

# Subnational Income, Growth, and the COVID-19 Pandemic

M. Ali Choudhary, Ilaria Dal Barco, Ijlal A. Haqqani, Federico Lenzi, and Nicola Limodio 

## ABSTRACT

Using real-time data and machine-learning methods, we produce monthly aggregates on gross national income (GNI) for 147 Pakistani districts between 2012 and 2021. We use them to understand whether and how the COVID-19 pandemic affected the growth and subnational distribution of income in Pakistan. Three findings emerge from our analysis. First, districts experienced a sizable decline in income during the pandemic, as their monthly growth rate dropped on average by 0.133 percentage points. Second, a larger income drop took place in districts with a higher COVID-19 incidence, which correspond to urban areas characterized by a higher population density. Third, COVID-19 caused a decline in income inequality across districts, with richer districts experiencing more negative income growth during the pandemic.

JEL classification: E02, O11, O47

Keywords: COVID-19 pandemic, growth, satellite data

## 1. Introduction

The COVID-19 pandemic has produced dramatic changes around the globe, as the effects of the virus and government containment policies have disrupted our societies and economies starting from February

M. Ali Choudhary is a professor of Economics and Public Policy at Loughborough Business School and the State Bank of Pakistan, I.I. Chundrigar Road, Karachi, Pakistan, and the Centre for Economic Performance, 32 Lincoln's Inn Fields, WC2A 3PH, London, UK, and the Loughborough Business School, Sir Richard Morris Building, Loughborough University, Epinal Way, Loughborough, Leicestershire, LE11 3TU; his email addresses are [ali.choudhary@sbp.org.pk](mailto:ali.choudhary@sbp.org.pk) and [ali.choudhary@lboro.ac.uk](mailto:ali.choudhary@lboro.ac.uk). Ilaria Dal Barco is Predoctoral Associate at Bocconi University, Via Roentgen 1, 20136 Milan, Italy; her email address is [ilaria.dalbarco@unibocconi.it](mailto:ilaria.dalbarco@unibocconi.it). Ijlal A. Haqqani is a economist at the State Bank of Pakistan, I.I. Chundrigar Road, Karachi, Pakistan; her email address is [ijlal.ahmad@sbp.org.pk](mailto:ijlal.ahmad@sbp.org.pk). Federico Lenzi is a PhD student from the Northwestern University, Kellogg School of Management, 2211 Campus Drive, Evanston, IL 60208; his email address is [federico.lenzi@kellogg.northwestern.edu](mailto:federico.lenzi@kellogg.northwestern.edu). Nicola Limodio (corresponding author) is a associate Professor of Finance at Bocconi University, Department of Finance, BAFFI CAREFIN and IGIER, Via Roentgen 1, 20136 Milan, Italy; his email address is [nicola.limodio@unibocconi.it](mailto:nicola.limodio@unibocconi.it). The research for this article was financed by the International Growth Center and the State Bank of Pakistan. The authors would like to thank the members of the Monetary Policy Committee of the State Bank of Pakistan, namely, Governor Reza Baqir, Murtaza Syed, Jameel Ahmed, Asad Zaman, Naved Hamid, Azam Faruqee, Hanid Mukhtar, and Tariq Hassan, for the encouragement to find innovative data sources to measure economic growth during the COVID-19 pandemic. The authors also express their gratitude to Minister Asad Umar, Ijaz Nabi, Brigadier Saeed, Lieutenant Colonel Adnan, Major Sami and his team, Imtiaz Ahmed, Kashif Sahazad, and Nadeem Hanif. The authors would also like to acknowledge the useful suggestions of Giorgia Barboni, Johannes Boehm, Nicola Gennaioli, Thiemo Fetzter, Federico Rossi, Chris Roth, Nicolas Serrano Velarde, Tom Schmitz, and participants at various conferences, seminars, and workshops. A supplementary online appendix is available with this article at *The World Bank Economic Review* website.

2020. While there is ample knowledge on how high-income countries and their subnational units reacted (Chetty et al. 2020; Woloszko 2020; Chen et al. 2020; Delle Monache, Emiliozzi, and Nobili 2021; Giannone, Paixão, and Pang 2022), the same level of analysis and evidence is lacking in low-income countries because of the scarce availability of recent data.

This paper examines the effect of the COVID-19 pandemic on subnational income in a low-income country, Pakistan, and its districts. However, one of the main challenges was the lack of real-time data on gross national income (GNI). To compensate for this, we gathered real-time data from a variety of sources that could, a priori, have a potential relationship with economic activity. We therefore developed a machine-learning algorithm to now-cast the GNI covering 147 districts in Pakistan from 2012 to 2021 by combining traditional administrative data with night-lights and other satellite data. Our approach builds on the frontier in this literature by relying on multiple satellite products (Asher et al. 2021; Ch, Martin, and Vargas 2021; Beyer, Hu, and Yao 2022), integrating our prediction exercise with a robust empirical inference (Athey 2017), and connecting micro and macro data (Vavra 2021).

Three novel findings contribute to the current debate on the effects of the COVID-19 pandemic and containment policies on the economy. First, we observe that Pakistani income slowed down during the pandemic, when the growth rate dropped on average by 0.133 percentage points. Second, there exists a robust and negative correlation between the incidence of COVID-19 (cases, deaths, and recoveries), which hit urban and densely populated areas more aggressively, and income growth. Third, we observe that the previous two effects led to lower income inequality across districts due to higher-income districts experiencing higher negative income growth during the pandemic. This creates a sort of “convergence to the bottom”—albeit only temporary. Our findings bring a perspective to the latest literature on convergence (Patel, Sandefur, and Subramanian 2021; Kremer, Willis, and You 2022; Pande and Enevoldsen 2021; Acemoglu and Molina 2021), and in particular to the economic effects of COVID-19 on growth and inequality in low- and middle-income countries. Our monthly data set suggests that conditional convergence has been taking place in Pakistan since before the pandemic. In fact, districts with a lower level of income in 2012 have been growing faster than districts with higher incomes before the pandemic. During the COVID-19 period, the gap between the growth trajectories of districts seems to have further diminished. However, a key distinction exists between pre-pandemic and pandemic-induced periods. While the hypothesis of convergence before the pandemic is supported by higher growth across low-income districts and may represent a permanent move toward a new steady state, during the pandemic, growth dynamics are mainly governed by high-income and urban districts slowing down the most. This suggests it may only be a temporary shock rather than an effective convergence to the bottom. These results parallel the findings of Gupta, Malani, and Woda (2021) in India, who use representative panel data on household finance and consumption instead of satellite now-casting. Our work is also connected to the literature on growth in regions and regional convergence (Gennaioli et al. 2013, 2014; Ganong and Shoag 2017; Lessmann and Seidel 2017; Giannone, Paixão, and Pang 2022; Giannone unpublished manuscript; Hsieh and Moretti unpublished manuscript) showing two key determinants of the recent pandemic-induced recession: urbanization and COVID-19 incidence. We highlight the differential effect of the pandemic across urban and rural districts, and we find that this heterogeneous incidence is mainly due to the high density of population in cities. Our results are also in line with Moeen et al. (2021), who show that the service sector was the most hit by COVID-19, followed by industry, while agriculture was only lightly affected. Moreover, they find that richer households lost more than poorer ones and that urban districts were more affected than rural ones. These findings corroborate the ones in this paper and, in particular, highlight that shocks to manufacturing and highly productive districts can create long-term effects on investment and productivity, in line with the work of Choudhary and Limodio (2022). Second, our results are in line with the work of Saez and Zucman (2016) showing that inequality declines during recessions, though this specific case may be due to a decline in contact-intensive activities in urban centers (Koren and Pető 2020) rather than financial returns. In this respect, our results are aligned with the findings of Deaton (2021) on

the lower pandemic-induced within-country inequality and the recent World Bank report suggesting that years of poverty eradication vanished in a few months.<sup>1</sup> Finally, this paper contributes to an emerging literature in macro-development assessing the effects and costs of COVID-19 on low- and middle-income countries (Alfaro, Becerra, and Eslava 2020; Alon et al. 2020; Gottlieb et al. 2021b,a).

The remainder of the paper is organized as follows: The next section introduces some key papers in this literature; in the Section “Data and Methodology”, we illustrate the data gathering procedures and methodology; in the Section “Results”, we present the main results; and a technical guide on the employed algorithms can be found in the supplementary online appendix. Finally, the Section “Conclusions” offers some concluding remarks.

## 2. Related Literature

As stressed by the Bank for International Settlements (Tissot et al. 2020), the current crisis has called into question the traditional statistical aggregates. The constant mutations of the virus result in a rapidly escalating framework where the economic impact varies heterogeneously among sectors and geographic areas. Standard statistical aggregates are often available at the national level, with several months of delays. For this reason, the literature exploring novel sources of data is rapidly expanding.

Chetty et al. (2020) is an important contribution in this field. Exploiting real-time and granular data on American companies, it tracks the crisis’ impact on consumption and the labor market. Through a different approach, Woloszko (2020) proxies them from Google Trend and now-casts the national GDP for 46 OECD and G20 countries. A wider approach is proposed by Chen et al. (2020), integrating search queries with electricity and unemployment data. Following a similar approach, Delle Monache, Emiliozzi, and Nobili (2021) build a weekly economics index for Italy through granular administrative data. Using a social accounting matrix multiplier, Moeen et al. (2021) assess the impact of COVID-19 on macroeconomic variables in Pakistan. Their study reveals a 26.4 percent decline in GDP from mid-March to the end of June 2020, with services experiencing the most significant losses (17.6 percent).

Similar studies are not reproducible in emerging markets with a structural deficiency of administrative data and low Internet penetration. To overcome this obstacle, a growing number of researchers are referring to satellite data (see Donaldson and Storeygard (2016) and Nagaraj and Stern (2020)). This novel source of information is available at a very granular level for the entire globe and almost in real time. Following this literature, Beyer, Franco-Bedoya, and Galdo (2021) combine VIIRS night-lights and electricity consumption to monitor the pandemic impact in India. This study shows that the drop in habitual activities persists after the restrictions’ lifting. It also suggests that the pandemic particularly affects the manufacturing and in-migration areas, while the out-migration areas seem to experience a reduced decline. Also, the work of Roberts (2021) obtains similar results, using night-lights to study COVID-19’s impact on Morocco.

In this literature, the work of Henderson, Storeygard, and Weil (2012) has popularized the use of night-lights as a popular proxy for economic development in emerging markets. Nevertheless, the recent findings of Asher et al. (2021) cast some shadow on their effectiveness in time-series analysis, since their elasticity with the local output varies according to the level of aggregation and the context. In other words, night-lights can be correlated to several development indicators, and discerning what they are proxying in different regions is difficult. Some papers overcome this issue by adopting different and more detailed proxies for local economic output. Engstrom, Hersh, and Newhouse (2017) proves that the extraction of daytime features from satellite data explains 60 percent of average log consumption in emerging markets;

1 Refer to “Updated Estimates of the Impact of COVID-19 on Global Poverty: Turning the Corner on the Pandemic in 2021?” by D. G. Mahler, N. Yonzan, C. Lakner, R. A. Castaneda Aguilar, and H. Wu, published on 24 June 2021, on the World Bank Data blog and available at <https://blogs.worldbank.org/opendata/updated-estimates-impact-covid-19-global-poverty-turning-corner-pandemic-2021>.

Jain (2020) shows that all the satellite data hide implicit biases (for example, clouds, saturation, non-random misclassification, meteorological variables) in the realization process, whereas Burke et al. (2021) specify that the errors attributed to these models tend to be overestimated and related to the low-quality administrative data used as reference.

Our work includes insights from this literature: we use satellite lights in line with Henderson, Storeygard, and Weil (2012), but consider different moments of these series and include other real-time data sets as discussed by Asher et al. (2021). In addition to these data sets, we also partnered with local electricity providers to build a district-level electricity data set, in line with Beyer, Franco-Bedoya, and Galdo (2021), and add a finer split between electricity for domestic, commercial, and industrial use.

### 3. Data and Methodology

This section provides an overview of the data and illustrate the methods used in the analysis.

#### 3.1. Data

For our analysis, many different databases have been used. In particular, we collected data from various sources to account for as many variables as possible that might be relevant predictors of economic activity.

Pakistan is among the low- and middle-income countries offering the most detailed and extensive administrative data. Most of these resources are available in traditional wide economic macro-aggregates, while micro-aggregates are often available only upon request. The principal statistical publications are released by the State Bank of Pakistan, the Ministry of Finance, and the Pakistani Bureau of Statistics. From the latter, we used three sources: (a) the annual Pakistan Economic Survey (PES), which contains an extensive set of variables such as wages, doctor fees, and import/export of cargo; (b) the Monthly Bulletins of Statistics, which report price indexes for over 400 items at the city-month level; and (c) the Survey on COVID-19, which provides information on migratory movements during the pandemic. Additionally, the National Electric Power Regulatory Authority produces granular data on electricity consumption at the tehsil-month level, disaggregated by various destination uses (commercial, domestic, industrial, and others).

In accordance with the growing literature on satellite imagery, we included VIIRS night-lights as a potential predictor of economic activities. In particular, we opted for the high-frequency and less pre-processed VNP46A1-VIIRS/NPP Daily Gridded Day Night Band 500m Linear product. Furthermore, to capture economic activities beyond those directly associated with artificial lights, such as agriculture, we incorporate a far wider range of potential economic activity detectors. Among these, there are weather-related variables such as sun hours, temperature, and humidity, which are significant inputs in agricultural production. Simultaneously, we included a wide range of satellite data from NASA that encompass measurements related to vegetation health, fires, and cloud characteristics, including presence, thickness, and water content, as well as the detection of pollutants such as ozone, CO<sub>2</sub>, and NO<sub>2</sub>. All these variables could predict the quality of the harvest, influence agricultural yields, or be a good proxy for industrial activity. Considering the results of Asher et al. (2021), we also measure the primary destination use of land using the yearly landcover map produced by the European Space Agency (ESA) under the Climate Change Initiative (CCI).

