqbal, Z. (1999). *Social Accounting Matrix of Pakistan for 1989-90*. Pakistan Institute of Development Economics.

Siddiqui, R., & Iqbal, Z. (2001). *Tariff reduction and functional income distribution in Pakistan: A CGE model.* Pakistan Institute of Development Economics.

Siddiqui, R., Kemal, A. R., Siddiqui, R., & Ali Kemal, M. (2008). *Tariff reduction, fiscal adjustment and poverty in Pakistan: A CGE-based analysis.*

Waheed, A., & Ezaki, M. (2008). Aggregated and compact disaggregated financial social accounting matrices for Pakistan. *Journal of Economic Cooperation,* 29(4), 17–36.

Wawire, N. (2017). Determinants of value added tax revenue in Kenya. *Journal of Economics Library,* 4(3), 322–344.

Yusuf, A. A., Djoni, H., & Wawan Hermawan, Y. (2007). AGEFIS: Applied general equilibrium for fiscal policy analysis. Working Paper in Economics and Development Studies.

Zeshan, Muhammad (2020). 'Chapter: Pakistan', in Aguiar,A., Chepeliev, M., Corong, E. L., McDougall, R. and van der Mensbrugghe, D., 2019. The GTAP database: version 10. *Journal of Global Economic Analysis,* 4(1), pp. 1 – 27.

APPENDIX

Information about SAM 2017

The social accounting matrix (SAM) developed for this study was based on 2017 information about the economy. It is a 110 × 110 matrix including one row and column each for labels and total.

It consists of 34 commodities produced by the following 34 activities:

I. Agriculture, hunting, forestry, and fishing
II. Mining and quarrying
III. Food, beverages, and tobacco
IV. Textiles and textile products
V. Leather, leather products, and footwear
VI. Wood and products of wood and cork
VII. Pulp, paper, paper products, printing, and publishing
VIII. Coke, refined petroleum, and nuclear fuel
IX. Chemicals and chemical products
X. Rubber and plastics
XI. Other nonmetallic minerals
XII. Basic metals and fabricated metal
XIII. Machinery,
XIV. Electrical and optical equipment
XV. Transport equipment
XVI. Manufacturing and recycling
XVII. Electricity, gas, and water supply
XVIII. Construction
XIX. Sale, maintenance, and repair of motor vehicles and motorcycles; retail sale of fuel
XX. Wholesale trade and commission trade, except for motor vehicles and motorcycles
XXI. Retail trade, except motor vehicles and motorcycles; repair of household goods
XXII. Hotels and restaurants
XXIII. Inland transport
XXIV. Water transport
XXV. Air transport
XXVI. Other supporting and auxiliary transport activities; activities of travel agencies
XXVII. Post and telecommunications
XXVIII. Financial intermediation
XXIX. Real estate activities
XXX. Renting of M&Eq and other business activities
XXXI. Public administration and defence; compulsory social security
XXXII. Education

XXXIII. Health and social work

XXXIV. Other community, social, and personal services

The factors of production include

I.	Pun-r-low skilled	XIII.	Bal-r-low skilled
II.	Pun-r-high skilled	XIV.	Bal-r-high skilled
III.	Pun-u-low skilled	XV.	Bal-u-low skilled
IV.	Pun-u-high skilled	XVI.	Bal-u-high skilled
V.	Sin-r-low skilled	XVII.	Pun-r-cap
VI.	Sin-r-high skilled	XVIII.	Pun-u-cap
VII.	Sin-u-low skilled	XIX.	Sin-r-cap
VIII.	Sin-u-high skilled	XX.	Sin-u-cap
IX.	KP-r-low skilled	XXI.	Kp-r-cap
X.	KP-r-high skilled	XXII.	Kp-u-cap
XI.	KP-u-low skilled	XXIII.	Bal-r-cap
XII.	KP-u-high skilled	XXIV.	Bal-u-cap

Pun, Sin, KP, and Bal refer to Punjab, Sindh, Khyber-Pakhtunkhwa, and Balochistan, respectively, whereas r and u refer to rural and urban. Cap represents capital.

Eight household categories introduced in the SAM 2017 are: as

hh-pun-r	hh-kp-r
hh-pun-u	hh-kp-u
hh-sin-r	hh-bal-r
hh-sin-u	hh-bal-u

Pun, sin, kp, bal, u, and r have similar meanings as mentioned above.

Furthermore, in Tables 3.2, 3.3 and 3.4, we aggregated some sectors, such as agriculture & food (I & III), textile, leader & rubber (IV, V, X), mining, metals & chemical (II, IX, XI, XII), construction & energy (VIII, XIV, XVII, XVIII), other manufacturing (VI, VII, XIII, XVI), trade (XIX, XX, XXI), transportation & communication (XV, XXIII, XXIV, XXV, XXVI, XXVII), and services (XXII, XXVII – XXXII).

Parameter Values for Commodities/Activities

The parameter values for commodities/activities or industries are given using the same order (and numbering) as given above:

Table 1A: Parameter Values

Commodities/Activities	SIGMA1	SIGMA2	SIGMA3	EXP_ELAST
I	2.97	2.9	2.95	3.5
II	2.89	2.85	2.89	2.3
III	2.7	2.7	2.7	2.3
IV	2.25	2.25	2.25	8.25
V	2.25	2.25	2.25	5.34
VI	1.7	1.65	1.75	5.25
VII	1.7	1.65	175	2.5
VIII	1.7	1.65	1.75	2.3
IX	2	1.8	1.9	2.3
X	2	1.8	1.9	5.3
XI	1.5	1.5	1.5	3.1
XII	1.5	1.5	1.5	2.8
XIII	2.5	2	2.25	3.87
XIV	2.5	2	2.25	2.3
XV	2	2	2	2.3
XVI	2.5	2.5	2.5	2.3
XVII	1.9	1.9	2	2.3
XVIII	1.9	1.9	2	2.75
XIX	1.25	1.2	1.2	2.75
XX	1.25	1.2	1.2	2.60
XXI	1.25	1.2	1.2	2.50
XXII	1.25	1.25	1.25	2.45
XXIII	2.1	2	2	2.45
XXIV	2.3	2.2	2.2	2.73
XXV	1.7	1.7	1.65	2.73
XXVI	1.5	1.5	1.55	2.25
XXVII	1.5	1.5	1.55	2.25
XXVIII	2.7	2.5	2.6	2.10
XXIX	1.5	1.3	1.5	2.10
XXX	1.5	1.3	1.5	2.2
XXXI	2.8	2.8	2.8	2.13
XXXII	1.7	1.5	1.6	2.13
XXXIII	1.7	1.5	1.6	2.13
XXXIV	1.7	1.5	1.6	2.15

Note: These parameter values are based on the statistical database used and also adapted from various studies such as Zeshan (2020), Debowicz et al., (2012), and Dorosh, Niazi and Nazli (2004) with some adjustments for the number of sectors, whereas SIGMAPRIM and SIGMAOUT are adapted from Amir et al. (2013).

REVISITING URBAN IMMOVABLE PROPERTY VALUATION: AN APPRAISAL OF SPATIAL HETEROGENEITIES IN PUNJAB USING BIG DATA

Shoaib Khalid and Fariha Zameer

ABSTRACT

This study undertook the urban immovable property valuation in two major cities of Punjab, namely, Lahore and Faisalabad, using big data and advanced spatial analysis techniques to explore the significant impact of location-specific parameters on urban immovable property prices. To compute the immovable property values, big data analytics were employed in Geographic Information System (GIS). The traditional hedonic price models give little importance to the spatial characteristics of individual housing units and revolve around the structural attributes of houses. However, spatial heterogeneity should be considered while appraising residential property prices since the house characteristics may vary spatially. To address this issue, we established different valuation models based on the ordinary least square regression and the Fast Geographic Weighted Regression (FastGWR) model, a scalable open-source implementation of Python and Message Passing Interface (MPI) that can process millions of observations. Using these valuation models, the total net worth of the residential real estate market in both study areas was estimated. The results demonstrated the excellent performance of our valuation models and displayed spatial heterogeneity with higher accuracy. The valuation models explained the relationship of explanatory variables with response variables up to 75 per cent for Faisalabad and around 85 per cent for Lahore. Results show that the floor area, proximity of health facilities, recreational sites, and marketplaces add a premium to prices, while the nearness of educational institutions, worship places, and solid waste transfer stations or dumping sites lessen the property values in both cities. However, the proximity of industrial units and graveyards affects property values negatively in Lahore but positively in Faisalabad.

1. INTRODUCTION

Background

The economic value of some commodity is the value that is estimated based on its benefits for an individual. In other words, the degree of utility of a certain product for people is known as its economic value. Although there are certain methods to quantify this pretend value, it is difficult to measure the exact economic value of something. Market prices are set using estimates of economic values and by taking into consideration the perceptible and imperceptible features of the commodities (Fisher et al., 2015; Gabrielli & French, 2020; Lovett, 2019). A hedonic valuation is a widely-used approach to quantifying the economic value of a commodity, which uses statistical regression analysis of various attributes of that commodity based on its past transactions. Attributes of a commodity determine the level of its utility for an individual and, in turn, influence its market price. Thus, hedonic price models are devised based on the impact of each attribute that contributes to the total market price of the commodity. The basic arrangement of a hedonic price model is a functional relationship between the price of heterogeneous goods and their qualitative characteristics (Baranzini et al., 2008, 2010; Bateman et al., 2001).

Certain structural and locational features can be identified that significantly influence the value of urban immoveable properties in a specific area using the hedonic pricing approach. There is evidence that certain locational factors strongly influence the value of the real estate, such as the area of the property, covered space, structural characteristics of the building, walkability, security, and the provision of urban amenities (electricity, natural gas, solid waste collection, sewerage system, paved streets, clean drinking water, distances to the markets, schools, workplaces, recreational sites and health facilities). The hedonic pricing models for the valuation of real estate properties can be built using non-spatial techniques, such as ordinary least squares (OLS) method, or spatial techniques, such as geographically weighted regression (Boza, 2015; Erickson et al., 2011; Gilderbloom et al., 2015; Machin, 2011; Pace & Gilley, 1998; Pagourtzi et al., 2003; Shabana et al., 2015). A hedonic valuation of real estate properties is usually based on four sets of explanatory variables, namely structural characteristics, locational attributes, environmental features, and, lastly, neighbourhood traits, whereas selling prices are considered a response variable in multiple regression (Freeman, 1981).

Mapping the values of urban immovable properties is important to understand the dynamics of the real estate market and to monitor unrestrained prices, which is critical for housing market appraisals (Brown et al., 2020; Gaffney, 2009). It is also needed to review the master plans of cities for sustainable planning in future (Barreca et al., 2020). Valuation maps may assist sellers and buyers (Goix et al., 2019; Wang et al., 2020) as well as government agencies whenever they need value assessment of immovable properties as in the case of real property taxation (Chapman et al., 2009; Larson & Shui, 2020). Taxes levied on immovable properties based on their fair market values enhance the fairness of the tax because urban services provided by the government in a particular area maximise the capital benefit for the owners. The scale of the provision of urban services positively contributes to the market value of real estate properties, while the provision of urban amenities requires government funds.

Housing Market and Policies in Pakistan

A well-operational housing market is one of the key features of a robust economy but, unluckily, the housing market in Pakistan is not functioning well. The housing demand in urban centres is increasing due to the rapid urbanisation and rural-to-urban migration in Punjab. On the other hand, the supply side is not performing well due to various factors. Large tracts of precious land in the heart of the city cores are not being utilised wisely and locked up unnecessarily. In the absence of master planning, the policies and processes of urban planning prove counterproductive most of the time. The pace of the urban infrastructure service provision is inadequate because the authorities run short of funds due to an inefficient revenue collection from property taxes, which is the main source of funds for improvements. Property rights are weak and there is a mess of regulations that hinder the land development and construction processes. The supply of finance for land acquisition and development is limited by the public as well as the government. The supply of land for housing is becoming scarce in urban centres of Pakistan, like Karachi and Lahore, due to obsolete and poor regulations. Moreover, the rate of housing construction has not kept pace with the population growth rate and urbanisation. In Lahore, house prices have increased by 6.5 times in 20 years from 1992 to 2012, but the average household income has not improved at the same rate. Pakistani cities have higher incremental ratios of house prices and house rents to household income as well as construction cost to income compared to other Asian countries. This quick upsurge in the market has made housing unaffordable for a huge number of families living in urban centres giving rise to informal housing and slums. More than one-third of the urban population of Lahore and around half of the residents of Karachi are forced to live in katchi abadis (Dowall & Ellis, 2009; Haque, 2015; Wani et al., 2020; Yuen & Choi, 2012).

Housing markets in Punjab do not operate well and the case is even worse in the big cities, including Lahore and Faisalabad where affordable housing is scarce, the housing deficit is wide, and a large proportion of the population is destined to live in informal housing (Malik et al., 2020; Wajahat, 2012). On the other hand, around 75 per cent of the residential plots in nearly 50 per cent of housing colonies are owned by certified speculative investors who never construct houses but only bid up prices. Therefore, affordable housing has become a big challenge for low-income masses (Zaman & Baloch, 2011). The reasons behind this practice include the safe investment opportunities in the real estate sector, handsome returns on investment, non-levying of any taxes on vacant residential plots by the authorities, and above all, easy prospects to evade taxes through under-invoicing at the time of sale-purchase because of the flawed valuation system. The price per marla of vacant residential plots has risen at a rate of 41 to 140 per cent per annum in Lahore (Gul et al., 2018), but the increases in official values have not matched the fair market prices.

Residential property prices are highly explosive even in the short run, such as even monthly, in Pakistan. The main reason behind the sharp upward trend of house prices is high demand and short supply due to speedy population growth and quick urbanisation. This situation suits the investors to reap higher profits but burdens the public at large. Authorities seem not serious to address the issue as housing policies are inconsistent and not published regularly. Launching of affordable and large-scale housing schemes by the public sector can only stabilize the ever-increasing house prices (Ahmed et al., 2021).

Immovable Property Valuation by FBR

FBR issues circulars annually, directing each Regional Tax Office (RTO) across the country to constitute committees comprising a Chief Commissioner (CC), three officers from the respective RTO, a representative of property dealers, and a representative of the Association of Builders and Developers of Pakistan (ABAD). The representatives of property dealers and ABAD are nominated by the concerned CC. The committee assesses the valuation tables and notifies revised valuation tables for tax imposition within the jurisdiction of the RTO. A representative of the CC may physically visit the areas to observe major valuation changes (FBR, 2020). The procedure seems an exercise based on the individual discretion of the CC. Moreover, no scientific method of calculation is considered. Table 1 shows the amount by which the market rates have remained higher than the FBR valuation table rates.

Table 1. Price per Marla (PKR) of FBR and the Property Portal of Zameen.com

DHA Lahore	Aug-2016			Feb-2019			Jul-2019			Jan-2020		
	FBR	Zameen	Difference	FBR	Zameen	Difference	FBR	Zameen	Difference	FBR	Zameen	Difference
Phase I	672000	1908900	184%	806400	2140200	165%	880000	2190375	149%	860000	2144250	149%
Phase II	552000	1935900	251%	662400	2240325	238%	880000	2124000	141%	800000	2183400	173%
Phase III	552000	3243600	488%	662400	2909250	339%	960000	3082950	221%	800000	3176325	297%
Phase IV	525525	2095875	299%	630630	2482650	294%	1080000	2591550	140%	850000	2478150	192%
Phase V	420000	2783250	563%	504000	3072150	510%	1240000	3221100	160%	900000	3313350	268%
Phase VI	405000	2184975	440%	486000	2400750	394%	1100000	2461950	124%	850000	2450025	188%
Phase VIII	-	-	-	378000	2522700	567%	840000	2577600	207%	500000	2557350	411%
Rahbar	-	-	-	405600	2403450	493%	510930	2403450	370%	400000	2395125	499%

Source: Zameen.com and FBR

Immovable Property Valuation by District Collector (DC Rates)

District Collector's valuation (DC rates) have been even poorer than FBR valuation rates as DC rates were based on just the average of the past disclosed property transaction prices by the taxpayers of the respective localities till 2019. Sellers and buyers do not reveal the actual transaction amounts but get registered minimal values to avoid taxes and scrutiny by the authorities (Dowall & Ellis, 2009). DC rates notified in 2020 for Lahore appear to be somewhat better than in the past. Perhaps the DC office has adopted a method similar to the one adopted by the FBR. Although the details of the recent method are not known yet, the rates are still far lower than the fair market prices. The following table may tell the whole story. The percentage difference in the table shows that fair market rates have remained significantly higher than the DC rates.

Table 2. Price per Marla (PKR) of DC Rates and the Property Portal of Zameen.

DHA Lahore	2016			2017			2020		
	DC Rate	Zameen	Difference	DC Rate	Zameen	Difference	DC Rate	Zameen	Difference
Phase I	560000	1773000	217%	600000	1778400	196%	962500	2144250	123%
Phase II	460000	1967850	328%	520000	2026800	290%	962500	2183400	127%
Phase III	460000	3138975	582%	520000	3086550	494%	110000	3176325	189%
Phase IV	420000	2249325	436%	490000	2255400	360%	825000	2478150	200%
Phase V	280000	2606175	831%	450000	2801250	523%	962500	3313350	244%
Phase VI	270000	2141325	693%	320000	2230875	431%	687500	2450025	256%
Phase VIII	-	-	-	-	2293425	617%	550000	2557350	365%
Rahbar	-	-	-	-	-	-	378000	2395125	534%

Source: Zameen.com and BOR

Significance of the Study

The real estate market of Pakistan can be a vibrant source of economic growth as around 60-70 per cent of the total wealth of the country is stored in its real estate assets. The real estate market contributes around 2 per cent to the GDP of the country, while the combined contribution of THE housing and construction sectors is nearly 9 per cent. The Federal Board of Revenue valued the real estate sector of the economy of Pakistan at about US$700 billion. The return on Investment in the real estate market of Pakistan is exceptional, which could be over 100 per cent in some cases (Ouattara et al., 2018). Pakistan is home to over 225 million people and the population is growing at a high rate with an average annual growth rate of 2.4 per cent. The average household size in Pakistan is 6.5 persons per household (Pakistan Bureau of Statistics, 2019). The year-on-year housing requirement of the country is 700,000 units per year but only about half of this demand is being met. Around 1 million housing units per year are needed to meet the prevailing need only. The urban housing deficit is estimated to be around 4 million units, while it is around 6 million units in rural and peri-urban areas. Hence, the overall housing shortfall is around 10 million units and it is growing. Around 32 million households in Pakistan have no shelter or live in very poor conditions (S. B. of Pakistan, 2019; Rizvi, 2018). Launching of affordable housing projects is need of the hour in Pakistan. The current study can help planners and policymakers identify the locations for affordable housing projects in urban centres like Lahore.

Whenever the government acquires land for new infrastructure building or some particular project, the compensatory payments to the owners are made according to DC valuation tables, which are always far lower than fair market values, which result in mass protests, demonstrations, and unrest (Sabir et al., 2017). Thus, a more realistic valuation of real properties close to the fair market price is essential also for fair compensation to the landowners and not only for tax purposes (Malaitham et al., 2020).

A precise housing property valuation is important to facilitate real estate market stakeholders, such as sellers, buyers, property agents, investors, financer banks, and insurance companies. An accurate valuation cannot be carried out by ignoring the locational attributes and considering only the physical characteristics of houses. The location of a house sometimes carries more weight for its price than structural features (Mankad, 2021).

Both the land use change in cities and the urban planning by the authorities are symbiotic (Liang et al., 2018), therefore, a precise prediction of the urban land use growth in rapidly expanding cities, such as Lahore and Faisalabad, is important for an operative, adaptable, and sustainable urban planning. Many public properties, such as the Government Officers Residences (GORs), railway land, and properties held by the Evacuee Trust, the Auqaf Department and other institutions are lying idle or are not being utilised to their full potential in terms of commercial usage. The public sector alone owns urban properties worth trillions of rupees in Pakistan that are either underutilised or not used at all. These properties are dead capital since they cannot be used for wealth generation. All these insights are useful for policymakers.

Objectives

The general objective of this study is to build a valuation model for urban immovable properties based on spatial attributes by utilising big data ($n \geq 1.2$ million) in two big cities of Punjab, i.e., Lahore and Faisalabad. The specific objectives are:

1. To examine the dynamics of urban immovable property values based on location-specific parameters using big data.

2. To investigate the spatial variations in urban immovable property values.

2. LITERATURE REVIEW

Hedonic Real Estate Valuation and Market Characteristics

Hedonic models are employed to estimate the impact of various factors that affect the market price of real estate properties. There are two key assumptions of a hedonic valuation model. The first assumption is that the market is perfectly competitive and that both buyers and sellers are perfectly informed. The second assumption is second that there are no discontinuities in the product range available to buyers. There are certain extensions of the hedonic pricing model which relax these assumptions applicable to the data-generating process (DGP). The hedonic valuation approach is important not only to estimate price function in the housing market but also to explore the underlying preferences of the people regarding housing characteristics in a certain vicinity (Freeman, 1981; Taylor, 2008). A housing market is somewhat different from a perfectly competitive market because the products in a housing market are distinguished to varying degrees, which means that there are discontinuities present in the product range available to buyers. The information about the quantity and quality of the features that constitute the price of the product in a housing market is difficult and costly to obtain. The selling transection in a housing market becomes more of a process rather than an event. This selling transection process may influence the market price of the houses. The influence of the selling transection process should be estimated through the inclusion of time-on-market variable into the hedonic house price model (Knight, 2008).

