Economics Working Paper Series
Faculty of Business, Economics and Law, AUT

Forecasting GDP growth from outer space

Jaqueson K. Galimberti

2020/02

Forecasting GDP growth from outer space*

Jaqueson K. Galimberti[†]
Auckland University of Technology

Abstract

We evaluate the usefulness of satellite-based data on night-time lights for forecasting GDP growth across a global sample of countries, proposing innovative location-based indicators to extract new predictive information from the lights data. Our findings are generally favorable to the use of night lights data to improve the accuracy of model-based forecasts. We also find a substantial degree of heterogeneity across countries in the relationship between lights and economic activity: individually-estimated models tend to outperform panel specifications. Key factors underlying the night lights performance include the country's size and income level, logistics infrastructure, and the quality of national statistics.

Keywords: night lights, remote sensing, big data, business cycles, leading indicators.

JEL codes: C55, C82, E01, E37, R12.

Links to supplementary files:

- ▷ Country-specific reports (25 MB, Dropbox): http://bit.ly/2Jx26VM
- ▷ Documented cases of light changes (52 MB, Dropbox): http://bit.ly/2OnyPN8
- ▷ Night lights time series indicators: http://bit.ly/2IAN0v4
- ▷ Individual predictive tests results: http://bit.ly/2qFE5Ce

*The research summarized in this paper was presented previously to the audience of several meetings: the 36th International Symposium on Forecasting (Santander, 2016), the 23rd International Conference Computing in Economics and Finance (New York, 2017), the 70th European Meeting of the Econometric Society (Lisbon, 2017), the SECO/SNB/KOF Meeting (Zurich, 2017), the Econometrics and Business Statistics Department at Monash University (Melbourne, 2017), the 1st Vienna Workshop on Economic Forecasting (Vienna, 2018), the JRC-CAS Workshop on Big Data and Macroeconomic Forecasting (Ispra, 2018), and the 34th Conference of the Centre for International Research on Economic Tendency Surveys (Rio de Janeiro, 2018). I thank the comments and suggestions made by several participants of these meetings, my former colleagues at the Chair of Applied Macroeconomics and KOF/ETH Zurich, the editor and two anonymous referees. Financial support from the SAS-IIF grant is gratefully acknowledged; I also thank the corresponding grant committee for the feedback.

[†]Corresponding author: School of Economics, Auckland University of Technology, Auckland, New Zealand.
Personal website: https://sites.google.com/site/jkgeconoeng/.
Contact e-mail: jaqueson.galimberti@aut.ac.nz.

1 Introduction

Forecasts of economic activity are crucial to the decision-making process of policymakers and market participants in general. A premise for informed economic decisions is to have a proper expectation of the future state of the market of interest. In practice the decision-maker is then continuously faced with an intricate forecasting challenge of finding leading indicators for the variables that are relevant to her/his business. In this paper we propose and evaluate the usage of satellite-based data on night-time lights for the prediction of GDP growth across a global sample of countries. Our main contribution is the design of innovative measures for the extraction of predictive signals of macroeconomic activity from the richness of information provided by the night lights dataset.

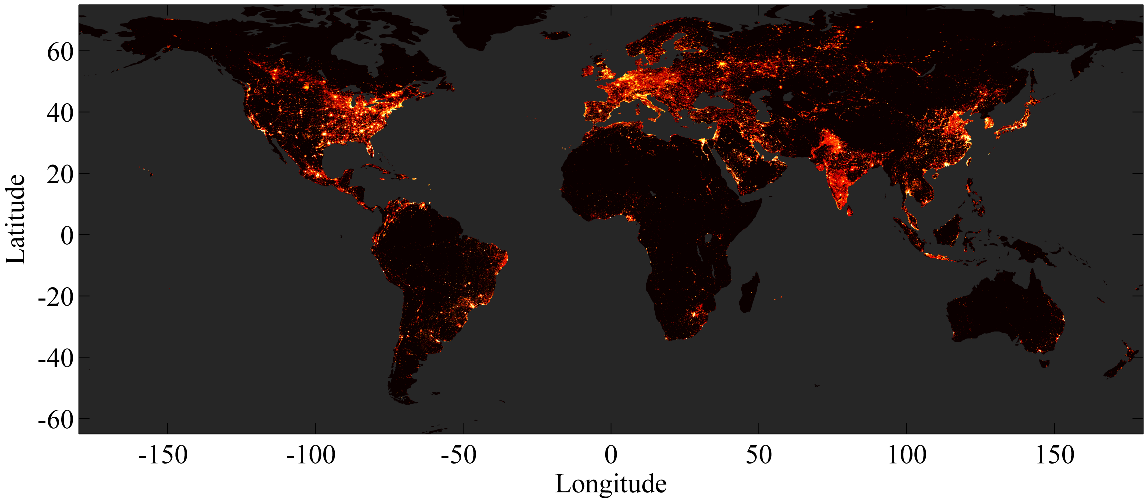
The night lights data consist of gridded observations of light intensities captured across the globe during night time. An illustrative snapshot of this global dataset is presented in figure 1.1. In order to evaluate the usefulness of these data for economic forecasting we construct Sum of Lights (SoL) measures, aggregating the intensities of lights observed within the borders of each country. One key innovation in this paper is the development of alternative location-based SoL indicators, designed to focus on the lights emitted from selected areas instead of the entire country's territory; particularly, we propose focusing on areas showing (significant) positive/negative correlations with the country's history of GDP growth rates. As our results reveal, a substantial portion ($\approx 85\%$ on average) of the lights signals observed over a country's territory are not significantly correlated with the country's aggregate production, and would therefore only add noise to an indiscriminate aggregation of the country's lights. We show how a proper classification of these geo-located signals can lead to a substantial improvement of the accuracy of the night lights-based forecasts.

Another important contribution of this paper is an examination of alternative assumptions on the cross-country specification of the relationship between light emissions and economic growth. Namely, we question the common practice of assuming that this relationship is homogeneous across countries, and show that there are substantial accuracy improvements to be achieved as well by allowing for individual country or partially pooled specifications. We also find that the heterogeneity of performances of the night lights-based forecasts can be associated with some country-specific factors, such as the country's size, income level, expenditure and production composition, logistics infrastructure, and quality of national statistics.

1.1 Motivation and relation to literature

The use of night lights data has been prominent in the recent economic literature, with applications that range from the geographical mapping of economic activity (Sutton and Costanza, 2002; Doll et al., 2006; Ghosh et al., 2010), to regional development analysis (Michalopoulos and Papaioannou, 2013a,b), to the evaluation of the accuracy of national income accounts

Figure 1.1: Snapshot of world stable night lights averaged over the year of 2013.



Notes: Based on OLS stable lights data obtained from NOAA-NGDC, as detailed in Section 3. The averaged night lights intensity measures are depicted in logarithmic scale after satellite intercalibration (see Appendix A.2) and normalization of grid cell areas to Earth's curvature across the latitudinal dimension.

(Chen and Nordhaus, 2011; Henderson et al., 2012; Nordhaus and Chen, 2015; Pinkovskiy and Sala-i Martin, 2016); see also Donaldson and Storeygard (2016) for a more general review of applications of satellite-based data in economics. In order to construct comparative measures of living standards across countries and regions, these studies have focused either on time-averaged relationships, hence taking advantage mainly of the geographical dimension of the luminosity data variability, or on the contemporaneous relationship between light emissions and economic activity.

Here, in contrast, we focus on the (lagged) time variations in the intensity of night lights within a country, evaluating their usefulness to improve the accuracy of forecasts of economic activity. To the best of our knowledge this is the first application of the night lights data to economic forecasting at a global scale[1], and this gap in the literature seems to be associated with the difficulty in providing a rationale for seeing night lights as a leading indicator for the aggregate economy. Thus, in order to challenge this view, we motivate our exercise by discussing some potential channels through which lagged changes in light emissions can anticipate current changes of a country's GDP.

One possible mechanism determining the usefulness of night lights data for GDP forecasting is related to the measurement error hypothesis, also explored by Henderson et al. (2012) for (contemporaneous) improvement of GDP measures. Namely, because GDP statistics are subject to measurement errors, due to, for example, informal economic activity or less developed statistical agencies, the night lights data can provide an alternative proxy to economic activity.

[1]Focusing on the case of China, Zhao et al. (2017) used the lights data for forecasting economic activity at different levels of regional aggregation.

Using a similar analytical framework, we show that lagged observations of lights growth also provide predictive information on GDP growth rates as long as economic activity is serially correlated; that is the case even after accounting for the persistence in measured GDP growth rates with an autoregressive benchmark model.

Another observation that motivates the use of the night lights data for economic forecasting comes from a time to build argument, along the lines of the seminal contribution of Kydland and Prescott (1982) to macroeconomic modeling. Namely, because productive capital, like factories and infrastructure, takes time to build and, importantly, generates lights during that time, tracking the variation of geo-located lights can signal future increases of production that will be realized when the construction is complete. In fact, assisted by an automated algorithm, we have managed to document several cases around the world where light changes anticipated broader regional economic development. Besides, one could also argue that the geographical spread of production chains can favor the use of location-based signals as predictors of final output. Namely, changes in economic activity at one place can signal further changes to come at different geographical locations that will end up affecting future aggregate measures of economic activity.

1.2 Approach summary

We process the night lights indicators for a sample of 167 countries at an annual frequency over the period from 1992 to 2013.[2] We then construct one-step-ahead GDP growth forecasts based on a benchmark first-order autoregressive (AR) model, and on AR(1) models augmented by the lagged values of the night lights indicators.[3] Here we also distinguish between two alternative specifications with respect to the estimation of these models, namely, a panel and country-individual specifications. As a robustness check we also consider alternative partial pooling specifications in an attempt to fine-tune the observed trade-off between heterogeneity and estimation uncertainty in the relationship between the night lights indicators and GDP growth.

We then proceed with the comparative evaluation of the accuracy of the night lights-based forecasts relative to the benchmark model forecasts. We consider two main exercises with respect to the sample used for the estimation of the models' parameters. First, an in-sample

[2]One major limitation of the night lights data available and employed here and in most of the literature is the time dimension, particularly because the satellite images need to be averaged over the annual frequency in order to reduce weather and other seasonal effects to provide more accurate measurements (there are exceptions, as, for example, Ishizawa et al., 2017, who used monthly lights data to measure economic recovery after natural catastrophes); here we attempt to compensate for the short time dimension with a broad sample in the cross-section dimension.

[3]In spite of their simplicity, univariate autoregressive models are commonly used as benchmark for GDP growth forecasts in the literature and are often found to be hard to beat with more sophisticated models (see, e.g., Chauvet and Potter, 2013); our focus on first-order autoregressions is due to the limited sample of annual night lights observations available, as explained in the previous footnote, although we obtain similar results using an AR(2) model as benchmark.

evaluation, where the full sample of data is used for both the estimation and the evaluation of the conditional predictions. Second, a recursive out-of-sample evaluation, where only data up to the forecasts base periods are used for estimation, starting with forecasts for 2001. Our main measure of evaluation is the night lights-based forecasts' root mean square error (RMSE), particularly relative to those obtained by the benchmark forecasts. We also evaluate the statistical significance of our results by conducting bootstrapped tests for the comparison of the predictive accuracy of nested models.

1.3 Results summary

Overall, we find evidence favorable to the use of night lights data for GDP growth forecasting, particularly with individually-estimated models, which achieve in-sample RMSE accuracy improvements ranging from 3% to 9% (cross-country GDP-weighted averages) relative to the benchmark model. Among the night lights indicators, we find that those based on the location of lights provide the greatest improvements to the accuracy of the GDP growth forecasts. In fact, with a relaxed assumption on the timing of the classification of lights, namely, when the gridded correlations are calculated using the full-sample, the average relative improvement rises to a remarkable 37% rate.