The provincial-level aggregates for gross national income in Pakistan are taken from the United Nations' Sustainable Development Goals data set. This data set combines country-level information with periodic household surveys. The compilation of these data is carried out by the Global Data Lab, hosted by the Nijmegen Center for Economics (NiCE) at Radboud University in the Netherlands. Figures are available in 2011\$ PPP up until 2018. We decompose the provincial aggregate at the tehsil level using the share of population obtained from the United Nations' WorldPop platform. To further validate this decomposition, we use an indicator for living standards provided by the United Nations Development

Program (UNDP) in 2017. We also collect the sectoral composition of GDP for the main four provinces of Pakistan (Punjab, Sindh, Balochistan, and Khyber Pakhtunkhwa), provided by the Institute for Policy Reforms.

Finally, statistics on COVID cases, deaths, and recoveries were produced at the district level by provincial authorities and released upon our request. Due to institutional factors, it was not possible to obtain COVID data at the tehsil level; hence, we decided to conduct the analysis at the district-month level. Lastly, the number of available health facilities is retained from the Humanitarian Data Exchange of the United Nations Office for the Coordination of Humanitarian Affairs (OCHA).

Supplementary online appendix S1 offers a more schematic and complete list of all the specific data sets used and their respective sources.

### 3.2. GNI Prediction

In many low- and middle-income countries, such as Pakistan, data are usually scarce and slow to become available. The main challenge we had to face for this analysis was the lack of real-time disaggregated data on GNI. To remedy these shortcomings, we assigned to each tehsil a fraction of the province's GNI proportional to its population and we gathered real-time data from a variety of sources that could a priori have a potential relationship with economic activity. We therefore obtained a data set at the tehsil-year level that we can use to produce monthly estimates for GNI after 2018. These are then aggregated at the district-year level to produce a statistical analysis of income growth in relation to the local incidence of the COVID-19 pandemic.

We do not assume a specific *ex ante* relationship between income and any of the potential predictors; instead, we employ a set of machine-learning (ML) algorithms that operate through supervised learning. Supervised learning implies that algorithms need to be trained on a set of already-existing data before being able to make predictions. GNI data in our setting are available yearly up until 2018; therefore, for the training we use a data set with observations at the tehsil year level from 2012 to 2018. The potential real-time GNI predictors in our data set exhibit significant heterogeneity in their ranges. To mitigate the potential dominance of certain variables solely based on their scale, we employ a standard technique called min-max normalization. This normalization allows us to rescale all the variables between 0 and 1, ensuring a fair comparison and preventing undue influences based on the original ranges. We opt for this normalization over standardization, as we do not have any prior knowledge of the underlying distributions of the variables.

We focused on a total of five classes of algorithms: elastic net, random forest, bagging, boosting, and support vector machines. In order to assess which one performs better, it is necessary to split the data set into a train sample (comprising 75 percent of observations) and a test sample (representing the remaining 25 percent of observations). This division allows the algorithm to learn from the training sample, identifying which variables should be considered and their respective predictive powers. Subsequently, we can use the trained algorithm to make predictions using the test sample, enabling a comparison between the predicted values and the actual data. The best result is achieved by the bagging algorithm, with an overall mean square error of 0.01506. Supplementary online appendix S2 provides supplementary details on the methodology for more in-depth reference, including the mean square error values for all the algorithms.

Once we had identified the best-performing algorithm, we employed it to generate predictions for the years following 2018. Furthermore, as we wanted to carry out the analysis at the month-year level, to produce the estimates, we employed a data set in which predictors were disaggregated at the tehsil-month-year level. With this approach, we were able to obtain GNI predictions for each tehsil-month, accomplishing two objectives simultaneously: on the one hand, we were able to obtain real-time figures for GNI in Pakistan; on the other hand, we could develop a data set that is both geographically and temporally disaggregated.

Once estimates at the tehsil level were produced, we collapsed the data set at the district level to carry out the analysis. To verify that our data are correctly attributed to each district, despite initially obtaining the geographical disaggregation using the proportion of population, we use the Human Development Index provided by the United Nations Development Program (UNDP). This indicator is available for 2017 for each Pakistani district, and one of its components is a measure of living standards, calculated through a survey on the living conditions of households. Therefore, we decomposed provincial GNI in 2017, taking into account the different living conditions in each district. The correlation between this newly disaggregated measure of GNI and our estimated GNI is very high, 0.98, and statistically indistinguishable from unit, as graphically represented by [fig. S2.2](#) in Supplementary online appendix S2.

### 3.3. Empirical Analysis

The primary aim of our analysis is to investigate the role of COVID-19 on GNI growth in Pakistan. To explore this, we employ the following empirical model:

$$\text{growth}_{dmy} = \alpha + \beta \text{Covid19}_{dmy} + \varepsilon_{dmy}, \quad (1)$$

where  $\text{growth}_{dmy}$  represents the GNI growth rate of district  $d$  in month  $m$  and year  $y$  and is regressed on four different COVID-19 indicators. The four measures of COVID incidence are (a)  $\text{Covid19}_{my}$ , a dummy variable which takes value 1 from May 2020 onward for all districts in Pakistan; (b)  $\text{Cases}_{dmy}$ , the natural logarithm of the number of COVID-19 cases in district  $d$  during month  $m$  of year  $y$ ; (c)  $\text{Deaths}_{dmy}$ , the natural logarithm of the number of COVID-19 deaths in district  $d$  during month  $m$  of year  $y$ ; (d)  $\text{Recoveries}_{dmy}$ , the natural logarithm of the number of COVID-19 recoveries in district  $d$  during month  $m$  of year  $y$ .

In addition, we augment equation (1) by including an interaction with a dummy variable for urban districts,  $\text{Urban}_d$ , which takes unit value for districts containing at least one of the 20 largest Pakistani cities. The augmented model is therefore as follows:

$$\text{growth}_{dmy} = \alpha + \beta \text{Covid19}_{dmy} + \gamma \text{Covid19}_{dmy} \times \text{Urban}_d + \delta \text{Urban}_d + \varepsilon_{dmy}. \quad (2)$$

The purpose of this specification is to extrapolate and analyze the differential impact of COVID-19 between urban and rural districts. Even in this case, we use the four different measures of COVID incidence and standard errors are clustered at the district level.

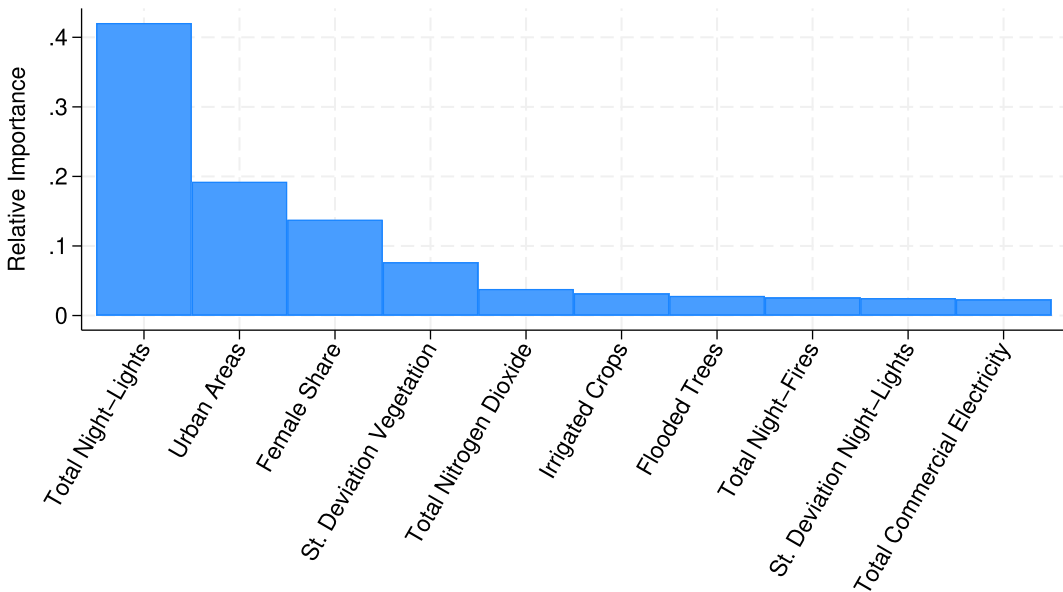
Finally, to further explore the dynamics of income growth in Pakistani districts, we examine the differential effects of COVID-19 based on the starting economic condition. We explore the following empirical model:

$$\text{growth}_{dmy} = \alpha + \beta \text{Income}_{d2012} + \gamma \text{Covid19}_{my} + \delta \text{Income}_{d2012} \times \text{Covid19}_{my} + \varepsilon_{dmy}. \quad (3)$$

The income growth of district  $d$  in month  $m$  and year  $y$  is regressed on the dummy for COVID that takes value 1 from May 2020 onwards,  $\text{Income}_{d2012}$ , which represents the level of income of district  $d$  in 2012, and the interaction between these two variables. As before, standard errors are clustered at the district level. Across all of these specifications, the standard errors are clustered at the district level.

## 4. Results

[Figure 1](#) presents the variables selected as those with the biggest predictive power for district GNI by the bagging algorithm. Night-lights emerge as the most significant predictor in our analysis, further reinforcing the strong evidence that they serve as a good proxy for economic activity. The second most influential indicator is the presence of urban areas, indicating that cities in Pakistan generally exhibit higher levels of wealth and concentrate a significant portion of the country's production. Another noteworthy predictor is

**Figure 1.** Main GNI Predictors

Source: Authors' analysis using satellite data to produce GNI estimates.

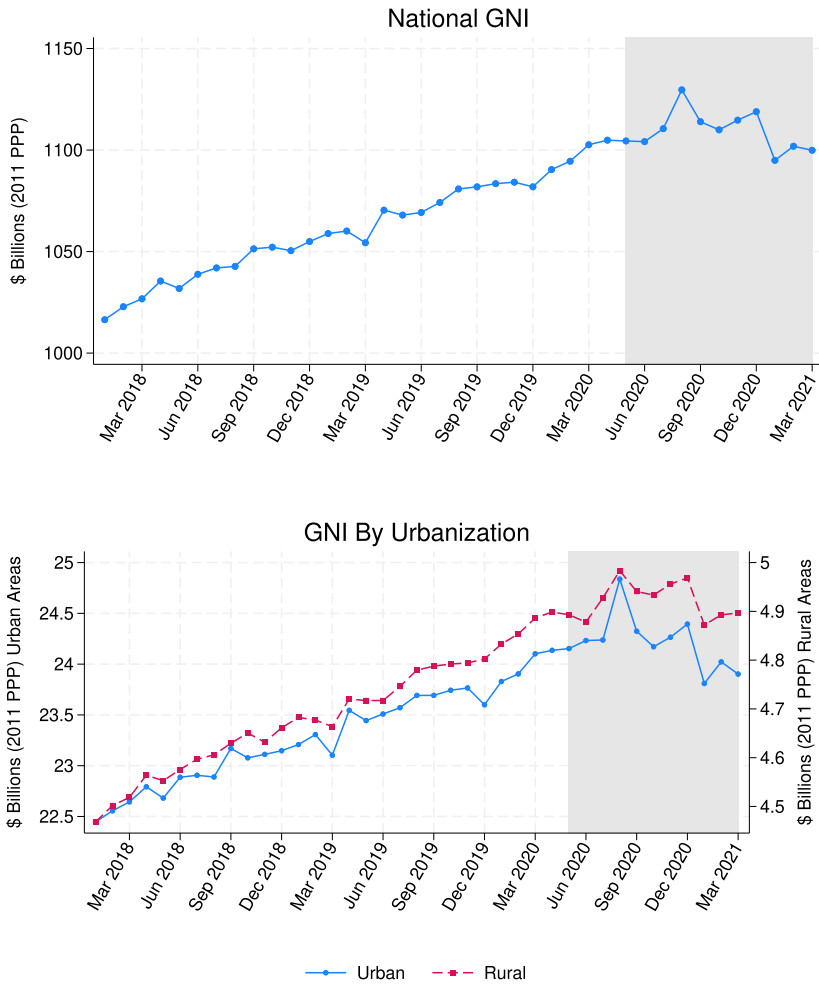
Note: The figure proposes the relative importance of the 10 main predictors of our bagging model. Their influence is rescaled on a 0 to 1 basis for better graphical visualization.

the female share of the population, which ranks third in terms of relevance. This suggests that the gender composition of the population has an impact on the local income dynamics.

The results of our estimation are reported in fig. 2. The top panel displays GNI in billions of US\$ aggregated at country level from 2018 until 2021. It shows that there is a steady growth path from early 2018 to August 2020. The COVID-19 outbreak halts this trend in September 2020, leading to a gradual decline. Figure S3.3 in the supplementary online appendix shows a similar picture, with the time range extended from 2012 to 2021. To understand the drivers of this decline, we define a district as being “urban” if it contains one of the top 20 cities by size, as defined by the Pakistani Bureau of Statistics in its 2017 census.<sup>2</sup> As a result, 20 districts are classified as “urban” and the remaining 127 as “rural.” The bottom panel of fig. 2 depicts the GNI trend divided by urban and rural areas: the blue line with dots represents the average GNI in urban districts, while the dashed red line with squares represents the average GNI in rural districts. It highlights that the average GNI of urban districts (on the left y-axis) is four times bigger than the average GNI of rural districts (on the right y-axis). It is important to note that while urban districts exhibit a steep decline in income as the pandemic begins, this decline is much milder for rural districts. This smaller loss might be due to the suspension of high-interaction activities mostly concentrated in cities, or to the strategy of smart lockdown, promptly imposed by the Pakistani government only on certain hot spots across the country. Figure S3.4 in the supplementary online appendix offers a version of fig. 2 in which urban districts are classified differently: in the top panel, only districts containing one of the top 10 cities are classified as urban, while in the bottom panel, urban districts are those with one of the top 50 cities. The results are qualitatively similar, with urban districts exhibiting a steeper decline than rural districts.

