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

Mohammad Ahmad

(CGP #02-098)

2ND RASTA CONFERENCE

Wednesday 1st & Thursday 2nd June 2022

Marriott Hotel, Islamabad

RESEARCH FOR SOCIAL TRANSFORMATION & ADVANCEMENT

Pakistan Institute of Development Economics

Islamabad

ABSTRACT

This project uses 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, the RDP can be a sufficient statistic for tracking economic growth and development at lower administrative levels where identification and preparation of development projects takes place. The author intends to build a unique panel dataset of city-level RDP estimates for Pakistan from 2012-2020, using satellite imagery on night-time 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.

The goal is this project is to provide an open access repository of decentralized economic growth data for Pakistani cities which can be used for evidence-based research on Pakistan. In its conclusion, the project will produce a panel dataset of spatial economic growth for Pakistani cities from 2012-2020 using highly granular satellite data and a well-defined, academically sound estimation methodology.

TABLE OF CONTENTS

INTRODUCTION

The government of Pakistan currently does not have a methodology for estimating Regional Domestic Product (RDP) at the provincial, district or city level. The need for disaggregated estimates of economic growth has become even more salient after the 18th amendment to the constitution (2010), under which 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 in public finance.

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

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 of subnational GDP estimation can be derived.

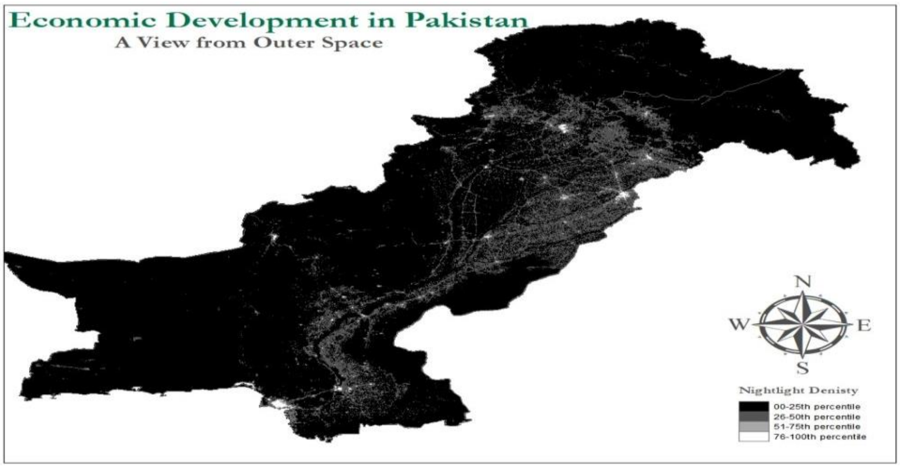
Moreover, while this proposed top-down approach to RDP estimation may leave certain supranational economic activities unaccounted which is a natural problem in subnational income accounting, it nevertheless has the advantage of being a consistent estimator of growth across all districts, which is an important requirement for distributive purposes under study. Economists 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 can't benefit from this affordable and near-real time data source.

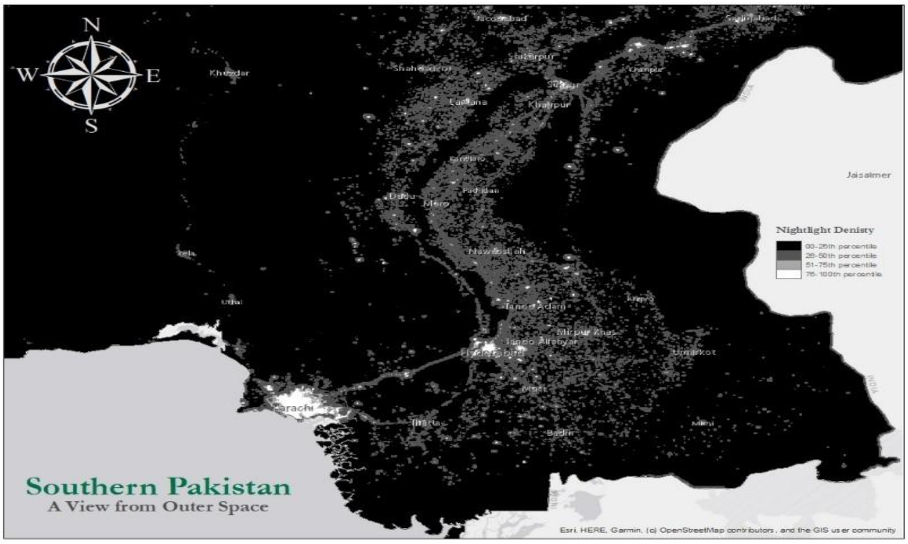
The estimation of regional RDP is an important exercise for growth-oriented regimes that devise part of whole of their economic policies at the subnational/provincial level. There are three broad reasons why estimation of district level GDP may not only be a valuable indicator of economic growth, but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