Floor area and structural characteristics of the residential property also play a key role in determining its value (Xiao et al., 2017) but locational attributes are also equally important. Therefore, the inclusion of this set of variables may give a true assessment. Spatial hedonic valuation models work well in this regard (Helbich et al., 2013; Koschinsky et al., 2012). The location of residential properties strongly affects their value. House buyers intend to pay more for easy access to transport facilities, for example (Seo et al., 2014; Tian et al., 2017). The relevant literature suggests that house values and proximity distances to certain amenities and urban services are strongly correlated but the results are inconclusive. In some cases, the relationship is positive, such as the nearness of shopping facilities, gyms, metro stations, workplaces and educational institutions, while in some cases it is also negative, such as the nearness of supermarkets and solid waste dumping and waste transfer stations. Houses in suburban vicinities fetch lower prices compared to the inner-city areas. Moreover, government policies, street and road conditions, water supply, and security affect house prices significantly (De & Vupru, 2017; Xiao et al., 2017; Yang et al., 2018b; Zhang et al., 2018). Real estate valuation methods can be categorised into two classes, i.e., one class of models deals with spatial dependence, while the second class deals with spatial heterogeneity (L. Krause & Bitter, 2012).

Real Estate Valuation and Spatial Dependence

The usage of the spatial dependence approach in immovable property valuation research has a long tradition and a vast literature can be traced in this regard. The degree of spatial autocorrelation among independent observation values in geographic space is known as spatial dependence (Crawford, 2009). Spatial autocorrelation means the presence of a tendency between the observation values to be more similar when they are situated closer and more dissimilar when they are positioned at distant sites in geographic space. It is considered positive when high observation values cluster near highs and low observation values found near lows while it would be negative when the case is opposite (Griffith, 2004; Hubert et al., 1981). According to the proposition of the hedonic price theory, many utility-bearing features contribute to the value of heterogeneous commodities and these are the quality features that increase or decrease the utility of the users, not the commodities themselves. Thus, econometrically, hedonic price indices are constructed by regressing the commodity price on quality features to estimate the implicit prices. Dissimilar combinations of these features affect the preference and utility of the users differently, so the impact estimation of all possible attributes is important for a good valuation appraisal (Lancaster, 1966; Rosen, n.d.). A user's willingness to pay for a certain commodity feature denotes the hedonic price of that feature under the assumption of consumer utility maximisation. The price function between attributes of a commodity and its price can be created using historical transactions to estimate the implicit hedonic prices of specific features of the commodity. As the attributes of houses vary significantly, a hedonic price model can be established to value their characteristics separately. Landscape characteristics of houses affect their prices remarkably, so an appraisal of housing valuation based on location-specific parameters is imperative (Goodman, 1978).

Can (1990) reported that the inclusion of spatial attributes spillover yielded a more accurate appraisal of the urban immoveable properties. Can & Megbolugbe (1997) maintained that the spatial dependence, hidden in the residential real estate data sets, affected the accuracy of value estimation considerably. Liao & Wang, (2012) noted a U-shaped pattern of spatial dependence among the house prices in the city of Changsha, China, which meant that proximity of high- and low-priced houses tended to contribute more positively to the implicit house prices, while the proximity of medium priced houses tended to contribute comparatively less. They informed that the house prices tended to be lower in densely built-up surroundings which was a surprising result because it contradicted the results of the western city market appraisals where housing units in central business districts were always found to be more expensive than the houses at distant locations.

Expansion Method and Spatial Heterogeneity

The occurrence of spatial heterogeneity is a recognised issue in real estate datasets and a common solution to this issue is to segregate the geographic area of interest into homogeneous regions with specified functions while undertaking value estimation (Kauko, 2003). Spatial heterogeneity can be defined as the variability of a qualitative or measurable value in a distribution in space (Dutilleul & Legendre, 1993). So, for this study, the phenomenon of spatial heterogeneity can be referred to when real estate properties with similar characteristics are valued differently in different localities across the space of interest. Examining the spatial heterogeneity in real estate valuation has also a long history that started with the development of expansion techniques by Casetti, (1972). The expansion method contains a procedure for reassembling a simple initial mathematical model to broaden its scope by redefining a few or all of its parameters. A spatial version of a non-spatial model can be created through this method by redefining the parameters into functions of spatial variables that index a fundamentally pertinent framework of that initial model (Casetti, 1997).

Can (1990) employed the spatial expansion method to construct a hedonic regression model for the appraisal of spatial variation of residential property prices with respect to the influence of neighbourhood attributes. Anselin (1995) used local indicators of spatial association (LISA) to account for spatial heterogeneity along with spatial dependence and concluded that LISA was a statistic capable of explaining any observation value with respect to nearby spatial clustering of similar values where the sum of LISAs for all observation values might be proportionate to an indicator of global spatial association. Brunsdon et al. (1996) designed a geographically weighted regression (GWR) to capture spatial non-stationarity using the expansion technique and redesigning the parameters of Kernel Regression (KR). It is a local type of spatial statistic that estimates and maps the actual parameters for each observation location in space and, thus, allows the visual analysis instead of estimating the parameters fitted to a trend surface as in the case of global techniques. Fik et al. (2003) used an interactive variable approach to capture spatial heterogeneity in house prices and argued that the magnitude of the effect contributed to house prices by different externalities is always unique at different locations. Hence, the absolute location should always be regarded as interactive with other contributing variables. Bitter et al. (2007) inspected the spatially varying relationship between house prices and housing attributes in Tucson, Arizona, USA. They compared the performance of the global spatial expansion method with that of GWR and concluded that the GWR had more prognostic accuracy and explanatory power than that of the spatial expansion method. The results of their empirical study strengthened the notion that the degree of the effect of housing attributes on house price differs over space. The matter of spatial heterogeneity can only be resolved through models with spatially varying coefficients such as submarket-based dummy variable models, spatial expansion methods or GWR models. The

GWR offers higher accuracy and has the exclusive benefit of providing the visual-spatial distribution of implicit values (Wen et al., 2017).

Capturing spatial heterogeneity is critical for precise house price appraisals because real estate property prices differ from one place to another within urban centres and are affected significantly by the provision of urban facilities. However, the magnitude of the effect of these urban facilities is not the same at every location in space (Redfearn, 2009; Yuan et al., 2020). Wu et al. (2020) reported that housing submarkets can be identified better using the spatial heterogeneity approach instead of the spatial dependence approach as the spatial dependence approach overlooks the complexity of the urban space. Jiang (2018) advocated the superiority of the spatial heterogeneity approach, which is based on the scaling law, over the spatial dependence approach, which is based on Tobler's law. He argued for a paradigm shift from the use of Euclidean geometry to fractal geometry for geospatial analyses, especially in cases of working with big data.

Using Big Data for Real Estate Market Appraisal

Data that is very large, complex, and continuously piling up over time at some media platform, which is difficult to process and interpret using the usual traditional techniques is referred to as 'big data,' such as the data from social media, transactional data of organisations, and machine-generated data. When this data is also geo-referenced, it can be called spatial-big data or geo-big data (Dalton & Thatcher, 2015; Gao et al., 2017; Goodchild, 2013; Guo et al., 2014). Wu et al. (2016) used big data in the form of recorded check-in data from the Sina Visitor System (a social media platform) to analyse the willingness of home purchasers to pay for different urban amenities in the Chinese city of Shenzhen. Yang. Chu, et al., (2020) scrutinised the capitalisation effect of proximity as well as accessibility to a bus rapid transit (BRT) system on residential property value within a 1.5-km corridor on both sides of BRT lanes in Xiamen Island, China. They extracted big data using a web crawling programme from Fang.com, which is a mainstream property portal in China. Singh et al. (2020) extracted big data on housing sale prices and other variables from various internet platforms using five software packages for data mining, manipulation, and classification, such as gdata, tidyverse, stringr, lubridate, and caret. Ma et al. (2020) argued that the use of big data for real estate market appraisal is advantageous as it can provide a large number of factors, offering the researcher more variety of features to select from, and a chance to mine more veiled information that might be disregarded while using traditional datasets and methods. Furthermore, the analysis of big data helps to model the issues more accurately because the field-specific algorithms seem robust enough in identifying the relationships as compared to the traditional statistical techniques.

Real Estate Valuation Models

The OLS is a type of linear regression that is used to estimate the unknown parameters and to test the nature of the relationship between a dependent and explanatory variable(s) in hedonic valuation models (Dismuke & Lindrooth, 2006; Frost, 2019). The OLS assumes that the relationship between the regressand and the regressors is stationary and constant for all the observed locations across the whole area under study. It uses a single equation to estimate the relationship (Wooldridge, 2016). Several hedonic pricing studies have used OLS for parameter estimation but quite often, the regression residuals have spatial autocorrelation, which is against the assumption of randomness (Pace & Gilley, 1998). Measuring spatial association is very important while working on a spatial model because spatial autocorrelation may lead to critical errors in model interpretation. Two global techniques, i.e., Moran's I and GI*, are the best choices among many other tools to identify spatial patterns in the dataset and the degree of spatial association. These techniques also enable a researcher to detect local "pockets" of dependence when used jointly. A combination of both techniques (Moran's I and GI*) may better help a researcher working with a spatial model (A. Getis, 2008; A. Getis & Ord, 2010a).

Furthermore, an asymptotically normally distributed statistic, such as Oi, can be employed to test for a local spatial autocorrelation and hot spot analysis in the presence of global autocorrelation because type-1 errors may occur when the global autocorrelation structure in the dataset is ignored. Type 1 error results in false positives, which happens in hypothesis testing when the null hypothesis is rejected even if it is true. Therefore, a local spatial statistic must be interpreted according to the degree of global spatial association present in the data. Oi is the most suitable tool for hotspot analysis with larger datasets but it assumes spatial stationarity (Ord & Getis, 1995, 2001). Spatial non-stationarity within spatial data can be explored using spatial cluster analysis with the help of a combination of two tools, which are GI* and K-means. K-means is a multivariate cluster identification technique that divides data into homogeneous groups, taking into account the geographical location of features and their spatial relationships. Spatial cluster analysis assesses the degree of spatial autocorrelation between features and quantifies the statistical significance of identified clusters (Peeters et al., 2015).

The presence of spatial dependence is unavoidable in any real estate dataset, which can be modelled using a spatial weight matrix and then incorporated into either a Spatial Lag Model (SLM) or a Spatial Error Model (SEM) (Kim et al., 2003). Spatial Lag Model (SLM), also known as Spatial Autoregressive Model (SAM), and Spatial Error Model (SEM) are two widely used basic models to capture spatial dependence in real estate market analysis. SLM or SAM considers a linear combination of real estate property values in nearby space, while SEM considers the impact of location-specific variables on the dependent variable and the resulting spatial autocorrelation in the error terms. A General Spatial Model (GSM), which is capable of yielding better estimation results, may be constructed by combining both the SLM and SEM models of spatial dependence that can be developed by imposing restraints on the Spatial Durbin (SD) model (Brasington & Hite, 2005; LeSage, 2008).

The GWR is being employed extensively in hedonic valuation modelling to explore spatial heterogeneity. It is a computation-intensive tool which estimates location-specific parameters. The GWR is a weighted least-squares regression that assigns more weight to some observations than others in computing the regression coefficients. The weights change at each point location, which is why the GWR is known as a local model technique. The GWR is a "distance-decay" model of spatial association, which follows Waldo Tobler's first law of geography. The influence of the calibrated data points on weights keeps decreasing with the increasing distance around a certain location of interest. Most of the time, the distances are taken in a straight line from one location to another, known as the Euclidean distance (ED) in the GWR (Brunsdon et al., 1996; A. S. Fotheringham, 1997; A. S. Fotheringham et al., 2015; Scott & Janikas, 2010). The coefficients are taken as functions of a specific spatial location in the GWR. A coefficient surface (as opposite to a trend surface in global models) of the whole area under study can be created with a set of continuously varying estimated coefficients at every location to explore the spatial heterogeneity using the GWR where each set of regression coefficients is estimated by weighted least squares (Brunsdont et al., 1998; Lu et al., 2014). Gollini et al., (2015) introduced the GW model and its implementation in R-package to calibrate and map the localised relationships of covariates using a moving window weighting (MWW) technique.

Traditional GWR techniques seem to fail in the case of larger datasets with millions of observations. The processing time increases exponentially with an increasing number of calibration locations and GWR scales may take more than two weeks to compute results on a model with one hundred thousand observations (Harris et al., 2010). Yu (2007) and Feuillet et al. (2018) had to split their datasets into several subsets of geographic units due to the computational limitation of the traditional GWR tools compromising the quality of results and reducing the utility of their study because splitting of the created bias under boundary and zoning effects which posed a hurdle in capturing the true spatial heterogeneity. Fotheringham et al. (2017) attempted to improve the scalability power of the traditional GWR by relaxing its assumption that all the coefficients being estimated in the model function at a solo spatial scale and, instead, assigning each relationship of covariates a unique spatial scale to minimise concavity and mitigate over-fitting. They named the new improved extension as multiscale geographically weighted regression (MGWR). It is based on the Bayesian spatially varying coefficients (SVC) model and uses a back-fitting algorithm in bandwidth vector selection and model calibration. Oshan et al. (2019) introduced a Python-based mgwr software package for the deployment of the MGWR model as an alternative to the original R-based package with enhanced efficiency in terms of kernel selection. They incorporated an adaptive kernel option whose width changes according to the spatial distribution of observations automatically. Z. Li & Fotheringham (2020) upgraded the computing ability of the MGWR to make it applicable to a dataset with observations up to one hundred thousand by extending the parallelisation framework through a method of splitting the calibrations block-wise. They compared this new version with the original one and found the improved version five hundred times faster in calibrating the model.

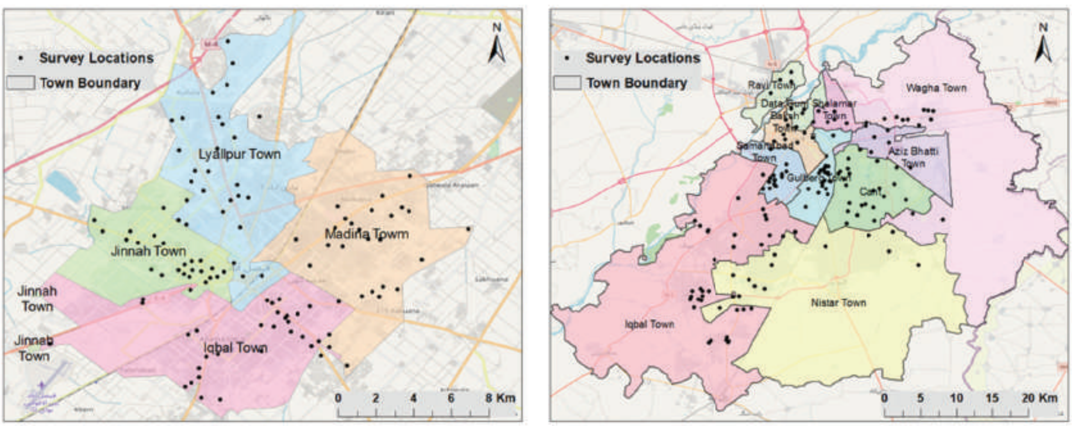
Li et al. (2019) developed FastGWR, which was reported to be a more sophisticated extension of the GWR, capable of computing and calibrating data with millions of observations. They compared the computation time performance of four open-source GWR software packages – FastGWR, MGWR (PySAL), GWmodel, and spgwr – with subsets of different sizes from a total of 1.28 million house price observations. They found FastGWR, 12 to 3,400 times faster than the rest of the implementations when applied to a dataset of 10 thousand to 15 thousand observations, while on a dataset with 20 thousand observations, only FastGWR was able to successfully calibrate a GWR model. FastGWR reduces the memory constraint significantly from O (n2) to O (nk). Here, 'n' denotes the number of observation locations and 'k' denotes the number of explanatory variables. Furthermore, it uses parallel model diagnostic calculation procedures. which considerably decreases the required computation time for a GWR model calibration.

3. RESEARCH METHODOLOGY

Study Area

This study covered two major cities of Punjab, i.e., Lahore and Faisalabad.

Figure 1. The Study Area: Faisalabad and Lahore



Data and Sources

The data used in this study were collected from different sources including governmental and non-governmental organizations and web scraping programmes. A variety of data types were used such as property price data, house parcels, parcel types, road network, urban land use, and location of important places. There are several online portals for property buying and selling in Pakistan including www.prop.pk, www.zameen.com, www.homespakistan.com, www.realproperty.pk, www.pakrealestate.com, www.aarz.pk, www.realproperty.pk, www.graana.com, and www.propertylink.pk. Among these property portals, zameen.com is the biggest online portal, which provides services for buying, selling, and renting all types of residential and commercial properties. The services of this online portal are available in more than ninety cities in Pakistan. It records the location, price, and structural attributes of houses including the area, number of bedrooms, bathrooms, and number of storeys.

The information about the house price and the location allows the analysis of the distribution of property values across the city. We obtained this information from zameen.com through a web scraping programme since the website does not have any Application Programming Interface (API) to download the information. The web scraping programme can download the required information from a browser at a low cost (Li et al., 2017). The data was downloaded for 3,526 houses in Faisalabad and geocoded the properties on a map. It is noteworthy that these are not the prices at which the properties are bought or sold but the prices of the residential properties advertised by homeowners or real estate agents. We used those property listings that are usually 3 to 5 per cent higher than the actual transactions as a proxy of actual market prices since there is no mechanism to record real estate transactions in Pakistan. Although the selling price is disclosed at the time of property transfer for tax payments, the level of under-invoicing is very high and the official records only account for 5 to 50 per cent of the fair market prices of properties (Wahid et al., 2021). We acquired the property listing data through web scraping for the year 2019. It is noted that though the inclusion of the temporal stability of the transactions would have been an ideal situation, such information is typically not available in most developing countries and Pakistan is no exception. The geometry for house parcels was generated through various methods such as digitising detailed property maps and satellite images, and a part of the parcel dataset was obtained from the Urban Unit Faisalabad, an organisation dedicated to providing urban sector planning and management services. The dataset contains information about the type of property, taxable or exempted, and the location of the property. The locational and environmental attributes were generated in the study area using GIS and Google Maps.

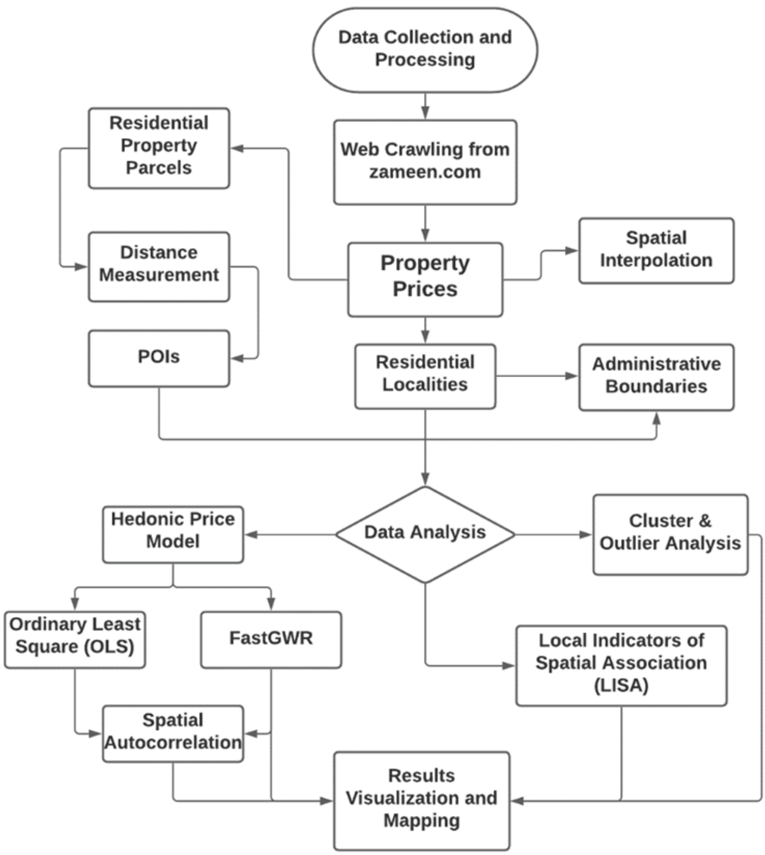
Property data, as advertised on zameen.com, containing asking prices of 26,031 geo-referenced real estate properties in Lahore and 3,526 properties in Faisalabad, were selected for price surface generation, The data were retrieved through a web scraping program (Thomas & Mathur, 2019; Yang, Chu, et al., 2020). We used this big data comprising residential real estate listings as a proxy of fair market prices (Wu et al., 2016) because the recorded transaction data from the FBR and DC tables are under-invoiced and are only 5 to 50 per cent of the market prices. Therefore, the data from government agencies may not produce appropriate estimations. Although the listings contain seller prices, they may be assumed to be very close to actual transaction values since the sellers do not decrease the prices by more than 3 to 5 per cent, on average.

Previous relevant literature suggests that the actual transactional prices of real estate properties are always significantly correlated with the seller asking prices. Thus, using the seller asking prices as a proxy of the actual transactional price in a hedonic appraisal does not distort the estimates considerably unless one or more independent variables are systematically associated with one of the two prices. Moreover, when the demand for real estate properties exceeds supply, the real estate market turns into a seller-oriented market and sellers become price setters. For this reason, actual prices come very close to the seller asking prices in such markets (Ibeas et al., 2012; Salon et al., 2014; Yang, Chau, et al., 2020). Moreover, there is a possibility that a real estate listed on a property portal for sale does not necessarily mean it would be sold as it may be later withdrawn from the listings. However, it may give an idea of the seller's perception. Relevant literature suggests that these types of listings may be used in situations where actual transaction records are not readily available. Examples include a study conducted by Katsalap (2008) in Ukraine and Yang, Chu, et al. (2020) in China. Land use parcels datasets of Lahore and Faisalabad were acquired from the Urban Unit Lahore and deficiencies in property parcels dataset of the Urban Unit were removed through digitisation and geo-referencing.