2 The list of principal cities established with the 2017 census, is available at <https://www.pbs.gov.pk/content/provisional-summary-results-6th-population-and-housing-census-2017-0> and also [https://en.wikipedia.org/wiki/List\\_of\\_cities\\_in\\_Pakistan\\_by\\_population](https://en.wikipedia.org/wiki/List_of_cities_in_Pakistan_by_population).

Figure 2. Gross National Income 2018–2021

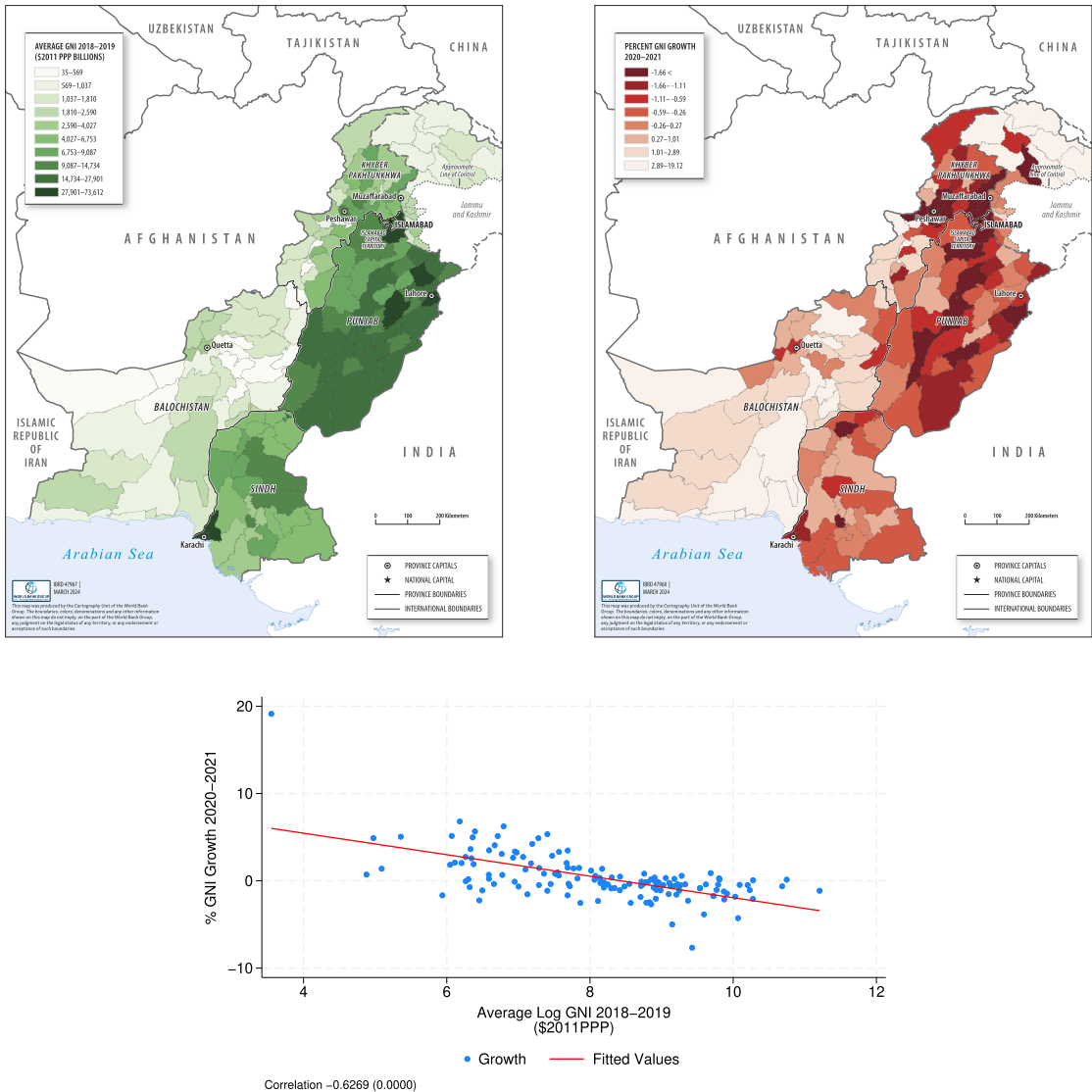


Source: Authors' analysis based on GNI estimates and major Pakistani cities.

Note: The first graph reports the Pakistani Gross National Income from January 2018 to March 2021. The second plot shows the mean income of rural districts in red (right-hand side vertical axis) and urban districts in blue (left-hand side vertical axis). All the values are expressed in billions of 2011 PPP dollars. The gray area indicates the temporal framework covered by our "Dummy COVID": May 2020–March 2021.

In order to explore the spatial distribution of income, fig. 1 reports three pictures. The top-left panel presents a map with the average income per district between 2018 and 2019. In this map, high-income districts are indicated with dark green colors, and it is notable how the concentration of economic activities takes place mainly along the Indus River and the metropolitan areas (Islamabad, Karachi, Lahore, Peshawar, and Quetta). The arid and sparsely populated lands of Balochistan appear to be the poorest, followed by the mountain regions of Gilgit-Baltistan and Khyber Pakhtunkhwa. The top-right panel shows the average income growth between 2020 and 2021. In this case, darker colors indicate a stronger decline (or a smaller increase) in growth. The darkest areas are once again in the densely populated Punjab districts and in major urban areas. By comparing these two maps, it is already clear that districts with high incomes before the pandemic were the most severely hit after the outbreak. The bottom panel of fig. 3 shows exactly this negative correlation between GNI growth during 2020–2021 (on the y-axis) and the log level of GNI in the previous years (on the x-axis). The correlation is  $-0.67$  and statistically different

**Figure 3. GNI and GNI Growth by District**



Source: Authors’ analysis based on GNI estimates

Note: The upper-left panel illustrates the average income of districts during the period between 2018 and 2019. Darker colors represent districts with higher income levels. The upper-right panel displays the percentage variation of income between 2020 and 2021. For 2021, only the first three months are considered due to data availability. Darker colors represent districts with lower growth. The bottom panel presents a graph depicting the district average GNI in 2018–2019 on the horizontal axis and its percentage variation between 2020 and 2021 on the vertical axis. The linear relationship between these variables is shown in red, and the correlation is noted below the graph. All values are expressed in billions of 2011 PPP dollars.

from zero below the 1 percent significance threshold. Figure S3.5 in the supplementary online appendix reports the same descriptive evidence in terms of income per capita: results are similar, including the negative and significant correlation between the pre-pandemic level of income and the growth of per capita income during the pandemic. We prefer to present the analysis with income levels rather than in per capita terms, given that the numbers for the population may not be adjusted based on the incremental COVID-19 mortality. Figure S3.6 in the online appendix shows the overall income growth during the pre-pandemic

**Table 1.** Summary Statistics

Variables	(1) Observations	(2) Mean	(3) St. deviation	(4) 50th p.tile	(5) 5th p.tile	(6) 95th p.tile
Log income	5,733	21.93	1.368	22.03	19.91	23.94
Income growth	5,733	0.297	2.556	0.195	-2.505	3.324
Dummy COVID	5,733	0.282	0.450	0	0	1
COVID cases	5,733	785.3	6,565	0	0	2,073
COVID deaths	5,733	19.91	158.4	0	0	56
COVID recoveries	5,733	681.2	5,946	0	0	1,721

*Source:* Summary statistics of the main variables used for this study.

*Note:* The variable “Log income” represents the logarithm of the district gross income, while “Income growth” is the percentage variation between months. The “Dummy COVID” assumes the value 1 from May 2020 to March 2021. “COVID cases,” “COVID deaths,” and “COVID recoveries” are set to zero for the months preceding the pandemic. The data set follows 147 districts from January 2018 to March 2021.

period from 2012 to 2019. [Table 1](#) presents the baseline summary statistics for most variables included in our empirical analysis.

After these descriptive figures, we want to explore the relationship between income growth and COVID-19 incidence. Panel A of [table 2](#) presents the empirical results of equation (1). Column (1) shows that COVID-19 had a negative and statistically significant effect on income growth between 2018 and 2021. After the COVID outbreak, districts’ growth was on average 0.133 percentage points lower. Similarly, the remaining three columns of panel A show that districts exhibiting a higher incidence of COVID-19 cases, deaths, or recoveries experience lower income growth throughout the period. Columns (2), (3), and (4) show that a 100 percent increase in COVID-19 cases, deaths, and recoveries implies a 0.0216, 0.0485, and 0.0297 percentage-point decline in income growth, respectively.

Panel B of [table 2](#) further investigates whether and how the results of panel A differ across urban and rural districts using the specification of equation (2). Column (1) shows a key result of our analysis. The COVID-19 dummy has a negative effect on growth on average, but this effect is much bigger in urban districts than in rural ones. Before COVID, the difference in income growth between rural and urban districts appeared not to be statistically different from 0. During the COVID period, instead, the growth rate of rural districts declined on average by 0.104 percentage points, while the growth rate of urban districts declined by an additional 0.216 percentage points. Columns (2), (3), and (4) show that when controlling for COVID-19 cases, deaths, and recoveries, there is no difference between urban and rural areas. In other words, for a given number of cases, deaths, or recoveries, the effects do not differ based on the degree of urbanization.

The most straightforward explanation for this is given by the first three columns of [table 3](#), which show the regressions of the logarithm of COVID cases, deaths, and recoveries on the Urban dummy. Looking at the results, it becomes apparent that COVID incidence is much higher in urban districts. This implies that the underlying explanation for the bigger decline in GNI growth that urban districts experience is that they have been hit the hardest by the pandemic. These findings are also in line with the policy of smart lockdown adopted by the Pakistani government, which imposed a partial lockdown only on selected hot spots nationwide. To have a better understanding of the factors that led to cities being impacted the most, we look at the potential drivers of this heterogeneous incidence. In the last three columns of [table 3](#), the COVID variables are regressed on the standardized population density in each district. The findings demonstrate how population concentration has a significant role in the spread of COVID.

In supplementary online appendix [S3](#) it is possible to find a battery of additional tables that verifies the robustness of these results. [Tables S3.2, S3.3, S3.4, S3.5, and S3.6](#) replicate [table 2](#) with some modifications. [Table S3.2](#) includes district fixed effects: while the results employ a different source of variation, namely within-districts only, the magnitudes, signs, and significance of the coefficients are very similar.

**Table 2.** COVID-19 and GNI Growth, 2018–2021

Variables	(1)	(2)	(3)	(4)
	<b>Income growth</b>			
<b>Panel A—Overall</b>				
Covid <sub>my</sub>	−0.133*** (0.0415)	−	−	−
Cases <sub>dmy</sub>	−	−0.0216*** (0.00378)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0485*** (0.00566)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0297*** (0.00416)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.000378	0.000914	0.00168	0.00177
Mean dep. var.	0.297	0.297	0.297	0.297
S.D. dep. var.	2.556	2.556	2.556	2.556
<b>Panel B—Urban</b>				
Covid <sub>my</sub>	−0.104** (0.0472)	−	−	−
Covid <sub>my</sub> × Urban <sub>d</sub>	−0.216*** (0.0595)	−	−	−
Cases <sub>dmy</sub>	−	−0.0189*** (0.00471)	−	−
Cases <sub>dmy</sub> × Urban <sub>d</sub>	−	−0.00952 (0.00659)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0464*** (0.00774)	−
Deaths <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−0.00326 (0.0106)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0280*** (0.00514)
Recoveries <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−	−0.00481 (0.00802)
Urban <sub>d</sub>	−0.0339 (0.0361)	−0.0653 * (0.0350)	−0.0448 (0.0338)	−0.0588* (0.0354)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.000360	0.000696	0.00137	0.00150
Mean dep. var.	0.297	0.297	0.297	0.297
S.D. dep. var.	2.556	2.556	2.556	2.556

Source: Authors' analysis based on Gross National Income (GNI) estimates, major Pakistani cities and data on the COVID-19 pandemic.

Note: Panel A estimates the impact on districts' income growth rates of the pandemic period (column 1), the logarithm of COVID cases (column 2), the logarithm of COVID deaths (column 3), and the logarithm of COVID recoveries (column 4). Panel B repeats the same analysis but decomposes the impact between rural and urban districts. The sample includes all 147 Pakistani districts from January 2018 to March 2021. No fixed effects are included in the analysis. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

**Table 3.** COVID, Urban Areas, and Population Density

Variables	(1) Cases	(2) Deaths	(3) Recoveries	(4) Cases	(5) Deaths	(6) Recoveries
Urban <sub><i>d</i></sub>	1.038*** (0.113)	1.091*** (0.155)	1.146*** (0.209)	–	–	–
Density <sub><i>dmy</i></sub>	–	–	–	0.167*** (0.00906)	0.172*** (0.0101)	0.194*** (0.0141)
Obs.	5,733	5,733	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.00811	0.0270	0.0105	0.0148	0.0588	0.0501
Mean dep. var.	0.144	–1.159	–0.189	0.144	–1.159	–0.189
S.D. dep. var.	3.909	2.270	3.798	3.909	2.270	3.798

Source: Authors' analysis based on population density, major Pakistani cities, and data on the COVID-19 pandemic.