1. RDP allows provinces with economic growth as 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 as to be replicated, and what policies can be shelved as ineffective. A realization of the DDP can lead to an institutionalized data-driven dialogue between interlinked provincial departments, i.e. Planning & Development and Ministry of Finance.
2. RDP can aid Intergovernmental resource distribution bodies such as the National Finance Commission (NFC) and Provincial Finance Commissions (PFCs) in developing a formulaic distribution criteria based on the RDP and provincial GDP. This also creates a performance yard stick for provincial governments to bargain with the center for an increase in the share of the financial award.
3. RDP provides disaggregated market signals to private investors, especially to Foreign Direct Investment (FDI). A system of RDP is indicative of a serious government reform effort towards attractive private finance. In Pakistan's case, and in the case of all other

middle income, high poverty (MIHP) countries, the same logic also applies to foreign aid.

Lack of economic data at lower administrative levels can potentially constrain the efficient planning and utilization of public finances, impeding economic development. On the other hand, an equitable and accountable distribution of public resources can improve democratic practices and provision of public goods in Pakistan. The remainder of this report will proceed as follows: a review of literature will be presented to qualify the exercise at hand, followed by a description of the data used and acknowledgement of its sources; this will be followed by a discussion of the key methodological issues in calculating the RDP for cities in Pakistan for the period 2012-2020.





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 source is night-time luminosity, also known as nightlight density. Henderson, Storeygard and Weil (2012) show that night-lights data are an accurate proxy for economic activity. Lee (2016) used nightlights density to predict economic activity in North Korea. Kudamatsu, Persson, and Stromberg (2016) merge 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 night-lights data for measuring RDP comes from Perez-Sindin et al. (2021), who benchmark the VIIRS nightlights data against municipal RDP measures prepared by Colombia's national statistics agency data. With a standardized regression coefficient of 0.827 when regressing official RDP on night-lights measures, the authors conclude that night-lights can serve as a good indicator of municipal RDP.

Another source of remotely sensed data involves the mapping of human settlement and urbanization 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 ones.

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 in order to help governments around the world 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-style research, and has been leveraged in many areas of public policy. Kudamatsu, Persson, and Stromberg (2016) use 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) present the relationship between crop choice patterns and land ownership. Maue et al., (2020) use satellite data to estimate yields of smallholder farmers. In urban development literature, Turner, Haughwout, and van der Klaauw (2014) study the impact of land use regulations on urban sprawl using satellite data.

Another innovate application of satellite data is done by Casaburi and Troiano (2016), who use satellite data to identify buildings in an Italitan city that have 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) document the new geography of poverty and show that 49.3% of the world's extremely poor live in eight middle-income high-poverty countries, including Pakistan. The authors argue 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 posits that a key pillar of this invisible infrastructure involved building feedback loops and data systems that make public way 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.

DATA

At the time of authoring this report (12/21), this project uses 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.

Dataset #1 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 uses processed annual composites of the VIIRS NTL for 2012-2020 (Elvidge et al., 2021). A masked annual average of VNL V2 data is 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 use.

Dataset #2 is the World Settlement Footprint (WSF 2019), a global binary raster dataset of on 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 the year 2019. WSF 2019 was released at the United Nations Climate Change Conference (COP26) in November 2021.

Dataset #3 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. WSF Evolution was conjointly released with WSF 2019 at the United Nations Climate Change Conference (COP26) in November 2021.

Dataset #4 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 are sourced from the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA), among other sources.

METHODOLOGY

The methodology for calculating city-level RDP can be divided into two broad categories: The first 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 upon 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, and would simply look at these facto spread of human settlement.