Out-of-sample, the performance of the individual specifications deteriorates substantially under the recursive estimation approach, a result that we attribute to the estimation biases caused by the use of too small samples at the country-individual level. Notwithstanding, the night lights approach still attain out-of-sample improvements for a substantial fraction of countries: about 46% on average across the indicators, most of which being statistically significant at the 20% significance level; focusing on the full-sample correlation-based indicator, statistically significant improvements are obtained for more than 76% of the countries in our sample. The magnitude of these out-of-sample improvements also vary quite substantially across the countries, 47 of which achieve gains in the range between 20-61%, 42 in the range between 10-20%, and 39 in the range between 0-10%.

What explains all this variation in the usefulness of the night lights across countries? We attempt to answer this question by evaluating several country-specific factors that could be associated with the performance of the night lights. Interestingly, our analysis indicates that the night lights appear more useful for economic forecasting in bigger countries, but to a lesser extent in low income countries. Besides, we also find that countries with lower consumption expenditure, smaller agriculture sector (both as share of total GDP), better logistics infrastructure, and better national statistics, tend to obtain better forecasts with the night lights data.

1.4 Paper organization

Beyond this introduction, and some concluding remarks by the end, the paper proceeds into six sections: Section 2 discusses the potential channels that can turn lights data into useful leading indicators of economic activity; Section 3 describes the night lights data and the construction of the associated leading indicators; Section 4 outlines the forecasting model specifications, their estimation, and the forecasting evaluation approach; Section 5 presents the main forecast evaluation results; Section 6 provides supplementary results on statistical significance and partial pooling alternatives; and Section 7 presents an analysis of potential explanatory factors of the night lights cross-country performance.

2 Lights as leading indicators

Underlying the use of night lights data to predict GDP is the hypothesis that the emission of lights indicates the presence of economic activity.[4] Clearly, the direction of causality between these variables goes from GDP to lights, namely, it is the human activity on the ground that generates the lights that are captured by the satellites. Nevertheless, for forecasting purposes our main interest is to uncover potential channels through which lagged changes in light emissions anticipate current changes of a country's GDP. In this context we discuss two main possible mechanisms that can turn the night lights data into useful leading indicators: (i) a measurement errors hypothesis; and, (ii) a time to build argument.

2.1 Measurement error hypothesis

The use of night lights data for forecasting economic activity can be motivated in the context of a GDP measurement error statistical framework along the lines of Henderson et al. (2012). Particularly, let a country's real GDP growth statistics be affected by measurement errors according to

$$y_t = z_t + u_t, \tag{2.1}$$

where y_t stands for measured real GDP growth, z_t for true real GDP growth, and u_t for the measurement error due to, for example, informal economic activity or mismatches between the national accounts and the changing structure of the underlying economy (see Landefeld et al., 2008). Furthermore, assume lights are generated with economic activity according to

$$x_t = \beta z_t + e_t, \tag{2.2}$$

where x_t stands for measured lights growth, β for the elasticity of lights with respect to true real GDP, and e_t for a measurement error in this relationship due to, for example, satellite sensor's noise (see Section 3.1) or changes in how production is translated into lights within a

[4] Whereas this relationship is on the basis of this paper and the previous applications in the literature, it is important to note that the relationship between night-time emission of lights and economic activity is not guaranteed so as to make of the former a substitute for traditional sources of macroeconomic data (see, e.g., Mellander et al., 2015, for an assessment of the night lights data as a proxy for economic activity). Throughout this paper we argue, and present evidence, in favor of the night lights data as a complement to other sources of data.

country; as discussed in Henderson et al. (2012, ps. 1006, 1021) the latter could be caused by changes in the sectoral composition of a country's GDP, where some industries may generate more lights than others, or a nonlinear relationship between development and lights emission due to urbanization and technological effects.

Using the framework above, Henderson et al. (2012) showed how an improved estimate of true GDP growth can be obtained by combining national statistics with measured lights growth. The argument follows from a well known result in measurement error analysis according to which different error-prone measures can be combined to recover the true value of the common target variable. Following a similar rationale, we show how lagged light measurements can provide useful predictions of GDP growth. However, in contrast to Henderson et al. (2012), we are interested in the use of lights data for the prediction of *measured* GDP growth data, namely, y_t, which is the observable we have available for forecast evaluation purposes.

In the context of our empirical application, the lights-based forecasts are obtained by augmenting a benchmark AR(1) model with lights growth indicators. Abstracting from country-specific factors and intercepts (the complete forecasting model specifications will be described in Section 4), the benchmark forecasts are given by the fitted values of an AR(1) model,

$$\hat{y}_t = \hat{\rho} y_{t-1}, \tag{2.3}$$

and the lights-based forecasts by

$$\tilde{y}_t = \hat{\varrho} y_{t-1} + \hat{\theta} x_{t-1}, \tag{2.4}$$

where $\hat{\rho}$, $\hat{\varrho}$, and $\hat{\theta}$ stand for estimated parameters. In fact, under the framework of equations (2.1) and (2.2), we can derive the theoretical OLS estimates of these parameters and compare the implied accuracy of these models' forecasts (see online appendix A.1 for these derivations). As expected, we find that the lights-based forecasts outperform the benchmark model as long as $\beta \neq 0$; it is also important to note that true GDP is required to be serially correlated for the lagged specifications to be relevant. The improvement obtained with the lights indicator, as measured by a ratio between the forecasts mean squared errors, is then given by

$$\frac{\beta^2 \sigma_u^2}{\beta^2 \sigma_z^2 + \sigma_e^2 \left(1 + \sigma_z^2 / \sigma_u^2\right)}, \tag{2.5}$$

where σ_z^2, σ_u^2, and σ_e^2 stand for the variances of z_t, u_t, and e_t, respectively. Intuitively, the usefulness of the lights data increases with $|\beta|$ and the magnitude of the measurement error in GDP statistics, σ_u^2, and decreases with the magnitude of the measurement error in the growth vs. lights relationship, σ_e^2.

2.2 Lights and time to build examples

It is well known that productive capital, like factories and infrastructure, takes time to build (Kydland and Prescott, 1982). Whereas this suggests the investment component of GDP as an important predictor of future growth, it is not always the case that such decomposition in national statistics is readily available.[5] In this context, the night lights data can provide an alternative predictor, particularly considering that such investments often generate intense lights during construction time. That is, tracking the variation of geo-located lights can signal future increases of production that will be realized when the construction of such production facilities is complete.

In order to further substantiate this argument we document some cases where light changes anticipated regional economic development. With the assistance of an automated algorithm to select locations with substantial light changes over time, we have documented a few dozen such cases across the world, including the development of industrial/economic zones, oil/gas extraction and processing plants, hydroelectric and mining projects, and urban sprawls. Animated snapshots of the observed night lights and Google Earth images (Gorelick et al., 2017) around these locations are available in a supplementary file. Here, for illustrative purposes, we focus on two of these cases.

Gravatai Automotive Industrial Complex (Brazil): figure 2.1 shows the case of the construction of a General Motors factory in the city of Gravatai, part of the metropolitan region of Porto Alegre, Brazil. The plant location decision was announced in 1997, which was also when its construction started. The factory opened and started producing in 2000.

Zhengzhou Airport Economy Zone (China): figure 2.2 shows the case of the development of an industrial zone next to the Zhengzhou Xinzheng International Airport, in the Henan Province, China. The development was approved in October 2007, and expanded into a comprehensive bonded zone, which are special customs areas providing favorable taxation policies to bonded processing and logistics of goods for trade, in October 2010. Foxconn's "iPhone city" is located in the area.

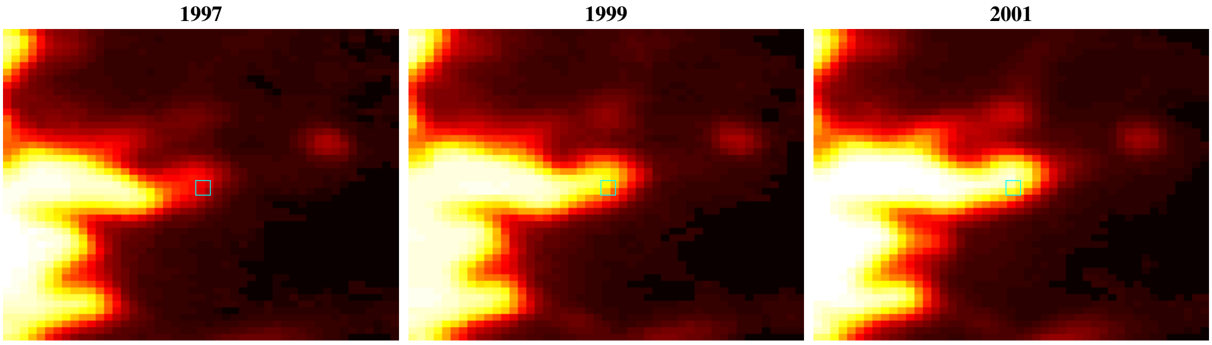
2.3 Other potential channels

Considering the richness of the night lights data there are certainly other plausible channels through which lagged lights can turn into useful leading indicators for economic activity. Informal economic activity may be one important factor that is more promptly captured by the lights data than other sources. That may be particularly relevant for forecasting purposes in

[5] Also notice that there is some ambiguity over how in-progress capital formation and inventories are recorded in national statistics due to issues in determining the timing of assets ownership (see United Nations, 2009, p. 108).

Figure 2.1: Gravatai Automotive Industrial Complex, Brazil.

(a) Snapshots of averaged night lights.

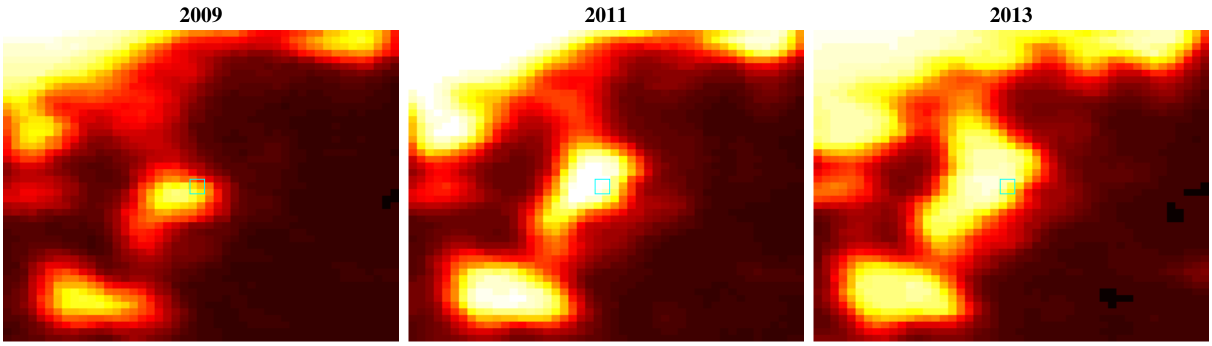


(b) Corresponding Google Earth Zoomed Images.

Notes: The selected case is located at 29.9373° South and 50.9147° West, and is marked in the night lights snapshots with a square symbol. Link to Google maps: https://www.google.com/maps/@-29.9373,-50.9147,10000m.

Figure 2.2: Zhengzhou Airport Economy Zone, China.

(a) Snapshots of averaged night lights.



(b) Corresponding Google Earth Zoomed Images.

Notes: The selected case is located at 34.5406° North and 113.8546° East, and is marked in the night lights snapshots with a square symbol. Link to Google maps: https://www.google.com/maps/@34.5406,113.8546,10000m.

countries where businesses tend to start operations informally, but end up entering the formal economy (and GDP statistics) only after succeeding to mature.[6] Another potential channel comes from the geographical spread of production chains, which can favor the use of location-based signals as predictors of final output. Namely, changes in economic activity at one place can signal further changes to come at different geographical locations that will end up affecting future aggregate measures of economic activity.