Note: This table estimates the different COVID incidences between urban and rural areas. The dependent variables are the logarithm of COVID cases (columns 1 and 4), the logarithm of COVID deaths (columns 2 and 5), and the logarithm of COVID recoveries (columns 3 and 6). Columns (1), (2), and (3) include, as independent variable, a dummy that takes value 1 in an urban district, while columns (4), (5), and (6) include as independent variable the standardized population density of each district. The sample includes all 147 Pakistani districts from January 2018 to March 2021. No fixed effects are included in the analysis. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

Table S3.3 investigates the determinants of the economic recession; its main message is that districts with a high income share in the service sector appear to be the most hit by the incidence of the COVID-19 pandemic. Table S3.4 includes in the regressions a control for the number of health facilities in each district. This serves to check that results are not driven by the difference in administrative capacity between urban and rural districts. This may be an issue if the allocation of doctors, or health inputs, or bureaucratic skills may be strategically allocated more to poorer-performing districts, as highlighted by Limodio (2021). However, despite this potential threat to identification, we do not observe a change in the key results. Table S3.5 uses two alternative definitions for the urban dummy: in panel A, districts classified as urban are those including one of the top 10 biggest cities, while in panel B, urban districts are those with one of the biggest 50 cities in the country. Even with this different definition of urban, the interpretation of the results does not considerably differ. Finally, table S3.6 uses as a dependent variable the growth rate of income per capita, leading once again to similar results.

Table 4 presents five versions of equation (3): Column (1) provides an indication of conditional convergence in the last decade in Pakistan, by showing that districts with a 1-standard-deviation-lower income in 2012 are growing by 0.03 percentage points more between 2012 and 2021. Column (2) introduces in the regression the COVID-19 dummy and its interaction with the standardized level of income in 2012. With this specification, Income<sub>2012*d*</sub> has a lower but nonetheless negative and significant coefficient. The COVID-19 dummy is negative and statistically different from zero. What is most relevant is, however, the coefficient of the interaction term, as it expresses the effects of the pandemic on the growth trajectories of districts. Districts with a 1-standard-deviation-higher income in 2012 were already growing 0.016 percentage points less. During the pandemic period, they slowed down by an additional 0.14 percentage points. Columns (3), (4), and (5) include district and/or time fixed effects, showing how the result remains quite unchanged regardless of which fixed effects are included in the specification.

To further validate these findings, supplementary online appendix S3 offers two additional tables. First, table S3.7 uses the natural logarithm of the mean income per district in 2012 rather than the standard deviation. The findings remain the same; in particular, the interaction term's coefficient is still negative and significant. Second, table S3.8 adds dummies that, in 2012, assigned each district to a particular income tercile. According to the findings, richer districts (third tercile) are those experiencing a steeper decline in growth, especially during COVID. This result underlines how the already existing differences in growth trajectories across Pakistani districts have been further reinforced by the pandemic, which P hit

**Table 4.** COVID-19 and Growth Trajectories, 2012–2021

Variables	(1)	(2)	(3)	(4)	(5)
			<b>Income growth</b>		
Income 2012 <sub><i>d</i></sub>	−0.0303*** (0.00885)	−0.0164** (0.00731)	−0.0164** (0.00731)	–	–
Covid <sub><i>my</i></sub>	–	−0.137*** (0.0380)	0.00250 (0.0635)	−0.137*** (0.0380)	0.00250 (0.0635)
Income 2012 <sub><i>d</i></sub> × Covid <sub><i>my</i></sub>	–	−0.139*** (0.0449)	−0.139*** (0.0449)	−0.139*** (0.0449)	−0.139*** (0.0449)
District FE	No	No	No	Yes	Yes
Month FE	No	No	Yes	No	Yes
Year FE	No	No	Yes	No	Yes
Obs.	16,170	16,170	16,170	16,170	16,170
Adj. R <sup>2</sup>	7.38e−05	0.000458	0.0118	−0.00627	0.00509
Mean dep. var.	0.324	0.324	0.324	0.324	0.324
S.D. dep. var.	2.600	2.600	2.600	2.600	2.600

Source: Authors' analysis based on Gross National Income (GNI) estimates and data on the COVID-19 pandemic.

Note: Column (1) estimates the impact of the standardized mean income in 2012 on the income growth rate of districts, without controlling for fixed effects. The remaining columns explore how the COVID pandemic influences this relation, controlling for different fixed effects: column (2) includes no fixed effects, column (3) controls for time fixed effects, column (4) controls for district fixed effects, and column (5) controls for both time and district fixed effects. The sample includes all 147 Pakistani districts from January 2012 to March 2021. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

hardly richer districts. It is crucial to emphasize that this evidence does not necessarily signify a decrease in overall inequalities during the pandemic period. While horizontal inequalities, which refer to disparities across districts, may have decreased, vertical inequalities may have persisted. In essence, the correlation between income growth and inequality is not straightforward, as the adverse effects of the pandemic may have disproportionately affected the most vulnerable individuals, particularly those residing in wealthier districts.

## 5. Conclusions

In this paper, we apply a method at the frontier of the machine-learning literature to calculate monthly aggregates on gross national income (GNI) for 147 Pakistani districts between 2012 and 2021 using machine learning and real-time satellite data. Our work shows that Pakistani districts experienced a decline in income growth during the COVID-19 pandemic, as the average monthly growth rate dropped by 0.133 percentage points. We verify that the incidence of COVID-19, measured through cases, deaths, and recoveries, was higher in cities and appears to have a negative and sizable effect on income. Finally, we show that COVID-19 induced a sizable within-country difference in growth patterns, as districts with high pre-pandemic income experienced negative and strong growth during the pandemic. While, on the one hand, this may reduce district inequality and the prominence of urban centers, on the other hand, this process may lower the long-term prospects of the most dynamic Pakistani districts and harm long-term growth.


## Data Availability Statement

Data can be accessed at <https://www.dropbox.com/scl/fo/t2f4rvxwvm9hbbm3pquth/h?rlkey=tk6bs3ml505jf1dnha25v7or0&dl=0>.

## References

- Acemoglu, D., and C. A. Molina. 2021. "Converging to Converge? A Comment." Technical report, National Bureau of Economic Research.
- Alfaro, L., O. Becerra, and M. Eslava. 2020. "EMES and COVID-19: Shutting Down in a World of Informal and Tiny Firms." Technical report, National Bureau of Economic Research.
- Alon, T., M. Kim, D. Lagakos, and M. VanVuren. 2020. "How Should Policy Responses to the COVID-19 Pandemic Differ in the Developing World?" Technical report, National Bureau of Economic Research.
- Asher, S., T. Lunt, R. Matsuura, and P. Novosad. 2021. "Development Research at High Geographic Resolution: An Analysis of Night-Lights, Firms, and Poverty in India Using the Shrug Open Data Platform." *World Bank Economic Review* 35(4): 845–71.
- Athey, S. 2017. "Beyond Prediction: Using Big Data for Policy Problems." *Science* 355(6324): 483–85.
- Beyer, R. C., S. Franco-Bedoya, and V. Galdo. 2021. "Examining the Economic Impact of COVID-19 in India through Daily Electricity Consumption and Nighttime Light Intensity." *World Development* 140: 105287.
- Beyer, R. C., Y. Hu, and J. Yao. 2022. *Measuring Quarterly Economic Growth from Outer Space*. Number 9893. International Monetary Fund.
- Burke, M., A. Driscoll, D. B. Lobell, and S. Ermon. 2021. "Using Satellite Imagery to Understand and Promote Sustainable Development." *Science* 371(6535): eabe8628.
- Ch, R., D. A. Martin, and J. F. Vargas. 2021. "Measuring the Size and Growth of Cities Using Nighttime Light." *Journal of Urban Economics* 125: 103254.
- Chen, S., D. O. Igan, N. Pierri, and A. F. Presbitero. 2020. "Tracking the Economic Impact of COVID-19 and Mitigation Policies in Europe and the United States." *IMF Working Papers* 2020(125).
- Chetty, R., J. N. Friedman, M. Stepner *et al.* 2020. "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data." Technical report, national Bureau of economic research.
- Choudhary, M. A., and N. Limodio. 2022. "Liquidity Risk and Long-Term Finance: Evidence from a Natural Experiment." *Review of Economic Studies* 89(3): 1278–313.
- Deaton, A. 2021. "COVID-19 and Global Income Inequality." Technical report, National Bureau of Economic Research.
- Delle Monache, D., S. Emiliozzi, and A. Nobili. 2021. "Tracking Economic Growth during the COVID-19: A Weekly Indicator for Italy." *Bank of Italy Note COVID-19, January*.
- Donaldson, D., and A. Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30(4): 171–98.
- Engstrom, R., J. S. Hersh, and D. L. Newhouse. 2017. "Poverty from Space: Using High-Resolution Satellite Imagery for Estimating Economic Well-Being." *World Bank Policy Research Working Paper*(8284).
- Ganong, P., and D. Shoag. 2017. "Why Has Regional Income Convergence in the US Declined?" *Journal of Urban Economics* 102: 76–90.
- Gennaioli, N., R. La Porta, F. Lopez-de Silanes, and A. Shleifer. 2013. "Human Capital and Regional Development." *Quarterly Journal of Economics* 128(1): 105–64.
- . 2014. "Growth in Regions." *Journal of Economic Growth* 19: 259–309.
- Giannone, E., N. Paixão, and X. Pang. 2022. "Jue Insight: The Geography of Pandemic Containment." *Journal of Urban Economics* 127: 103373.
- Gottlieb, C., J. Grobovšek, M. Poschke, and F. Saltiel. 2021a. "Lockdown Accounting." *BE Journal of Macroeconomics* 22(1): 197–210.
- . 2021b. "Working from Home in Developing Countries." *European Economic Review* 133: 103679.
- Gupta, A., A. Malani, and B. Woda. 2021. "Inequality in India Declined during COVID." Technical report, National Bureau of Economic Research.
- Henderson, J. V., A. Storeygard, and D. N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102(2): 994–1028.
- Jain, M. 2020. "The Benefits and Pitfalls of Using Satellite Data for Causal Inference." *Review of Environmental Economics and Policy* 14(1): 157–169.
- Koren, M., and R. Pető. 2020. "Business Disruptions from Social Distancing." *Plos one* 15(9): e0239113.
- Kremer, M., J. Willis, and Y. You. 2022. "Converging to Convergence." *NBER Macroeconomics Annual* 36(1): 337–412.

- Lessmann, C., and A. Seidel. 2017. "Regional Inequality, Convergence, and Its Determinants—A View from Outer Space." *European Economic Review* 92: 110–32.
- Limodio, N. 2021. "Bureaucrat Allocation in the Public Sector: Evidence from the World Bank." *Economic Journal* 131(639): 3012–40.
- Moemen, M. S., Z. Haider, S. H. Shikoh, N. Rizwan, A. Ejaz, S. Davies, and A. W. Rana. 2021. *Estimating the Economic Impacts of the First Wave of COVID-19 in Pakistan Using a SAM Multiplier Model*, 2001. Intl Food Policy Res Inst.
- Nagaraj, A., and S. Stern. 2020. "The Economics of Maps." *Journal of Economic Perspectives* 34(1): 196–221.
- Pande, R., and N. T. Enevoldsen. 2021. "Growing Pains? A Comment on 'Converging to Convergence'." Technical report, National Bureau of Economic Research.
- Patel, D., J. Sandefur, and A. Subramanian. 2021. "The New Era of Unconditional Convergence." *Journal of Development Economics* 152: 102687.
- Roberts, M. 2021. "Tracking Economic Activity in Response to the COVID-19 Crisis Using Nighttime Lights—The Case of Morocco." *Development Engineering* 6: 100067.
- Saez, E., and G. Zucman. 2016. "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data." *Quarterly Journal of Economics* 131(2): 519–78.
- Tissot, B., and B. De Beer. 2020. *Implications of COVID-19 for Official Statistics: A Central Banking Perspective*. Irving Fisher Committee on Central Bank Statistics, Bank for International.
- Vavra, J. 2021. "Tracking the Pandemic in Real Time: Administrative Micro Data in Business Cycles Enters the Spotlight." *Journal of Economic Perspectives* 35(3): 47–66.
- Woloszko, N. 2020. "Tracking Activity in Real Time with Google Trends." OECD Economic Department Working Papers 1634

**Supplementary Online Appendix**  
**Subnational Income, Growth, and the COVID-19 Pandemic**  
**M. Ali Choudhary, Ilaria Dal Barco, Ijlal A. Haqqani, Federico Lenzi, and**  
**Nicola Limodio **

## **S1. Data**

In this section of the supplementary online appendix are presented all the data sets employed in the analysis, their sources, the time span, and their level of disaggregation.