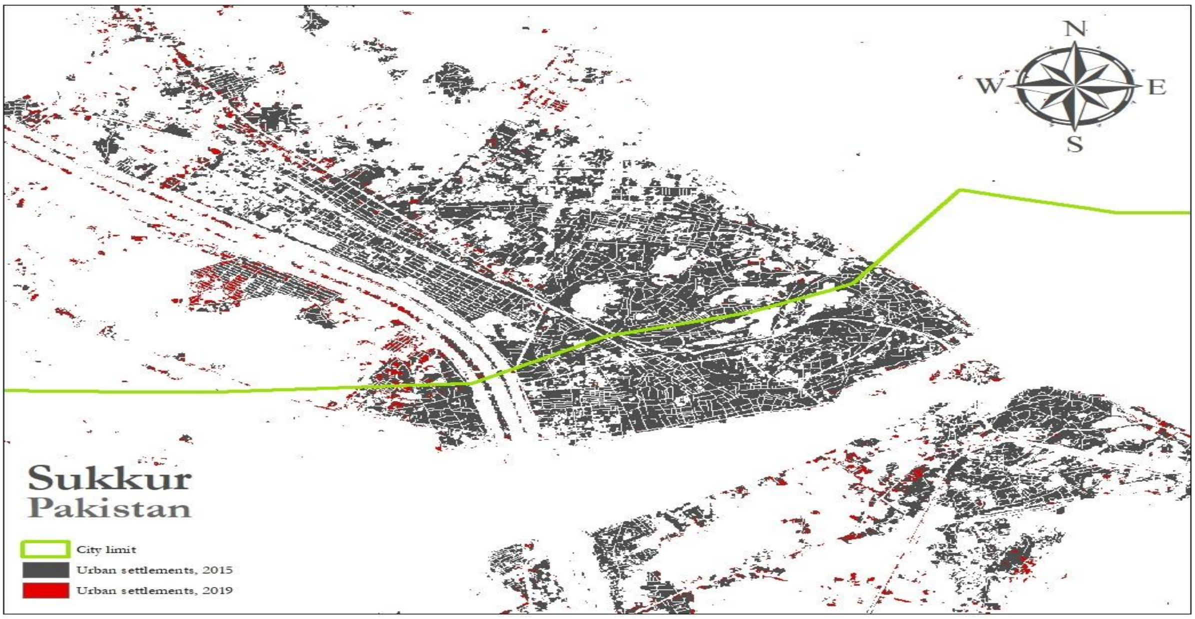
Since both definitions of city limits are valuable for research and policymaking, choosing one over the other would not be pareto optimal. With this in mind, this study develops growth estimates for two definitions. The first is the de jure method, which looks at growth in economic activity within legal city boundaries, as defined in dataset #4. The second definition uses the richness of WSF 2019 and 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 a number of complex tradeoffs along quality, coverage and time dimensions. Satellite data is 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 time series, and can allow the researcher 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 provides data 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 years at very granular levels.

Figure 1: De Jure Vs De Facto City Limits of Gujranwala City (WSF 2019)

Figure 2: Evolution of Urban Development between 2015 and 2019 in Sukkur, Sindh



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. Prior to 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, I conduct the benchmarking exercise by comparing Pakistan to a bundle 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. In order to benchmark these to national accounts, the table also presents GDP per capita (World Bank WDI) figures for all countries. The third column presents share of NTL per million residents, which normalizes NTL data by population and hence allows for a more analogous estimate to GDP per capita.

Table 1 provides key insights about 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 estimate (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 scase of Egypt and Mexico, both of which have large per capita differences in national accounts data, 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 centers in developing agglomeration effects, which are captured both in national accounts data as well as NTL-based measures. The large elasticity of NTL- based measures in Mexico and Egyptmay be capturing urban agglomeration effects that industrial cities like Cairo or Mexico city enjoy in 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 twelve and twenty four, 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	1,353.20	1	6126.1	1	532.5	1
Afghanistan	2,078	0.43	89	0.07	2285.9	0.37	185.4	0.35
Bangladesh	5,138	1.07	264.4	0.2	1605.4	0.26	242.3	0.45
Colombia	14,931	3.1	943.4	0.7	18541.1	3.03	2842	5.34
Egypt	12,607	2.62	3,576.90	2.64	34953.3	5.71	6,601.2	12.4
India	6,503	1.35	11,600.00	8.57	8393.3	1.37	636.2	1.19
Mexico	18,444	3.83	5,039.10	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 the purposes of this report, Table 1 provides sufficient support for NTL-based measures of 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 of GDP contribution by each province as estimated by the NTL- based methodology.