Altogether, we take these potential channels as a motivation for the evaluation of the usefulness of the night lights data for GDP growth forecasting that follows in the remainder of this paper.

3 Night lights data and indicators

3.1 Sources and issues

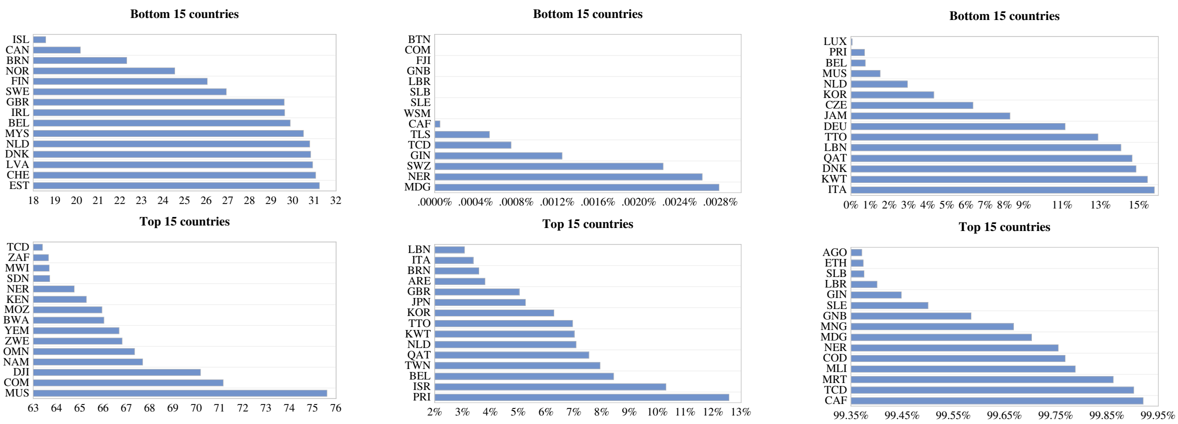
Satellite imagery data on night lights are obtained from the Earth Observation Group (EOG) at the National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center (NGDC), and come in the form of annual composite images representing the intensity of lights captured by sensors on-board the Operational Linescan System (OLS). These images are produced by EOG scientists by averaging cloud-free observations of night lights and cover Earth's surface between 75 degrees north and 65 degrees south latitude. The intensity of the night lights radiance are converted into 6-bit digital number (DN) values, ranging between 0 and 63, and allocated over a global grid of 30 arc second cells according to their geographic location. We use the stable lights version of the data, which focus on persistent lighting sources obtained through the application of a background noise filtering algorithm (Elvidge et al., 2003).

The night lights annual composites cover the period between 1992 and 2013, and are based on data from a total of six satellites, some operating simultaneously; hence, for some years two composite images have been produced, in which cases we adopt their weighted average after intercalibration, using each cell's number of cloud-free observations as weights. The intercalibration is necessary because the OLS has no on-board calibration of the visible band while sensor performance degrades over time, not to mention the evolution in sensor specifications across the launched instruments. These factors are particularly important for the comparison of night light emissions over time, and in order to account for them we adopt the regression-based intercalibration procedure proposed by Elvidge et al. (2009), which takes an area with low variation in emitted lights over time (Sicily) as reference to estimate re-scaling parameters across the satellite-year composites. Details on this procedure are provided in Appendix A.2.

[6] Although we have not found a statistically significant relationship between our results and third-party estimates of informality, we note that according to the World Bank Enterprise Surveys (2017), 11% of firms surveyed globally state to have started operations without being formally registered, while this statistic rises to about 40% for countries such as Nigeria and Indonesia; in Bolivia, the average firm reports to have operated in the shadow economy for over four years (0.7 years for the global average).

Figure 3.1: Averaged statistics of night lights data for selected countries.

(a) Cloud-free observations. (b) Fraction of top-coded pixels. (c) Fraction of unlit pixels.



Notes: The statistics are averaged over the sample period 1992-2013 for each country in our sample (see text for excluded countries). Countries are denoted by their ISO alpha-3 code, listed in Appendix A.5.

There are a few other important issues that are known to affect the night lights data. First, the annual composite images are based on averaged cloud-free observations,[7] the availability of which can vary substantially across countries depending on weather conditions: from a minimum average (across the country's cells) of 3.05 cloud-free data points, observed in Iceland during the year of 1999, to a maximum of 103.24, observed in Mauritius during 2010. This restriction is particularly relevant for Nordic countries, as the averaged statistics in figure 3.1a indicate. Second, sensor saturation, caused by signals exceeding the sensor's detection range, interfere in the measurement of brighter sources of light. These signals are recorded with the highest DN value in the OLS scale (=63), or "top-coded",[8] and tend to be more frequently observed in the more densely populated countries; see figure 3.1b. Third, as evidenced in figure 3.1c, the focus on stable lights leads to a substantial increase in the fraction of unlit pixels, particularly in the more sparsely populated countries, which can affect the signals quantity-quality trade-off on the construction of location-based measures of lights.

Another important aggregation issue relates to the area underlying each cell in the gridded dataset. Due to the Earth's curvature, the area covered by each pixel depends on its latitude, for example: $0.85km^2$ at equator, $0.37km^2$ at S65º, and $0.22km^2$ at N75º. That is important for the aggregation of night lights at the country level because the closer the detected lights (and their changes) are to equator, the bigger their amplitude on the ground; to make these pixels

[7]Other than cloud coverage, data points are also discarded when any of the following features are present: sunlight and glare (scattered sunlight penetration into the telescope), moonlight, and lighting from the aurora (see Elvidge et al., 2003).

[8]Because the annual DNs are averages of daily observations, which, in turn, have been averaged from higher resolution images, it is possible that sensor saturation also affects signals coded at lower values (see Hsu et al., 2015; Bluhm and Krause, 2018). For that reason, our statistics on top-coded pixels are based on a threshold DN value of 90% the maximum value on the scale of each satellite's intercalibrated DNs.

comparable (and aggregable), we re-scale the gridded light intensity measures by multiplying them by their latitude-implied area.[9]

Before the computation of the night lights indicators (detailed below), the global composite images need to be processed for the extraction of light intensity measures within the countries borders. For that purpose we use the Database of Global Administrative Areas, GADM version 2.8 (http://gadm.org/), which contains definitions of 256 countries/territories borders across the globe. This sample reduces to 190 countries after matching the records to those of the International Monetary Fund's (IMF) World Economic Outlook (WEO) database (October 2017 vintage), which is our source of data on GDP. Notice the GDP data are unbalanced in the time dimension, with samples varying by country. Our sample of countries is finally reduced to 167 countries after excluding those with a population smaller than 100,000, a land area smaller than 1,000km^2, South Sudan for lack of earlier GDP data, and Equatorial Guinea for having most of its lights coming from gas flares. A list of the countries included in our sample is presented in Appendix A.5.

3.2 Night lights indicators

The geo-located time series data on night lights provide a potentially rich source of predictive information on economic activity. Naturally, there are several possible ways to extract this information, and different measures can be constructed on the basis of the night lights data to capture the evolution and geographical spread of economic activity. Here we distinguish between two types of indicators:[10] (i) aggregate indicators; and, (ii) location-based indicators.

Aggregate indicators have been the focus of most of the past literature looking at the relationship between economic activity and night light emissions (e.g., Ghosh et al., 2010; Chen and Nordhaus, 2011; Henderson et al., 2012; Pinkovskiy and Sala-i Martin, 2016). Here we focus on the country's **Sum of Lights** (SoL), which is obtained by simply summing up the light intensity DNs observed within that country's borders. Under the hypothesis that more (less) lights means more (less) production, here we use SoL growth rates (log changes for every growth rate throughout the paper) as a predictor for GDP growth.

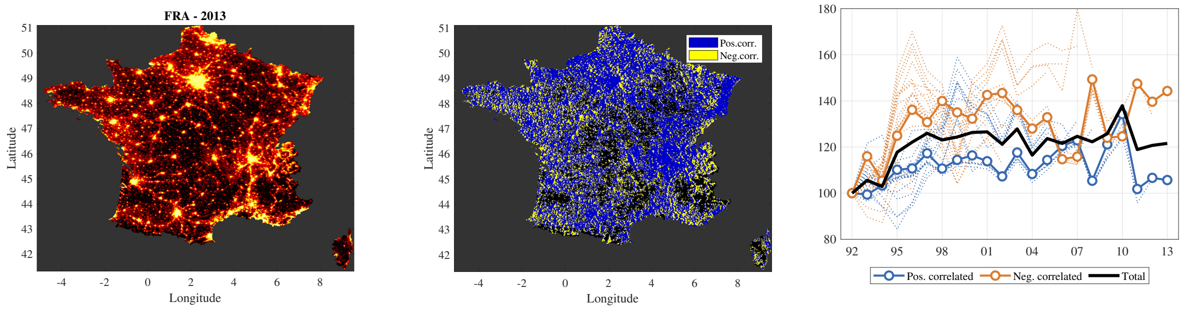
One potential weakness of the aggregate SoL indicator is that it does not account for the quality of the signals coming from different locations within the country's territory. Namely, by pooling all the country's lights the SoL indicator can be affected by noisy signals from locations that have little correlation with economic activity, potentially missing the predictive content

[9]Further issues are known to affect the spatial resolution of the night lights data, though of secondary importance for our purposes: the spatial precision of the night lights data is affected by "blooming" effects, that is, a tendency to overestimate the true extent of lit area on the ground (see Doll, 2008); also, there is some overlap between pixels because the value assigned to each of them is based on an on-board smoothing algorithm that averages blocks of pixels from a finer resolution image (see Elvidge et al., 2004).

[10]In a previous version of this paper we have also considered distribution-based indicators (for example, the DNs median, the DNs kurtosis, etc), which, in spite of some merit on the cross-country characterization of light emission patterns, were found to have too limited time variation to be useful for forecasting purposes.

Figure 3.2: Correlation-based classification of lights for France.

(a) Snapshot of night lights. (b) Correlated pixels. (c) Time series.



Notes: The snapshot image in panel (a) is produced according to figure 1.1 notes, though "turning off" the lights outside the country's borders. The pixels' classification in panel (b) is based on the correlations (regardless of significance) between GDP growth and lagged lights growth using observations up to 2013. The colored lines in panel (c) depict the evolution of the sum of lights (1992=100) over pixels classified according to the past correlation of their lagged lights growth and the country's GDP growth, in this case summing up only those significant at the 25% level; solid lines refer to the classification obtained using observations up to 2013, whereas the dashed lines refer to previous vintages.

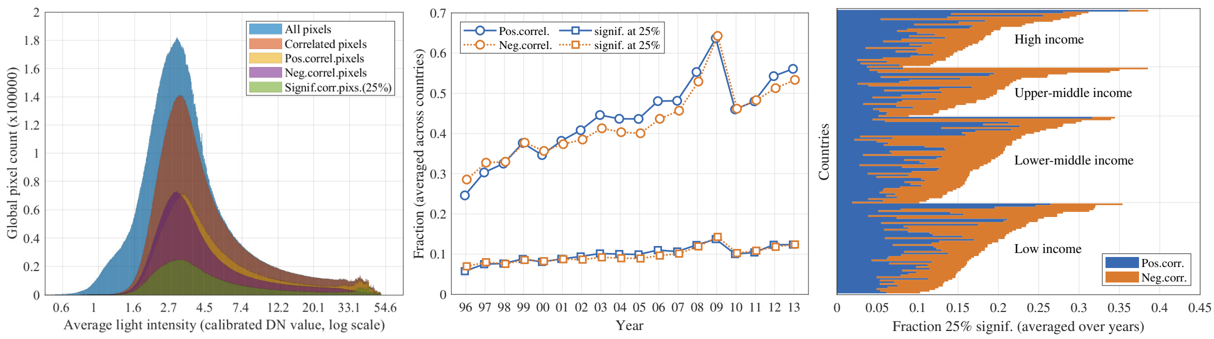
from relevant locations because of opposite lights variation from locations with less informative signals. In an attempt to circumvent this issue we propose the use of location-based measures. Here the idea is to decompose a country's SoL by dividing its pixels according to a given criterion. Particularly, we propose a classification of the country's pixels according to their past correlation with economic activity. This is done by constructing pixel-by-pixel time series of light intensity changes, and then calculating their correlations with the country's aggregate times series of GDP changes. We then distinguish between two sets of pixels leading to two new indicators: **Positive/Negatively correlated pixels SoL**. An illustration of this correlation-based classification is presented in Figure 3.2 for the case of France.