Processing Operations

All the following processing operations were executed using ArcGIS 10.4.1 software. The selected properties were displayed using coordinate values for the sake of creating a price surface for whole cities. Raster price surfaces were generated using geo-referenced property points as input through Inverse Distance Weighted (IDW) interpolation with barriers. The IDW is a deterministic method of randomly distributed points, which computes the values of unknown points with the weighted average of known values (Hu et al., 2013; L. Li & Revesz, 2004; S. Li et al., 2017). The points close to the centre of the cell being estimated gets more weight in the averaging process and it assumes that the values closer to each other are similar than the values at larger distances. The raster price surface layers are then converted to point layers. The price field was shifted from point feature price surface layers to property parcels and the resultant parcels were converted to points. Near tables were generated within attribute tables of property point layers for all the selected spatial amenities. The near tables contain Euclidean distance in metres from a property point to the nearest relevant spatial amenity point. We used the planar method as our study areas were not extended to a large geographic space. Moreover, the data layers were projected. No search radius was defined to give every property point some value of the distance because the selected spatial amenities were well distributed all over and not confined to a specific part of the cities.

Figure 2. Methodological Framework of the Research Project



The Spatial Hedonic Valuation Model

A hedonic house property valuation model based on various attributes was constructed as follows:

$$y= \beta_0+ \beta_1 (A)+\beta_2 (d.WP)+\beta_3 (d.HR)+\beta_4 (d.Rec)+\beta_5(d.Mar)+\beta_6 (d.Ind)+\beta_7 (d.HF)+\beta_8 (d.Gy)+\beta_9 (d.Edu)+\beta_{10} (d.Ban)+ \beta_{11} (d.Com)+\beta_{12} (d.SCom)+\beta_{13} (d.SW)+\beta_{14} (d.AF)+\epsilon \qquad 3.1$$

In Equation 3.1 the variables are defined as follows:

- y is the estimated value of a house;
- β_0 is the intercept;
- A is the floor area of the house;
- d.WP is the proximity distance to the nearest worship place;
- d.HR is the proximity distance to the nearest hotel or restaurant;
- d.Rec is the proximity distance to the nearest recreational site such as a park, playground or other recreational site;
- d.Mar is the proximity distance to the nearest market, a shopping centre, or a supper store;
- d.Ind is the proximity distance to the nearest industrial unit;
- d.HF is the proximity distance to the nearest health facility such as a hospital or clinic;
- d.Gy is the proximity distance to the nearest graveyard;
- d.Edu is the proximity distance to the nearest educational institute such as a school, college, university or technical training institute;
- d.Ban is the proximity distance to the nearest bank or automated teller machine (ATM);
- d.Com is the proximity distance to the nearest commercial building;
- d.SCom is the proximity distance to the nearest semi-commercial building;
- d.SW is the proximity distance to the nearest solid waste dumping site/collection or transfer station;
- d.AF is the proximity distance to the nearest animal farm; and
- ϵ represents the error term.

Variable Selection

Based on an extensive literature review, we initially selected fourteen covariates to be included in the analysis. We included the prime structural attribute of housing properties, the floor area, which explains the value of the properties the most (Gluszak & Zygmunt, 2018), and the rest of the regressors were spatial. The variables were eliminated one by one until no multicollinearity was detected in the data. The spatial variables, such as commercial places, semi-commercial buildings, banks and ATMs, restaurants, and animal farms were found to be multicollinear and, therefore, were excluded from the model. This exclusion resulted in nine explanatory variables being included in the final model.

Table 3. Description of the Variables for the Spatial Hedonic Valuation Model

Category	Features	Description	Faisalabad	Lahore
Parcel counts	Parcels	Number of total parcels in the study area	416,168	808,710
House counts	House Parcels	Number of residential properties in the study area	268,911	780,178
House Attributes	Area (m^2)	Total area of the residential properties in square metres	30.22	116.04
	Area (Marla)	Total area of the residential properties in Marlas	1.20	5.55
Valuation	Total worth	Total worth of residential properties	PKR 2.97 trillion ($ 17.52 billion)	PKR 11.36 trillion ($ 66.83 billion)
	Average Price	Average price per square meter (per Marla)	PKR 98,279 (PKR 2.47 million)	PKR 97,897 (PKR 2.04 million)
Amenities	Solid Waste	Number of solid waste facilities and transfer stations	70	1,091
	Graveyard	Number of graveyards	72	166
Cultural	Worship Places	Number of worship places (i.e., mosques, churches, and temples)	1,409	2,281
Education and Health Facilities	Institutes	Number of educational institutions (schools, colleges, and universities)	1,705	4,329
	Health Facility	Number of health facilities (hospitals, clinics, and dispensaries)	318	2,902
Recreation	Parks and Recreation	Number of public parks and recreational sites	368	1,381
Industrial and Commercial	Industries	Number of industrial units	9,319	8,901
	Market Places	Number of market places	141	5,439
	Commercial	Number of commercial buildings	66,769	96,685

	Semi-Commercial	Number of semi-commercial buildings	2,679	68,967
	Bank and ATMs	Number of banks and automated teller machines	229	2,153
	Restaurants	Number of restaurants and cafes	612	2,793
	Animal Farms	Number of animal farms (poultry and dairy farms)	1,224	2,015

Table 4. Explanatory variables

Variables	**Faisalabad**	**Lahore**
Mean Area (Marla)	4.46	7.11
Mean Area (m^2)	111.87	148.62
Mean Distance to Worship places (Metres)	127.15	352.23
Mean Distance to Solid Waste Facilities (Metres)	798.70	696.50
Mean Distance to Parks (Metres)	520.96	397.55
Mean Distance to Markets (Metres)	1,769.26	495.77
Mean Distance to Institutions (Metres)	129.66	287.90
Mean Distance to Industrial Units (Metres)	165.34	408.30
Mean Distance to Health Facilities (Metres)	388.75	440.37
Mean Distance to Graveyards (Metres)	860.15	978.85
Mean Distance to Commercial buildings (Metres)	31.18	1,464.49
Mean Distance to Semi Commercial buildings (Metres)	1343.56	1654.77
Mean Distance to Banks & ATMs (Metres)	1,004.45	811.83
Mean Distance to Hotel / Restaurant / Café (Metres)	434.17	443.11
Mean Distance to Animal farms (Metres)	283.94	2332.94

Interpolation of Property Values

Before computing the property valuation model, the inverse distance weighting (IDW) interpolation was performed to obtain the predicted surface of property prices. The IDW interpolation with barriers and without barriers was applied to the data to avoid under- or over- prediction. The results of the interpolation with barriers and without barriers for property prices are shown in Figure 4 and Figure 6.

Figure 3. Interpolation of property prices (A) IDW interpolation (B) IDW interpolation with barriers (Faisalabad)

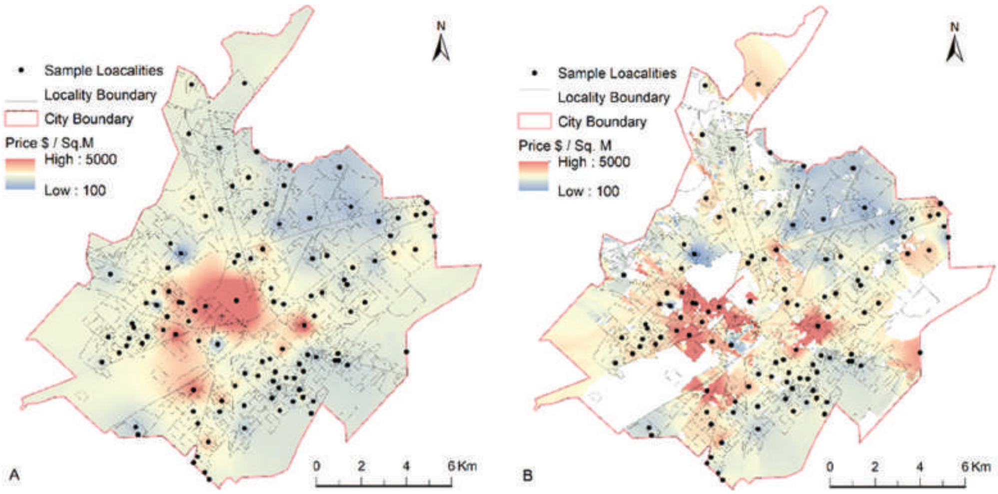


Figure 4. Transferring Interpolated Values of House Prices to the Locality Boundary and the Individual House Parcel

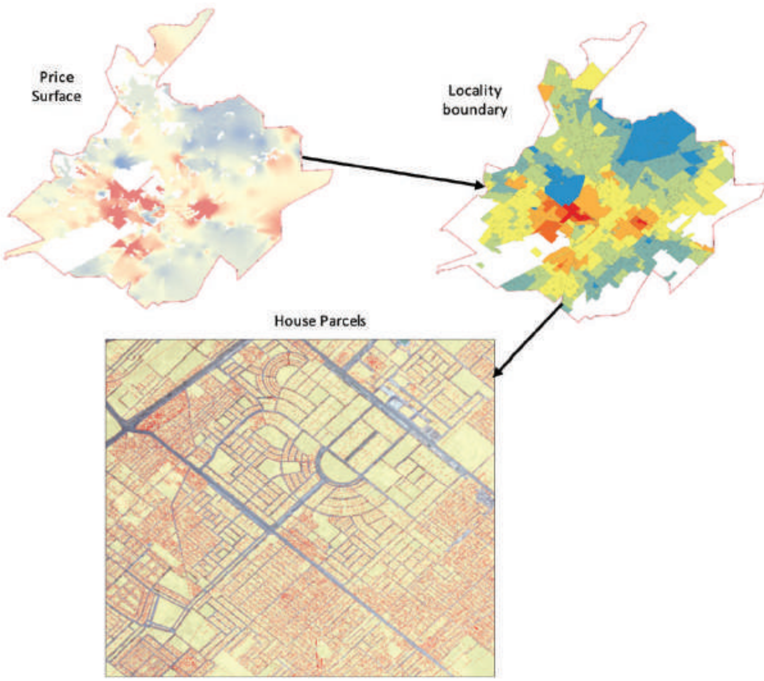
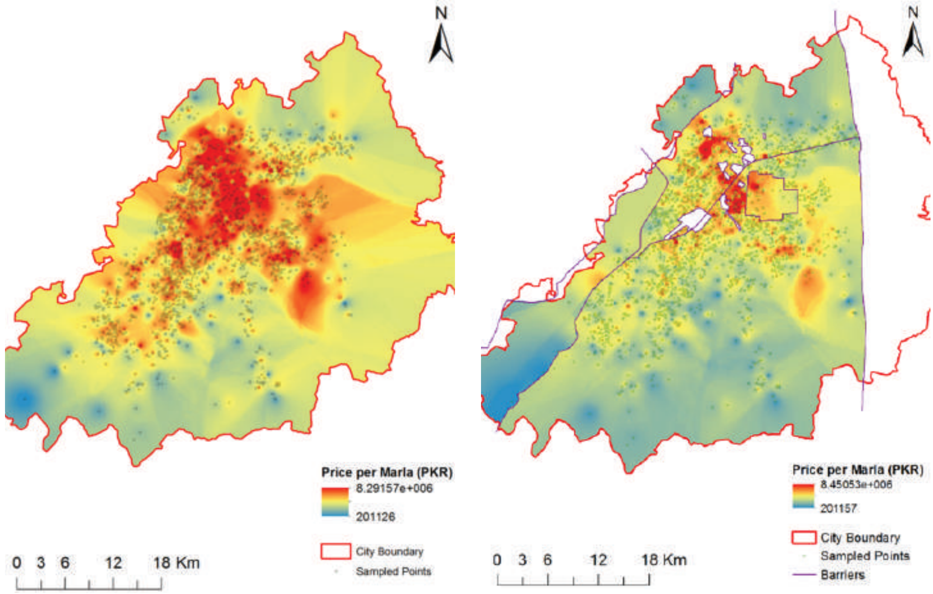


Figure 5. Interpolation of property prices (A) IDW interpolation (B) IDW interpolation with barriers (Faisalabad)



Analyses

Running the OLS- a Linear Global Model

We first performed the hedonic valuation analysis using the OLS model. The OLS is considered the best technique among all the regression methods and is used as a proper initial procedure before conducting any other regression-based spatial analysis. The OLS in the spatial statistics toolset of ArcGIS desktop can be used to discover, inspect, and model the linear spatial relationship between a dependent and one or more explanatory variables with a global approach. This means that OLS computes the relationship between the variables using a single equation for the whole area under study and assumes that the relationship remains consistent and stationary at all locations. The OLS tool in ArcGIS Desktop also computes statistics other than calculating the coefficients of regression by default. These statistics include the t-test to compare the means of variables, R2 and adjusted R2 to assess the model performance, variance inflation factor (VIF) of each explanatory variable to detect multicollinearity, Akaike's information criterion (AIC) to measure the model's effectiveness, joint F-test and joint Wald test to show the model's significance and measure the effectiveness of regressors, the Koenker (BP) test for the level of stationarity of the relationship, and the Jarque-Bera test to assess the bias or validity of the specified model.

The value of R2 and adjusted R2 could be any number between 0 and 1. R2 value usually increases if an additional explanatory variable is added to the model, while the adjusted R2 considers the overall complexity of the data and its value does not increase when an additional regressor is added. Therefore, adjusted R2 is always lower than R2. A VIF value of any regressor in the model must be less than 7.5 to keep the redundancy within an acceptable limit. A model specification with the comparatively smallest AIC value is considered the most effective among all the considered versions. The null hypothesis for both the joint F and joint Wald tests is that none of the regressors in the model are effective and test results are interpreted based on the significance of the associated p-values.

However, the joint F test is considered reliable when the p-value of the Koenker test is not statistically significant, otherwise, the joint Wald test is preferred. The Koenker test determines the consistency of the spatial relationship and if the associated p-value is significant, there is sufficient evidence to reject the null hypothesis that the relationship is consistent and static and the alternative hypothesis accepted that there is spatial heterogeneity or heteroscedasticity in the data. The associated p-value of the Jarque-Bera test must not be significant statistically to confirm that the data is normally distributed. However, if the result is the opposite, it can be concluded that the assumption of data normality does not hold (Mitchell, 2005; Scott & Janikas, 2010).

Checking the level of Spatial Autocorrelation

The Global Moran's I test is applied to check the degree of spatial autocorrelation of regression residual values in a global environment. Using residual values, a hotspot analysis was carried out to discover the significant high- and low-value clustering. Due to the spatial nonstationarity found in the data distributed across the geographic space, global models such as the OLS cannot explain the relationship between some sets of variables correctly and produce biased estimates (Fotheringham, et al., 1996). The nature of the model must alter over space to reflect the structure within the data. One of the main objectives of spatial analysis is to identify the nature of relationships that exist between variables. Typically, this is undertaken by calculating statistics or estimating parameters with observations taken from different spatial units across a study area.

The Geographically Weighted Regression (GWR) was run, which allows the actual parameters of each location in space to be estimated and mapped as opposed to having a trend surface fitted to them. Variations in relationships over space, such as those described above, are referred to as spatial nonstationarity (Hu et al., 2016). Currently, the global regression models are the leading methods to study the relationships among the geographical or environmental factors and the spatial distribution of a component. These models estimate the average regression variable for each explanatory variable in the entire study area, regardless of location and orientation. Because of complex urban systems and the strong interaction between different elements, the assumption of spatial stationary relationships may be violated, especially at large geographic scales. Therefore, the local regression model can effectively complement the global model for inferring the relationship at a more precise level (A. Fotheringham et al., 2002).

Cluster and Outlier Analysis

Cluster and outlier analysis was used to identify the clustering pattern of property prices across the study area. This analysis identifies the spatial clustering of features with high/low values and spatial outliers using Anselin local Moran's I statistic, if any (Anselin, 1995; Anselin & Le Gallo, 2006). It also computes the z-score, p-values, and a code representing the type of the cluster for the features (house parcels in our case) that are statistically significant (p = .05). While the z-score tells about the strength of spatial autocorrelation, the p-values express its statistical significance. The positive value of 1 means that a feature has a nearby feature with similarly high values and this feature is part of a cluster, while the negative value shows that the feature has neighbours with dissimilar values and the feature is an outlier. The p-values for the feature must be small enough for the cluster or outlier to be considered statistically significant. The cluster and outlier analysis is given as:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^{n} w_{i,j} \left(x_j - \bar{X}\right) \qquad 5$$

where x_i is an attribute for feature i,$\bar{X}$ is the mean of the corresponding attribute, S_i^2 is the variance of feature i, and $w_{i,j}$ is the spatial weight between features i and j.

Running the GWR- Local Model

The GWR is a commonly used tool for studying the spatial variability of phenomena in a geographic area. It estimates site-specific parameters, making the calibration process more extensive. The GWR explores spatial non-stationarity, a condition where a linear regression model cannot explain the relationship effectively between variables over a geographical area (A. Getis & Ord, 2010b). To check the spatial dependence, the local geographically weighted regression analyses the residential property value for each locality. This allows us to incorporate the regional variation in the regression model and explore the relationship in much more detail (A. Getis & Ord, 2010b; Mitchel, 2005). The GWR works by allowing model coefficients to vary regionally. Essentially, the regression model is run for each location rather than the whole study area.

A major limitation of the current open-source GWR software is that the maximum number of data points such software can process is around 15,000 values on a standard desktop computer. In this age of big data, it places severe restrictions on the use of the GWR (Li et al., 2019) for larger areas (i.e., entire cities) such as in the present study. To overcome this restriction, an open-source Python-based application, namely FastGWR, developed by Li et al., 2019 was used. This application utilises the Message Passing Interface (MPI), a standard communication protocol, for parallel computers whose goal is high performance, scalability, and portability. Through parallelisation, the FastGWR optimises the memory usage to boost the computing performance significantly which can process millions of observations. It also outperforms the existing software packages available for GWR computations.

To overcome the memory restrictions, the FastGWR algorithm improves the linear algebra within the GWR calibration. The memory requirements are reduced from $O(n^2)$ to $O(nk)$, where n is the number of observations and k is the number of covariates. Given that k is far smaller than n in most GWR applications, this approach saves a significant amount of memory. The FastGWR offers parallel model diagnostic computation methods that can be applied to very large datasets consisting of millions of observations and it considerably decreases the time required for GWR calibration by a factor of up to a thousand times than the existing techniques (Li et al., 2019). While there are various MPI implementations available, for this study, the OpenMPI was used, which is an open-source implementation. Besides the OpenMPI, a Python wrapper "mpi4py" is used to integrate the MPI into the FastGWR algorithm as described by Li et al., (2019). The call to the FastGWR programme is as follows:

$$mpiexec - np\ 32\ python\ fastgwr - mpi.py - data\ input.csv - out\ gwr.csv - a \\ - bw\ 1000 \tag{4}$$

where mpiexec is the command to execute an MPI-based programme; the argument – np 32 indicates the number of processors to allocate; – data input.csv is the name of the input data table containing coordinates and associated dependent and independent variables; -out gwr.csv is the file containing the GWR outputs that include an ID for each calibration location, predicted values, residuals, local parameter estimates, and local standard errors; -a (-f) indicates the use of an adaptive or fixed bandwidth using a bisquare (Gaussian); and – bw 1000 indicates a user-defined bandwidth for calibrating the GWR model. In this model, the adaptive kernel is used with 1,000 metres of bandwidth.

4. FINDINGS AND DISCUSSION

Global Model Implementations

Linear model implementation produced adjusted R2 values of 0.85 for Lahore and 0.75 for Faisalabad. The values were slightly lower than R2 values and there were no big differences, which indicates that the models were properly specified. The joint-F test result is not interpretable because the associated p-value of the Koenker BP test was statistically significant. Therefore, the joint-Wald test statistic was used to determine the overall model significance. Since the p-value of the joint Wald statistic was statistically significant and smaller than 0.05 at a 95 per cent confidence level, there was enough evidence to reject the null hypothesis and accept the alternative hypothesis that the regressors in the model were effective. The P-value associated with the Koenker BP test was also much lower than 0.05 at a 95 per cent confidence level, which points toward the conclusion that the relationship between the response and explanatory variables was not consistent and stationary implying that spatial heterogeneity exists. That result was expected as the relevant literature confirms that the occurrence of spatial non-stationarity in the housing data is a common real-life phenomenon because the degree of the effect contributed to the house prices by different externalities is always unique at different locations.

Spatial heterogeneity can be reported using a global model like OLS but could further be measured using some local models, such as GWR (Bitter et al., 2007; Brunsdon et al., 1996; Fik et al., 2003; Wen, Jin, et al., 2017). A statistically significant p-value of the Jarque-Bera test is the indication of the presence of non-normality in the distribution of the residuals and that the model is not unbiased and could not be trusted fully. However, the relevant literature suggests that a model with a significant p-value of the Jarque-Bera test can be trusted when working with a large dataset because it is a proven fact that the assumption of normality for response as well as explanatory variables is not required in an OLS linear model when the sample size is too large. Since the distribution of regression residuals depends on the distribution of regression variables, the normality assumption can be ignored for residual distribution as well in case of large sample size (Lumley et al., 2002). When a regression model has non-normally distributed residuals, one needs to check robust standard errors and robust p-values to check if the variable coefficients are significant statistically instead of standard errors and probabilities (Mitchell, 2005; Scott & Janikas, 2010). Since Koenker statistics were statistically significant in the model diagnostics, only those robust p-values (probabilities) were checked that were smaller than 0.05 for all explanatory variables. There was sufficient evidence to reject the null hypothesis and accept the alternative hypothesis that all the coefficients were significant. The coefficient values reflect the nature and strength of the relationship of each regressor with the response variable.