Table 2: 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	1,234	1	6126.1	1	532.5	1	38000	182000
Punjab	671	0.54	6098.1	0.99	319.2	0.6	20656	188000
Sindh	332	0.27	6927.2	1.13	532.5	1	10209	213000
KP	101	0.08	2847.6	0.46	196.5	0.37	3113	102000
Balochistan	61	0.05	4944.7	0.81	237	0.45	1878	152000
Islamabad CT	41	0.03	20249.7	3.31	75.1	0.14	1249	624000
AJK + GB	29	0.02	6236.6	1.02	10.6	0.02	893	190000

Punjab:

As can be seen in the first set of results, Punjab province leads with an estimated 54 percent contribution to Pakistan's GDP. When normalized by population, per million estimate of GDP is 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 province's highest unit of NTL per cell is 60 percent of the highest national NTL per cell, which is located in Karachi.

Spatial disaggregation of economic activity by province also allows me to impute the per capita Regional Domestic Product (RDP) at the provincial level. With an estimated 54 percent contribution to national GDP of Pakistan (PKR 38 trillion, 2019), I impute a per capita income of PKR 188,000 in Punjab province. This is very close to the national per capita income, which is PKR 182,000.

Sindh:

The second largest contributor to Pakistan's GDP based on NTL-based GDP estimates, Sindh contributes to 27 percent of Pakistan's overall GDP. When normalized by population, Sindh's NTL per million is higher than the national level by thirteen percent. This implies that a segment of Sindh's population is 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 is PKR 213, 000.

Khyber Pakhtunkhwa:

As per NTL-based RDP estimates, Khyber Pakhtunkhwa province contributes 8 percent to the national GDP of Pakistan. After normalizing for population, the NTL per million estimate is 46 percent of the national NTL per million estimates. This 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 include 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 district in KP have one of the fastest growing young populations in the country. Moreover, the biggest low hanging fruit in terms of income gains in KP lie in the newly merged tribal districts which currently have a largely subsistence economy.

Balochistan:

NTL-based estimates show that Balochistan contributes approximately eight percent to the national economy of Pakistan. At 0.81 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. 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 Islamabad Capital Territory show the first empirical evidence so far of the high importance of cities and urban centers for the purpose of economic development. Despite having a very small industrial base and an economy largely centered around the seat of government, Islamabad's NTL per million estimate is more than three times the overall national mean. Moreover, at PKR 624, 000 the imputed per capita income for the city is also more than three times the national per capita income. This estimate provides motivation for a NTL-based City Development Product (CDP) to compare how cities fare in the distribution of economic activity and its growth over time.

DISTRICT RDP: UNDERSTANDING VARIATION WITHIN PROVINCES

A similar exercise can be conducted for all districts of Pakistan in order to understand the regionality of economic activity within provinces. Table 3 presents descriptive statistics of the most productive districts across Pakistan. Not surprisingly, all are districts where an overwhelming majority of the population lives in urban settings.

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 have significantly higher per capita incomes. That the same is the case is true for cities like Lahore and Karachi is no surprise, the significant income differences for smaller cities points to the important agglomeration effects induced by smaller cities. It is worth remembering that compared to their larger counterparts, local governments for 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.