The vintage of data used for the calculation of the pixels' correlations is an important determinant of the predictive quality of the correlation-based SoL indicators. Here we consider two possibilities: (i) a *real-time* classification (dashed lines in Figure 3.2c), where the pixel correlations are (re-)calculated recursively according to the data availability, that is, the classification used in a given year is based only on the night lights and GDP data from the previous years; and, (ii) a *full-sample* classification (solid lines in Figure 3.2c), where the correlations are calculated on the basis of data available at the forecast base year of 2013. Naturally, these alternative classifications are motivated from different forecast evaluation goals. Whereas the former classification simulates the information restrictions of a real-time forecaster, the latter can be informative about the predictive value of the quality of the pixels classification. We shall return to this important distinction in our analysis of the forecasting results.

Finally, with the purpose of reducing the inclusion of noisy lights signals, the correlation-based indicators can be further specialized to account only for pixels with statistically signifi-

Figure 3.3: Correlated pixels statistics.

(a) Distribution by intensity. (b) Fractions by year. (c) Fractions by country.



cant correlations. Namely, before summing up the lights in each classification (positive/negative) we also calculate the associated significances of the pixel correlations and discard those pixels where the correlation significance was above a given threshold value. For that purpose we have found that a p-value threshold of 25% yields the best results on average without affecting the feasibility of the indicator across countries.

Figure 3.3 presents some statistics about the distribution of correlated pixels underlying the construction of this new indicator. First, on a global scale, we can see from Figure 3.3a that the distribution of correlated pixels tends to be concentrated on pixels with higher average light intensity: this is reasonable as it indicates that places with low levels of light emissions (left tail of the distribution) are not related to economic activity; it is also a good sign that our approach is able to filter out so many noisy signals. Another interesting finding is that the negatively correlated pixels tend to be skewed towards less intensively lit locations, while most pixels with a higher light intensity (right tail of the distribution) show a positive correlation with aggregate economic activity.

Furthermore, panels 3.3b and 3.3c present statistics on the fractions of correlated pixels at the country level. First, over time, we can see there is an increasing trend on the average fraction of pixels identified as correlated with economic activity, although this trend is less pronounced when focusing on the significant correlations. While this may be attributed to the augmenting samples used for these calculations as time goes on, there is a marked decline between 2009-10, possibly a lagged effect of the global financial crisis of 2007-08. Finally, figure 3.3c shows there is substantial heterogeneity in these pixel classifications across countries. Interestingly, neither the total fractions of correlated pixels, nor their balance between positive and negative correlations, appear to be related to the countries levels of development, as the fractions are also dispersed within each classification of income per capita according to the World Bank.

4 Forecasting approach

4.1 Model specifications

In order to construct forecasts for GDP growth we estimate both pooled and individual countries model specifications. As a benchmark we adopt a simple AR(1) model, as given by

$$y_{i,t} = \alpha_i + \rho y_{i,t-1} + \varepsilon_{i,t}, \tag{4.1}$$

for the pooled specification, and

$$y_{i,t} = \alpha_i' + \rho_i y_{i,t-1} + \epsilon_{i,t}, \tag{4.2}$$

for the individual specification, where $y_{i,t}$ stands for country i's $(= 1, \ldots, 167)$ GDP growth rate for year t $(= 1993, \ldots, 2014)$, and $\alpha_i^{(\prime)}$ for country fixed effects.[11] The use of univariate AR models as a benchmark for GDP growth forecasts is common in the literature. Chauvet and Potter (2013), for example, select an AR(2) model to evaluate the accuracy of more sophisticated multivariate models, such as vector autoregressions and dynamic factor models, for forecasting US quarterly GDP growth rates.[12] They find that the simpler AR(2) model is hard to beat, particularly during non-recessionary phases of the business cycle.

The night lights-based forecasts are obtained by augmenting the benchmark models with the night lights indicators discussed in the previous section.[13] Letting $\mathbf{x}_{k,i,t}$ stand for a vector containing indicator k, the augmented models are given by

$$y_{i,t} = \alpha_{k,i} + \varrho_k y_{i,t-1} + \boldsymbol{\theta}_k \mathbf{x}_{k,i,t-1} + \varepsilon_{k,i,t}, \tag{4.3}$$

$$y_{i,t} = \alpha_{k,i}' + \varrho_{k,i} y_{i,t-1} + \boldsymbol{\theta}_{k,i} \mathbf{x}_{k,i,t-1} + \epsilon_{k,i,t}, \tag{4.4}$$

for the pooled and individual specifications, respectively, where the parameter vectors $\boldsymbol{\theta}_{k(,i)}$ have dimensions conformable to the number of combined indicator measures. For example, for the standard SoL indicator, $\mathbf{x}_{k,i,t}$ is univariate, that is, containing only one indicator series at a time; for the cases of the location-based indicators, two SoL growth measures are produced

[11] We have also experimented with the inclusion of period fixed effects in all model specifications but have found that, whereas their inclusion can improve the robustness of parameter estimates to cross-correlated disturbances (for example, global shocks), it also tends to deteriorate the models' forecasting performance, particularly for the in-sample evaluation exercise, where the period fixed effects cannot be used for computation of conditional forecasts.

[12] Our focus on the AR(1) specification is due to the limited availability of annual night lights data and the effects that adding an extra parameter on the second lag of GDP growth rates can have on estimation uncertainty. Results taking the AR(2) model as a benchmark are reported in Appendix A.3 and show similar night lights improvement figures as those obtained under the AR(1) specification presented in the main text.

[13] Models including only the night lights indicators, namely, without the AR(1) term, yield poor forecasting performance relative to the benchmark, which is not surprising considering the relevance of persistence in the GDP series.

according to the decomposition of a country's pixels in a given year into positive and negatively correlated pixels. Also notice that the lights indicators are introduced with a lag so as to reflect our interest in 1-year-ahead forecasts; these specifications can be easily adjusted to instead use the contemporaneous relationship between lights and growth, more in the spirit of a nowcasting exercise.[14]

One important issue in the estimation of these models using a global sample of countries is the likely presence of outliers in the estimated relationships, mostly due to country-specific disruptive events such as wars and armed conflicts. Such outliers can introduce substantial biases in the estimation of the model parameters. To deal with this issue we adopt a two-stage estimation approach for outliers detection. First, we estimate the benchmark panel model specification with all available observations and derive the corresponding studentized residuals, obtained by dividing the raw residuals by an independent estimate the residual standard deviation. Outliers are then detected based on the statistical significance (p-value smaller than 1% under Student's t-distribution) of each disturbance; a total of 67 outliers are detected according to this procedure (these are listed in the Appendix A.5). In the second stage we obtain the final estimates of the models, both panel and individual specifications, excluding the detected outliers from the sample.

4.2 Model estimates

The estimates of the panel models, (4.1) and (4.3), as well as averages of their individual counterparts, (4.2) and (4.4), are reported in Table 4.1. Interestingly, the panel estimate for the lagged values of the standard total SoL indicator (All pixels) is not statistically significant, while the average of its individual estimates is only marginally significant at the 10% level of significance. In contrast, the estimates are more favorable to the correlation-based indicators, which are found to have statistically significant relationships with GDP growth, and reasonable signs on the estimated coefficients, that is, positive (negative) for the indicator reflecting the positively (negatively) correlated pixels.

The models using the correlated pixels indicators are also clearly superior in terms of explanatory power, as evidenced by the adjusted R^2 statistics. Besides, the individual specifications obtain even greater improvements to explanatory power by allowing for heterogenous relationships between the lights indicators and GDP growth. Nevertheless, this improvement comes at the cost of an increase in the number of estimated parameters, hence leading to less parsimonious models, which, in turn, can have a negative effect on the accuracy of the estimates and their implied forecasts. This tradeoff between parsimony and fit is captured by the reported information criteria, namely the Bayesian and the Akaike IC, which only differ in terms of how

[14] Although the data we use in this paper are available for free only at the annual frequency, there is an obvious potential for more timely products on the basis of the raw daily images used to construct the annual composites. Besides, monthly data have been produced since 2013 from the more recently launched Suomi National Polar-orbiting Partnership satellite's Visible Infrared Imaging Radiometer Suite (VIIRS) sensors.

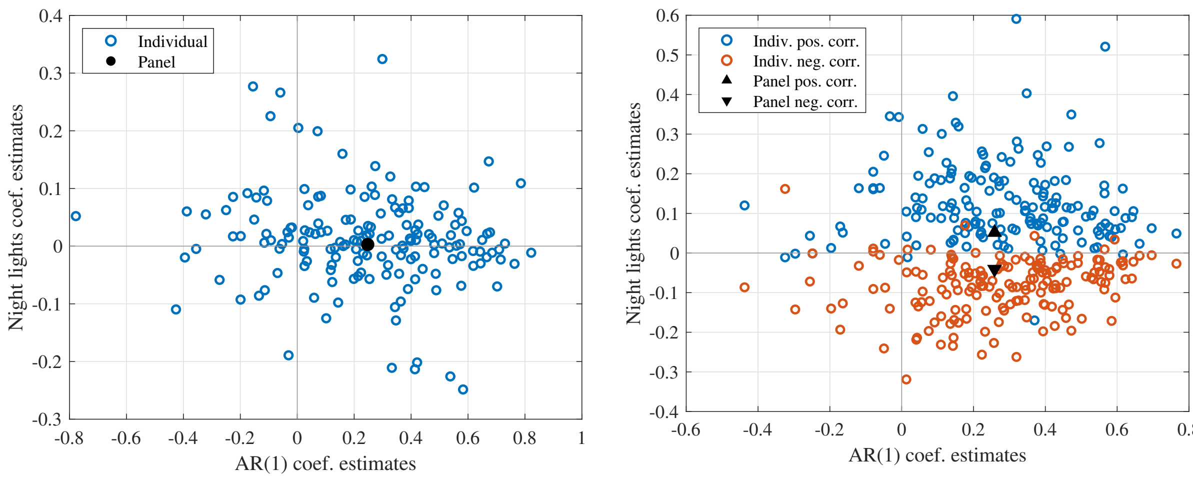
Table 4.1: Models estimates.