### **DATA SET COVID 19 Data**

**AGENCY** Provincial Authorities

**FREQUENCY** Monthly

**LEVEL** District

**TIME** March 2020–May 2021

**URL** Available on request

**VARIABLES** Cases, Deaths, Recoveries

### **DATA SET Earth Observations Group**

**AGENCY** NASA – National Aeronautics and Space Administration

**FREQUENCY** Monthly

**LEVEL** Tehsil

**TIME** January 2012–March 2021

**URL** <https://neo.sci.gsfc.nasa.gov/>

**VARIABLES** Aerosol Thickness, Clouds Fraction, Clouds Optical Thickness, Clouds Particle Radius, Clouds Water Content, Night Fires, NO<sub>2</sub>, Ozone, Sea Chlorophyll, Sea Surface Temperature, Snow on the Ground, Surface Temperature Anomaly Day, Surface Temperature Anomaly Night, Surface Temperature Day, Surface Temperature Night, Vegetation Index (NDVI), Water Vapor

### **DATA SET Electricity**

**AGENCY** National Electric Power Regulatory Authority and K-Electric

**FREQUENCY** Monthly

**LEVEL** Tehsil

**TIME** July 2011–March 2021

**URL** Available on request

**VARIABLES** Commercial Consumption, Domestic Consumption, Industrial Consumption, Other Consumption

### **DATA SET Global Gas Flares Observed from Space**

**AGENCY** NASA – National Aeronautics and Space Administration

**FREQUENCY** Yearly

**LEVEL** Tehsil

**TIME** 2012–2020

**URL** [https://eogdata.mines.edu/download\\_global\\_flare.html](https://eogdata.mines.edu/download_global_flare.html)

**VARIABLES** Average Temperature, Clear Observations, Detection Frequency, Total Volume (for both upstream and downstream flames)

### **DATA SET Gross National Income**

**AGENCY** Global Data Lab

**FREQUENCY** Yearly

**LEVEL** Province

**TIME** 2010–2018

**URL** [https://globaldatalab.org/areadata/view/gnic/PAK/?levels=1%2B2%2B3%2B5%2B4&interpolation=0&extrapolation=0&nearest\\_real=0](https://globaldatalab.org/areadata/view/gnic/PAK/?levels=1%2B2%2B3%2B5%2B4&interpolation=0&extrapolation=0&nearest_real=0)

**VARIABLES** Gross National Income per Capita (\$ 2011 PPP)

### **DATA SET Land Cover**

**AGENCY** ESA – European Space Agency

**FREQUENCY** Yearly

**LEVEL** Tehsil

**TIME** 1992–2020

**URL** <https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=overview>

**VARIABLES** Bare Areas, Crop Irrigated, Crop Rainfed, Grassland, Lichens and Mosses, Mosaic Crop, Mosaic Herbaceous Cover, Mosaic Tree and Shrubs, Mosaic Vegetation, Shrubland, Shrubs or Herbaceous Flooded, Sparse Vegetation, Tree Broadleaved Deciduous, Tree Broadleaved Evergreen, Tree Cover Flooded Fresh, Tree Cover Flooded Saline, Tree Mixed Leaf Type, Tree Needleleaved Deciduous, Tree Needleleaved Evergreen, Permanent Snow and Ice Urban Areas, Water Bodies

**DATA SET** Monthly Bulletins of Statistics

**AGENCY** Pakistan Bureau of Statistics

**FREQUENCY** Monthly

**LEVEL** City

**TIME** January 2012–March 2021

**URL** <https://www.pbs.gov.pk/publications/par>

**VARIABLES** Price in Pakistani Rupees for 486 Items

**DATA SET** Pakistan Economic Survey

**AGENCY** Government of Pakistan—Finance Division

**FREQUENCY** Yearly

**LEVEL** City

**TIME** 2006–2021

**URL** [https://www.finance.gov.pk/survey\\_2021.html](https://www.finance.gov.pk/survey_2021.html)

**VARIABLES** Container Imported, Container Exported, Doctors Consulting Fees, Total Container in Terminals

**DATA SET** Survey on COVID-19

**AGENCY** Pakistan Bureau of Statistics

**FREQUENCY** One-off

**LEVEL** District

**TIME** COVID Period

**URL** <https://www.pbs.gov.pk/content/survey-covid-19>

**VARIABLES** Return Migration

**DATA SET** VIIRS Boat Detection (VBD)

**AGENCY** NASA – National Aeronautics and Space Administration

**FREQUENCY** Monthly

**LEVEL** Territorial waters

**TIME** July 2016–May 2021

**URL** [https://eogdata.mines.edu/map\\_selector/](https://eogdata.mines.edu/map_selector/)

**VARIABLES** Number of Blurry Lights, Number of Boats, Number of Gas Flares, Number of Glow Lights, Number of Platform lights, Number of Recurring Lights, Number of Weak Lights, Number of Weak and Blurry Lights

**DATA SET** VIIRS—Night-lights

**AGENCY** NASA – National Aeronautics and Space Administration

**FREQUENCY** Monthly

**LEVEL** Tehsil

**TIME** January 2012–June 2021

URL <https://ladsweb.modaps.eosdis.nasa.gov/search/order/2/VNP46A1--5000>

VARIABLES Night-lights

DATA SET Weather

AGENCY State Bank of Pakistan

FREQUENCY Daily

LEVEL City

TIME January 2012–March 2021

URL Available on request

VARIABLES Cloud Cover, Dew Point, Feels Like Temperature, Heat Index, Humidity, Max Temperature, Min Temperature, Moon Illumination, Pressure, Sun Hours, Total Snow, Total Precipitations, UV Index, Visibility, Wind Chill Temperature, Wind Degree, Wind Gust Speed, Wind Speed

DATA SET Population

AGENCY WorldPop

FREQUENCY Yearly

LEVEL Tehsil

TIME 2000–2020

URL <https://www.worldpop.org/geodata/listing?id=30>

VARIABLES Population by Age and Sex

### SAMPLED DAYS VIIRS NIGHT-LIGHTS VNP46A1

20 January 2012; 22 February 2012; 15 March 2012; 22 April 2012; 20 May 2012; 20 June 2012; 22 July 2012; 26 August 2012; 15 September 2012; 21 October 2012; 15 November 2012; 15 December 2012; 15 January 2013; 17 February 2013; 7 March 2013; 16 April 2013; 16 May 2013; 8 June 2013; 10 July 2013; 4 August 2013; 14 September 2013; 9 October 2013; 28 November 2013; 28 December 2013; 25 January 2014; 24 February 2014; 29 March 2014; 23 April 2014; 23 May 2014; 29 June 2014; 21 July 2014; 15 August 2014; 15 September 2014; 21 October 2014; 15 November 2014; 15 December 2014; 15 January 2015; 21 February 2015; 18 March 2015; 20 April 2015; 15 May 2015; 18 June 2015; 15 July 2015; 15 August 2015; 17 September 2015; 20 October 2015; 15 November 2015; 18 December 2015; 10 January 2016; 14 February 2016; 9 March 2016; 13 April 2016; 12 May 2016; 6 June 2016; 10 July 2016; 12 August 2016; 24 September 2016; 25 October 2016; 26 November 2016; 23 December 2016; 29 January 2017; 15 February 2017; 23 March 2017; 26 April 2017; 22 May 2017; 18 June 2017; 24 July 2017; 26 August 2017; 15 September 2017; 15 October 2017; 15 November 2017; 21 December 2017; 15 January 2018; 15 February 2018; 21 March 2018; 21 April 2018; 15 May 2018; 15 June 2018; 20 July 2018; 10 August 2018; 15 September 2018; 17 October 2018; 16 November 2018; 17 December 2018; 14 January 2019; 8 February 2019; 14 March 2019; 28 April 2019; 27 May 2019; 27 June 2019; 2 July 2019; 4 August 2019; 23 September 2019; 24 October 2019; 24 November 2019; 21 December 2019; 18 January 2020; 24 February 2020; 15 March 2020; 23 April 2020; 15 May 2020; 15 June 2020; 18 July 2020; 21 August 2020; 15 September 2020; 15 October 2020; 15 November 2020; 18 December 2020; 20 January 2021; 16 February 2021; 8 March 2021; 8 April 2021; 21 May 2021.

## S2. Machine Learning and Satellite Data

### S2.1. Additional Elements on Machine Learning and Income

The Global Data Lab provides data on per capita income for each of the Pakistani provinces from 2010 to 2018. Starting from this aggregate, we redistribute the provincial gross national income at tehsil level, using population data, as described in the Data and Methodology section. We require data on income for

years beyond 2018, and satellite observations only begin from 2012, when VIIRS sensors for night-lights started transmitting. Thus, in our research we consider the period of time from 2012 to 2021.

To estimate the GNI figures for years after 2018, we need to train the algorithms. Hence, we produce a data set with local income at the tehsil-year level from 2012 to 2018. We also include all the potential predictors with the same level of aggregation: Electricity;<sup>3</sup> Landcover;<sup>4</sup> Local Prices;<sup>5</sup> Natural Observations;<sup>6</sup> Night-lights; Population; Weather.<sup>7</sup>

To handle the correlated predictors, we decide to maintain only the most influential on local income. Additionally, we incorporate a set of dummy variables for tehsils, districts, and provinces to capture time-invariant geographical characteristics and absorb the impact of varying data production practices. It is important to highlight that we do not include an indicator for years, as we want our algorithm to learn directly from the data without following a specific temporal pattern.

Lastly, we perform a min-max normalization to rescale our data between 0 and 1. This step is essential due to the heterogeneous scales of the observed variables and the lack of prior information about their final distributions. Additionally, we convert our categorical variables (district, tehsil, province, electric company) into dummy variables.

In the prediction exercise, we use Python to compare the performance of the following machine-learning algorithms:

- elastic-net: It adopts a classical linear regression, coping with the presence of multiple predictors through a shrinkage parameter. While the lasso and the ridge refer to two opposite reduction formulas, this model finds the best combinations among these criteria. After this step, it uses another cross-validation to identify the optimal weight for the final predictors.
- random forest: Based on trees, this model assumes a series of sequential decisions using the most predictive variable at each step. The random forest proposes only a set of the predictors in every decision node, performing its task on the entire data set.
- bagging: This is another model based on decision trees, but it differs from the random forest by proposing all the predictors and a bootstrap of the original data at every decision node.
- boosting: It is the last model based on decision trees. In every node, it chooses among all the predictors and attributes a greater weight in the sample constitution to the observations wrongly classified in precedent steps.
- support vector machine: It draws hypothetical hyperplanes in variables' space to categorize the observations into final predictions. The criteria behind the hyperplane realization define its type: the linear opts for a parametric approach, the polynomial switches to a non-parametric one, and the radial basis function opts for a  $k$ -nearest neighborhood algorithm.

3 Commercial Consumption, Domestic Consumption, Industrial Consumption, Other Consumption.

4 Bare Areas, Crop Irrigated, Crop Rainfed, Grassland, Lichens and Mosses, Mosaic Crop, Mosaic Herbaceous Cover, Mosaic Tree and Shrubs, Mosaic Vegetation, Shrubland, Shrubs or Herbaceous Flooded, Sparse Vegetation, Tree Broadleaved Deciduous, Tree Broadleaved Evergreen, Tree Cover Flooded Fresh, Tree Cover Flooded Saline, Tree Mixed Leaf Type, Tree Needleleaved Deciduous, Tree Needleleaved Evergreen, Permanent Snow and Ice Urban Areas, Water Bodies.

5 Price in Pakistani Rupees for 486 Items.

6 Aerosol Thickness, Clouds Fraction, Clouds Optical Thickness, Clouds Particle Radius, Clouds Water Content, Night Fires, NO<sub>2</sub>, Ozone, Sea Chlorophyll, Sea Surface Temperature, Snow on the Ground, Surface Temperature Anomaly Day, Surface Temperature Anomaly Night, Surface Temperature Day, Surface Temperature Night, Vegetation Index (NDVI), Water Vapor.

7 Cloud Cover, Dew Point, Feels Like Temperature, Heat Index, Humidity, Max Temperature, Min Temperature, Moon Illumination, Pressure, Sun Hours, Total Snow, Total Precipitations, UV Index, Visibility, Wind Chill Temperature, Wind Degree, Wind Gust Speed, Wind Speed.

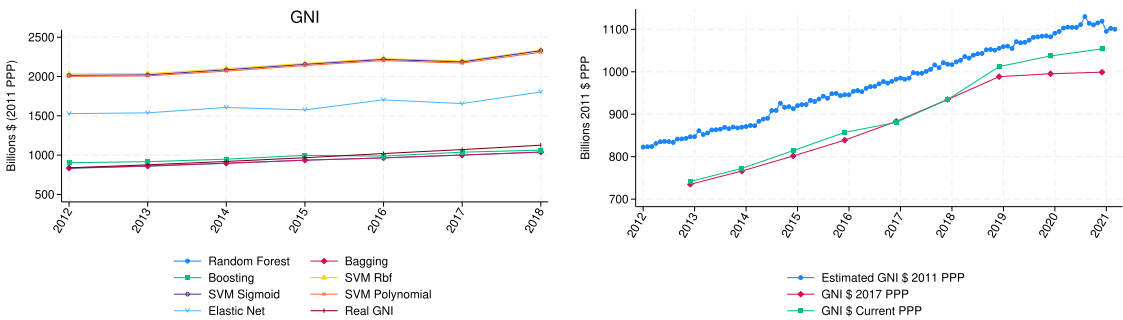
**Table S2.1.** Mean Square Error

Models	(1) Mean square error
bagging	0.01506
boosting	0.01532
elastic net	0.01606
random forest	0.01509
svm—sigmoid kernel	0.01566
svm—polynomial kernel	0.01562
svm—rbf kernel	0.01568

Source: Authors' analysis based on results of the estimation algorithms.

Note: This table reports the mean square error of the algorithms employed in our study. The abbreviation “svm” stands for “support vector machine,” while “rbf” means “radial basis function.”

**Figure S2.1.** Algorithms Comparison



Source: Authors' analysis based on Gross National Income (GNI) estimates using different algorithms and World Bank data.

Note: The graph on the left plots the estimated yearly GNI from 2012 to 2018, together with the reference value of “Real GNI.” The abbreviation SVM stands for “support vector machine,” while Rbf indicates “radial basis function.” The graph on the right compares the GNI estimated with our best performing model (bagging) with the yearly data provided by the World Bank. For these time series, we report the value in the last month of the year.

We do not implement more complex algorithms to permit an easy and fast replication of our policy toolkit. In addition, we tune every algorithm through a randomized grid search. Table S2.1 reports the mean square error of all our models, while fig. S2.1 plots their estimates and our real data.

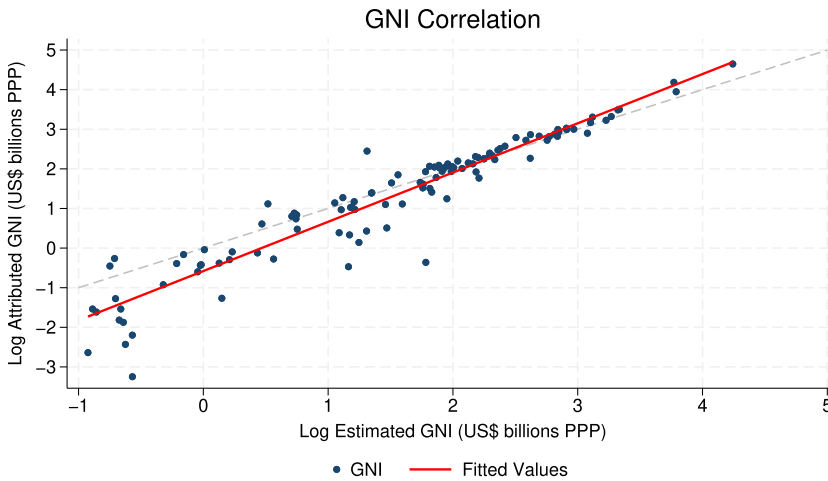
The bagging algorithm achieves the lowest mean square error and obtains the best replication of the original data. As a result, it is the one we adopt for the final estimations in our study. To generate our final estimates, we employ a distinct data set from the one used for the training of the algorithms. In particular, we use a data set disaggregated at the monthly level. This allows us to leverage the disaggregated nature of our predictors and derive estimates at a more granular level, specifically on a monthly basis.