The results of the linear model for the entire cities of Faisalabad and Lahore are presented in Table 3. The adjusted R2 value explained the relationship of explanatory variables to the house prices up to 75 per cent for Faisalabad and around 85 per cent for Lahore in the linear model. However, the robustness of the model was improbable since residential property prices were less normally distributed. Most of the coefficients for the explanatory variables were as expected and explained the relation between the response variable and explanatory variables.

Total floor area (in Marlas for Lahore and m2 for Faisalabad) showed a positive relationship with the house price, which indicates that every additional Marla in the floor area increased the house price by PKR 2,202,216.09 in Lahore, while every additional m2 in floor area increased the house price by PKR 86,256.45 in Faisalabad. Spatial variables, i.e., distance to the nearest health facility, marketplace, and recreational facility, such as a park, lowered the house price by PKR 2,614.52, 1,149.4, 1,105.2, respectively, with a one-meter increase in the Euclidean distance. It implies that people value these facilities and like to live near them. On the other hand, the results show that as the distance between the house and the nearest industrial unit, educational institution, graveyard, solid waste dumping site/transfer station, and worship place increased, the house price also increased by PKR 1,921.57, PKR 647.83, PKR 563.58, PKR 429.79 and PKR 234.92 per metre, respectively. It indicates that the residents do not prefer to reside near these features in Lahore. In Faisalabad, the dynamics, however, are somewhat different from Lahore.

In Faisalabad, as the Euclidean distance to the nearest health facility, park, marketplace, industrial unit and graveyard decreased by one metre, the house price increased by PKR 1,495.64, PKR 1460.89, PKR 517.08, PKR 130.66, and PKR 45.87, respectively. On the other hand, as the distance to the nearest educational institution, solid waste dumping site/transfer station and worship place decreased by one metre, the house price also decreased by PKR 504.57, PKR 393.37, PKR 355.84, respectively. Unlike Lahore, the proximity to industrial units and graveyards contributed positively to house prices in Faisalabad. A possible reason for this contradiction may be the presence of small industrial units, such as power looms, and graveyards within residential areas, especially in the central parts of Faisalabad. This pattern is not present on such a large scale in Lahore.

Table 5. Results of the Linear Model

Variable	Lahore				Faisalabad			
	Adjusted R^2 =0.85				Adjusted R^2 =0.75			
	Coeff.	**t-stat**	**p-Value**	**VIF**	**Coeff.**	**t-stat**	**p-Value**	**VIF**
Intercept	356,219.99	20.36	0.000*	-----	3,280,398.6	199.14	0.000*	----
Floor Area	2,202,216.09	2,050.22	0.000*	1.04	86,256.45	881.22	0.000*	1.04
D_Worship Places	234.92	6.95	0.000*	3.14	355.84	5.99	0.000*	1.11
D_SolidWaste Site	429.79	24.80	0.000*	2.82	393.37	34.48	0.000*	1.62
D_Parks	-1105.20	-48.86	0.000*	1.85	-1,460.89	-119.19	0.000*	1.18
D_Market	-1149.40	-46.33	0.000*	4.68	-517.08	-94.11	0.000*	1.55
D_Institutes	647.83	15.85	0.000*	3.39	504.57	8.30	0.000*	1.20
D_Industries	1,921.57	73.14	0.000*	1.54	-130.66	-3.04	0.002*	1.23
D_Hospitals	-2,614.52	-109.91	0.000*	3.10	-1,495.64	-73.75	0.000*	1.30
D_Graveyards	563.58	39.45	0.000*	1.26	-45.87	-4.51	0.000*	1.51

The results are per our expectations and in line with the findings of relevant previous studies. Some of the supporting references are discussed here. The floor area of a real estate property is the most important factor that determines its price (Ma et al., 2020), while the proximity of shopping facilities adds a premium to house prices (De & Vupru, 2017; Xiao et al., 2017; Yang et al., 2018b; Yang, Chau, et al., 2020; Zhang et al., 2018). The presence of worship places in the vicinity of every religion contributes to house prices positively. However, some housing properties that are near worship places may experience price devaluation because of noise and a higher number of visitors that create disturbances for the residents (Brandt et al., 2014; Thompson et al., 2012). On the other hand, urban green spaces, public parks, playgrounds, and other recreational sites add a premium to the housing properties (Crompton & Nicholls, 2020; Liao & Wang, 2012; Shabana et al., 2015). Residential properties close to graveyards fetch lower values due to superstitions linked to the burial grounds (Hassan et al., 2021). Similarly, an industrial neighbourhood is a negative influence on residential property prices in most of the research findings (Grislain-Letrémy & Katossky, 2014; Munshi, 2020).

Local Model Implementation

Table 6. Results of the FastGWR Model for Faisalabad

Predictor	Faisalabad	
	$R^2 = 0.78$	
	Coeff.	**SE**
Area	0.00480	0.00015
D_Worship Places	0.00052	0.00041
D_Solid Waste Facilities	-0.00038	0.00089
D_Parks	0.00094	0.00064
D_Markets	0.00443	0.00065
D_Institutes	0.00042	0.00043
D_Industry	0.00057	0.00049
D_Hospitals	0.00039	0.00062
D_Graveyards	0.00283	0.00076

The FastGWR model estimation produced an adjusted R squared of 78 per cent for Faisalabad, which shows a strong relationship between the house value and predictors. Table 4 presents the results of the FastGWR model estimation for the entire city. The coefficients of the FastGWR are positively correlated except for the distance to a solid waste facility, which deceptively indicates that the house values decreased as the distance increased from the solid waste facility. Since these coefficients are the average values that are affected by the high negative values in the results, we also examined the results locally. Figure 5c depicts the significant parameter estimates for the distance to a solid waste facility. The map shows that the house parcels coloured in blue had a converse effect of distance to solid waste facility, meaning that the values of these houses decreased as the distance from solid waste facility increased. One possible reason for this inverse coefficient could be that earlier the solid waste facilities were established away from the settlements but with time, the settlements have grown around these facilities and the land prices near these facilities have also increased. The low variance inflation factor (less than 7.5 for each regressor) indicates that there is no multicollinearity as we had already eliminated the muiltcollinear explanatory variables, while the regression residuals are random and not spatially autocorrelated.

Figure 10 shows the distribution of regression residuals with very low R2 values. In the OLS results table, we need to understand the t-statistics value that evaluates the statistical significance of the explanatory variables. The higher the t-statistic, the more significant the variable is. This value explains that the area of the house is the most important structural variable for house price in the entire city, while the other significant variables are the distance to a solid waste facility, distance to worship places, and distance to educational institutes, respectively. These accessibility variables have positive coefficients indicating that the house price increased as the distance from these features increased. The other significant variables with negative coefficients are the distance to parks, distance to markets, distance to hospitals, distance to graveyards, and distance to industries respectively. The negative coefficients suggest that residential property prices decreased as the distance from locational features increased. These findings are similar to that of Li et al., (2019), who studied the city of Los Angeles, California.

The results of the OLS model for the FD-I rating area are presented in Table 5. The semi-log model explains the relationship up to 77 per cent, while the linear model explained it up to 80 per cent. The coefficients are as expected but the distances to worship places, parks, markets, educational institutes, and hospitals were negative. This indicates that the house price decreased as the distance from these variables increased. The worship places are key cultural features in the city that appear to impact the prices of residential houses positively (Brandt et al., 2014; De & Vupru, 2017). In the FD-I zone, the average price per square metre was US$875 and the average house price was US$81,320 with an average area of 95 square meters. The t-statistics are suggestive of the order of significance for these negative coefficients, which indicates that the distance from places of worship, distance from parks, distance from health facilities, and distance from the market were the most significant locational features influencing prices, respectively. The coefficient of distance to a solid waste facility is negative in the linear model, while in the semi-log model, this coefficient is positive. However, they are not statistically significant.

Model Implementation for Different Rating Areas in Faisalabad

Table 7. Results of the Linear Model for Different Rating Areas in Faisalabad

Rating Area FD-I	Semi-Log OLS Model R^2 =0.77			Linear Model R^2 =0.80			
	Coeff.	t-stat	p-Value	Coeff.	t-stat	p-value	VIF
Intercept	10.35575	1307.87	0.0000*	14,688.14	25.88	0.0000*	----
House Area	0.00985	301.19	0.0000*	793.34	338.31	0.0000*	1.06
D_Worship Places	-0.00063	-30.30	0.0000*	-47.65	-31.82	0.0000*	1.12
D_SolidWaste Facility	0.000002	0.27	0.7859	-0.24	-0.41	0.6831	1.22
D_Parks	-0.00038	-30.17	0.0000*	-30.39	-33.18	0.0000*	1.13
D_Market	-0.00005	-10.15	0.0000*	-6.38	-17.09	0.0000*	1.18
D_Institutes	-0.00012	-5.19	0.0000*	3.95	2.35	0.0186*	1.12
D_Industries	0.00028	18.32	0.0000*	25.08	22.32	0.0000*	1.14
D_Hospitals	-0.00016	-13.58	0.0000*	-16.28	-18.45	0.0000*	1.10
D_Graveyards	0.00016	30.30	0.0000*	12.22	31.22	0.0000*	1.30
Rating Area FD-II	Semi-Log Model R^2 =0.78			Linear Model R^2 =0.81			
	Coeff.	t-stat	p-value	Coeff.	t-stat	p-value	VIF
Intercept	10.16367	852.36	0.0000*	-10,913.75	-10.54	0.0000*	----
House Area	0.01060	187.93	0.0000*	1,015.18	207.08	0.0000*	1.07

D_Worship Places	**-0.00007**	**-1.82**	**0.0680**	**-2.52**	**-0.72**	**0.4705**	**1.11**
D_SolidWaste Facility	-0.00023	-27.30	0.0000*	-21.46	-29.28	0.0000*	1.53
D_Parks	0.00015	7.21	0.0000*	9.03	4.98	0.0000*	1.44
D_Market	0.00005	7.27	0.0000*	0.72	1.20	0.2305	1.54
D_Institutes	0.00016	4.77	0.0000*	19.52	6.70	0.0000*	1.31
D_Industries	-0.00032	-11.34	0.0000*	-22.13	-8.90	0.0000*	1.59
D_Hospitals	0.00026	16.44	0.0000*	16.89	12.07	0.0000*	1.32
D_Graveyards	0.00047	49.54	0.0000*	38.72	46.48	0.0000*	1.69
Rating Area FD-III	**Semi-Log Model R2 =0.71**			**Linear Model R2 =0.79**			
	Coeff.	**t-stat**	**p-value**	**Coeff.**	**t-stat**	**p-value**	**VIF**
Intercept	10.39261	5,323.04	0.0000*	17,254.89	135.38	0.0000*	----
House Area	0.00761	726.77	0.0000*	612.54	896.17	0.0000*	1.04
D_Worship Places	0.000141	22.37	0.0000*	8.88	21.66	0.0000*	1.12
D_SolidWaste Facility	0.000048	40.00	0.0000*	2.66	34.25	0.0000*	1.60
D_Parks	-0.00012	-98.16	0.0000*	-8.76	-106.34	0.0000*	1.15
D_Market	-0.00002	-32.97	0.0000*	-1.72	-42.69	0.0000*	1.53
D_Institutes	0.000011	1.79	0.0741	1.03	2.47	0.0134*	1.21
D_Industries	-0.00005	-10.94	0.0000*	-1.25	-4.15	0.00003*	1.24
D_Hospitals	-0.00008	-37.37	0.0000*	-7.11	-50.95	0.0000*	1.31
D_Graveyards	-0.000045	-41.54	0.0000*	-1.77	-24.89	0.0000*	1.53

Table 8. Results of the FastGWR Model for Rating Area FD-I, FD-II, and FD-III

Predictors	FD-I (R^2=0.61)		FD-II (R^2=0.59)		FD-III (R^2=0.80)	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Area	0.00511	0.00020	0.00524	0.00021	0.00475	0.00015
D_Worship Places	0.00072	0.00053	0.00181	0.00056	0.00049	0.00039
D_Solid Waste Facility	0.00516	0.00043	0.00025	0.00065	-0.00101	0.00093
D_Parks	0.00044	0.00062	0.00515	0.00043	0.00078	0.00064
D_Market	0.00498	0.00050	0.00205	0.00057	0.00449	0.00066
D_Institutes	0.00047	0.00054	0.00086	0.00053	0.00038	0.00041
D_Industry	0.00220	0.00061	0.00109	0.00063	0.00038	0.00048
D_Hospitals	0.00165	0.00050	0.00442	0.00048	0.00006	0.00064
D_Graveyards	0.00539	0.00053	0.01000	0.00049	0.00209	0.00079

These values indicate that in the FD-I rating area, the distance to solid waste facilities did not influence the prices of residential properties, while all other explanatory variables had an impact on the house prices positively or negatively. The results of the FastGWR are presented in Table 6. The R-squared value for FD-I is 0.61, indicating that the model explained a 60 per cent variance based on the explanatory variables. As suggested by the coefficients of all the predictive variables, there existed a positive correlation. This zone comprises the central business district (CBD) where most of the properties are commercial and semi-commercial and there are only 28,090 residential properties.

The FD-II rating area is characterised by small- and medium-scale industries, timber market, and wholesale businesses. Although the average house area in this zone was smaller, the average price per square metre was 22.6 per cent higher than the FD-I rating area and the average house price was also 20.32 per cent higher (i.e., US$102,063). The higher prices of small houses in this area are due to the ease of access to workplaces and proximity to the city centre. The results of the valuation model for this region are dissimilar to the results of the FD-I rating area. All the explanatory variables are statistically significant except the distance from worship places. Although the average distance to the places of worship was much smaller (93.4 metres), worship places did not seem to influence the residential property prices in this particular zone. This effect is possibly due to the socio-economic conditions of the area since the average house size and the income level were lower than other residential districts. The worship places give the impression of being less important for the residents possibly due to the degree of adherence to religion and the level of noise from the loudspeakers of mosques. Researchers have found the negative effects of places of worship on adjacent house prices but this effect declines with the increasing distance and diminishes after 300 metres (Brandt et al., 2014; Do et al., 1994).

Figure 6. Maps of Significant Parameter Estimates for the Predictive Variables: Area of the House (a), Distance To Worship Places (b),Distance to Solid Waste Sites (c) Distance to Parks (d)

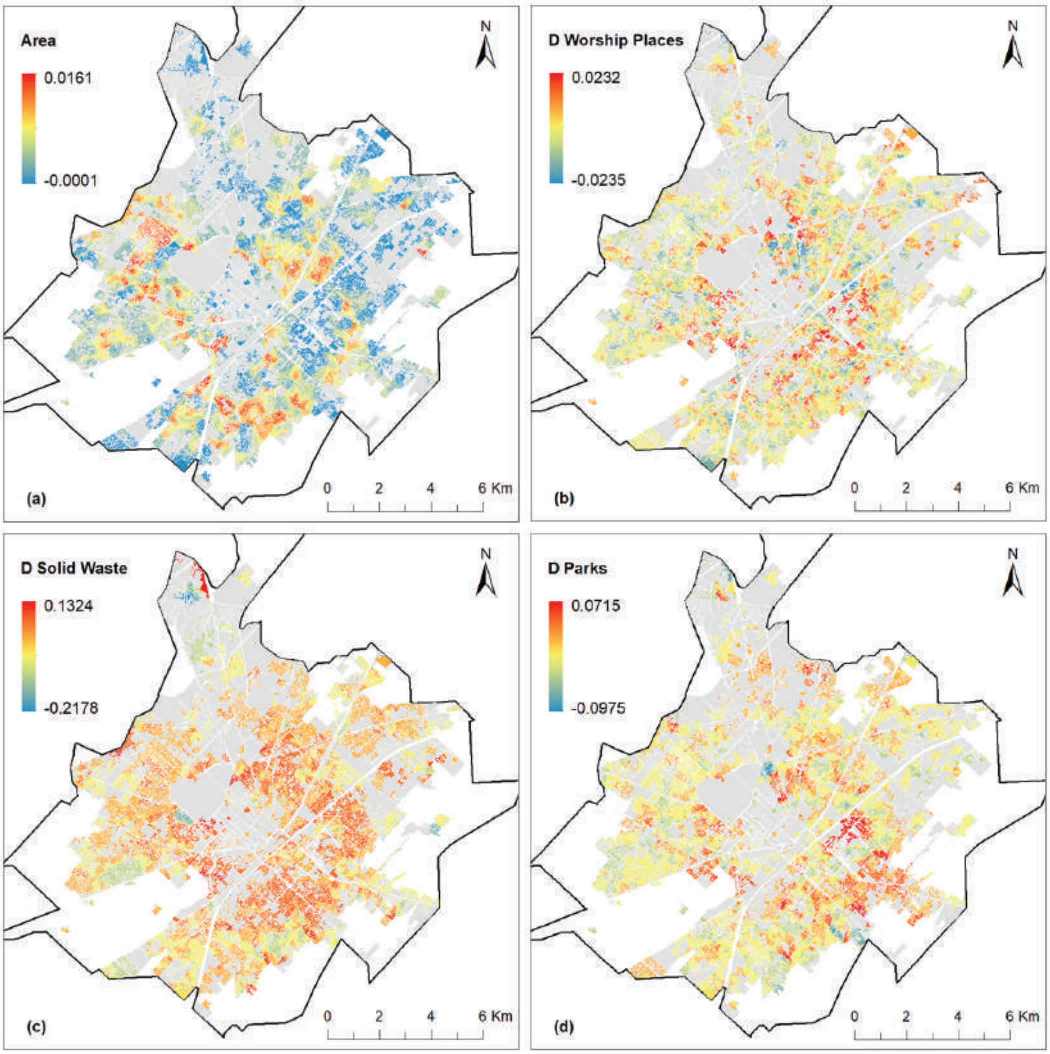
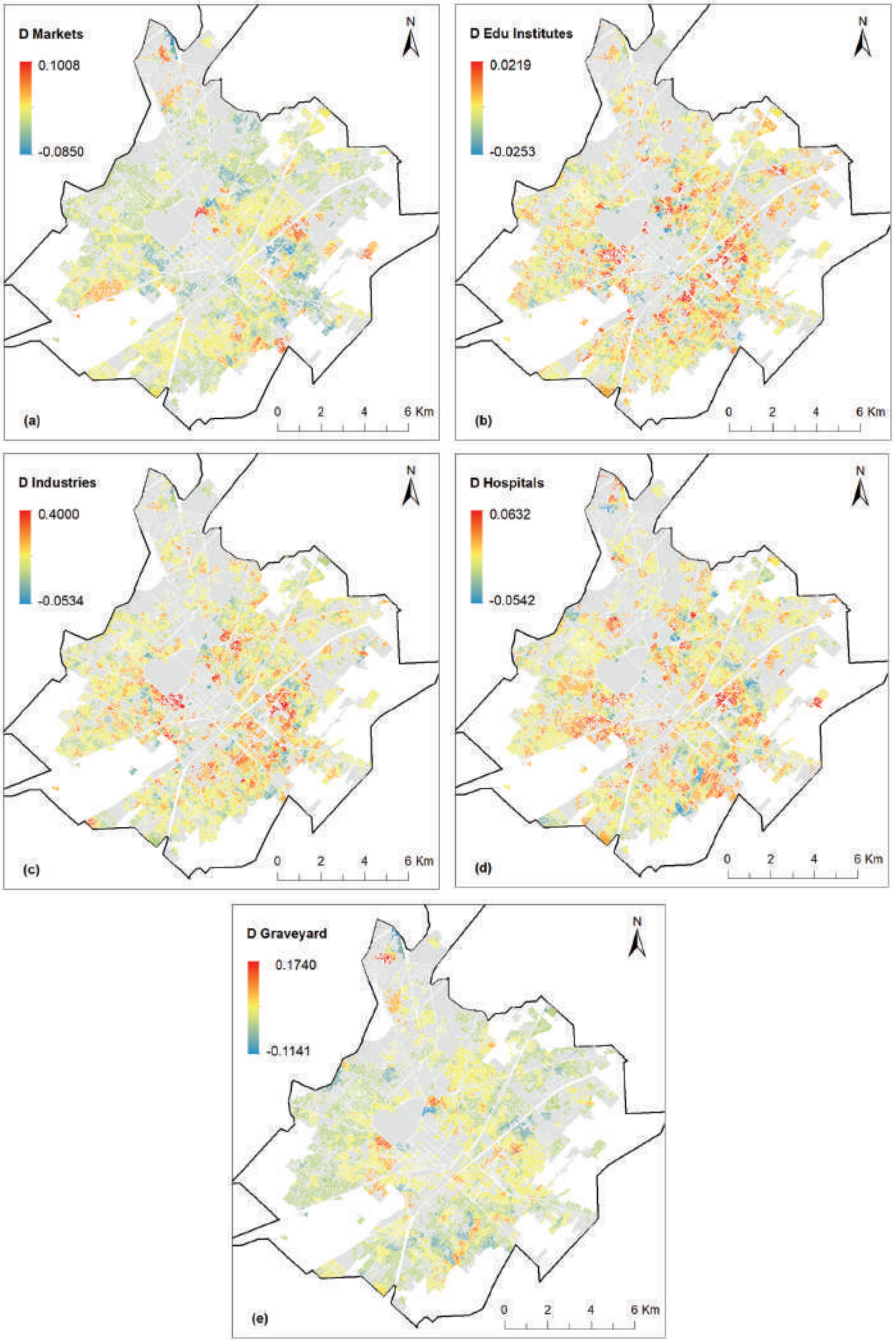