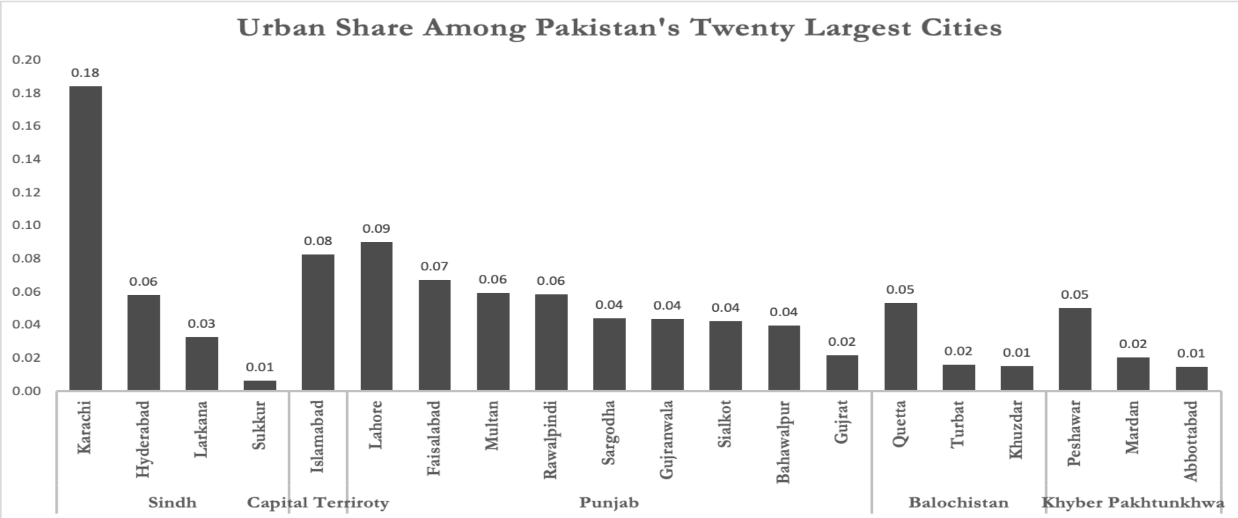
Table 3: Districts of Pakistan with highest VIIRS VNL V2 nightlights, 2019 (Elvidge et al., 2021). Population data is from PBS Census, 2017. GDP data taken from World Bank's World Development Indicators (WDI)

Rank	District	NTL (Aggregate)	NTL Per Million	Imputed RDP	
		Abs. (000s)	Abs.	RDP (PKR, Bil.)	P.C. RDP (PKR)
1	Lahore	87,185	7840.4	2685	242000
2	Rawalpindi	51,504	9533.6	1586	294000
3	Faisalabad	47,776	5934.2	1471	187000
4	Karachi (N=10)	45,238	5934.2	1393	724000
5	Islamabad	40,567	5934.2	1249	624000
6	Kasur	29,481	8533.245	908	263000
7	Rahim Yar Khan	28,165	5858.304	867	180000
8	Sheikhupura	27,772	8026.621	855	247000
9	Multan	27,423	5777.856	845	178000
10	Attock	26,207	13892.85	807	428000

CITY-LEVEL DEVELOPMENT PRODUCT: THE TWENTY LARGEST CITIES

In this section, I present 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 contributes 18 percent of the urban economic output. By magnitude, Karachi is by far the largest urban center of economic activity, followed by Lahore, Faisalabad and Islamabad at approximately 7 percent each. After accounting for the mega cities 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 of small and medium sized manufacturing clusters concentrated in each of these cities.

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 neighboring 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 my analysis.

CDP Growth, 2013 to 2020

So far, we have developed an understanding of the spatial distribution of economic activity across Pakistan's provinces, districts and cities. I now explain if these patterns result in differential rates of growth. The case presented here is for the same twenty cities presented in the previous section. As the largest urban population centers, 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 contemporary economic growth in Pakistan. Figure 4 below presents the growth rates for each city, along with its share of CDP among the pool of twenty cities.