Predictors	Panel estimates			Individual estimates (averaged)		
	(1)	(2)	(3)	(1')	(2')	(3')
Lag GDP growth	0.256***	0.248***	0.259***	0.246***	0.245***	0.275***
	(0.027)	(0.029)	(0.026)	[0.022]	[0.022]	[0.017]
Lag SoL growth:						
All pixels		0.003			0.011*	
		(0.004)			[0.006]	
Positively corr. pixels			0.052***			0.127***
			(0.009)			[0.008]
Negatively corr. pixels			-0.041***			-0.083***
			(0.005)			[0.005]
N. observations	3598	3334	3316	3598	3334	3316
N. estim. params.	168	169	170	334	501	668
R^2 (adjusted)	0.277	0.279	0.403	0.338	0.348	0.608
Bayesian IC	5.526	5.508	5.313	5.767	6.103	5.937
Akaike IC	5.237	5.198	5.000	5.193	5.185	4.707

Notes: Estimates obtained regressing real GDP growth on the corresponding predictors and country fixed effects. Estimation by least squares in two stages: first, with all available observations; second, excluding outliers (67 in total; see text for details). Values inside parentheses are country-clustered robust standard errors, while values inside square brackets are standard errors of the means of the individual estimates. ***, **, and * stand for 1%, 5%, and 10% levels of statistical significance, respectively. The Bayesian and Akaike information criteria are calculated as $(-2\log\mathcal{L}+p\log N)/N$ and $(-2\log\mathcal{L}+2p)/N$, respectively, where $\mathcal{L}$ is the value of the normal likelihood function, p is the number of estimated parameters, and N is the number of observations.

Figure 4.1: Individual estimates of forecasting models.

(a) Total SoL. (b) Correlation-based SoL.



severe the number of estimated parameters is penalized. Here there is mixed evidence for the individual models relative to the panel estimates: whereas the more severe BIC deteriorates (increases), the AIC indicates the increase in the number of parameters is worthwhile. Also notice the AIC always improves (decreases) with the inclusion of additional night lights indicators within both the panel and the individual specifications.

The assumption of a common relationship between the night lights indicators and GDP growth across countries, as implied by the panel specifications, is further put into question when we look at the individual estimates for the country-individual model specifications, as depicted in Figure 4.1. Namely, we observe a wide range of individual parameter estimates: the coefficients associated with the total SoL indicator, for example, range from -0.25 to 0.33. Although the estimates for the correlation-based indicators tend to be consistent with their expected signs, it is clear that even in this case the individual estimates are too dispersed to justify pooling. Interestingly, the individual AR(1) coefficient estimates are also found to be widely dispersed in relation to their panel estimates. Hence, in spite of the likely higher estimation uncertainty in the individual specifications (eqs. 4.2 and 4.4), due to the use of smaller samples of data, it seems important to give full consideration to this alternative on the evaluation of the predictive performance of the night lights-based forecasts.

4.3 Evaluation exercises

In order to evaluate the quality of the night lights indicators as predictors of annual GDP growth we conduct two main exercises, differing mainly with respect to the sample used for the estimation of the model parameters and evaluation of the forecasts. More formally, under the model specifications described above, the construction of one-step-ahead conditional GDP growth forecasts, $\hat{y}_{k,i,t+1}$, can be generically expressed as given by

$$\hat{y}^{\mathcal{S}}_{k,i,t+1} = \hat{\alpha}^{\mathcal{S}}_{k,i} + \hat{\varrho}^{\mathcal{S}}_{k(,i)} y_{i,t} + \hat{\boldsymbol{\theta}}^{\mathcal{S}}_{k(,i)} \mathbf{x}_{k,i,t}, \tag{4.5}$$

where the superscript $\mathcal{S}$ is introduced to denote the sample used in the estimation of the model parameters, and the $(,i)$ subscript distinguishes between the panel and individual countries specifications.

First, we evaluate the models' **in-sample** predictive performance. To that purpose we construct GDP growth forecasts for every year in the period from 1993 to 2014, namely, with $t = \{1992, \ldots, 2013\}$ in equation (4.5), using model parameters estimated with our full sample of data, that is, with $\mathcal{S} = \{1992, \ldots, 2014\}$ in (4.5). Naturally, this is not a realistic real-time forecasting exercise considering that data beyond the forecast base period are normally not available to a forecaster estimating the forecasting model. To approach this issue we propose a second exercise to evaluate **recursive out-of-sample** (OS) forecasts. Namely, we restrict our evaluation to annual forecasts for the period from 2001 to 2014, constructed with model estimates based on an augmenting recursive sample; under the notation of equation (4.5), while $t = \{2000, \ldots, 2013\}$, $\mathcal{S}_t = \{i\}^{t}_{i=1992}$. Importantly, for the OS exercise we have found that the effects of estimation uncertainty can be reduced for the night lights-augmented models by fixing the AR(1) and intercept parameters to the recursive benchmark estimates, while (re-)estimating only the night lights relationships.[15]

Our main measure of evaluation is the forecasts' root mean squared errors (RMSE), calculated as usual for each country and model specification. We then construct the night lights RMSE ratios in relation to the AR(1) benchmark RMSE, where values below one indicate the former outperformed the benchmark, and vice versa. Considering that we have a total of 167 countries in our sample, we synthesize our evaluation by averaging the RMSEs across countries, using country GDP (in PPP terms) as weights. A similar weighted averaging is applied to summarize the RMSE ratios, except that these are averaged geometrically.

5 Forecast evaluation

5.1 Averaged statistics

The averaged results for the main evaluation exercises are presented in Table 5.1. Starting with the in-sample results, in panel (i), we observe that the usefulness of the night lights data depends on whether the forecasting model was estimated pooling all the countries together or individually. Here the evidence is in favor of the individual estimation, yielding in-sample predictions between 5% and 27% more accurate than the panel estimated models. Across the night

[15] Using equation (4.5)'s notation this is equivalent to setting $\hat{\alpha}^{\mathcal{S}_t}_{k,i} = \hat{\alpha}^{\mathcal{S}_t}_{i}$, $\hat{\varrho}^{\mathcal{S}_t}_{k(,i)} = \hat{\rho}^{\mathcal{S}_t}_{(i)}$, for all k, while $\hat{\boldsymbol{\theta}}^{\mathcal{S}_t}_{k(,i)}$ is obtained by regressing $\left(y_{i,j+1} - \hat{y}^{\mathcal{S}_t}_{i,j+1}\right)$ on $\mathbf{x}_{k,i,j}$ recursively with $j = \mathcal{S}_t$.

Table 5.1: Forecast evaluation statistics.

Model	Panel RMSE	Panel Ratio	Panel SSR	Individual RMSE	Individual Ratio	Individual SSR	Sample
(i) In-sample evaluation.							
AR(1) benchmark	2.203	-	65%	2.079	-	64%	3465
AR(1) + Lagged SoL indicators:							
+ Total SoL growth	2.201	1.00	66%	2.026	0.97	67%	3335
+ Correlated pixels SoL growth							
Real-time classification	2.216	0.99	68%	1.935	0.91	69%	2845
Full-sample classification	1.802	0.80	73%	1.322	0.63	77%	3316
(ii) Out-of-sample evaluation.							
AR(1) benchmark	2.400	-	64%	2.524	-	63%	2232
AR(1) + Lagged SoL indicators:							
+ Total SoL growth	2.401	1.00	64%	2.555	1.01	60%	2220
+ Correlated pixels SoL growth							
Real-time classification	2.531	1.07	63%	2.709	1.07	60%	2180
Full-sample classification	2.003	0.82	70%	1.998	0.78	69%	2188

Notes: All statistics are weighted cross-country averages using the countries GDPs (in PPP terms) as weights and are based on forecasts over the period from 1993 to 2014, for the in-sample evaluation, and from 2001 to 2014, for the out-of-sample evaluation. RMSE stands for root mean squared errors. SSR stands for the sign success ratio of predicted growth change. Ratios are first computed individually, relative to the corresponding benchmark specification over identical forecasting samples, and then geometrically averaged using countries GDPs as weights. The out-of-sample forecasts are constructed recursively with models estimated using only past observations available at the forecast base period. The real-time classification of pixel correlations are re-calculated for every forecast base period using only past data, while the full-sample classification is based on data up to 2013.

lights indicators, the correlation-based ones stand out, with accuracy improvements reaching up to 37% (full-sample classification) in relation to the benchmark model. An averaged measure of the predicted sign success (SSR) is also presented in Table 5.1, where the differences in performance suggest that part of the improvements brought by the night lights is due to better sign predictions too.

Interestingly, a greater improvement is obtained with the use of the full-sample instead of real-time vintages for the calculation of the pixel correlations. In spite of being infeasible for a real-time forecaster, the results obtained under this full-sample classification indicate the relevance of locations' classification for the accuracy of the night lights-based forecasts. Particularly, an understanding of the spatial evolution of light intensities at locations with higher predictive content for economic activity can be an interesting avenue for future research.

Turning to the out-of-sample results, presented in panel (ii) of Table 5.1, we observe a deterioration of the individual model's performances relative to their panel estimated counterparts. Whereas this could put into question our in-sample conclusions, favoring the individually-

Table 5.2: Frequencies of night lights improvements over benchmark model across countries.

Statistics / specifications	In-sample Panel	In-sample Individual	Out-of-sample Panel	Out-of-sample Individual
(i) Indicator-specific improvement rates:				
(a) Total SoL growth	63.5%	92.2%	49.7%	34.7%
	(13.2%)	(27.5%)	(18.0%)	(25.1%)
Correlated pixels SoL growth				
(b) Real-time classification	64.7%	97.6%	25.7%	21.6%
	(25.1%)	(37.7%)	(15.0%)	(25.1%)
(c) Full-sample classification	86.8%	97.6%	88.6%	80.8%
	(80.8%)	(70.1%)	(84.4%)	(76.6%)
Average: (a) + (b) + (c)	71.7%	95.8%	54.7%	45.7%
	(39.7%)	(45.1%)	(39.1%)	(42.3%)
(ii) Max-t frequency of rejections at 20% significance:				
Models set: (a) + (b)	18.6%	38.9%	17.4%	22.2%
Models set: (a) + (b) + (c)	75.4%	64.7%	61.1%	64.7%

Notes: The frequencies measure the percentage of countries for which the night lights augmented model outperformed the benchmark. Statistics in parenthesis refer to the cross-country frequency of rejections of the Clark and West (2007) (CW) one-sided test for equal predictive accuracy in nested models, at the 20% level of significance computed using 1,000 bootstrap replications according to the Clark and McCracken (2012) method for nested model reality checks. The Max-t reality check refers to multiple-model variant of the CW test.

estimated models, these results seem to be driven mainly by parameter estimation errors, due to the small samples available for the first recursive estimations. For example, the first individual recursive forecasts, for the year 2001, are based on model estimates obtained using, at best, merely 7 data points, from 1994 (the 1992-93 SoL changes are used as a lagged value) to 2000, for each individual country; compare that to the more than 1,000 observations used under the panel estimation and it is not surprising that the recursive OS exercise favored the latter. Nevertheless, relative to their corresponding benchmark, the OS performances of the night lights-augmented models are more uniform within the panel and individual specifications: here only the full-sample classification of pixels is capable of providing an average improvement, equal to 18% and 22%, for the panel and individual specifications, respectively.

5.2 Individual country performances

The averaged statistics can conceal the cross-country heterogeneity of performances. Particularly, the averages can be affected by large outliers that push the evaluation measures towards a direction that does not reflect the majority of the results. To approach this issue we now focus on the distribution of results across countries. In Table 5.2 we present statistics on the cross-country frequency within which the night lights indicators improved over the benchmark model's forecasting accuracy (test results will be discussed below in Section 6.1).

Overall, the results reported in panel (i) of Table 5.2 confirm our findings with the averaged statistics. Namely, the individual models achieve higher improvement rates than the panel ones under the in-sample exercise, whereas the opposite is observed under the recursive OS evaluation. Notwithstanding, these results also reveal that even in OS forecasting the night lights data can still be useful for a considerable portion of countries, providing improvements to about 55%, using the panel specification, and 46%, using individual specifications, of the countries in our sample, on average across the indicators. Focusing on the full-sample correlation-based indicator, as reported in panel (i)-(c), provides an even more favorable outlook for the night lights data with improvements being registered for more than 80% of our sample of 167 countries.