Figure 7. Maps of Significant ParameterEstimates for Predictive Variables:Distance to Markets (a), Distance to Educational Institutions (c), Distance to Industries (c), Distance to Hospitals (d), Distance to Graveyards (e)



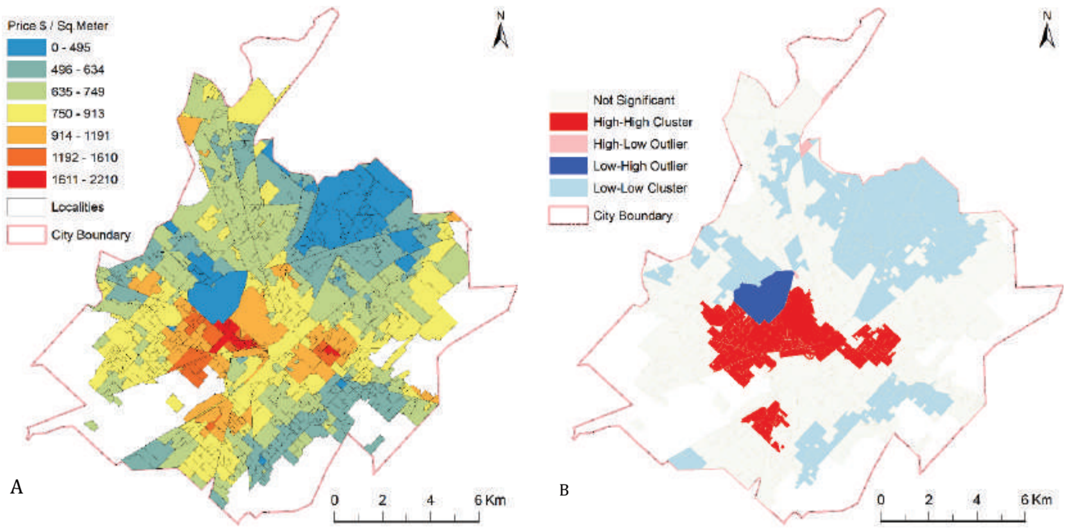
The coefficients for distance to solid waste facilities and distance to industries are negative and statistically significant (99 per cent confidence). Solid waste transfer stations provide opportunities for scavengers and scrap dealers to collect recyclable materials to earn their living. Small- and medium-scale industries, like power looms, garment factories, embroidery units, plastic products, leather factories, paper, and chemical factories exist in this area, thus, offering employment opportunities to the residents. Distance to market is statistically significant in the semi-log model but the results of the linear model were not statistically significant. There are only 4.13 per cent (n = 11,107) houses present in this zone, whereas 75.7 per cent of the houses in the FD-II are exempted from the property tax as per the policy of the revenue department and 24.27 per cent of the houses are responsible property tax collection. The results of the semi-log model explain the relationship between house price and explanatory variables by up to 78 per cent, while the linear model explained it by up to 80 per cent. In this zone, the results of the FastGWR are also positively correlated and the R-squared value is 0.59, which shows an intermediate to high performance by the model. Though the FastGWR model explains a relatively smaller variance as compared to the linear models (59 per cent vs. 80 per cent), the consideration of spatial aspects in the FastGWR makes it more reliable when applied to analyse geographical disparities. One possible reason for this weak relationship is that there are only 11,107 residential buildings situated in this zone, while most of the buildings are commercial and semi-commercial.

The third zone, the FD-III rating area, is the largest among all other zones in the city, which holds a total number of 339,420 properties of which 229,714 (85.42 per cent) are residential. In this zone, only 22 per cent of the residential properties are liable to pay the property tax and the rest 78 per cent are exempt from any kind of property tax. The results for the FD-III rating area are significant for all the explanatory variables except the distance from educational institutions, which is insignificant under the semi-log model but significant under the linear model. Several studies have demonstrated the effects of schools and educational institutions on property prices (Sah et al., 2016; Wen, Xiao, et al., 2017; Yang et al., 2018a). However, in this zone, the educational institutions did not impact the residential property prices. One explanation for this could be the fact that the schools do not have strict zonal boundaries and the students are willing to travel longer distances to study in educational institutions that offer better quality education. Distances to parks, hospitals, markets, graveyards, and industries have negative coefficients. The results of the FastGWR demonstrate that all the predictive variables have positive coefficients, while the distance to solid waste facilities had a negative effect, suggesting that the residential parcels closer to these facilities have higher values, while the residential properties away from a solid waste facility have lower values. The FastGWR model estimation for Lahore was not possible due to the unavailability of the required high computing power.

Cluster and Outlier Analysis

Figure 7 demonstrates the cluster and outlier analysis of property prices within the residential localities of Faisalabad. The localities highlighted in red colour are the high-high clusters, having high residential property values. One major high values cluster is around the city centre and the localities of Clock Tower, Jinnah Colony, Madina Town, Peoples Colony, Civil Lines, Susan Road, and Jaranwala Road. Another high-high cluster is in the locality of Samanabad. This high-high cluster is isolated from the main cluster of high residential property values. The localities coloured in light blue are the low-low clusters, while the grey-coloured localities do not have any significant clustering. One high-low outlier, coloured in pink, is also found in the area of Muslim Town, which indicates that one particular locality had higher property values and that locality is surrounded by the localities that had lower property values. The blue-coloured locality is the low-high outlier showing that this locality is surrounded by localities with relatively higher property values. This particular locality is the University of Agriculture, Faisalabad, a public university having an area of ~345 hectares.

Figure 8. Average Price (US$) Per Square Metre in Different Localities of Faisalabad: (A) Cluster and Outlier Analysis for the Property Values Within the Residential Localities (B)



The spatial autocorrelation analysis report of the hotspot analysis shows a statistically significant p-value. Thus, the null hypothesis suggesting the randomness of regression residuals stood rejected and the z-score of 767.76 points toward significant clustering. Furthermore, the hotspot analysis output map in Figure 9 tells the story of the nature and distribution of high- and low-value clusters. Most of the high-value clusters can be traced in and around the Walled City area (the CBD), posh areas of Model Town, Gulberg, DHA, and other elite-class residential colonies, while the low-value clusters are found in the peripheries.

Local Indicators of Spatial Association

The LISA demonstrates that the high-priced residential properties are clustered together and the low-priced houses are clustered together as well as one might expect. Figure 8 shows the results of cluster and outlier analysis (local Moran's I) in the study area (95 per cent significance). In the entire city, 21 per cent of the high-value residential houses were clustered together, while the cluster of low values accounted for 40 per cent of the total residential properties. This is a strong indication of the distribution of real estate values across the city and the associated income class. The costly residential houses are located in the east and southeastern parts of the city, including the localities of Madina Town and Iqbal Town. The localities in Lyallpur Town have a mixed type of values ranging from high-cost to low-cost residential houses since this town holds the older residential communities near the city centre and newly built communities in the north of the city. The higher cost of living in this area also gives rise to informal settlements.

The results regarding low-high outliers from LISA indicate that the low-cost houses were surrounded by the high-cost residential properties, while, in some areas, the high-cost houses were surrounded by the low-cost houses, which is a typical sign of topophilic adherence to the place of residents. For instance, if some of the residents become wealthier enough to buy a high-cost property, they prefer to build a luxurious house in the same poor community instead of buying a new house in the posh areas of the city, making it an outlier among other houses.

Figure 9; Distribution of High- and Low-Value Clusters

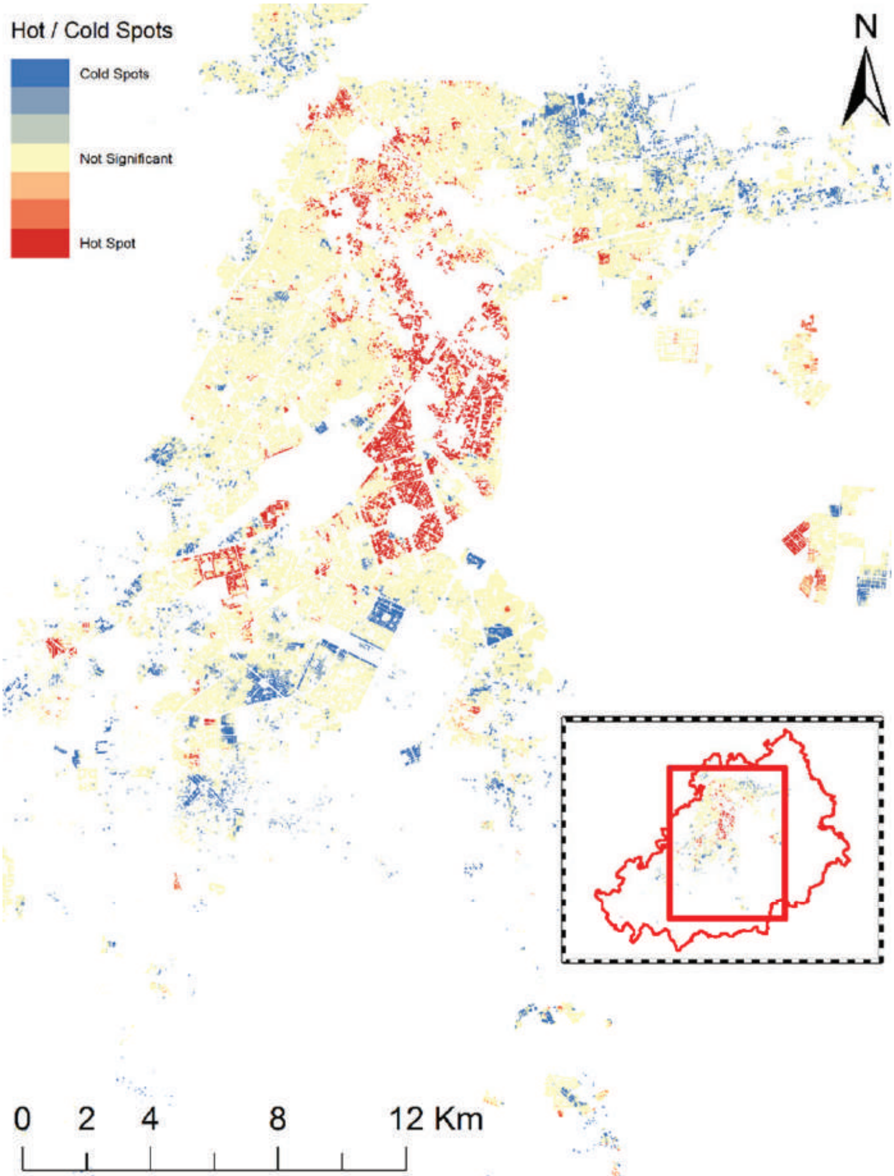


Figure 10. Cluster Map of Residential Properties in the Entire City (a) FD-I (b) FD-II (c) FD-III (d)

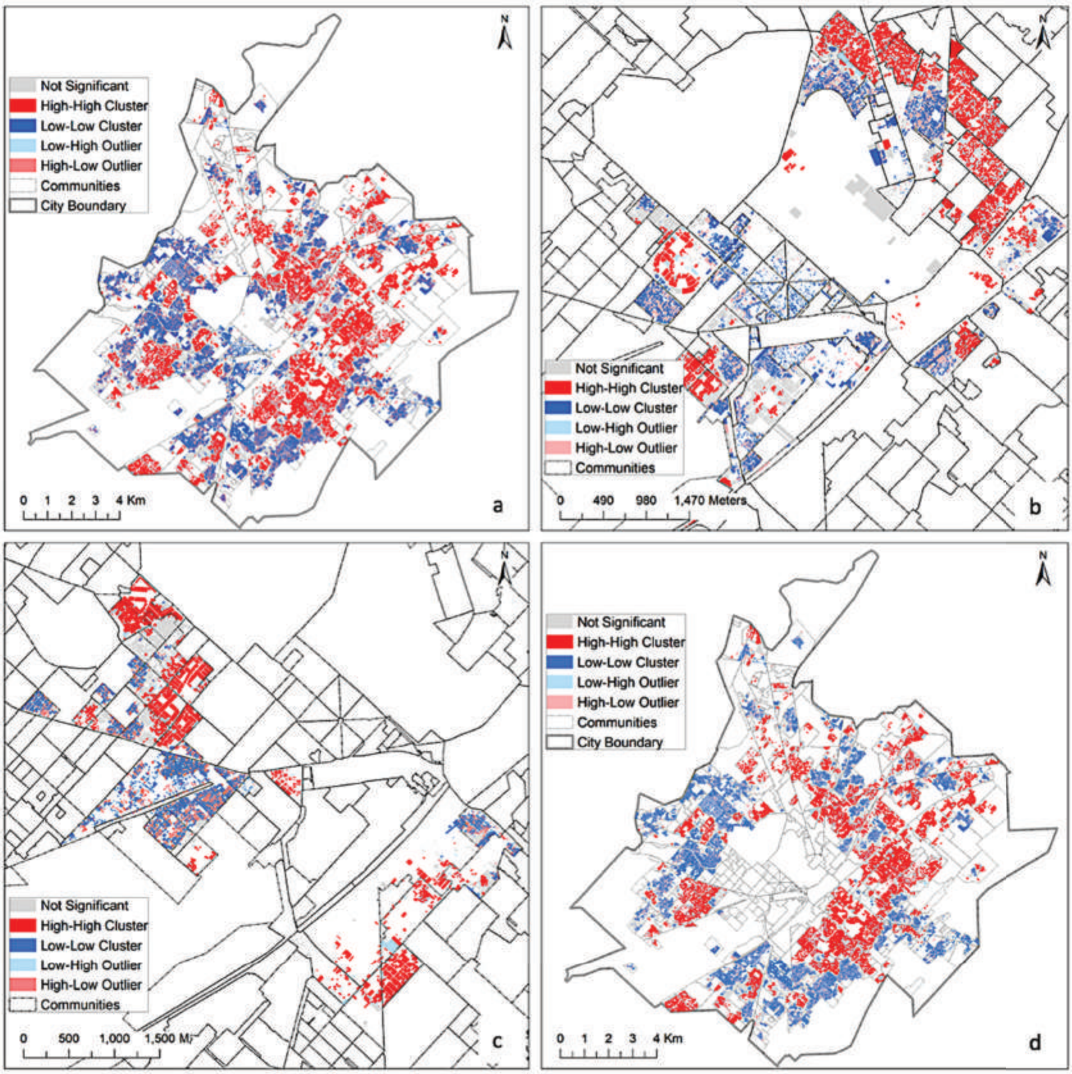


Figure 11. Percentage Distribution of Clustered Properties

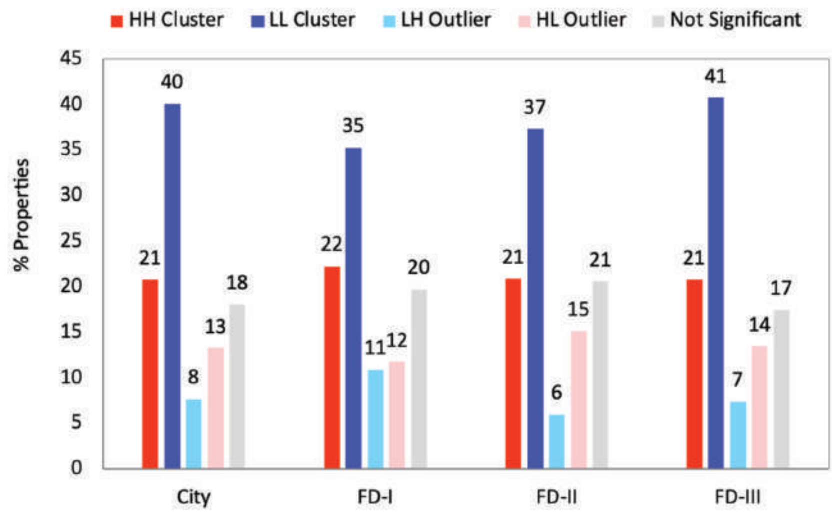


Figure 12. Estimated House Prices in Faisalabad

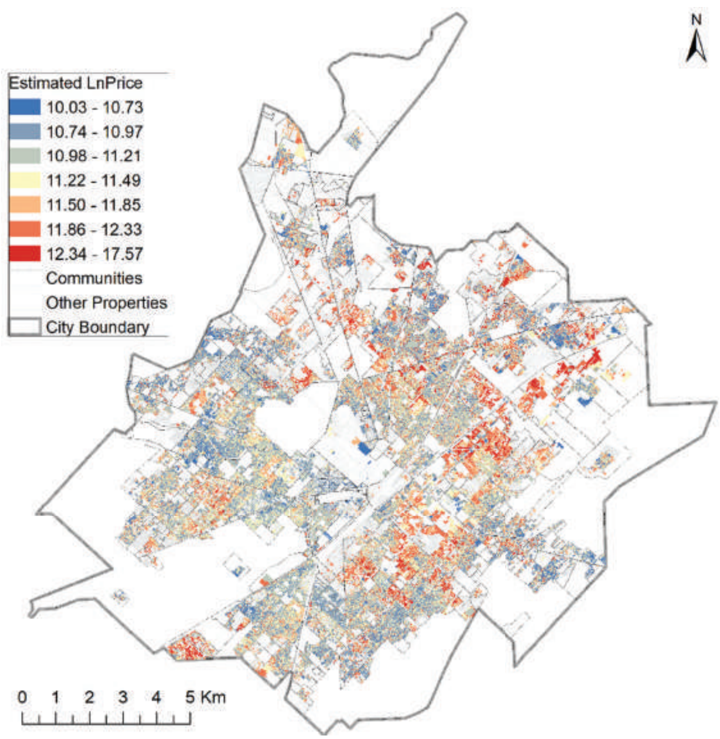


Figure 13. Residuals of the OLS Semi-log Model for the Entire City (a) FD-I (b) FD-II (c) FD-III (d)