To further verify that our initial geographical disaggregation of GNI based on the share of population did not significantly distort our predictions, we use an indicator of living standards provided for each district in 2017 by the United Nations Development Program (UNDP). We decompose provincial GNI of 2017 taking into account the different living conditions in each district. Figure S2.2 B2 graphically shows the high correlation between our estimates and this newly obtained measure of local GNI, which turns out to be as high as 0.98.

## S2.2. An Introduction to Satellite Data

Satellites are ever more present in academic and non-academic research as a massive amount of high-resolution and high-frequency data become available. Once in orbit, satellites can automatically

**Figure S2.2.** Correlation between Attributed and Estimated GNI in 2017



*Source:* Authors' analysis based on Gross National Income (GNI) estimates and the Human Development Index.

*Note:* The graph plots the correlation between the logarithm of our estimated yearly GNI in 2017 for each district (on the horizontal axis) and the logarithm of GNI attributed to each district based on a living standard indicator provided by the UNDP (United Nations Development Program) (on the vertical axis). The positive linear relationship between these variables is shown in red. All values are expressed in billions of 2011 PPP dollars.

collect and pre-process data through automated algorithms. This procedure permits bypassing the physical barriers of the traditional data-gathering operations (high costs, inaccessible areas, continuity in the project), although their innovative nature still poses some obstacles to their mainstream use. This short appendix provides the basic concepts behind this technology.

Almost all the satellite products come in raster format: simple images enclosing the globe in a georeferenced grid. Every quadratic cell of this grid is composed of pixels. The resolution is expressed as the side length of every cell. The smaller the side, the more pixels are contained in each meter (the resolution). There is no common standard for this unit of measure. Some companies report it in meters, while others report it in degrees. A good conversion rule is to assume  $0.1^\circ = 11.1$  km. Going a step deeper into the subject, we must understand how the pixel can reproduce a figure on our screen. This passage is essential for the storage and extraction of the collected data.

Each pixel contains three phosphors emitting red, blue, and green light. Being able to work under different intensities and combinations, these latter can recreate all of the existing palette of colors. When all of the phosphors are off, we visualize a black pixel. Conversely, when they are all at the maximum intensity, we observe a white element. The satellite recreates a picture of the planet in pixels, assigning to each the measured values in the form of pixel luminosity. The highest observations will correspond to white areas, while the null correspond to black ones. All the values left in the middle will assume several shades of gray.

Our Python algorithms exploit this technology to retrieve the data. Indeed, they overlaid the satellite rasters with the shapefile of the studied areas to produce summary statistics on the underlying pixels. This procedure is the standard for continuous variables, but the satellite can also assign arbitrary values to map categorical variables (like the landcover). In this framework, the Python algorithm counts the recurrence of a determined value in a given shapefile.

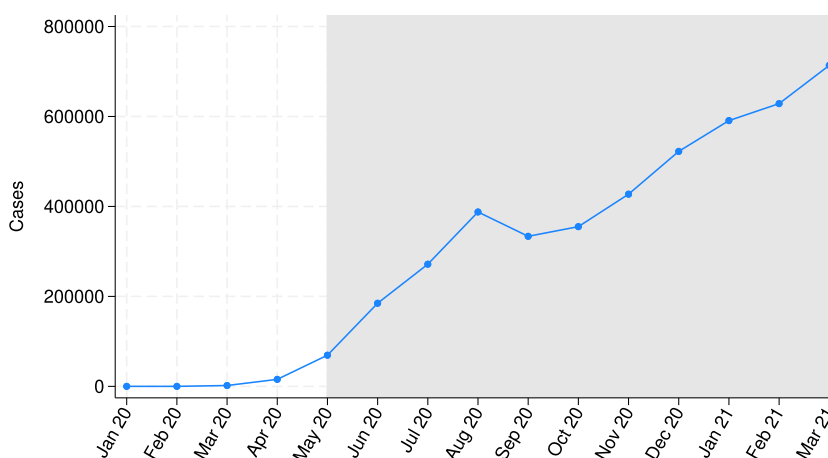
In addition, a common practice consists of cutting off the studied area from the entire raster. This operation has two main goals:

- producing illustrative material;
- reducing the computation time and the storage space.

The most popular format for raster files is the “GeoTIFF” extension. However, some satellite products use “NetCDF” files containing several rasters at once. Another popular format is the new “HDF5,” offering a wide range of enriched functions. It is also worth noting that some rasters represent only a fraction of the planet. This strategy allows better handling of this massive amount of data. Stata has no functions to compute these resources. This shortage can be related to the high-processing performances required by these files. Thus, we must resort to more advanced open-source software. Among them, the most employed by economists are Python, R, and Julia. It is also possible to work with QGIS or ArcGis. This software runs in Python, offering a more user-friendly interface. Unfortunately, it is more complex in organizing loops and storing the commands in reproducible code.

### S3. Additional Figures and Tables

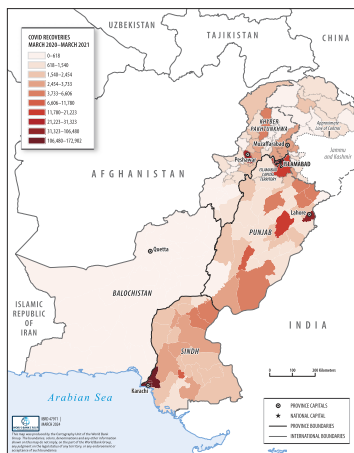
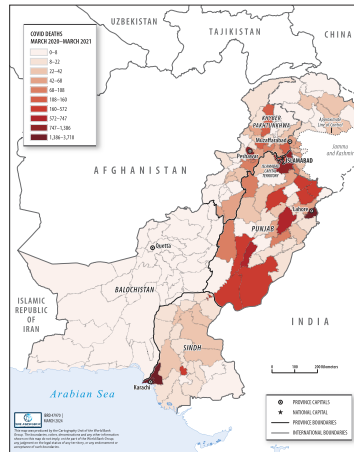
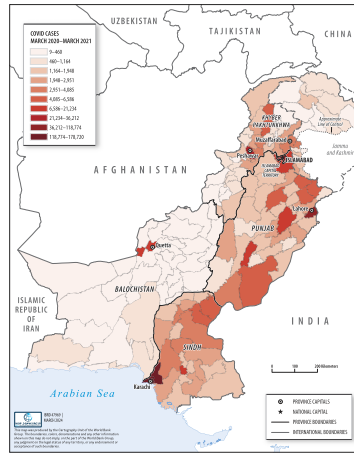
**Figure S3.1.** National Epidemic Curve



*Source:* Authors’ analysis based on data on the COVID-19 pandemic.

*Note:* The graph displays the monthly number of COVID-19 cases reported in each Pakistani district, from January 2020 to March 2021. The gray area indicates the temporal framework covered by our “dummy COVID”: May 2020–March 2021.

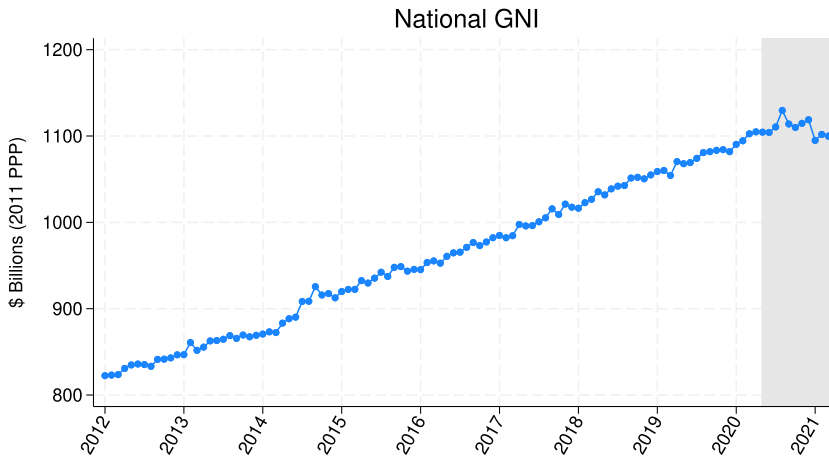
**Figure S3.2. COVID Statistics by Districts**



Source: Authors' analysis based on data on the COVID-19 pandemic.

Note: These three maps report the total number of COVID-19 cases (left), deaths (central), and recoveries (right) for the Pakistani districts between March 2020 and March 2021.

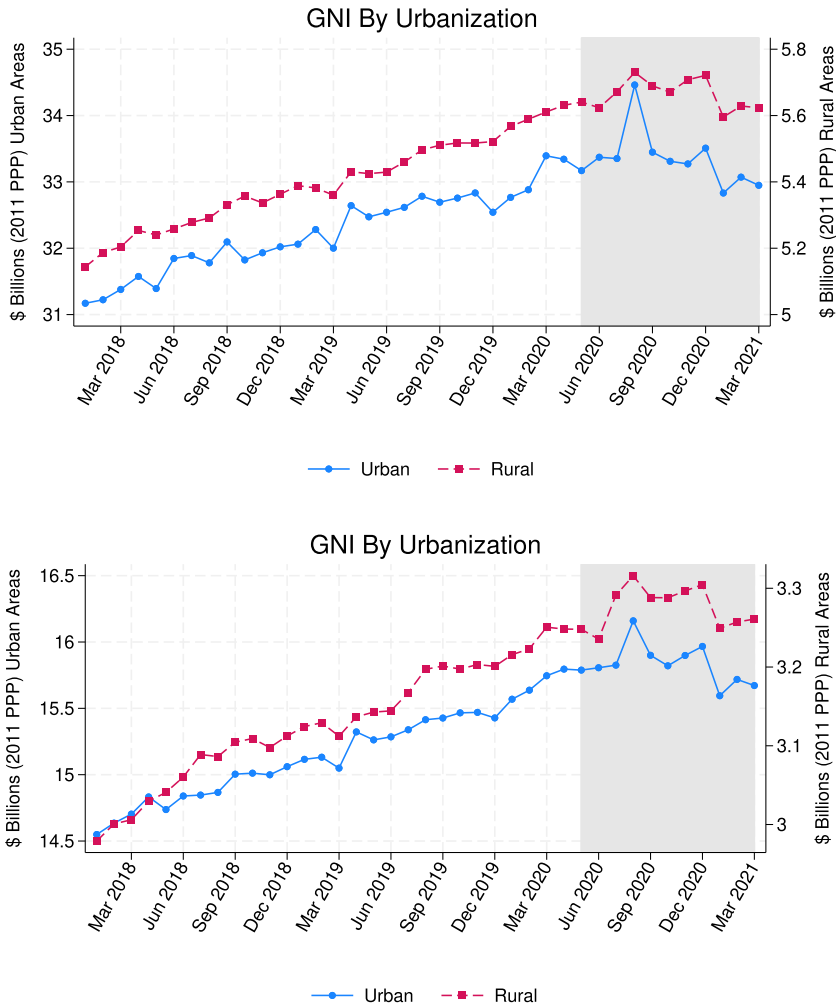
**Figure S3.3.** Gross National Income 2012–2021



*Source:* Authors' analysis based on estimated GNI.

*Note:* This plot reports the Pakistani Gross National Income from January 2012 to March 2021. All the values are expressed in billions of 2011 PPP dollars. The gray area indicates the temporal framework covered by our "dummy COVID": May 2020–March 2021.