Figure 4: CDP Growth Rates for Pakistan's Twenty Largest Cities

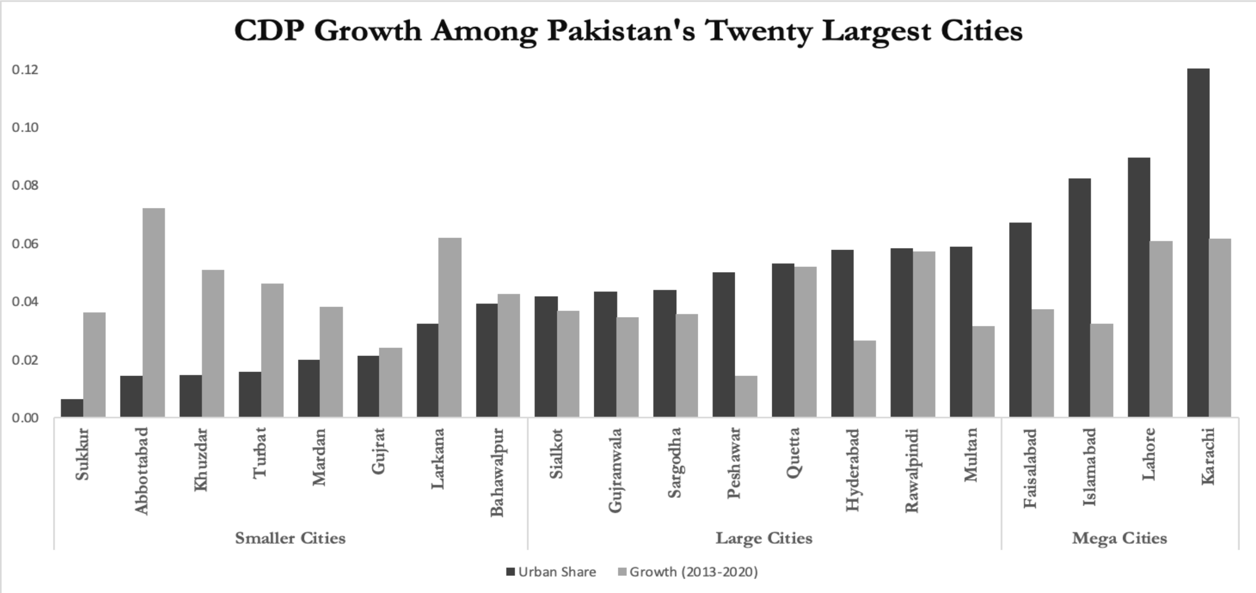


Figure 4 shows has many important insights. First, Pakistan's cities are growing at a fast pace. Second, this growth is not restricted to the "mega" cities like Karachi and Lahore. In fact, growth rates among smaller cities and medium-large cities are much higher than some of the mega cities. In order 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 census of Pakistan.

Next Steps

At the time of this report, data processing of NTL data has nearly been completed. Shapefiles for de facto city limits have been developed for 17 cities, and the set of de jure city limit shapefiles has been acquired. For NTL data, the PI has fully processed the VIIRS annual composites for all available years. Once the two datasets have been prepared, the process of compiling city level RDP estimates will be carried out, leading to the CDP dataset.

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 centered around macro-development in Pakistan. Over the course of this project, a number of natural extensions of this project have emerged. These include merging the dataset geographically with survey data from PSLM and from Census 2017. In order 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 the Pakistan's research dataverse.

BIBLIOGRAPHY

Asher, Sam, et al. "The socioeconomic high-resolution rural-urban geographic dataset on India (SHRUG)." Harvard Dataverse (2019).

Casaburi, Lorenzo, and Ugo Troiano. "Ghost-house busters: The electoral response to a large anti–tax evasion program." The Quarterly Journal of Economics 131.1 (2016): 273-314.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. "Measuring economic growth from outer space." American economic review 102.2 (2012): 994-1028.

Holmes, Thomas J., and Sanghoon Lee. "Economies of density versus natural advantage: Crop choice on the back forty." Review of Economics and Statistics 94.1 (2012): 1-19.

Kudamatsu, Masayuki, Torsten Persson, and David Strömberg. "Weather and infant mortality in Africa." (2012).

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

Maue, Casey C., Marshall Burke, and Kyle J. Emerick. Productivity dispersion and persistence among the world's most numerous firms. No. w26924. National Bureau of Economic Research, 2020.

Page, Lucy, and Rohini Pande. "Ending global poverty: Why money isn't enough." Journal of Economic Perspectives 32.4 (2018): 173-200.

Turner, Matthew A., Andrew Haughwout, and Wilbert Van Der Klaauw. "Land use regulation and welfare." Econometrica 82.4 (2014): 1341-1403.

Pérez-Sindín, Xaquín S., Tzu-Hsin Karen Chen, and Alexander V. Prishchepov. "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 (2021): 100647.

Marconcini, Mattia, et al. "Outlining where humans live, the World Settlement Footprint 2015." Scientific Data 7.1 (2020): 1-14.

Elvidge, C.D, Zhizhin, M., Ghosh T., Hsu FC, Taneja J. Annual time series of global VIIRS nighttime lights derived from monthly averages:2012 to 2019. Remote Sensing 2021, 13(5), p.922, doi:10.3390/rs13050922

Marconcini, M., Metz-Marconcini, A., Esch, T., Gorelick, N. (2021). Understanding Current Trends in Global Urbanisation – The World Settlement Footprint suite. GI_Forum, 1, 33-38.