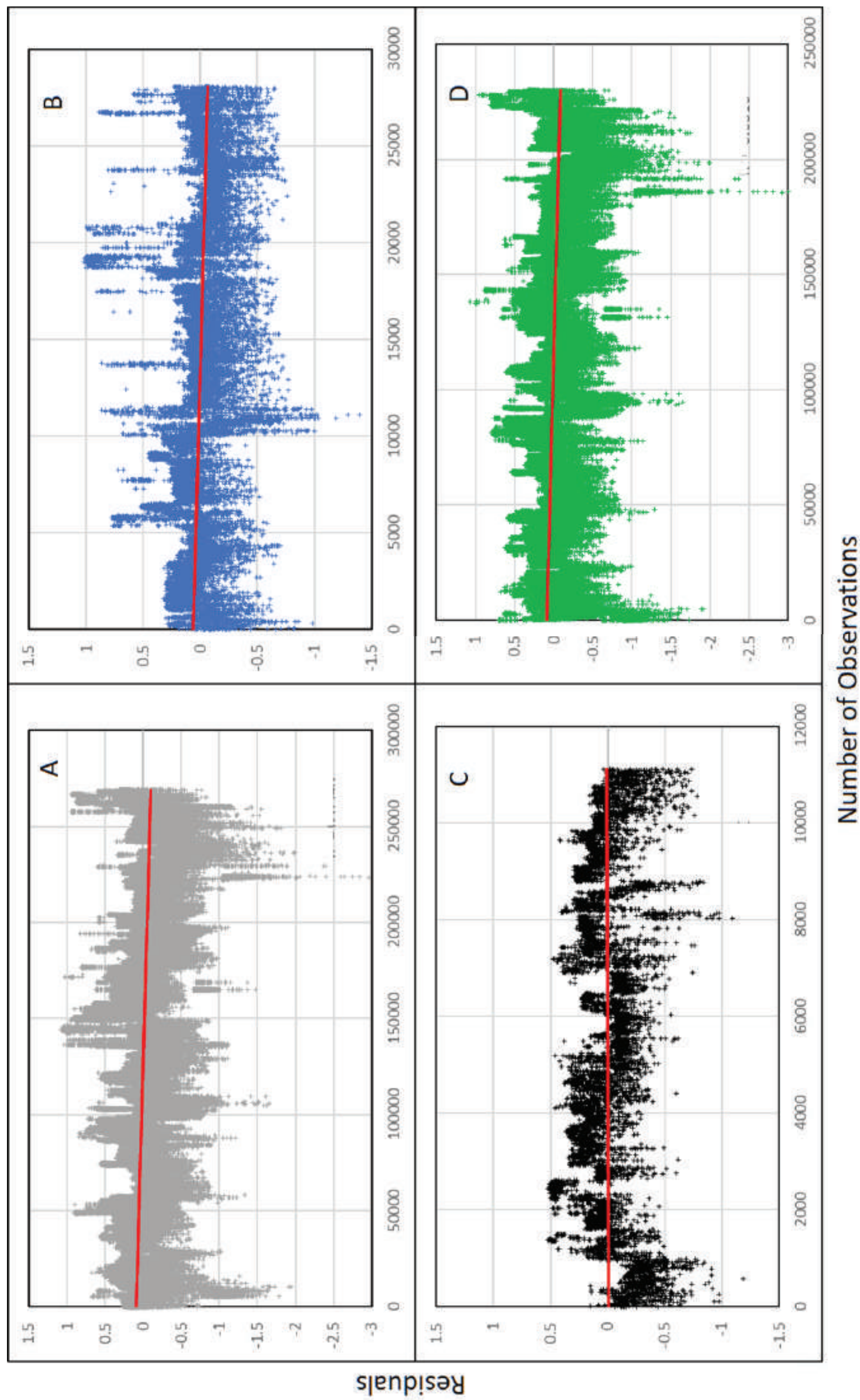
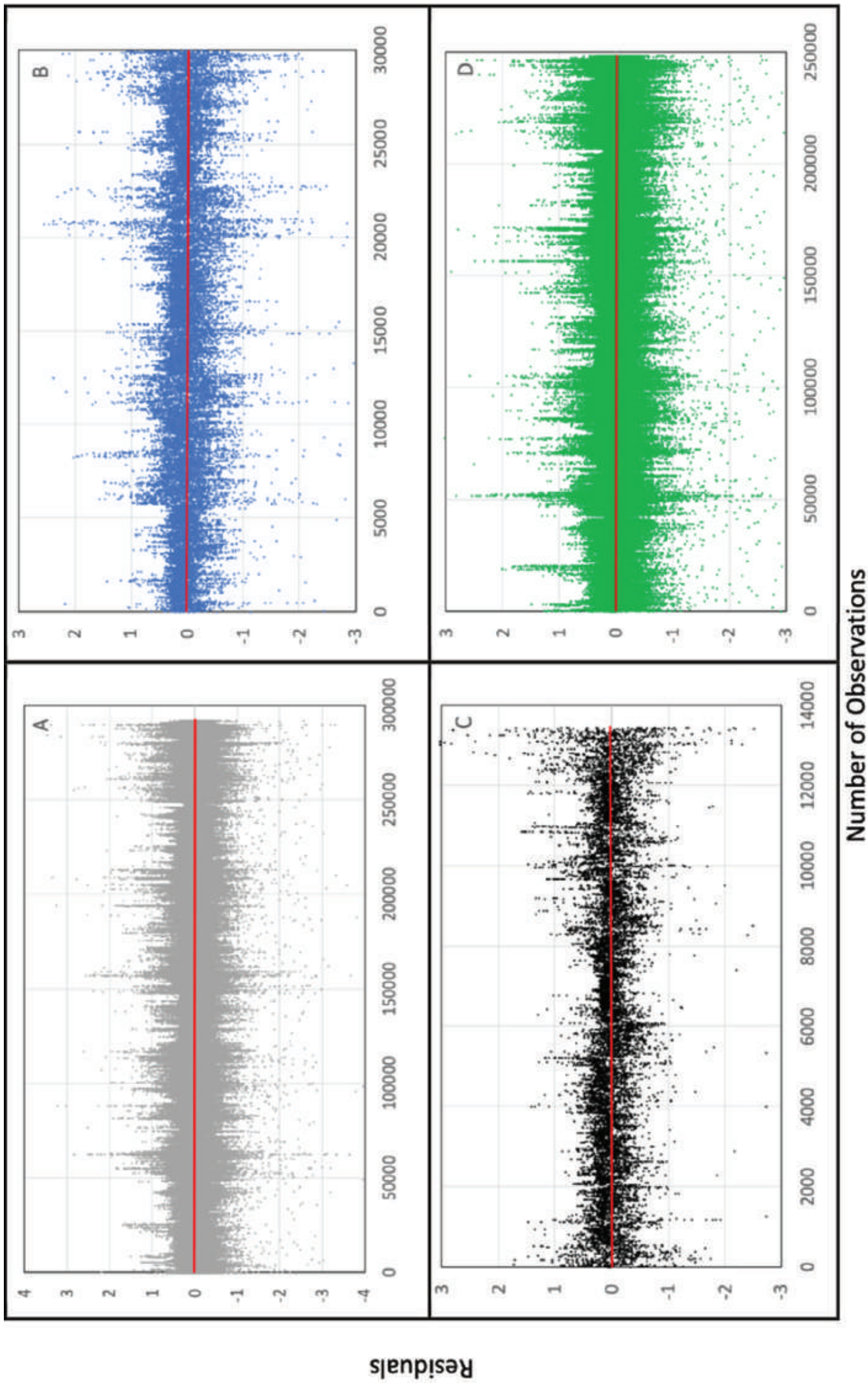


Figure 14. Residuals of the FastGWR Semi-Log Model Entire City (a) FD-I (b) FD-II (c) and FD-III (d)



5. CONCLUSION

This paper examined the spatial determinants of prices of residential properties in the cities of Lahore and Faisalabad using a spatial hedonic approach. The spatial hedonic models, the OLS, and the FastGWR regression models were used to analyse the association between several explanatory variables and the urban immoveable property prices. Nine locational features within four categories (amenities, cultural, educational and health facilities, and recreation) were selected to explore the correlations between the spatial determinants and housing prices. While the performance of the two models varied slightly, the results showed positive and negative statistically significant correlations between different locational features and residential property prices. The principal contributing factor to the price was the area of the house, which had a large positive coefficient in all the cases. Other positively correlated variables were distance to worship places, distance to a solid waste facility, and distance to educational institutions. The distance to public parks, markets, hospitals, graveyards, and distance to industries had a negative association with house prices.

The ongoing development and accelerated urban expansion have changed the cities in Punjab from a single centre to multiple centre pattern and there is a need for a reappraisal of properties to increase revenues from property taxes. The spatial determinants of housing valuation are significantly important as shown by this study and, hence, must be integrated into policy formulation and future urban planning and design.

There are several limitations of the study. For instance, it focused on locational features even though there also exist socio-economic determinants that might influence housing prices at different spatial scales. We could not include them due to two reasons. First, the housing level data on these socio-economic variables is not only a huge constraint in Pakistan but also in many other developing countries. Therefore, we did not include them in the model. Second, this study aimed to analyse particularly the locational features and highlight their influence on housing prices, which has been neglected in the previous studies in Pakistan's context. Third, the FastGWR model for Lahore was not possible due to the unavailability of the required high computing power.

6. RECOMMENDATIONS AND POLICY IMPLICATIONS

The results of this study have some general implications for policymakers, investors, real estate developers and urban planners. Our model has produced meaningful and reliable estimates, which explain residential property values. The spatial diversity of the coefficients is very important for decision-makers, requiring explicit knowledge of the local or regional housing markets. This helps to refine policies and have a better understanding of the local house price variations. The approach applied is flexible and can be applied to different geographical locations in Pakistan. Since the record of past market transactions plays a key role in the factual valuation of immovable properties, we suggest the formulation of a system to record the fair market prices of real estate properties. For this purpose, the following steps may be taken.

Firstly, the property transfer fee may be waived to attract the sellers and buyers towards disclosing the factual deal prices of properties and the deficit created by this relinquishment may be bridged through the increase in annual property tax. Secondly, an immovable property valuation desk may be established in revenue departments to serve the masses by assessing the market price of their properties at a nominal fee. This desk can record the geo-locations and structural attributes of properties, told by the assessment seekers as well as the assessed values and further, it may collect a reasonable earning for the public exchequer. A uniform and scientific method of immovable property valuation as demonstrated in this study, may be adopted by the Federal Board of Revenue as well as provincial revenue departments considering not only the structural attributes but also the spatial amenities that must be updated in real time at regular intervals. The residential property data should be accessible to the researchers to explore the different aspects of the real estate market, which may help in building suitable policies.

REFERENCES

Ahmed, R., Jawaid, S. T., & Khalil, S. (2021). Bubble detection in housing market: Evidence from a developing country. *SAGE Open,* 11(2). https://doi.org/10.1177/21582440211006690

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis,* 27(2), 93–115.

Anselin, L., & Le Gallo, J. (2006). Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects. *Spatial Economic Analysis,* 1(1), 31–52. https://doi.org/10.1080/17421770600661337

Attarwala, F. S. (2020). Indicators pointing south for real estate prices. Dawn, *The Business and Finance Weekly*.

Baranzini, A., Schaerer, C., Ramirez, J., & Thalmann, P. (2008). Basics of the hedonic price model. In A. Baranzini, C. Schaerer, J. Ramirez, & P. Thalmann (Eds.), *Hedonic methods in housing markets, pricing environmental amenities and segregation*. Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA. https://doi.org/10.1007/978-0-387-76815-1

Baranzini, A., Schaerer, C., Ramirez, J., & Thalmann, P. (2010). Feel it or measure it-perceived vs. measured noise in hedonic models. *Transportation Research Part D: Transport and Environment,* 15D(8), 473–482.

Barreca, A., Curto, R., & Rolando, D. (2020). Urban vibrancy: An emerging factor that spatially influences the real estate market. *Sustainability (Switzerland),* 12(1), 346. https://doi.org/10.3390/su12010346

Bateman, I. J., Day, B., & Lake, I. (2001). The Effect of Road Traffic on Residential Property Values: A Literature Review and Hedonic Pricing Study. *Bateman, Ian Day, Brett Lake, Iain Lovett, Andrew, January,* 208.

Bitter, C., Mulligan, G. F., & Dall'erba, S. (2007). Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems,* 9(1), 7–27. https://doi.org/10.1007/s10109-006-0028-7

Boza, E. (2015). *Investigation of Housing Valuation Models based on spatial and non-spatial techniques.* Middle East Technical University (METU), Ankara, Turkey.

Brandt, S., Maennig, W., & Richter, F. (2014). Do houses of worship affect housing prices? Evidence from Germany. *Growth and Change*, 45(4), 549–570. https://doi.org/10.1111/grow.12066

Brasington, D. M., & Hite, D. (2005). Demand for environmental quality: a spatial hedonic analysis. *Regional Science and Urban Economics,* 35, 57–82. https://doi.org/10.1016/j.regsciurbeco.2003.09.001

Brown, G., Reed, P., & Raymond, C. M. (2020). Mapping place values: 10 lessons from two decades of public participation GIS empirical research. In *Applied Geography* (Vol. 116, p. 102156). Elsevier Ltd. https://doi.org/10.1016/j.apgeog.2020.102156

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis,* 28(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Brunsdont, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression - Modelling spatial non-stationarity. *Journal of the Royal Statistical Society Series D: The Statistician,* 47(3), 431–443. https://doi.org/10.1111/1467-9884.00145

Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. *Economic Geography,* 66(3), 254–272. https://doi.org/10.2307/143400

Can, A., & Megbolugbe, I. (1997). Spatial Dependence and House Price Index Construction. *Journal of Real Estate Finance and Economics,* 14(1–2), 203–222. https://doi.org/10.1023/A:1007744706720

Casetti, E. (1972). Generating Models by the Expansion Method: Applications to Geographical Research. *Geographical Analysis,* 4(1), 81–91. https://doi.org/10.1111/j.1538-4632.1972.tb00458.x

Casetti, E. (1997). The expansion method, mathematical modeling, and spatial econometrics. *International Regional Science Review,* 20(1–2), 9–33. https://doi.org/10.1177/016001769702000102

Chapman, J. I., Johnston, R. J., & Tyrrell, T. J. (2009). Implications of a land value tax with error in assessed values. *Land Economics,* 85(4), 576–586. https://doi.org/10.3368/le.85.4.576

Crawford, T. W. (2009). Scale analytical. In *International encyclopedia of human geography* (pp. 29–36). Elsevier. https://doi.org/10.1016/B978-008044910-4.00399-0

Crompton, J. L., & Nicholls, S. (2020). Impact on property values of distance to parks and open spaces: An update of U.S. studies in the new millennium. *Journal of Leisure Research,* 51(2), 127–146. https://doi.org/10.1080/00222216.2019.1637704

Dalton, C. M., & Thatcher, J. (2015). Inflated granularity: Spatial "Big Data" and geodemographics. *Big Data and Society,* 2(2). https://doi.org/10.1177/2053951715601144

De, U. K., & Vupru, V. (2017). Location and neighbourhood conditions for housing choice and its rental value: Empirical examination in an urban area of North-East India. *International Journal of Housing Markets and Analysis,* 10(4), 519–538. https://doi.org/10.1108/IJHMA-10-2016-0072

Dismuke, C. E., & Lindrooth, R. (2006). Ordinary least squares. In E. Chumney & K. N. Simpson (Eds.), *Methods and Designs for Outcomes Research* (pp. 93–104). American Society of Health-System Pharmacists.

Do, A. Q., Wilbur, R. W., & Short, J. L. (1994). An empirical examination of the externalities of neighborhood churches on housing values. *The Journal of Real Estate Finance and Economics,* 9(2), 127–136.

Dowall, D. E., & Ellis, P. D. (2009). Urban Land and Housing Markets in the Punjab, Pakistan. *Urban Studies,* 46(11), 2277–2300. https://doi.org/10.1177/0042098009342599

Dutilleul, P., & Legendre, P. (1993). Spatial Heterogeneity against Heteroscedasticity: An Ecological Paradigm versus a Statistical Concept. *Oikos,* 66(1), 152. https://doi.org/10.2307/3545210

Erickson, D. L., Lovell, S. T., & Méndez, V. E. (2011). Landowner willingness to embed production agriculture and other land use options in residential areas of Chittenden County, VT. *Landscape and Urban Planning,* 103(2), 174–184. https://doi.org/10.1016/j.landurbplan.2011.07.009

FBR. (2020). *Circular; C.No. 1(121)R&S/2017-15297-R, Constitution of committees for updation of valuation tables of immoveable property under sub-section (4) of Section 68 of the Income Tax Ordinance, 2001* (p. 1). Revenue Division, Federal Board Of Revenue, Government of Pakistan.

Feuillet, T., Commenges, H., Menai, M., Salze, P., Perchoux, C., Reuillon, R., Kesse-Guyot, E., Enaux, C., Nazare, J. A., Hercberg, S., Simon, C., Charreire, H., & Oppert, J. M. (2018). A massive geographically weighted regression model of walking-environment relationships. *Journal of Transport Geography,* 68, 118–129. https://doi.org/10.1016/j.jtrangeo.2018.03.002

Fik, T. J., Ling, D. C., & Mulligan, G. F. (2003). Modeling spatial variation in housing prices: A variable interaction approach. In *Real Estate Economics* (Vol. 31, Issue 4, pp. 623–646). John Wiley & Sons, Ltd. https://doi.org/10.1046/j.1080-8620.2003.00079.x

Fisher, B., Naidoo, R., & Ricketts, T. (2015). *A field guide to economics for conservationists*. Macmillan Learning.

Fotheringham, A., Brunsdon, C., & Charlton, M. (2002). Geographically weighted regression: The analysis of spatially varying relationships. In *John Wiley & Sons* (Vol. 13).

Fotheringham, A. S. (1997). Trends in quantitative methods I: Stressing the local. *Progress in Human Geography,* 21(1), 88–96. https://doi.org/10.1191/030913297676693207

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and temporal weighted regression (GTWR). *Geographical Analysis,* 47(4), 431–452. https://doi.org/10.1111/gean.12071

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers,* 107(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

Freeman, A. M. (1981). Hedonic prices, property values and measuring environmental benefits: A survey of the issues. In *Measurement in Public Choice* (pp. 13–32). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-05090-1_2

Frost, J. (2019). *Introduction to Statistics: An intuitive guide* (First Edit). Statistics by Jim Publishing: State College, PA, USA, 2019.

Gabrielli, L., & French, N. (2020). Pricing to market: property valuation methods – A practical review. *Journal of Property Investment and Finance*. https://doi.org/10.1108/JPIF-09-2020-0101

Gaffney, M. (2009). The role of land markets in economic crises. *American Journal of Economics and Sociology,* 68(4), 855–888. https://doi.org/10.1111/j.1536-7150.2009.00657.x

Gao, S., Li, L., Li, W., Janowicz, K., & Zhang, Y. (2017). Constructing gazetteers from volunteered Big Geo-Data based on Hadoop. *Computers, Environment and Urban Systems,* 61, 172–186. https://doi.org/10.1016/j.compenvurbsys.2014.02.004

Getis, A. (2008). A history of the concept of spatial autocorrelation: A geographer's perspective*. Geographical Analysis,* 40(3), 297–309. https://doi.org/10.1111/j.1538-4632.2008.00727.x

Getis, A., & Ord, J. K. (2010a). The analysis of spatial association by use of distance statistics. In *Advances in Spatial Science* (Vol. 61, pp. 127–145). Springer International Publishing. https://doi.org/10.1007/978-3-642-01976-0_10

Getis, A., & Ord, J. K. (2010b). The analysis of spatial association by use of distance statistics. In *Perspectives on Spatial Data Analysis* (pp. 127–145). Springer.

Gilderbloom, J. I., Riggs, W. W., & Meares, W. L. (2015). Does walkability matter? An examination of walkability's impact on housing values, foreclosures and crime. *Cities,* 42(PA), 13–24. https://doi.org/10.1016/j.cities.2014.08.001

Gluszak, M., & Zygmunt, R. (2018). Development density, administrative decisions, and land values: An empirical investigation. *Land Use Policy,* 70, 153–161. https://doi.org/10.1016/j.landusepol.2017.10.036

Goix, R. Le, Giraud, T., Cura, R., Le Corre, T., & Migozzi, J. (2019). Who sells to whom in the suburbs? Home price inflation and the dynamics of sellers and buyers in the metropolitan region of Paris, 1996–2012. *PLoS ONE,* 14(3), e0213169. https://doi.org/10.1371/journal.pone.0213169

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). Gwmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software,* 63(17), 1–50. https://doi.org/10.18637/jss.v063.i17

Goodchild, M. F. (2013). The quality of big (geo)data. *Dialogues in Human Geography,* 3(3), 280–284. https://doi.org/10.1177/2043820613513392

Goodman, A. C. (1978). Hedonic prices, price indices and housing markets. *Journal of Urban Economics,* 5(4), 471–484. https://doi.org/10.1016/0094-1190(78)90004-9

Griffith, D. A. (2004). Spatial autocorrelation. In *Encyclopedia of social measurement* (pp. 581–590). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00334-0

Grislain-Letrémy, C., & Katossky, A. (2014). The impact of hazardous industrial facilities on housing prices: A comparison of parametric and semiparametric hedonic price models. *Regional Science and Urban Economics,* 49, 93–107. https://doi.org/10.1016/j.regsciurbeco.2014.09.002

Gul, A., Nawaz, M., Basheer, M. A., Tariq, F., & Raheel Shah, S. A. (2018). Built houses as a tool to control residential land speculation - A case study of Bahria Town, Lahore. *Habitat International,* 71, 81–87. https://doi.org/10.1016/j.habitatint.2017.11.007

Guo, H., Wang, L., Chen, F., & Liang, D. (2014). Scientific big data and Digital Earth. *Chinese Science Bulletin,* 59(35), 5066–5073. https://doi.org/10.1007/s11434-014-0645-3

Haque, N. U. (2015). Flawed urban development policies in Pakistan. *PIDE Working Papers,* 1(119), 1–20.

Harris, R., Singleton, A., Grose, D., Brunsdon, C., & Longley, P. (2010). Grid-enabling geographically weighted regression: A case study of participation in higher education in England. *Transactions in GIS,* 14(1), 43–61. https://doi.org/10.1111/j.1467-9671.2009.01181.x

Hassan, M. M., Ahmad, N., & Hashim, A. H. (2021). Homebuyers Superstitious Belief: Feng Shui and Housing Property. *International Journal of Academic Research in Business and Social Sciences,* 11(7). https://doi.org/10.6007/ijarbss/v11-i7/10297

Helbich, M., Jochem, A., Mücke, W., & Höfle, B. (2013). Boosting the predictive accuracy of urban hedonic house price models through airborne laser scanning. *Computers, Environment and Urban Systems,* 39, 81–92. https://doi.org/10.1016/j.compenvurbsys.2013.01.001

Hu, S., Cheng, Q., Wang, L., & Xu, D. (2013). Modeling land price distribution using multifractal IDW interpolation and fractal filtering method. *Landscape and Urban Planning,* 110, 25–35. https://doi.org/10.1016/j.landurbplan.2012.09.008

Hu, S., Yang, S., Li, W., Zhang, C., & Xu, F. (2016). Spatially non-stationary relationships between urban residential land price and impact factors in Wuhan city, China. *Applied Geography,* 68(1111), 48–56. https://doi.org/10.1016/j.apgeog.2016.01.006

Hubert, L. J., Golledge, R. G., & Costanzo, C. M. (1981). Generalized procedures for evaluating spatial autocorrelation. *Geographical Analysis,* 13(3), 224–233. https://doi.org/10.1111/j.1538-4632.1981.tb00731.x

Ibeas, ángel, Cordera, R., Dell'Olio, L., Coppola, P., & Dominguez, A. (2012). Modelling transport and real-estate values interactions in urban systems. *Journal of Transport Geography,* 24, 370–382. https://doi.org/10.1016/j.jtrangeo.2012.04.012

Jiang, B. (2018). Spatial heterogeneity, scale, data character and sustainable transport in the big data era. In *ISPRS International Journal of Geo-Information* (Vol. 7, Issue 5, p. 167). Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/ijgi7050167

Katsalap, V. (2008). *Spatial dependency and heterogeneity in housing prices across Ukrainian cities*. National University, Kyiv-Mohyla Academy.