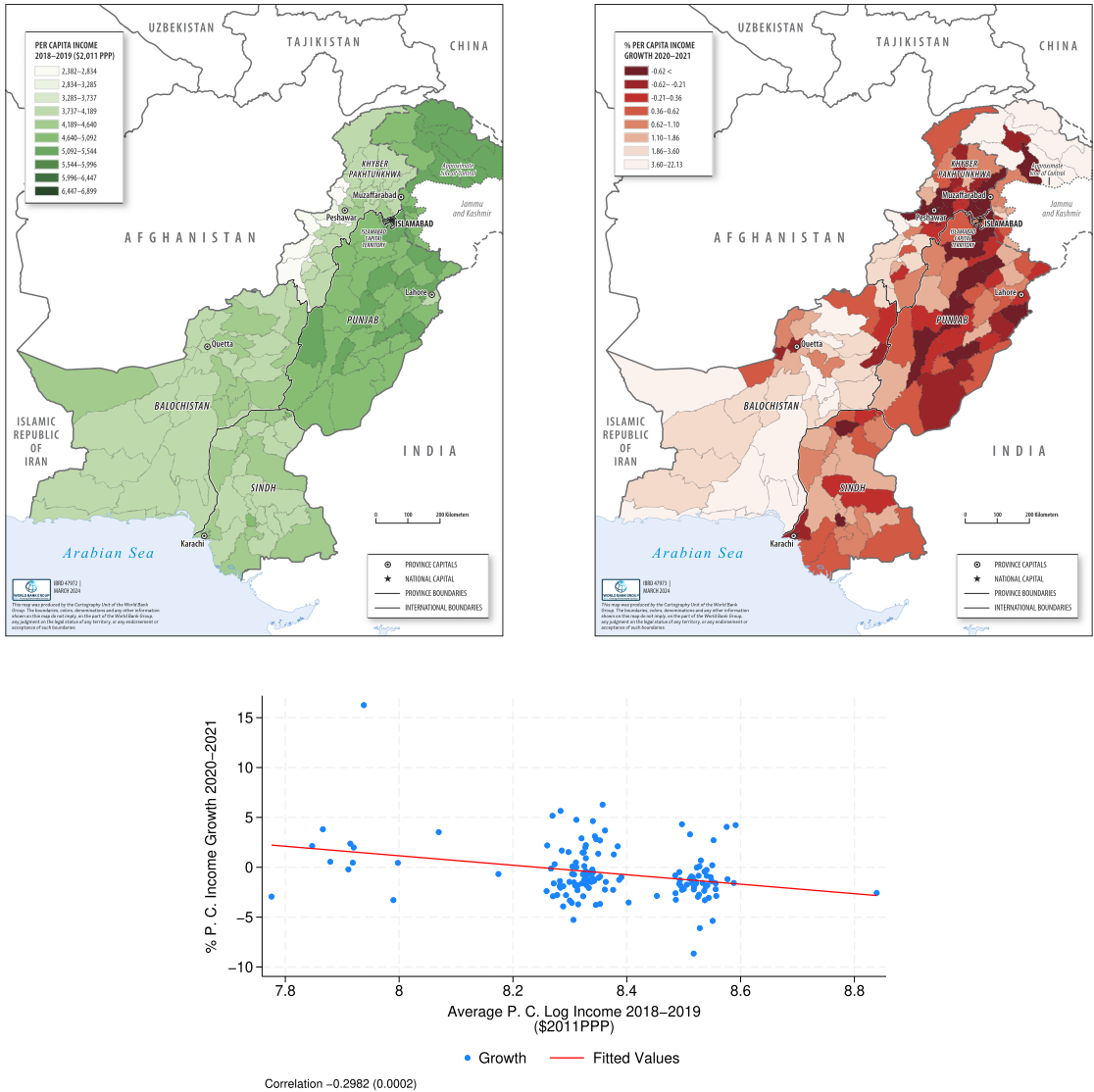
**Figure S3.4. Urban vs Rural GNI: Different Aggregates**



*Source:* Authors' analysis based on Gross National Income (GNI) estimates and major Pakistani cities.

*Note:* These plots decompose the average gross national income between rural districts in red (right-hand-side vertical axis) and urban districts in blue (left-hand-side vertical axis). In the top panel, districts are classified as urban if they include one of the 10 biggest cities, while in the bottom panel, districts are classified as urban if they include one of the 50 biggest cities. All the values are expressed in billions of 2011 PPP dollars. The gray area indicates the temporal framework covered by our "dummy COVID": May 2020–March 2021.

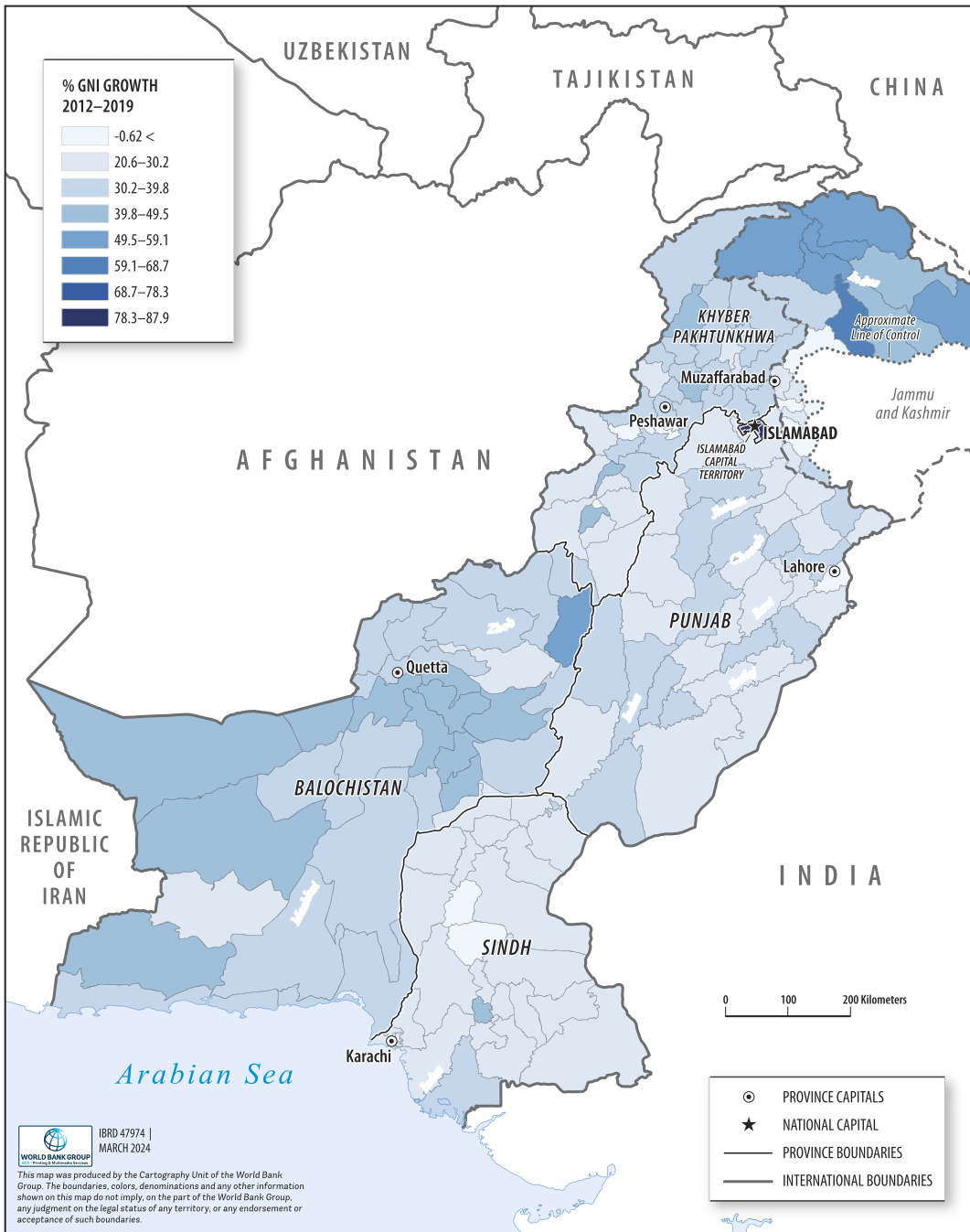
**Figure S3.5. Average Income per Capita and Income per Capita Growth by District**



**Source:** Authors' analysis based on Gross National Income (GNI) estimates.

**Note:** The upper-left panel illustrates the average income per capita of districts during the period between 2018 and 2019. Darker colors represent districts with higher income levels. The upper-right panel displays the percentage variation of income per capita between 2020 and 2021. For 2021, only the first three months are considered due to data availability. Darker colors represent districts with lower growth. The bottom panel presents a graph depicting the district average GNI per capita in 2018–2019 on the horizontal axis, and its percentage variation between 2020 and 2021 on the vertical axis. The linear relationship between these variables is shown in red, and the correlation is noted below the graph. All values are expressed in billions of 2011 PPP dollars.

**Figure S3.6.** Growth by District 2012–2019



Downloaded from <https://academic.oup.com/wber/article/39/2/362/7693247> by guest on 18 July 2025

*Source:* Authors' analysis based on Gross National Income (GNI) estimates.

*Note:* This map displays the GNI's growth rate between 2012 and 2019, for each of the 147 Pakistani districts. The yearly GNI is computed from the monthly average.

**Table S3.1.** Summary Statistics

Variables	(1) Observations	(2) Mean	(3) St. deviation	(4) 50th p.tile	(5) 5th p.tile	(6) 95th p.tile
Log income	5,733	21.93	1.368	22.03	19.91	23.94
Growth income	5,733	0.297	2.556	0.195	-2.505	3.324
Standardized income 2012	5,733	-4.34e-10	1.000	-0.358	-0.674	1.625
Log income 2012	5,733	21.66	1.389	21.77	19.63	23.64
Log income p.c.	5,733	8.374	0.170	8.358	7.949	8.571
Growth income p.c.	5,733	0.175	2.536	0.0306	-2.490	3.220
Log dom. electricity	5,733	11.54	7.577	14.90	-2.303	18.25
Growth dom. electricity	4,498	2,783	73,767	-0.108	-49.29	110.0
Log comm. electricity	5,733	10.11	6.923	13.24	-2.303	16.49
Growth comm. electricity	4,477	79.72	3,264	-0.495	-41.28	60.37
Log ind. electricity	5,733	10.61	7.326	13.92	-2.303	17.75
Growth ind. electricity	4,467	44.19	744.2	-0.546	-52.21	96.79
Log other electricity	5,733	11.26	7.180	14.48	-2.303	17.40
Growth other electricity	4,549	208.2	3,912	1.194	-70.38	271.3
Log night-lights sum	5,733	9.413	2.712	10.08	4.949	12.47
Growth night-lights sum	5,609	179.6	3,660	-1.288	-84.73	329.8
Log night-lights S.D.	5,733	1.654	1.528	1.894	-1.048	3.878
Growth night-lights S.D.	5,609	60.81	961.0	-1.254	-77.40	222.2
Log night-lights mean	5,733	0.213	1.745	0.295	-2.245	3.178
Growth night-lights mean	5,609	191.2	3,638	-0.632	-84.90	342.6
Log night-lights max	5,733	5.235	1.806	5.412	2.715	7.745
Growth night-lights max	5,609	95.95	1,442	0	-81.88	330
Dummy COVID	5,733	0.282	0.450	0	0	1
COVID cases	5,733	785.3	6,565	0	0	2,073
COVID deaths	5,733	19.91	158.4	0	0	56
COVID recoveries	5,733	681.2	5,946	0	0	1,721
Log cases	5,733	0.144	3.909	-2.303	-2.303	7.637
Log deaths	5,733	-1.159	2.270	-2.303	-2.303	4.027
Log recoveries	5,733	-0.189	3.798	-2.303	-2.303	7.451
Year	5,733	2,019	0.948	2,019	2,018	2,021
Month	5,733	6.154	3.534	6	1	12
District	5,733	74	42.44	74	8	140
Dummy urban	5,733	0.136	0.343	0	0	1

Source: Summary statistics of the variables used for this study.

Note: This table reports the summary statistics of all the variables considered in this study. The variable “Log income” represents the logarithm of the district gross income, while “Income growth” is the percentage variation between months. A similar approach is also adopted for reporting electricity consumption of domestic, commercial, industrial, and other users. The electricity growth has fewer observations, being constant to zero in districts subsequently connected to the grid. For the night-lights, we indicate the logarithm and the growth of several statistical aggregates observed at the district-month level (mean, sum, max, and standard deviation). The “Dummy COVID” assumes the value 1 from May 2020 to March 2021. “COVID cases,” “COVID deaths,” and “COVID recoveries” are set to zero for the months preceding the pandemic. We also report the logarithms of these variables. We use “Dummy urban” for indicating the districts containing the first 20 metropolitan areas of the country (Karachi, Lahore, Faisalabad, Gujranwala, Rawalpindi, Peshawar, Multan, Hyderabad, Sialkot, Bahawalpur, Islamabad, Quetta, Rahim Yar Khan, Sheikhpura, Sargodha, Attock, Sukkur, Larkana, Swat, Muzaffargarh). The data set follows 147 districts from January 2018 to March 2021.

**Table S3.2.** COVID Impact on Gross National Income (GNI) Growth—Robustness Check with District FE

Variables	(1)	(2)	(3)	(4)
	<b>Income growth</b>			
<b>Panel A—Overall</b>				
Covid <sub>my</sub>	−0.133*** (0.0415)	−	−	−
Cases <sub>dmy</sub>	−	−0.0203*** (0.00396)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0465*** (0.00667)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0287*** (0.00468)
District FE	Yes	Yes	Yes	Yes
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	−0.0181	−0.0177	−0.0171	−0.0170
Mean dep. var.	0.297	0.297	0.297	0.297
S.D. dep. var.	2.556	2.556	2.556	2.556
<b>Panel B—Urban</b>				
Covid <sub>my</sub>	−0.104** (0.0472)	−	−	−
Covid <sub>my</sub> × Urban <sub>d</sub>	−0.216*** (0.0595)	−	−	−
Cases <sub>dmy</sub>	−	−0.0181*** (0.00488)	−	−
Cases <sub>dmy</sub> × Urban <sub>d</sub>	−	−0.0103 (0.00684)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0445*** (0.00903)	−
Deaths <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−0.00613 (0.0121)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0273*** (0.00578)
Recoveries <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−	−0.00588 (0.00872)
District FE	Yes	Yes	Yes	Yes
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	−0.0181	−0.0178	−0.0173	−0.0171
Mean dep. var.	0.297	0.297	0.297	0.297
S.D. dep. var.	2.556	2.556	2.556	2.556

*Source:* Authors' analysis based on GNI estimates, major Pakistani cities, and data on the COVID-19 pandemic.