It is also interesting to look at how the usefulness of the night lights data varies across the countries in our sample. In Figure 5.1 we present the countries RMSE ratios, focusing on the forecasts obtained with the models augmented with the full-sample correlated pixels indicators, and grouped according to the World Bank income classification.[16] We note that the improvement rates reported above are found to be slightly skewed against the low income countries, with OS improvements with the individual (panel) models for only 58% (77%) of the countries in that group, compared to 94% (97%), 79% (93%), and 82% (88%), for the high, upper middle, and lower middle income groups of countries, respectively. The individual performances obtained with the panel model are also more stable across the high and upper middle income countries.

6 Supplementary evaluations

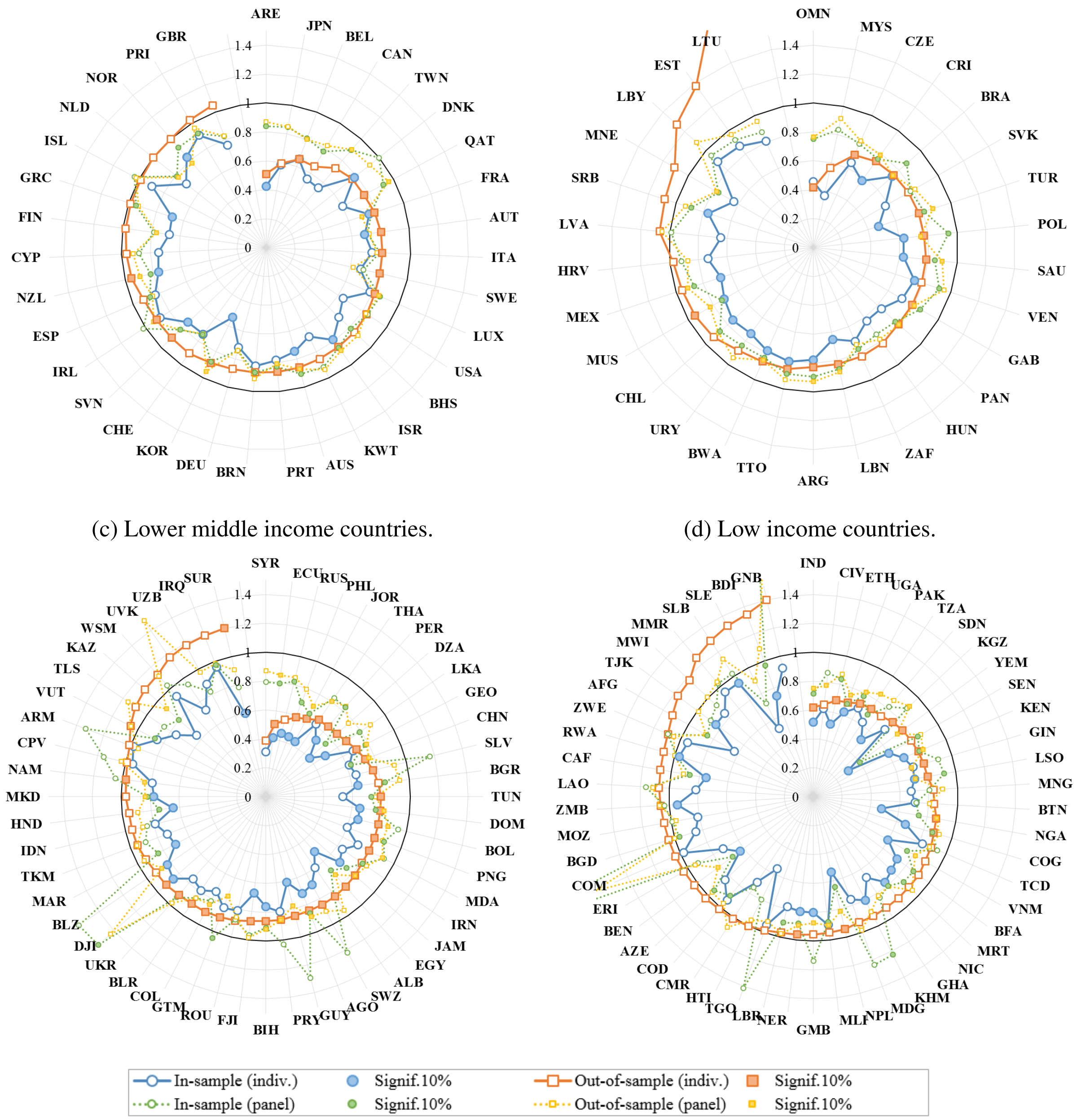
6.1 Statistical tests

Our analysis has so far been based on direct comparisons of sample accuracy measures of the forecasts derived from different modeling assumptions and night lights indicators. One important question is how these comparisons stand in statistical terms, namely, when a model is found to outperform (or not) the benchmark, how much confidence can we put on this being evidence that would transcend the sample used for the evaluation? To attempt to answer this question we conduct statistical tests comparing the predictive accuracy of the night lights-augmented forecasts to those obtained under the AR(1) benchmark model. For that purpose we follow the approach suggested by Clark and West (2007, CW) for comparison of nested[17] models, also adapting the inference to finite samples by simulating the empirical distribution of the test

[16]Country-specific evaluation reports are also available in a supplementary file.

[17]One traditional test for the hypothesis of equal predictive accuracy is the Diebold and Mariano (1995, DM) test, which is based on the mean difference in sample average losses (here assumed to be the squared forecast error). However, one important disadvantage of the DM test is that it does not account for parameter estimation uncertainty, an issue that is of great relevance in our application. Besides, our comparative evaluation involves nested models, that is, the night lights-augmented specifications, eqs. (4.3) and (4.4), would converge to the benchmark specifications, eqs. (4.1) and (4.2), under the null hypothesis that the night lights indicators are irrelevant to GDP growth predictions (for further discussion on these issues, see Elliott and Timmermann, 2016, Ch. 17).

Figure 5.1: Countries RMSE ratios grouped by income level.



(a) High income countries.
(b) Upper middle income countries.
(c) Lower middle income countries.
(d) Low income countries.

Notes: The RMSE ratios refer to forecasts obtained with the correlated pixels (full-sample classification) night lights model. Values below one indicate countries for which the night lights-based forecasts outperformed those from the benchmark model, and the opposite for the ratios above one. Statistical significance is computed from 1,000 bootstrap replications as detailed in the text. The countries are grouped according to their World Bank income classification and sorted within each group in ascending order according to their corresponding individual specification out-of-sample RMSE ratios.

statistics according to the bootstrap procedure proposed by Clark and McCracken (2012).[18]

The statistical tests for predictive improvement are again conducted separately for each country, model, pooling assumption, and evaluation exercise. The results are summarized in Table 5.2 in the form of cross-country frequencies of rejections of the null hypothesis that the night lights data bring no improvement to the accuracy of the benchmark AR(1) model.[19] As expected, these hit-rates tend to be smaller than the improvement rates observed above, based solely on the sample RMSEs. The only exception is the OS result with individual real-time correlation-based SoL forecasts, where the rate of CW rejections is slightly higher than the raw RMSE improvement rates; this is due to the fact that the CW test statistic corrects the MSE differentials for the estimation of additional parameters in the alternative model, which is another clear indication that these results are strongly affected by estimation uncertainty.

Whereas the lower rates of rejections of the null of no predictive improvement indicate caution should be taken in extrapolating our overall assessments beyond our sample, we note that the in-sample evidence may still provide some guidance on the applicability of our findings for particular countries. Indeed, Inoue and Kilian (2005) question the interpretation that in-sample evidence of predictability unaccompanied by similar out-of-sample evidence is likely to be spurious, showing that such an empirical regularity can be explained by the fact that in-sample tests tend to have higher power than out-of-sample ones. Country-specific significance results are also depicted in figure 5.1.

6.2 Partial pooling

The panel and individual specifications represent two extreme ends of a broad range of possibilities with respect to the grouping of countries for the estimation of the relationship between night lights and GDP growth. From one side, both our estimation and forecasting results have indicated the existence of too much cross-country heterogeneity in that relationship to endorse the full pooling assumption. The individual specifications, in turn, have been found to be severely affected by estimation uncertainty, particularly for the purpose of out-of-sample forecasting. Hence, we now explore some partial pooling alternatives in an attempt to improve this trade-off between heterogeneity and estimation uncertainty.

To approach the issue of partial pooling we adopt three different criteria. First, we evaluate the sub-grouping of countries according to their WB income classification. Second, we group

[18]In short, Clark and McCracken (2012) propose the use of a wild fixed regressor bootstrap procedure to approximate the asymptotic critical values in the comparison of forecasts based on nested models. There are only two differences in our application: (i) considering that we have a panel of countries, in order to preserve the cross-country correlations we use the same random resampling across the countries on each bootstrap replication; and, (ii) we use the benchmark (restricted) model to obtain the bootstrapped residuals instead of the unrestricted specification, including all night lights indicators, considering that this would be infeasible for the individually-estimated models; according to Clark and McCracken (2012), this makes little difference in practice.

[19]The individual tests underlying these rates are provided in an online supplement, including the results based on the DM test and the theoretical distribution of the tests statistics; the latter tend to show higher rejection rates (lower p-values), on average, than those obtained under the bootstrapped tests.

Table 6.1: Partial pooling RMSE ratios.

Specification	Panel	Clustering criterion (n.clusters)				Individual
		Income (4)	Region (7)	k-means (50)	k-means (100)	
(i) In-sample evaluation.						
Total SoL growth	1.00	0.99	1.00	0.97	0.97	0.97
Correlated pixels SoL growth:						
Real-time	0.99	0.99	0.97	0.90	0.91	0.91
Full-sample	0.80	0.79	0.74	0.66	0.63	0.63
(ii) Out-of-sample evaluation.						
Total SoL growth	1.00	1.00	1.00	1.00	1.00	1.01
Correlated pixels SoL growth:						
Real-time	1.07	1.06	1.05	1.06	1.04	1.07
Full-sample	0.82	0.81	0.81	0.81	0.82	0.78

Notes: See the notes to Table 5.1 for details about the construction of the RMSE ratio statistics. The partial pooling classifications by income (34 high, 29 upper middle, 51 lower middle, and 53 low income countries) and region (22 from East Asia & Pacific, 49 from Europe & Central Asia, 26 from Latin America & Caribbean, 18 from Middle East & North Africa, 2 from North America, 7 from South Asia, and 43 from Sub-Saharan Africa) are based on World Bank definitions. The k-means classifications are obtained separately for each model specification based on the Euclidean distance between the individual estimates of the AR(1) and the night lights indicator(s) coefficients.

countries according to their WB region. Third, we adopt a k-means clustering algorithm to group countries in a way that minimizes the Euclidean distance between the country individual estimates (see, e.g., Vahid, 1999; Sarafidis and Weber, 2015, for previous applications of clustering methods in panel data). The results of these exercises are presented in Table 6.1, which also reports the previous panel and individual results for comparative purposes.

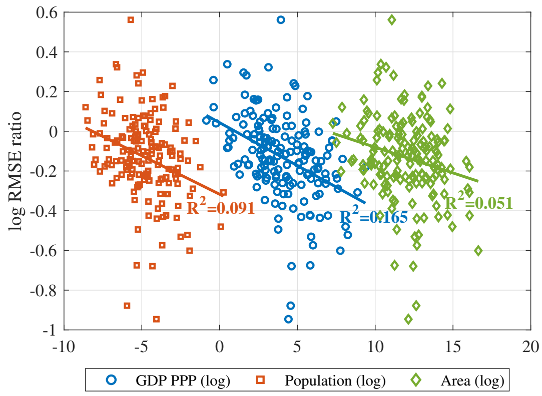
As the RMSE ratios in Table 6.1 indicate, there is little support for the grouping of countries according to their WB income class and regions. A slightly more favorable picture is obtained using the more agnostic approach based on the k-means clustering algorithm; however, the individual results are still favored in some of the exercises, particularly in the OS exercise with the full-sample correlation-based night lights indicators. Hence, whereas there is some space for improvement with the partial pooling alternative, more elaborate methods may be required to single out the optimal partitioning of countries with respect to the relationship between the night lights and GDP growth.