Kauko, T. (2003). On current neural network applications involving spatial modelling of property prices. *Journal of Housing and the Built Environment,* 18(2), 159–181. https://doi.org/10.1023/A:1023977111302

Kim, C. W., Phipps, T. T., & Anselin, L. (2003). Measuring the benefits of air quality improvement: A spatial hedonic approach. *Journal of Environmental Economics and Management,* 45(1), 24–39. https://doi.org/10.1016/S0095-0696(02)00013-X

Knight, J. R. (2008). Hedonic modeling of the home selling process. In P. T. Andrea Baranzini, José Ramirez, Caroline Schaerer (Ed.), *Hedonic methods in housing markets; Pricing environmental amenities and segregation* (pp. 39–54). Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA. https://doi.org/10.1007/978-0-387-76815-1_2

Koschinsky, J., Lozano-Gracia, N., & Piras, G. (2012). The welfare benefit of a home's location: An empirical comparison of spatial and non-spatial model estimates. *Journal of Geographical Systems,* 14(3), 319–356. https://doi.org/10.1007/s10109-011-0148-6

L. Krause, A., & Bitter, C. (2012). Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities,* 29(SUPPL.2). https://doi.org/10.1016/j.cities.2012.06.006

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy,* 74(2), 132–157. https://doi.org/10.1086/259131

Larson, W., & Shui, J. (2020). *Land valuation using public records and kriging: Implications for land versus property taxation in cities.*

LeSage, J. P. (2008). An introduction to spatial econometrics. In *Revue d'Economie Industrielle* (Vol. 123, Issue 3, pp. 19–44). https://doi.org/10.4000/rei.3887

Li, L., & Revesz, P. (2004). Interpolation methods for spatiotemporal geographic data. Computers, *Environment and Urban Systems,* 28(3), 201–227.

Li, S., Ye, X., Lee, J., Gong, J., & Qin, C. (2017). Spatiotemporal analysis of housing prices in China: A big data perspective. *Applied Spatial Analysis and Policy,* 10(3), 421–433. https://doi.org/10.1007/s12061-016-9185-3

Li, Z., & Fotheringham, A. S. (2020). Computational improvements to multi-scale geographically weighted regression. *International Journal of Geographical Information Science,* 34(7), 1378–1397. https://doi.org/10.1080/13658816.2020.1720692

Li, Z., Fotheringham, A. S., Li, W., & Oshan, T. (2019a). Fast Geographically Weighted Regression (FastGWR): a scalable algorithm to investigate spatial process heterogeneity in millions of observations. *International Journal of Geographical Information Science,* 33(1), 155–175. https://doi.org/10.1080/13658816.2018.1521523

Li, Z., Fotheringham, A. S., Li, W., & Oshan, T. (2019b). Fast Geographically Weighted Regression (FastGWR): a scalable algorithm to investigate spatial process heterogeneity in millions of observations. *International Journal of Geographical Information Science,* 33(1), 155–175. https://doi.org/10.1080/13658816.2018.1521523

Liang, X., Liu, Y., Qiu, T., Jing, Y., & Fang, F. (2018). The effects of locational factors on the housing prices of residential communities: The case of Ningbo, China. *Habitat International,* 81, 1–11. https://doi.org/10.1016/j.habitatint.2018.09.004

Liao, W. C., & Wang, X. (2012). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics,* 21(1), 16–27. https://doi.org/10.1016/j.jhe.2011.11.001

Lovett, A. A. (2019). Economic valuation of services. *In* C. von H. et al. (eds.) *Landscape*, pp. 315–326). Springer, Dordrecht. https://doi.org/10.1007/978-94-024-1681-7_20

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science,* 28(4), 660–681. https://doi.org/10.1080/13658816.2013.865739

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. In *Annual Review of Public Health* (Vol. 23, pp. 151–169). https://doi.org/10.1146/annurev.publhealth.23.100901.140546

Ma, J., Cheng, J. C. P., Jiang, F., Chen, W., & Zhang, J. (2020). Analyzing driving factors of land values on an urban scale based on big data and non-linear machine learning techniques. *Land Use Policy,* 94, 104537. https://doi.org/10.1016/j.landusepol.2020.104537

Machin, S. (2011). Houses and schools: Valuation of school quality through the housing market. *Labour Economics,* 18(6), 723–729. https://doi.org/10.1016/j.labeco.2011.05.005

Malaitham, S., Fukuda, A., Vichiensan, V., & Wasuntarasook, V. (2020). The hedonic pricing model of assessed and market land values: A case study in Bangkok metropolitan area, Thailand. In *Case Studies on Transport Policy* (Vol. 8, Issue 1, pp. 153–162). Elsevier. https://doi.org/10.1016/j.cstp.2018.09.008

Malik, S., Roosli, R., & Tariq, F. (2020). Investigation of informal housing challenges and issues: experiences from slum and squatter of Lahore. *Journal of Housing and the Built Environment,* 35(1), 143–170. https://doi.org/10.1007/s10901-019-09669-9

Mankad, M. D. (2021). Comparing OLS-based hedonic model and ANN in house price estimation using relative location. *Spatial Information Research*, 1–10. https://doi.org/10.1007/s41324-021-00416-3

Mitchel, A. (2005). The ESRI Guide to GIS analysis, Volume 2: Spatial measurements and statistics. *ESRI Guide to GIS Analysis.*

Mitchell, A. (2005). The ESRI guide to GIS analysis. Volume 2, *Spatial measurements and statistics* (First Edit). Esri Press.

Munshi, T. (2020). Accessibility, infrastructure provision and residential land value: Modelling the relation using geographic weighted regression in the city of Rajkot, India. *Sustainability (Switzerland),* 12(20), 1–16. https://doi.org/10.3390/su12208615

Ord, J. K., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis,* 27(4), 286–306. https://doi.org/10.1111/j.1538-4632.1995.tb00912.x

Ord, J. K., & Getis, A. (2001). Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of Regional Science,* 41(3), 411–432. https://doi.org/10.1111/0022-4146.00224

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Stewart Fotheringham, A. (2019). MGWR: A Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS International Journal of Geo-Information,* 8(6). https://doi.org/10.3390/ijgi8060269

Ouattara, K., Kim, Y., Raza, H., Hadi, Q. ul A., Rostom, A. M. T., Milyutin, A., Rashid, A. A., Khan, B. N., Haq, E., Grigoryeva, E., Funahashi, J., Chiquier, L., Zaheer, N., Dione, N. T., Hassler, O., Shaikh, S. A., Ahmad, S., Athar, S., & Conde, V. M. O. (2018). *PAKISTAN HOUSING FINANCE PROJECT (P162095)Report No: PAD2385.*

Pace, R. K., & Gilley, O. W. (1998). Generalizing the OLS and grid estimators. *Real Estate Economics,* 26(2), 331–347. https://doi.org/10.1111/1540-6229.00748

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: A review of valuation methods. In *Journal of Property Investment & Finance* (Vol. 21, Issue 4, pp. 383–401). MCB UP Ltd. https://doi.org/10.1108/14635780310483656

Pakistan Bureau of Statistics. (2017). Area & Population of Aadministrative Units. In *Population Census.*

Pakistan, S. B. of. (2019). *State Bank of Pakistan infrastructure, housing; SME finance policy for low-cost housing finance.* State Bank of Pakistan Infrastructure, Housing and SME Finance Department (Issue March).

Peeters, A., Zude, M., Käthner, J., Ünlü, M., Kanber, R., Hetzroni, A., Gebbers, R., & Ben-Gal, A. (2015). Getis-Ord's hot- and cold-spot statistics as a basis for multivariate spatial clustering of orchard tree data. *Computers and Electronics in Agriculture*, 111, 140–150. https://doi.org/10.1016/j.compag.2014.12.011

Rizvi, Z. M. (2018). *National affordable housing policy* (pp. 1–22).

Rosen, S. (n.d.). *Hedonic prices and implicit markets: Product differentiation in pure competition*. 34–55.

Sabir, M., Torre, A., & Magsi, H. (2017). Land-use conflict and socio-economic impacts of infrastructure projects: the case of Diamer Bhasha Dam in Pakistan. *Area Development and Policy,* 2(1), 40–54. https://doi.org/10.1080/23792949.2016.1271723

Sah, V., Conroy, S. J., & Narwold, A. (2016). Estimating school proximity effects on housing prices: The importance of robust spatial controls in hedonic estimations. *Journal of Real Estate Finance and Economics,* 53(1), 50–76. https://doi.org/10.1007/s11146-015-9520-5

Salon, D., Wu, J., & Shewmake, S. (2014). Impact of bus rapid transit and metro rail on property values in Guangzhou, China. In *Transportation Research Record* (Vol. 2452, pp. 36–45). National Research Council. https://doi.org/10.3141/2452-05

Scott, L. M., & Janikas, M. V. (2010). Spatial statistics in ArcGIS. In M. M. F. A. Getis (Ed.), *Handbook of applied spatial analysis: Software tools, methods and applications* (pp. 27–41). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03647-7_2

Seo, K., Golub, A., & Kuby, M. (2014). Combined impacts of highways and light rail transit on residential property values: A spatial hedonic price model for Phoenix, Arizona. *Journal of Transport Geography,* 41. https://doi.org/10.1016/j.jtrangeo.2014.08.003

Shabana, Ali, G., Bashir, M. K., & Ali, H. (2015). Housing valuation of different towns using the hedonic model: A case of Faisalabad city, Pakistan. *Habitat International,* 50*,* 240–249. https://doi.org/10.1016/j.habitatint.2015.08.036

Singh, A., Sharma, A., & Dubey, G. (2020). Big data analytics predicting real estate prices. *International Journal of Systems Assurance Engineering and Management,* 11(2), 208–219. https://doi.org/10.1007/s13198-020-00946-3

Stewart Fotheringham, A., Charlton, M., & Brunsdon, C. (1996). The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Systems,* 10(5), 605–627.

Taylor, L. O. (2008). Theoretical foundations and empirical developments in hedonic modeling. In P. T. Andrea Baranzini, José Ramirez, Caroline Schaerer (Eds.), *Hedonic methods in housing markets; Pricing environmental amenities and segregation* (pp. 15–37). Springer Science + Business Media, LLC, 233 Spring Street, New York, NY 10013, USA. https://doi.org/10.1007/978-0-387-76815-1_1

Thomas, D. M., & Mathur, S. (2019). Data Analysis by Web Scraping using Python. *Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019,* 450–454. https://doi.org/10.1109/ICECA.2019.8822022

Thompson, E., Butters, R. B., & Schmitz, B. T. (2012). The property value premium of a place of worship. *Contemporary Economic Policy,* 30(2), 215–222. https://doi.org/10.1111/j.1465-7287.2011.00255.x

Tian, G., Wei, Y. D., & Li, H. (2017). Effects of accessibility and environmental health risk on housing prices: a case of Salt Lake County, Utah. *Applied Geography,* 89*,* 12–21. https://doi.org/10.1016/j.apgeog.2017.09.010

Wahid, A., Mantell, E. H., & Mumtaz, M. Z. (2021). Under invoicing in the residential real estate market in Pakistan. *International Journal of Strategic Property Management,* 25(3), 190–203.

Wajahat, F. (2012). Perceptions of tenure security in a squatter settlement in Lahore, Pakistan. In *Transforming Asian Cities: Intellectual Impasse, Asianizing Space, and Emerging Translocalities* (pp. 137–147). https://doi.org/10.4324/9780203093894

Wang, W., van Noorloos, F., & Spit, T. (2020). Stakeholder power relations in land value capture: Comparing public (China) and private (U.S.) dominant regimes. *Land Use Policy,* 91, 104357. https://doi.org/10.1016/j.landusepol.2019.104357

Wani, S., Shaikh, H., & Harman, O. (2020). Urban property taxes in Pakistan's Punjab. International Growth Centre (IGC).

Wen, H., Jin, Y., & Zhang, L. (2017). Spatial heterogeneity in implicit housing prices: Evidence from Hangzhou, China. *International Journal of Strategic Property Management,* 21(1), 15–28. https://doi.org/10.3846/1648715X.2016.1247021

Wen, H., Xiao, Y., & Zhang, L. (2017). The school district, education quality, and housing price: Evidence from a natural experiment in Hangzhou, China. *Cities,* 66, 72–80. https://doi.org/10.1016/j.cities.2017.03.008

Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (Fifth Edit). Nelson Education, Ltd.

Wu, C., Ye, X., Ren, F., Wan, Y., Ning, P., & Du, Q. (2016). Spatial and social media data analytics of housing prices in Shenzhen, China. *PLoS ONE,* 11(10), e0164553. https://doi.org/10.1371/journal.pone.0164553

Xiao, Y., Chen, X., Li, Q., Yu, X., Chen, J., & Guo, J. (2017). Exploring determinants of housing prices in Beijing: An enhanced hedonic regression with open access POI data. ISPRS *International Journal of Geo-Information,* 6(11), 358. https://doi.org/10.3390/ijgi6110358

Yang, L., Chau, K. W., Szeto, W. Y., Cui, X., & Wang, X. (2020). Accessibility to transit, by transit, and property prices: Spatially varying relationships. *Transportation Research Part D: Transport and Environment,* 85, 102387. https://doi.org/10.1016/j.trd.2020.102387

Yang, L., Chu, X., Gou, Z., Yang, H., Lu, Y., & Huang, W. (2020). Accessibility and proximity effects of bus rapid transit on housing prices: Heterogeneity across price quantiles and space. *Journal of Transport Geography,* 88, 102850. https://doi.org/10.1016/j.jtrangeo.2020.102850

Yang, L., Wang, B., Zhang, Y., Ye, Z., Wang, Y., & Li, P. (2018a). Willing to pay more for high-quality schools? *International Review for Spatial Planning and Sustainable Development,* 6(1), 45–62. https://doi.org/10.14246/irspsd.6.1_45

Yang, L., Wang, B., Zhang, Y., Ye, Z., Wang, Y., & Li, P. (2018b). Willing to pay more for high-quality schools? A hedonic pricing and propensity score matching approach. *International Review for Spatial Planning and Sustainable Development,* 6(1), 45–62. https://doi.org/10.14246/irspsd.6.1_45

Yu, D. (2007). Modeling owner-occupied single-family house values in the City of Milwaukee: A geographically weighted regression approach. *GIScience and Remote Sensing,* 44(3), 267–282. https://doi.org/10.2747/1548-1603.44.3.267

Yuen, B., & Choi, S. (2012). *Making spatial change in Pakistan cities growth enhancing.* PK 11/12; World Bank Policy Paper Series on Pakistan.

Zaman, K. U., & Baloch, A. A. (2011). Urbanization of arable land in Lahore city in Pakistan: A case-study. *Canadian Social Science,* 7(4), 58–66.

Zhang, L., Chen, J., Hao, Q., & Li, C. Z. (2018). Measuring the NIMBY effect in urban China: The case of waste transfer stations in metropolis Shanghai. *Journal of Housing and the Built Environment,* 33(1), 1–18. https://doi.org/10.1007/s10901-017-9565-2

PART II

GROWTH & TAXATION

Policy Briefs

ESTIMATING THE DISTRIBUTIONAL BURDEN OF INDIRECT TAXES IN PAKISTAN

Iffat Ara

INTRODUCTION

The structure of federal taxes in Pakistan heavily relies on indirect taxes which constituted 62 per cent of total tax receipts in 2018-19, whereas direct taxes constituted 38 per cent. Of the 62 per cent indirect taxes, general sales tax (GST) dominated with a share of 38 per cent, whereas customs duties (CD) and federal excise duty (FED) constituted 18 per cent and 6 per cent, respectively.

Different studies in Pakistan have found indirect taxation to be either proportional or regressive (see, for example. Jamal and Javed, 2013; Wahid and Wallace, 2008; Refaqat, 2008; and SPDC, 2004). However, these studies have some issues. For example, Refaqat (2008) and Jamal and Javed (2013) considered taxes levied only on final consumption and did not incorporate taxes levied on intermediate inputs. Though Wahid and Wallace (2008) and SPDC (2004) accounted for taxes levied on intermediate inputs, they did not estimate the incidence by different commodity groups.

The study, on which this policy brief is based, assessed the incidence of federal indirect taxes and their distributional burden in Pakistan across deciles of households by filling these gaps. It used the most recent available Household Integrated Economic Survey (HIES) 2018-19 to observe household expenditures, and the latest available Input-Output Table 2010-11 to capture the cascading effect of indirect taxes. The paper examined the extent to which each component of indirect taxes (GST, CD and FED) can be considered progressive (i.e., placing a higher tax burden on higher income groups), regressive (i.e., placing a higher tax burden on lower income groups), or proportional (i.e., placing the same tax burden on each income groups). The paper also estimated incidence and its distribution for various commodity groups and heads of withholding tax (WHT) (subject to the availability of data), which is an indirect tax.

METHODOLOGY

The research followed an input-output model-based approach to estimate the incidence of indirect taxes that allows tracing the cascading effects of indirect taxes on intermediate inputs (see Ahmed and Stern, 1991). The following steps are required for the application of this approach:

- Computing the nominal rate of taxes that are based on revenue collection. Using nominal tax rates, instead of statutory rates, helps overcome the issue of tax compliance.
- Computating input-adjusted effective tax rates (ETRs).
- Taking households as a unit of analysis and their total expenditures to rank them by welfare level.
- Assuming the final burden of indirect taxes to be borne by consumers.
- Computing households' tax payments for each tax component.
- Computing the tax incidence which is the percentage share of tax payments in the respective household's total expenditures.

- Assessing progressivity/regressivity by comparing the average rate of tax payment across different income groups.

FINDINGS

Overall Incidence

The overall average incidence of all indirect taxes combined was 20.7 per cent in Pakistan. The distribution of incidence of all indirect taxes was regressive, ranging from 22 per cent in the lowest decile to 19 per cent in the highest decile. This suggests that for every PKR 100 expenditure, the poorest 10 per cent households, on average, devoted PKR 22 to paying indirect taxes, while the richest 10 per cent devoted PKR 19.

As for the distribution of the components of indirect taxes, except the FED-local, all components of indirect taxes were regressive. The magnitude and extent of regressivity (difference in the incidence for the bottom and top deciles) was the highest for GST-Imports where the bottom 10 per cent households, on average, paid 7.6 per cent, while the top 10 per cent paid 6.4 per cent of their total expenditures in taxes. The incidence of FED-Local was mildly progressive.

The incidence of all indirect taxes combined showed a regressive pattern in both rural and urban areas, where incidence was roughly one percentage point higher for each decile in rural areas compared to urban areas. In this case, as well, the incidence of all components of indirect taxes was regressive in both rural and urban areas except FED-Local, which exhibited a mild progressive pattern.

Incidence by Commodity Groups

Indirect taxes on basic food items indicated a highly regressive pattern across all deciles.[1] For example, 6.2 per cent of the poorest 10 per cent households' expenditures were on indirect taxes on basic food items compared to 2.3 per cent by the richest 10 per cent households. Other groups that showed regressive taxation patterns across all deciles included household items, pharmaceuticals, and tobacco and allied products. Some groups, such as personal items and transport services, depicted an overall regressive pattern but had proportional patterns for lower or middle-income groups.

Commodity groups that had progressive indirect taxation included non-basic food items, transport fuel, and durable goods. Though the magnitude of incidence was the highest for non-basic food items, the extent of progressivity was the highest for transport fuel. For example, the poorest 10 per cent households' 0.8 per cent expenditures were on indirect taxes on transport fuel, while the richest 10 per cent paid 2.3 per cent. Other groups, such as utilities and books & stationary, though, showed an overall progressive pattern, the incidence for the 50 per cent upper-income group was proportional. Taxes on communication services were proportional across all groups.

Incidence of Withholding Tax

The incidence of WHT on telephone and mobile usage portrayed a slightly regressive pattern, while the nature of WHT was proportional for middle deciles.[1] The incidence of WHT on petroleum products and electricity was progressive for all deciles. WHT on internet usage, air travel, and CNG stations was also progressive albeit minimally.

KEY POLICY RECOMMENDATIONS

Food inflation has often been a major public policy challenge for the governments in Pakistan and numerous measures are undertaken to control food prices to provide relief for the poor, for example, the exemption of major food items from indirect taxation. However, indirect taxes levied on inputs used to produce these items transfer to the final prices of these items and cause an increase in prices. Further, to raise revenues, governments often increase taxes on necessities that have inelastic demand, such as

[1] Items such as wheat flour, rice, pulses, vegetables, spices, fresh dairy, ghee, sugar, tea are considered as basic food items for this paperthis. The remaining food items are included in non-basic food group.

utilities, which puts a burden on the poor.

Regressivity affecting the poor segment needs to be addressed but without causing secondary distortions. For example, exempting selected essential items as well as their inputs from taxes would not only cause revenue losses but would also benefit the items not in the consumption basket of the poor.

An alternative way to avoid secondary distortions and support low-income groups is transfer payments, which can minimise the impact of taxes on them. Practices from other countries also demonstrate the use of transfer payments. For example, Karageorgas (1973) pointed out a decline in inequality after the initiation of transfer payments in Greece, with the highest benefit received by the lowest income groups.

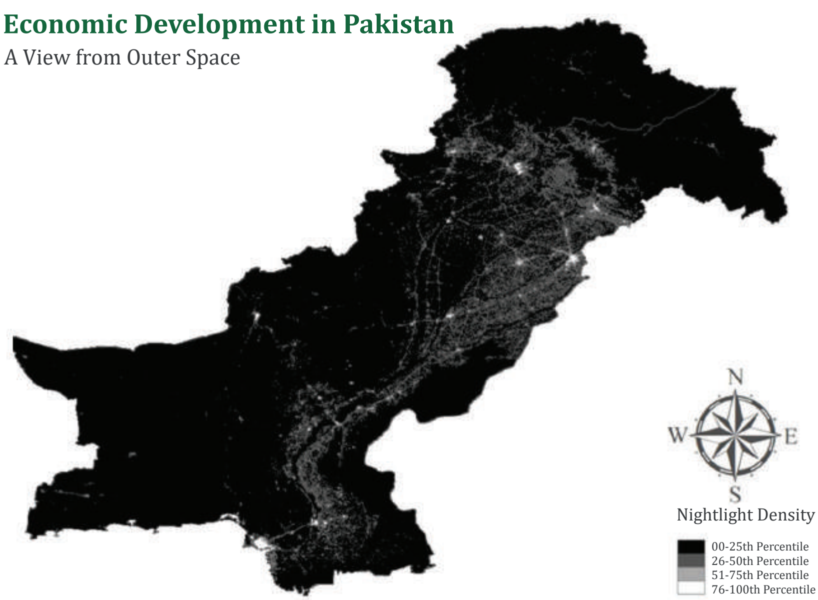
CITY DEVELOPMENT PRODUCT: DATA ARCHITECTURE FOR SUSTAINABLE ECONOMIC DEVELOPMENT IN PAKISTAN

Mohammad Ahmad

INTRODUCTION

The Government of Pakistan currently does not have a methodology for estimating Regional Domestic Product (RDP) at the provincial, district, or city levels. The need for disaggregated estimates of economic growth has become important after the 18th Amendment to the Constitution (2010), under which the key areas for economic development, including health, education, infrastructure, and industrial development, have been devolved to the provinces.