*Note:* Panel A estimates the impact on districts' income growth rates of the pandemic period (column 1), the logarithm of COVID cases (column 2), the logarithm of COVID deaths (column 3), and the logarithm of COVID recoveries (column 4). Panel B repeats the same analysis but decomposes the impact between rural and urban districts. The sample includes all 147 Pakistani districts from January 2018 to March 2021. District fixed effects are included in the analysis. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

**Table S3.3.** Determinants of Gross National Income (GNI) decline

Variables	Income growth			
	(1)	(2)	(3)	(4)
Covid <sub>my</sub>	-0.207*** (0.0336)	-	-	-
Covid <sub>my</sub> × Urban <sub>d</sub>	0.0569 (0.0523)	-	-	-
Covid <sub>my</sub> × ServiceShare <sub>p</sub>	-0.254*** (0.0599)	-	-	-
Covid <sub>my</sub> × ReturnMigration <sub>d</sub>	-0.0691*** (0.0216)	-	-	-
Cases <sub>dmy</sub>	-	-0.0216*** (0.00356)	-	-
Cases <sub>dmy</sub> × Urban <sub>d</sub>	-	0.00886 (0.00724)	-	-
Cases <sub>dmy</sub> × ServiceShare <sub>p</sub>	-	-0.0246*** (0.00615)	-	-
Cases <sub>dmy</sub> × ReturnMigration <sub>d</sub>	-	-0.00349* (0.00182)	-	-
Deaths <sub>dmy</sub>	-	-	-0.0380*** (0.00604)	-
Deaths <sub>dmy</sub> × Urban <sub>d</sub>	-	-	0.0197 (0.0125)	-
Deaths <sub>dmy</sub> × ServiceShare <sub>p</sub>	-	-	-0.0517*** (0.0112)	-
Deaths <sub>dmy</sub> × ReturnMigration <sub>d</sub>	-	-	-0.00362 (0.00285)	-
Recoveries <sub>dmy</sub>	-	-	-	-0.0230*** (0.00382)
Recoveries <sub>dmy</sub> × Urban <sub>d</sub>	-	-	-	0.0100 (0.00980)
Recoveries <sub>dmy</sub> × ServiceShare <sub>p</sub>	-	-	-	-0.0348*** (0.00790)
Recoveries <sub>dmy</sub> × ReturnMigration <sub>d</sub>	-	-	-	-0.00127 (0.00239)
Urban <sub>d</sub>	-0.0350 (0.0432)	-0.00175 (0.0385)	0.0420 (0.0327)	-0.00492 (0.0400)
ServiceShare <sub>p</sub>	0.112*** (0.0402)	0.0426 (0.0323)	-0.0171 (0.0270)	0.0473 (0.0329)
ReturnMigration <sub>d</sub>	0.0143 (0.0113)	-0.00371 (0.00979)	-0.00774 (0.00803)	-0.00483 (0.00984)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	2,769	2,769	2,769	2,769
Adj. R <sup>2</sup>	0.00310	0.00208	0.00312	0.00416
Mean dep. var.	0.297	0.297	0.297	0.297
S.D. dep. var.	2.556	2.556	2.556	2.556

*Source:* Authors' analysis based on GNI estimates, major Pakistani cities, data on the COVID-19 pandemic, and information on migration and the composition of the economy.

*Note:* This table wants to investigate the main factors determining GNI movements during the pandemic. It augments the specification of table 2 (in the main text) by including interactions between COVID-19 measures and two variables: the share of the service sector in each province and the share of people that returned home for COVID-related reasons in each district. Both measures are standardized. No fixed effects are included in the analysis. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

**Table S3.4. Health Facilities**

Variables	(1)	(2)	(3)	(4)
	Income growth			
Panel A—Overall				
Covid <sub>my</sub>	−0.133*** (0.0415)	−	−	−
Cases <sub>dmy</sub>	−	−0.0191*** (0.00389)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0409*** (0.00589)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0261*** (0.00435)
HealthFacilities <sub>d</sub>	−0.0859*** (0.0222)	−0.0825*** (0.0228)	−0.0778*** (0.0236)	−0.0793*** (0.0237)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.00310	0.00340	0.00384	0.00404
Mean dep. var.	0.297	0.297	0.297	0.297
Panel B—Urban				
Covid <sub>my</sub>	−0.104** (0.0472)	−	−	−
Covid <sub>my</sub> × Urban <sub>d</sub>	−0.216*** (0.0595)	−	−	−
Cases <sub>dmy</sub>	−	−0.0176*** (0.00471)	−	−
Cases <sub>dmy</sub> × Urban <sub>d</sub>	−	−0.0100 (0.00665)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0413*** (0.00740)	−
Deaths <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−0.00636 (0.0105)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0253*** (0.00518)
Recoveries <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−	−0.00672 (0.00808)
Urban <sub>d</sub>	0.182*** (0.0624)	0.146** (0.0575)	0.155*** (0.0555)	0.146** (0.0588)
HealthFacilities <sub>d</sub>	−0.0981*** (0.0252)	−0.0963*** (0.0255)	−0.0934*** (0.0259)	−0.0935*** (0.0264)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.00313	0.00336	0.00385	0.00399
Mean dep. var.	0.297	0.297	0.297	0.297

Source: Authors' analysis based on Gross National Income (GNI) estimates, major Pakistani cities, information on health facilities, and data on the COVID-19 pandemic.

Note: This table investigates how the number of health facilities in each district affects our results. Panel A estimates the impact on income growth of the pandemic months (column 1), the logarithm of COVID cases (column 2), the logarithm of COVID deaths (column 3), and the logarithm of COVID recoveries (column 4), controlling for the number of health facilities. Panel B repeats this analysis, decomposing the impact between rural and urban areas. The sample includes all 147 Pakistani districts from January 2018 to March 2021. No fixed effects are included in the analysis. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

**Table S3.5.** COVID Impact on GNI—Alternative Definitions of Urban Districts

Variables	(1)	(2)	(3)	(4)
			<b>Income growth</b>	
Panel A—Principal 10 urban areas				
Covid <sub>my</sub>	−0.117*** (0.0440)	−	−	−
Covid <sub>my</sub> × Urban <sub>d</sub>	−0.245*** (0.0703)	−	−	−
Cases <sub>dmy</sub>	−	−0.0198*** (0.00426)	−	−
Cases <sub>dmy</sub> × Urban <sub>d</sub>	−	−0.0118 (0.00722)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0471*** (0.00677)	−
Deaths <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−0.00637 (0.0131)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0278*** (0.00461)
Recoveries <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−	−0.0124 (0.0109)
Urban <sub>d</sub>	−0.00969 (0.0575)	−0.0399 (0.0566)	−0.0106 (0.0530)	−0.0236 (0.0588)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.000207	0.000627	0.00134	0.00147
Mean dep. var.	0.297	0.297	0.297	0.297
Panel B—Principal 50 urban areas				
Covid <sub>my</sub>	−0.0559 (0.0563)	−	−	−
Covid <sub>my</sub> × Urban <sub>d</sub>	−0.228*** (0.0738)	−	−	−
Cases <sub>dmy</sub>	−	−0.0160*** (0.00612)	−	−
Cases <sub>dmy</sub> × Urban <sub>d</sub>	−	−0.0108 (0.00726)	−	−
Deaths <sub>dmy</sub>	−	−	−0.0408*** (0.0103)	−
Deaths <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−0.0115 (0.0117)	−
Recoveries <sub>dmy</sub>	−	−	−	−0.0240*** (0.00697)
Recoveries <sub>dmy</sub> × Urban <sub>d</sub>	−	−	−	−0.0104 (0.00821)
Urban <sub>d</sub>	−0.0247 (0.0333)	−0.0714** (0.0313)	−0.0683** (0.0336)	−0.0618* (0.0315)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.00127	0.00120	0.00175	0.00199
Mean dep. var.	0.297	0.297	0.297	0.297

Source: Authors' analysis based on Gross National Income (GNI) estimates, major Pakistani cities, and data on the COVID-19 pandemic.

Note: This table replicates the specification of panel B of table 2, changing the definitions of urban districts. In panel A, districts are classified as urban if they include one of the 10 biggest cities, while in panel B districts are classified as urban if they include one of the 50 biggest cities. The sample includes all 147 Pakistani districts from January 2018 to March 2021. No fixed effects are included in the analysis. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

**Table S3.6.** COVID Impact on per Capita Income Growth, 2018–2021

Variables	(1)	(2)	(3)	(4)
	Per capita income growth			
Panel A—Overall				
Covid <sub>my</sub>	-0.0913** (0.0405)	-	-	-
Cases <sub>dmy</sub>	-	-0.0174*** (0.00373)	-	-
Deaths <sub>dmy</sub>	-	-	-0.0450*** (0.00571)	-
Recoveries <sub>dmy</sub>	-	-	-	-0.0276*** (0.00416)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	8.58e-05	0.000541	0.00144	0.00152
Mean dep. var.	0.134	0.134	0.134	0.134
S.D. dep. var.	2.548	2.548	2.548	2.548
Panel B—Urban				
Covid <sub>my</sub>	-0.0634 (0.0460)	-	-	-
Covid <sub>my</sub> × Urban <sub>d</sub>	-0.205*** (0.0587)	-	-	-
Cases <sub>dmy</sub>	-	-0.0145*** (0.00462)	-	-
Cases <sub>dmy</sub> × Urban <sub>d</sub>	-	-0.00924 (0.00652)	-	-
Deaths <sub>dmy</sub>	-	-	-0.0427*** (0.00776)	-
Deaths <sub>dmy</sub> × Urban <sub>d</sub>	-	-	-0.00102 (0.0106)	-
Recoveries <sub>dmy</sub>	-	-	-	-0.0261*** (0.00511)
Recoveries <sub>dmy</sub> × Urban <sub>d</sub>	-	-	-	-0.00275 (0.00798)
Urban <sub>d</sub>	-0.0713* (0.0373)	-0.105*** (0.0360)	-0.0829** (0.0339)	-0.0972*** (0.0352)
District FE	No	No	No	No
Year FE	No	No	No	No
Month FE	No	No	No	No
Obs.	5,733	5,733	5,733	5,733
Adj. R <sup>2</sup>	0.000194	0.000452	0.00121	0.00135
Mean dep. var.	0.134	0.134	0.134	0.134
S.D. dep. var.	2.548	2.548	2.548	2.548

Source: Authors' analysis based on Gross National Income (GNI) estimates, major Pakistani cities, and data on the COVID-19 pandemic.

Note: Panel A estimates the impact on districts' per capita income growth rates of the pandemic period (column 1), the logarithm of COVID cases (column 2), the logarithm of COVID deaths (column 3), and the logarithm of COVID recoveries (column 4). Panel B repeats the same analysis but decomposes the impact between rural and urban districts. The sample includes all 147 Pakistani districts from January 2018 to March 2021. No fixed effects are included in the analysis. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

**Table S3.7. COVID-19 and Growth Trajectories: Log of Income**

Variables	Income Growth				
	(1)	(2)	(3)	(4)	(5)
Income <sub>d2012</sub>	-0.0406*** (0.0114)	-0.0210* (0.0123)	-0.0210* (0.0123)	-	-
Covid <sub>my</sub>	-	4.114*** (1.199)	4.253*** (1.197)	4.114*** (1.199)	4.253*** (1.197)
Income <sub>d2012</sub> × Covid <sub>my</sub>	-	-0.196*** (0.0542)	-0.196*** (0.0543)	-0.196*** (0.0542)	-0.196*** (0.0543)
District FE	No	No	No	Yes	Yes
Year FE	No	No	Yes	No	Yes
Month FE	No	No	Yes	No	Yes
Obs.	16,170	16,170	16,170	16,170	16,170
Adj. R <sup>2</sup>	0.000410	0.00153	0.0129	-0.00553	0.00583
Mean dep. var.	0.324	0.324	0.324	0.324	0.324
S.D. dep. var.	2.600	2.600	2.600	2.600	2.600

Source: Authors' analysis based on Gross National Income (GNI) estimates and data on the COVID-19 pandemic.

Note: Column (1) estimates the impact of the logarithm of mean income in 2012 on the income growth rate of districts, without controlling for fixed effects. The remaining columns explore how the COVID pandemic influences this relation, controlling for different fixed effects: column (2) includes no fixed effects, column (3) controls for time fixed effects, column (4) controls for district fixed effects, and column (5) controls for both time and district fixed effects. The sample includes all 147 Pakistani districts from January 2012 to March 2021. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.

**Table S3.8. COVID-19 and Growth Trajectories: 2012 Income Terciles**

Variables	Income growth				
	(1)	(2)	(3)	(4)	(5)
Tercile2 Income <sub>d2012</sub>	-0.0851*** (0.0272)	-0.0492* (0.0267)	-0.0492* (0.0268)	-	-
Tercile3 Income <sub>d2012</sub>	-0.0895*** (0.0286)	-0.0371 (0.0283)	-0.0371 (0.0284)	-	-
Covid <sub>my</sub>	-	0.157 (0.0973)	0.297*** (0.104)	0.157 (0.0973)	0.297*** (0.104)
Tercile2 Income <sub>d2012</sub> × Covid <sub>my</sub>	-	-0.359*** (0.101)	-0.359*** (0.101)	-0.359*** (0.101)	-0.359*** (0.101)
Tercile3 Income <sub>d2012</sub> × Covid <sub>my</sub>	-	-0.524*** (0.102)	-0.524*** (0.102)	-0.524*** (0.102)	-0.524*** (0.102)
District FE	No	No	No	Yes	Yes
Year FE	No	No	Yes	No	Yes
Month FE	No	No	Yes	No	Yes
Obs.	16,170	16,170	16,170	16,170	16,170
Adj. R <sup>2</sup>	0.000127	0.000829	0.0122	-0.00595	0.00541
Mean dep. var.	0.324	0.324	0.324	0.324	0.324
S.D. dep. var.	2.600	2.600	2.600	2.600	2.600

Source: Authors' analysis based on Gross National Income (GNI) estimates and data on the COVID-19 pandemic.

Note: Column (1) estimates the impact of belonging to the second or third tercile of income in 2012 on the income growth rate of districts, without controlling for fixed effects. The remaining columns explore how the COVID pandemic influences this relation, controlling for different fixed effects: column (2) includes no fixed effects, column (3) controls for time fixed effects, column (4) controls for district fixed effects, and column (5) controls for both time and district fixed effects. The sample includes all 147 Pakistani districts from January 2012 to March 2021. Standard errors are clustered at the district level. The number of observations and adjusted R<sup>2</sup> (Adj. R<sup>2</sup>) for each regression are reported at the end of the table. The last row presents the mean of the dependent variable (Mean dep. var.). \*\*\*, \*\*, and \* indicate significance at the 1 percent, 5 percent, and 10 percent levels, respectively.