7 Potential explanatory factors

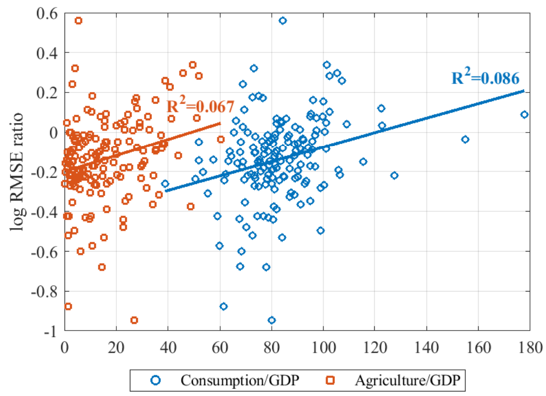
Our results show a non-negligible degree of heterogeneity in the performance of the night lights indicators in forecasting GDP growth across our global sample of countries, particularly when

Figure 7.1: Scatterplots between night lights performances and key country-specific factors.

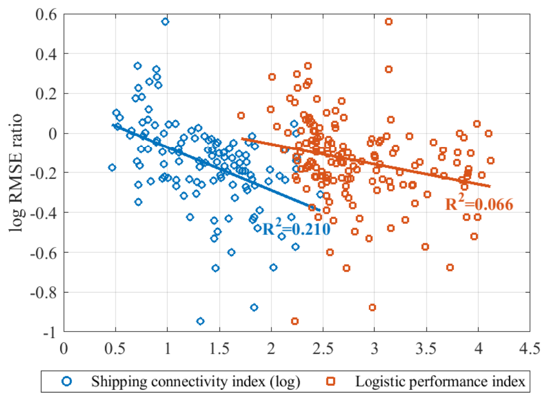
(a) Size-related factors.



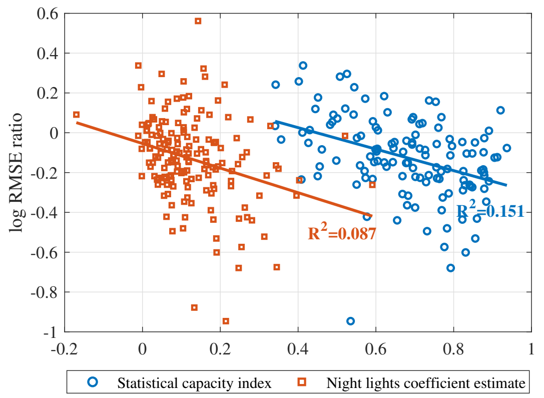
(b) Structure-related factors.



(c) Logistic-related factors.



(d) Data-related factors.



Notes: The countries (log) RMSE ratios (depicted in the vertical axis of each plot) are the same as those reported in Figure 5.1 obtained from individual model estimates.

forecasting out-of-sample. Whereas the countries income per capita classification seems to be associated with the stability of night lights performance (recall the analysis of Figure 5.1), we now attempt to uncover other key factors underlying the usefulness of the night lights data. For that purpose we examine the relationship between the night lights performances and some country-specific factors, including variables for the country's: economic development, infrastructure and sectoral composition; demographic and geographic characteristics; energy efficiency; informal sector size; and statistical capacity. The results are summarized in Figure 7.1, focusing on the four most important categories of factors for the performance of the forecasts based on the full-sample correlated pixels night lights indicator. Results for the other indicators, together with a complete list of all variables considered and their sources, are provided in Appendix A.4.

The first group of factors, presented in Figure 7.1a, refers to size-related measures, where bigger countries, in economic, population, and geographic terms, are found to obtain better night lights performances (notice lower log RMSE ratios represent better night lights performances relative to the benchmark). This is a quite sensible finding considering that the bigger the country the higher tends to be the number of night lights signals collected by the satellites. The second group of factors, presented in Figure 7.1b, relates to the structure of the countries expenditures and production sectors. Here, both the share of consumption and the share of agriculture in GDP are found to be negatively associated with the performance of the night lights forecasts. Considering that countries with higher consumption shares tend to present lower investment rates, we interpret this finding as consistent with the time to build argument outlined in Section 2.2. The finding on agriculture also seems consistent with the idea that the night lights data are more sensitive to concentrated industrial activities, which are likely to generate more luminosity than scattered countryside activity. The usefulness of geo-located signals is also evidenced by the results for the third group of factors, presented in Figure 7.1c, which indicate that the night lights tend to achieve better performance in countries with better logistics infrastructure and connectivity to global shipping networks.

Finally, the last group of factors, presented in Figure 7.1d, is related to the measurement error hypothesis outlined in Section 2.1. Intriguingly, the night lights forecasts tend to perform better for countries with higher scores in the World Bank Statistical Capacity assessments, a result in contrast with the past literature proposing the use of night lights to improve GDP measures in countries with less reliable statistics. One possible explanation for this result comes from the fact that, in contrast to the previous literature, we are estimating the lights-growth relationship individually for each country; thus, the better results we obtain for more developed countries can be associated with better estimates provided by the higher quality of GDP data for those countries.[20] This view is corroborated by the results relating the night lights performances with estimates of the relationship between GDP growth and (lagged) lights growth in the individual forecasting models, where stronger relationships are found to be associated with improved night lights performance.

8 Concluding remarks

In this paper we evaluated the usefulness of satellite-based data on night-time lights for the prediction of annual GDP growth across a global sample of 167 countries over the period from 1993 to 2014. We proposed innovative measures to improve the quality of the signals obtained from the lights data at different locations within a country, and evaluated their predictive content by augmenting an AR(1) GDP growth forecasting model with lagged values of the lights indicators. We have also considered alternative assumptions on the pooled estimation of the

[20]Recall from equation (2.5) that a lower measurement error in the lights-growth relationship, σ_e^2, can be expected to improve the performance of the lights data, while lower GDP measurement errors, σ_u^2, would lead to a deterioration of that improvement; hence, our results suggest that the effect of more precise individual lights-growth estimates dominate the decrease of relevance of the lights data for countries with better statistics.

relationship between lights and GDP growth across countries, ranging from country-individual specifications to partial and full panel specifications.

Overall, we have found evidence favorable to the use of night lights data for GDP growth forecasting. Importantly, our results indicate a substantial degree of heterogeneity across country estimates, and that these effects are relevant for the use of night lights as predictors of GDP growth. Namely, we have found that individually-estimated models tend to outperform pooled specifications, even though the former are subject to greater sampling variability due to the use of smaller samples. These biases have been particularly harmful under an out-of-sample forecast evaluation exercise, though we still find statistically significant improvements for over two thirds of our sample of countries. In spite of these estimation issues, we believe our analysis of innovative location-based night lights indicators provides an interesting framework for future applications of the night lights data for economic measurement and forecasting.

References

Bluhm, R. and M. Krause (2018). Top lights: Bright cities and their contribution to economic development. Mimeograph, Universities of Hannover and Hamburg.

Chauvet, M. and S. Potter (2013). Chapter 3 - forecasting output. In G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 2, Part A of *Handbook of Economic Forecasting*, pp. 141 – 194. Elsevier.

Chen, X. and W. D. Nordhaus (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences 108*(21), 8589–8594.

Clark, T. E. and M. W. McCracken (2012). Reality checks and comparisons of nested predictive models. *Journal of Business & Economic Statistics 30*(1), 53–66.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*(1), 291 – 311.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–63.

Doll, C. N. (2008). CIESIN thematic guide to night-time light remote sensing and its applications. Technical report, Center for International Earth Science Information Network, Palisades, NY, USA.

Doll, C. N., J.-P. Muller, and J. G. Morley (2006). Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics 57*(1), 75 – 92.

Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives 30*(4), 171–98.

Elliott, G. and A. Timmermann (2016). *Economic Forecasting*. Princeton University Press.

Elvidge, C. D., V. R. Hobson, I. L. Nelson, J. M. Safran, B. T. Tuttle, J. B. Dietz, and K. E. Baugh (2003). Overview of DMSP OLS and scope of applications. In V. Mesev (Ed.), *Remotely Sensed Cities*, pp. 281–300. Taylor & Francis.

Elvidge, C. D., J. Safran, I. L. Nelson, B. T. Tuttle, V. R. Hobson, K. E. Baugh, J. B. Dietz, and E. H. Erwin (2004). Area and positional accuracy of DMSP nighttime lights data. In *Remote Sensing and GIS Accuracy Assessment*, pp. 281–292. CRC Press.

Elvidge, C. D., D. Ziskin, K. E. Baugh, B. T. Tuttle, T. Ghosh, D. W. Pack, E. H. Erwin, and M. Zhizhin (2009). A fifteen year record of global natural gas flaring derived from satellite data. *Energies 2*(3), 595–622.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999). Geography and economic development. *International Regional Science Review 22*(2), 179–232.

Ghosh, T., R. L. Powell, C. D. Elvidge, K. E. Baugh, P. C. Sutton, and S. Anderson (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal 3*, 147–160.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017). Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.

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

Heston, A., R. Summers, and B. Aten (2002, 10). Penn world table version 6.1. Database, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Hsu, F.-C., K. E. Baugh, T. Ghosh, M. Zhizhin, and C. D. Elvidge (2015). Dmsp-ols radiance calibrated nighttime lights time series with intercalibration. *Remote Sensing 7*(2), 1855–1876.

Inoue, A. and L. Kilian (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews 23*(4), 371–402.

Ishizawa, O. A., J. J. Miranda, and H. Zhang (2017). Understanding the impact of windstorms on economic activity from night lights in Central America. Policy Research working paper WPS 8124, World Bank Group.

Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica 50*(6), 1345–70.

Landefeld, J. S., E. P. Seskin, and B. M. Fraumeni (2008, June). Taking the pulse of the economy: Measuring gdp. *Journal of Economic Perspectives 22*(2), 193–216.

Mellander, C., J. Lobo, K. Stolarick, and Z. Matheson (2015, 10). Night-time light data: A good proxy measure for economic activity? *PLOS ONE 10*(10), 1–18.

Michalopoulos, S. and E. Papaioannou (2013a). National institutions and subnational development in africa. *The Quarterly Journal of Economics 129*(1), 151.

Michalopoulos, S. and E. Papaioannou (2013b). Pre-colonial ethnic institutions and contemporary african development. *Econometrica 81*(1), 113–152.

Nordhaus, W. and X. Chen (2015). A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. *Journal of Economic Geography 15*, 217–246.

Pinkovskiy, M. and X. Sala-i Martin (2016). Lights, camera ... income! illuminating the national accounts-household surveys debate. *The Quarterly Journal of Economics 131*(2), 579–631.

Sarafidis, V. and N. Weber (2015, 4). A partially heterogeneous framework for analyzing panel data. *Oxford Bulletin of Economics and Statistics 77*(2), 274–296.

Sutton, P. C. and R. Costanza (2002). Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics 41*(3), 509 – 527.

United Nations (2009). *System of national accounts 2008*. New York.

Vahid, F. (1999). Partial pooling: a possible answer to pool or not to pool. In *Cointegration, Causality and Forecasting: Festschrift in Honor of Clive W. J. Granger*, Chapter 17, pp. 410–428. Oxford University Press.

Zhao, N., Y. Liu, G. Cao, E. L. Samson, and J. Zhang (2017). Forecasting China's GDP at the pixel level using nighttime lights time series and population images. *GIScience & Remote Sensing 54*(3), 407–425.