The availability of city-level data that is reliable and comparable can have multiple channels of impact on the state's planning, administrative, and public funds management capacity. Many developing countries, including Brazil, Colombia, and India, have developed provincial Systems of Regional Accounts (SRA) based on extensive surveys and data collection that are analogous to the national income accounting exercise that takes place each year to compute aggregate GDP measures. However, with the cheap availability of high-resolution satellite imagery, a more cost-effective estimation methodology for subnational GDP estimation can be derived.



NIGHTLIGHTS AS A RELIABLE ESTIMATOR OF ECONOMIC ACTIVITY

The availability of fine-grain, remotely-sensed data has become a new powerful tool for analysis within the economics discipline. Economists have found many applications for this source of data, and multiple rich streams of economic literature have developed as a result. One of the most commonly-used sources of satellite data is night lights (NTL), or the total visible light emitted from Earth's surface at night for any defined region.

FINDINGS

The results show that Punjab leads with an estimated 54 per cent contribution to Pakistan's GDP. When normalized by population, the per million estimate of GDP is almost identical to the national NTL-based GDP for all of Pakistan. The imputed per capita Regional Domestic Product (RDP) based on spatial disaggregation of economic activity by provinces is PKR 188,000 for Punjab. This number is very close to the national per capita GDP of PKR 182,000.

The second largest contributor to Pakistan's GDP based on NTL-based GDP estimates is Sindh. It contributes 27 per cent to Pakistan's overall GDP. When normalised by population, Sindh's NTL per million is higher than the national level by 13 per cent. This implies that a segment of Sindh's population is relatively more productive than the national mean. This is not a surprising result as Karachi is the economic capital of the country. As a result of Karachi's high productivity, the imputed per capita RDP for Sindh is PKR 213, 000.

As per NTL-based RDP estimates, Khyber Pakhtunkhwa province contributes 8 per cent to the national GDP of Pakistan. After normalising for population, the NTL per million estimate is 46 per cent of the national NTL per million estimates. Khyber Pakhtunkhwa's imputed RDP using the NTL-based approach shows a per capita income of PKR 102, 000. It is important to note that KP's estimates include the newly merged tribal districts (ex-FATA), which have one of the lowest levels of economic development in the country.

Estimates from this study show that Balochistan contributes approximately 8 per cent to the national economy of Pakistan. At 0.81 per cent of the national mean, the NTL per million score of Balochistan is higher than that of KP largely due to the smaller population base which goes into the denominator of the NTL per million results. The imputed per capita income for Balochistan is approximately PKR 152, 000, which, again, is higher than KP owing to a smaller population base.

Table : Cross-province comparisons of VIIRS VNL V2 nightlights, 2019 (Elvidge et al., 2021). Population data is from PBS Census, 2017. GDP data taken from World Bank's World Development Indicators (WDI).

Province	NTL (Aggregate)		NTL Per Million		Max		Imputed RDP	
	Abs. (000s)	Share	Abs.	Rel.	Abs.	Rel.	RDP (PKR, Bil.)	P.C. RDP (PKR)
Pakistan	**1,233.9**	**1**	**6126.1**	**1**	**532.5**	**1.00**	**38,000**	**182,000**
Punjab	670.7	0.54	6098.1	0.99	319.2	0.60	20,656	188,000
Sindh	331.5	0.27	6927.2	1.13	532.5	1.00	10,209	213,000
Kh. Pakhtunkhwa	101.1	0.08	2847.6	0.46	196.5	0.37	3,113	102,000
Balochistan	61.0	0.05	4944.7	0.81	237.0	0.45	1,878	152,000
Islamabad CT	40.6	0.03	20249.7	3.31	75.1	0.14	1,249	624,000
AJK + GB	29.0	0.02	6236.6	1.02	10.6	0.02	893	190,000

CITY DOMESTIC PRODUCT (CDP)

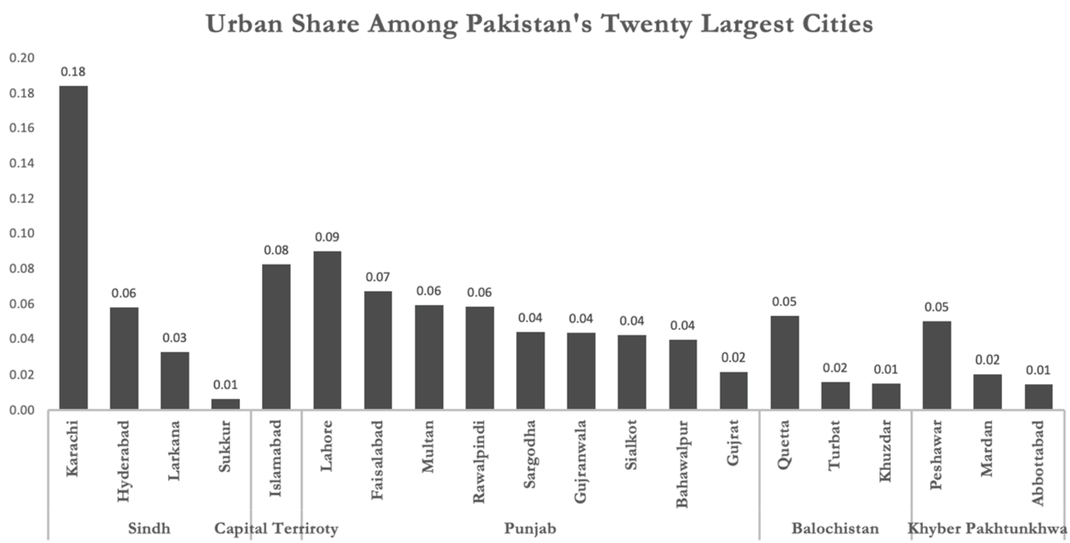
The NTL-based RDP estimation methodology can be used to estimate the City Domestic Product (CDP) at the city level. Figure 1 presents the urban share of economic activity among the 20 largest cities in Pakistan. Among these, the city of Karachi alone contributes 18 per cent of the urban economic output. By magnitude, Karachi is by far the largest urban centre of economic activity, followed by Lahore, Faisalabad, and Islamabad at approximately 7 per cent each. After accounting for the megacities, which are well understood as regions with high economic activity relative to the rest of the country, what shows up is the significant contribution of Pakistan's medium-sized cities. Cities like Hyderabad, Gujranwala, and Rawalpindi combined provide roughly the same economic output as the city of Karachi. This may be due to the presence of small- and medium-sized manufacturing clusters concentrated in each of these cities.

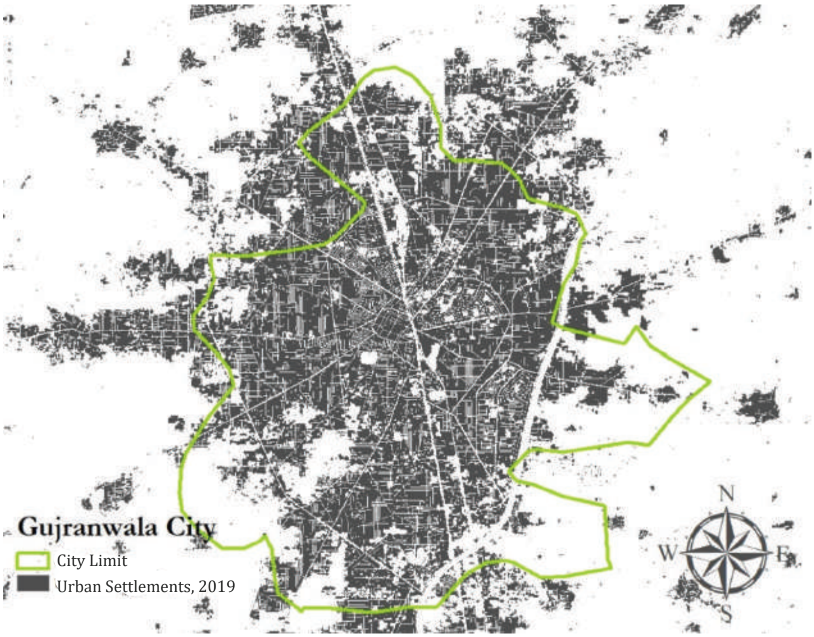
It is important to point out that the CDP estimates presented in Figure 2 are based on de jure city limits, and actual figures may be larger due to urban sprawl beyond the jurisdictional limit of the city. As an example, Figure 2 shows how Gujranwala city has outgrown its defined city limits. Since these de facto city limits are not factored into the above CDP estimates, the above results may be biased downwards. A natural next step would be to account for this bias by expanding the city limits in my analysis.

POLICY SIGNIFICANCE

The project to develop city-level RDP estimates is a promising avenue for building data architecture for economic and urban development in Pakistan. The CDP data allows researchers and policymakers alike to analyse the economic geography of Pakistan using a sound and consistent methodology. Preliminary findings highlight meaningful differences in economic activity across provinces, districts, and cities. These findings also show the promise of Pakistan's smaller cities, many of which are growing faster than their neighbouring cities. An exploration of the determinants of these growth differentials is a fruitful avenue for future research. The CDP dataset has the potential to be a significant value-add to Pakistan's research and policy space alike.

Figure : Economic activity in Pakistan's largest cities




Gujranwala City
City Limit
Urban Settlements, 2019
N
W
E
S

EVALUATION OF DIFFERENT TAX REFORM PROPOSALS

Muhammad Nadeem Sarwar

INTRODUCTION

To provide people with public goods and infrastructure, and to foster economic activities, governments need funds, which are mainly collected through taxes. The choice of tax structure directly affects tax revenues, economic growth, and income distribution. Therefore, while developing a comprehensive taxation system, governments must take a proper account of the taxation system's macroeconomic and distributional impacts.

This policy brief discusses the impact of various tax reform proposals on Pakistan's key macroeconomic indicators. The study on which this brief is based evaluated various tax rate reform proposals using the general equilibrium framework. Specifically, it identified and quantified the direction and magnitude of impacts of reducing the marginal income tax rate, decreasing the number of slabs, and introducing a flat income and corporate tax rate with a reduction in sales tax, customs duties, and other taxes on economic growth, consumption, exports, various sectors, and income.

METHODOLOGY

The computable general equilibrium (CGE) model was used to analyse the proposed tax reforms. The Input-Output (IO) Table 2017, developed by Asian Developed Bank, data from national income accounts, Labour Force Survey LFS, and the Household Integrated Economic Survey (HIES) were used to develop the Social Accounting Matrix (SAM). This was then used as the basic data source for the CGE model.

The SAM 2017 included 34 commodities produced by 34 activities and 24 factors owned by 8 categories of households. It was assumed that the government earns income mainly by collecting tax revenue, but also earns some capital income and receives aid and loans. The government provide spends this income on public goods administration, transfer payments, and subsidies to households and firms. Some of the government expenditures are on debt servicing.

RESULTS

The results show that the reduction in the personal income tax rate (Simulation 1) leaves households with more disposable income which finances increased consumption expenditures. As savings grow slowly, which is reflected by smaller growth in investment, domestic production fails to match with higher domestic demand so demand for imports increases. The aggregated demand is also fuelled by higher government expenditures and, therefore, in the case when balancing the budget is not binding, might lead to governments accumulating more debt and leaving less for the private sector. This together results in decreasing the GDP in the long run.

The long-run results of Simulation 2 show that the real GDP increases because of positive growth in private consumption, investment, government consumption, and higher trade. In this case, exports increase by 0.162 per cent compared to a decline of 0.389 per cent in the case of lower personal income tax only. This is primarily because of low financial and compliance costs owing to a simplified tax system, which encourages more investment and disencumberance from a complicated tax system.

In the short run, GDP growth is positive even in Simulation 1. The other difference is that there is a price increase even in the case when all the taxes are lower. This shows that a decrease in the cost of production due to lower taxes is not passed through to the consumers in the short run, which is an indication of firms' market power and frictions in the economic system.

SECTORAL IMPACTS

The results show that with a decrease in personal income tax, there is no significant positive impact on the output of firms. In the case of Simulation 2, a cut in rates of all kinds of taxes and reducing the number of taxes, the firms reap higher profits and, hence, can look forward to expanding their production capacity. This especially suits the export industry as it reduces its cost making the exports more compatible. Both these scenarios result in increasing take-home incomes of various categories of labour, i.e., rural and urban low-skilled and rural and urban high-skilled in all provinces.

POLICY RECOMMENDATIONS

This analysis leads to some simple but important policy recommendations. One of the policies that can be recommended is that simplifying the tax regime and lowering taxes will result in higher individual and corporate incomes bringing about a shift in favour of competitive and efficient sectors, which will result in higher economic growth. This higher growth will result in an increase in tax revenue without overburdening individuals and businesses.

Secondly, reducing rates of only one or a few taxes will not work as effectively as lowering rates of all the taxes, reducing the total number of taxes, and letting various sectors compete based on productivity and efficiency. The results of the study also show that reducing tax rates will result in increasing fiscal deficit. However, if the government is restricted to keeping budget balance or deficit under control, this may compel the government to cut down unnecessary expenditures and reduce its footprint on the economy, which will result in lowering labour demand in the public sector and releasing it for private firms and reducing market distortions.

REVISITING URBAN IMMOVABLE PROPERTY VALUATION: AN APPRAISAL OF SPATIAL HETEROGENEITIES USING BIG DATA IN PUNJAB

Shoaib Khalid and Fariha Zameer

INTRODUCTION

The current system of the valuation of immovable properties by government agencies (DC and FBR rates) is inefficient, non-scientific, and inconsistent as official valuation methods do not account for the spatial attributes of the real estate properties. As a result, the valuation of urban immovable properties remains highly incompatible with fair market values. Moreover, there is no mechanism to record actual market transactions. Due to a poor official property valuation system and low regulatory oversight, most of the gains go unreported, which, in turn, gives rise to black economy practices and loss of revenue for the national exchequer. Furthermore, whenever the government acquire lands for new infrastructure or any other project, the compensatory payments to the owners are made according to DC valuation tables, which are always far lower than fair market values resulting in mass protests, demonstrations, and unrest.

Owing to the above issues, there is a need for a more sophisticated system of valuation of immovable property based on spatial variables to bridge the gap between official rates and market rates for the extended revenue collection, help the sellers and buyers to avoid market speculation practices which make the property inflated, and appropriate compensation to the landowners other than tax purposes as well. Thus, to investigate the abovementioned problems in urban property valuation and suggest guidelines for policy, the study, on which this policy brief is based, examined the dynamics of urban immovable property values based on location-specific parameters using big data and investigated the spatial variations in the urban immovable property values.

METHODOLOGY AND DATA

The urban immovable property valuation in two major cities, i.e., Lahore and Faisalabad, was undertaken to explore the significant impact of location-specific parameters on the urban immovable property prices. To compute the immovable property values, big data analytics in Geographic Information System (GIS) was used. The traditional hedonic price models give little importance to the spatial characteristics of individual housing units and mainly use structural attributes of property for valuation. However, spatial heterogeneity should be considered while appraising residential property prices since the house characteristics may vary over space.

To address this issue, different valuation models based on the ordinary least square regression and the Fast Geographic Weighted Regression (FastGWR), a scalable open-source implementation of Python and Message Passing Interface (MPI), which can process millions of observations, were used. These valuation models estimated the total net worth of the residential real estate market in Lahore and Faisalabad.

The data were collected from different sources including the primary field survey, and governmental and private organizations. The data on property price data, house parcels, parcel types, road network, urban land use, and location of important places were used. This generated a big dataset of 1.2 million residential property parcels. Based on an extensive literature review, the prime structural attribute of housing properties, the floor area, which explains the value of the properties the most, and spatial regressors were selected as covariates This exclusion resulted in nine potential explanatory variables qualifying for the final model.

FINDINGS

The results show that floor area, and proximity to health facilities, recreational sites and marketplaces add a premium to the property values, while proximity to educational institutions, worship places, and solid waste facilities reduces the property values in both However, proximity to industrial units and graveyards affect the property values negatively in Lahore but positively in Faisalabad.

Nine locational features within four categories (i.e., amenities, cultural, educational and health facilities, and recreation) were selected to explore the correlation between the spatial determinants of housing prices. The results show positive and negative statistically significant correlations between different locational features and residential property prices. The main determinants of prices are the area of the house. Other positively correlated variables are distance to worship places, distance to solid waste facilities, and distance to educational institutions. The distance to public parks, markets, hospitals, graveyards, and distance to industries have a negative association with property values.

RECOMMENDATIONS AND POLICY IMPLICATIONS

The ongoing development and accelerated urban expansion have evolved the cities in Punjab from a single centre to multiple centres pattern, and there is a need for a reappraisal of properties to increase revenues from property taxes. Spatial determinants of housing valuation are significantly important and, hence, must be considered in policy formulation and future urban planning and design.

The results have some general implications for policymakers, investors, real estate developers, and urban planners. The estimates are appropriate to inform about the residential property values since the spatial diversity of the coefficients is important for decision-makers, who require explicit knowledge of the local or regional housing markets. This will help to refine policies and have a better understanding of the local house price variations.

Since the record of past market transactions plays a key role in the factual valuation of immovable properties, we suggest the formulation of a system to record the fair market prices of real estate properties. For this purpose, the following steps may be taken. Firstly, the property transfer fee may be waived to attract sellers and buyers towards disclosing the factual deal prices of properties and the decline in revenue due to this exemption may be bridged through the increase in annual property tax. Secondly, an immovable property valuation desk may be established in revenue departments to serve the masses by assessing the market price of their properties at a nominal fee. This desk can record the geo-locations and structural attributes of properties and assessed values. It may also bring in a reasonable earning for the public exchequer.

A uniform and scientific method of immovable property valuation demonstrated in this study may be adopted by the Federal Board of Revenue as well as provincial revenue departments. The valuation system should not only consider the structural attributes but also the spatial amenities, which must be updated in real-time at regular intervals. The residential property data should be accessible to the researchers to explore the different aspects of the real estate market which may help in building the suitable policies

RASTA Publications

2024 PIDE Reform Manifesto: Transforming Economy and Society

2024 Government Training & Development Endeavors

2023 The PDR 62 (3) RASTA Special Issue

2023 The PDR 62 (4) RASTA Special Issue

2023 The State of Commerce in Pakistan: International & Domestic

2023 PIDE Sludge Audit Vol – 2

2023 Power, Profits & Plans: The Political Economy of Housing in Pakistan

2023 Traffic Management & Congestion Mitigation Parking Policy for Islamabad Capital Territory

2023 BASIC NOTES 1 – 8,

2023 Identity, Inclusion and Social Capital in Upper and Southern Punjab

2022 Local Research, Local Solutions: Energy Issues, Volume I

2022 Local Research, Local Solutions: Urban Development, Volume II

2022 Local Research, Local Solutions: Technology & Public Service Delivery, Volume III

2022 Local Research, Local Solutions: Social Sector Development, Volume IV

2022 Local Research, Local Solutions: Markets & Regulations, Volume V

2022 Local Research, Local Solutions: Political Economy of Development Reform, Volume VI

2022 Regulatory Environment of the Professions in Pakistan

2022 Inflation Analytics

2022 Parallel Education Streams in the Public Sector

2022 The PDR 61 (2) RASTA Special Issue

2022 The PDR 61 (3) RASTA Special Issue

2022 Policy Research Methods: A Step-by-Step Guide from Start to Finish for Students and Practitioners

2022 Power Sector: An Enigma with No Easy Solution,

2022 PIDE Sludge Audit Vol – 1,

CGP Call for Research Proposals 1 – 6

RASTA Notes 1 – 3

About RASTA CGP

The Research for Social Transformation and Advancement (RASTA) at the Pakistan Institute of Development Economics (PIDE) is the largest economics and public policy research grants programme in Pakistan. Its mission is to promote research culture and develop extensive network of academia and think tanks across Pakistan producing high-quality, evidence-based policy research to inform public policy processes. The Competitive Grants Programme (CGP) is the flagship initiative of RASTA under which research proposals are invited bi-annually on specific themes/topics decided by the Research Advisory Committee (RAC). Applications from all around Pakistan and abroad are invited through open competition and awards are decided by the RAC after a rigorous and transparent review process. Anyone with a research focus on Pakistan's public policy issues relevant to the CGP themes/topics of each round can participate and win the CGP award.

For details, visit **www.pide.org.pk/rasta** and follow us on Twitter **@RASTA_PIDE**

RASTA

Pakistan Institute of Development Economics

RESEARCH FOR SOCIAL TRANSFORMATION & ADVANCEMENT
PAKISTAN INSTITUTE OF DEVELOPMENT ECONOMICS