A Online Appendix (not for publication)

A.1 Theoretical derivation of lights relative accuracy

Under the framework of equations (2.1) and (2.2), the OLS estimates of $\hat{\rho}$, $\hat{\varrho}$, and $\hat{\theta}$ are given by

$$\hat{\rho} = \frac{Cov\left(z_t, z_{t-1}\right)}{\sigma_z^2 + \sigma_u^2}, \tag{A.1}$$

$$\hat{\varrho} = \frac{Cov\left(z_t, z_{t-1}\right)\sigma_e^2}{\sigma_z^2\sigma_e^2 + \sigma_u^2\sigma_x^2}, \tag{A.2}$$

$$\hat{\theta} = \frac{\beta Cov\left(z_t, z_{t-1}\right)\sigma_u^2}{\sigma_z^2\sigma_e^2 + \sigma_u^2\sigma_x^2}, \tag{A.3}$$

respectively. Using these estimates, the mean squared forecast errors associated with the benchmark and the lights-based forecasts are given by

$$\begin{aligned}\hat{\Delta} &= E\left[\left(y_t - \hat{y}_t\right)^2\right] \\ &= \left(1 + \hat{\rho}^2\right)\sigma_y^2 - 2\hat{\rho} Cov\left(z_t, z_{t-1}\right), \\ &= \sigma_y^2 - Cov\left(z_t, z_{t-1}\right)^2 / \sigma_y^2, \end{aligned} \tag{A.4}$$

and

$$\tilde{\Delta} = \sigma_y^2 - Cov\left(z_t, z_{t-1}\right)^2 \frac{\left(\beta^2\sigma_u^2 + \sigma_e^2\right)}{\sigma_z^2\sigma_e^2 + \sigma_u^2\sigma_x^2}, \tag{A.5}$$

respectively. We can now compare the accuracy of the night lights-based forecasts to that of the benchmark model by focusing on their (normalized) MSFE ratio,

$$\begin{aligned}\frac{\sigma_y^2 - \tilde{\Delta}}{\sigma_y^2 - \hat{\Delta}} &= \frac{\left(\beta^2\sigma_u^2 + \sigma_e^2\right)\sigma_y^2}{\sigma_z^2\sigma_e^2 + \sigma_u^2\sigma_x^2}, \\ &= 1 + \frac{\beta^2\sigma_u^2}{\sigma_e^2\sigma_z^2/\sigma_u^2 + \sigma_x^2}, \\ &= 1 + \frac{\beta^2\sigma_u^2}{\beta^2\sigma_z^2 + \sigma_e^2\left(1 + \sigma_z^2/\sigma_u^2\right)}, \end{aligned} \tag{A.6}$$

where the second term of equation (A.6) reflects the (expected) improvement obtained with the use of the lights-augmented specification in relation to the benchmark autoregressive model.

A.2 Intercalibration of night lights data

The night lights data consist of a total of 34 global composite images coming from six different satellites operating over the period between 1992 and 2013. For comparative purposes, these data require intercalibration in order to adjust for varying sensor conditions. Here we follow the approach proposed by Elvidge et al. (2009), where a second order polynomial regression is estimated across the satellite-year composites over Sicily, and then used to adjust the global composites accordingly. The regression specification is given by

$$DN_{r,p} = \phi_0 + \phi_1 DN_{s,p} + \phi_2 DN_{s,p}^2, \tag{A.7}$$

where p stand for the pixel, s for the satellite-year composite to be re-scaled, and r (= F152006) for the reference satellite-year composite, which was selected so as to maximize the average fit of the regressions. The dispersion of the data used for estimation and the parameter estimates are presented in Figure A.1 and Table A.1, respectively.

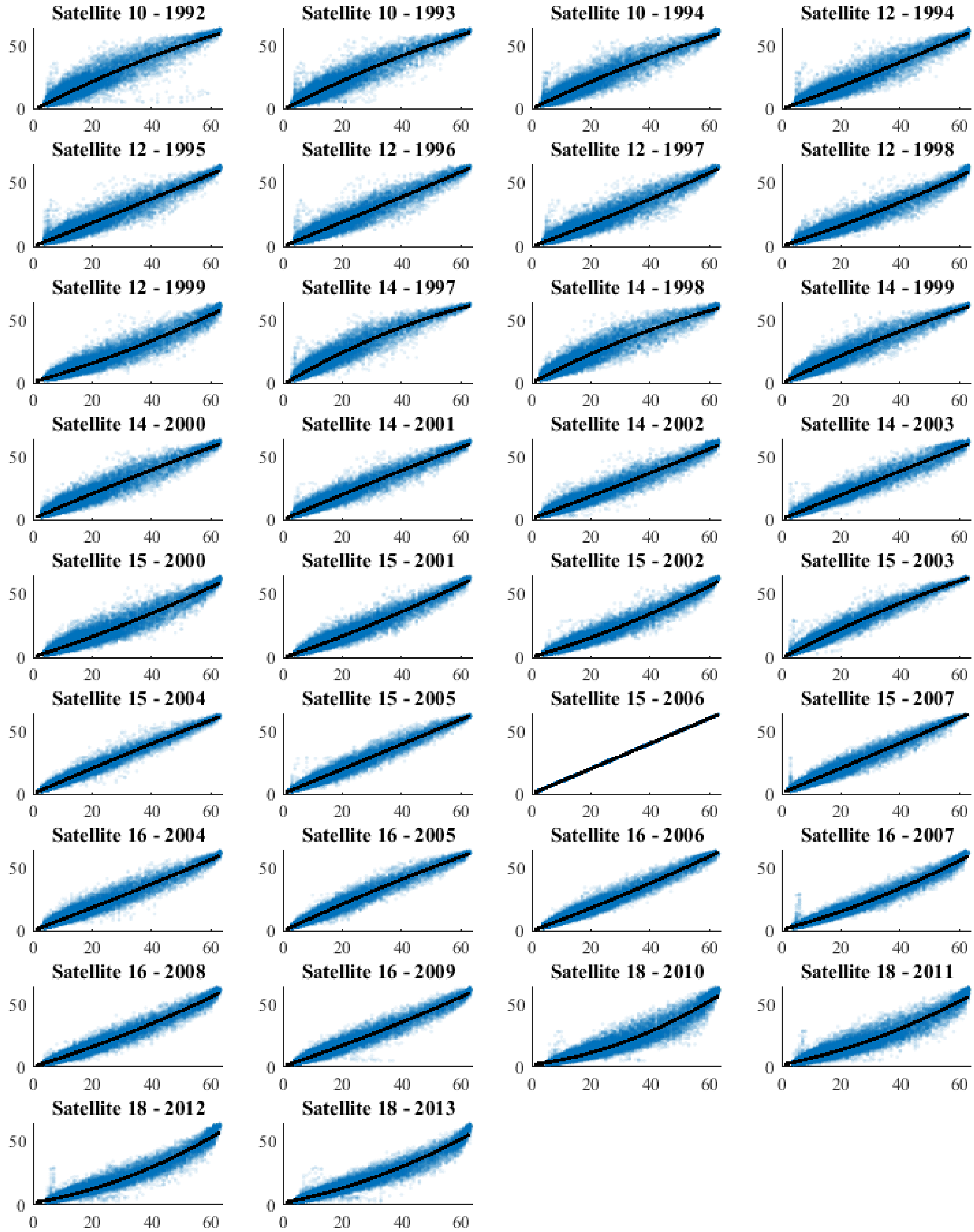
A.3 AR(2) benchmark results

Table A.2 presents the averaged forecast evaluation statistics taking an AR(2) model as the benchmark. Although the in-sample RMSEs are smaller than those observed under the AR(1) benchmark exercise presented in Table 5.1, the RMSE ratios of the night lights-based forecasts relative to the benchmark are roughly the same. A similar picture emerges in terms of RMSE ratios under the out-of-sample evaluation exercise, although in this case the AR(2) RMSEs are generally higher than those obtained with the AR(1) benchmark analysed in the main text.

A.4 Explanatory factors supplementary results and data sources

The data used for the analysis of explanatory factors come from many different sources, such as: (i) the International Monetary Fund (IMF) World Economic Outlook (WEO); (ii) the World Bank Development Indicators (WB-DI); (iii) the Gallup et al. (1999, GSM) physical geography dataset; (iv) the Penn World Tables (Heston et al., 2002, PWT); (v) OECD Real-Time and Revisions Database (RTRD); (vi) the International Labour Organization (ILO). A complete list of the variables considered is presented in Table A.3. The scatterplots relating all these variables to the out-of-sample performance of the night lights-based forecasts, individual models, are presented in Figures A.2-A.9.

Figure A.1: Scatter plots of Sicily's pixels DNs used for intercalibration of satellites.



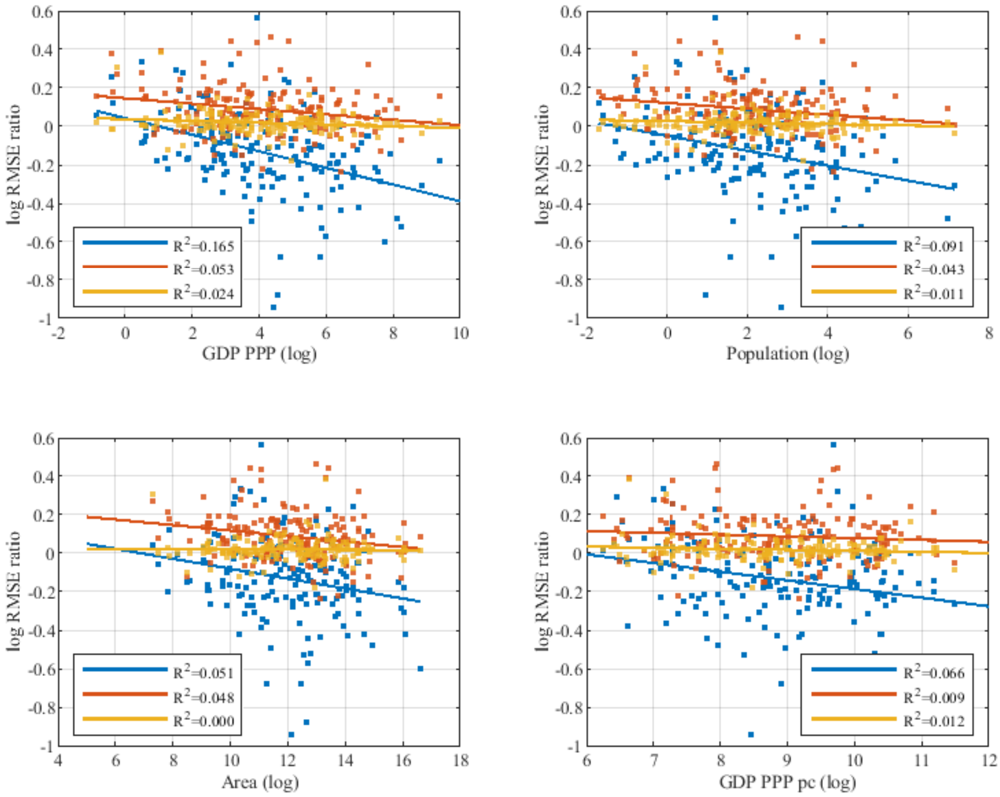
Notes: The DN values observed over Sicily for the reference satellite F152006 are plotted (x-axis) against Sicily's DN values from other satellite-year composites (y-axis). The black line depicts the fitted values according to equation (A.7).

Table A.1: Intercalibration regression estimates.

Sat.	Year	ϕ_0	ϕ_1	ϕ_2	R^2	N.Pixs.
10	1992	-1.8913	1.2378	-0.0042	0.904	29796
10	1993	-1.0148	1.2210	-0.0039	0.926	33413
10	1994	-0.4652	1.1417	-0.0031	0.932	30561
12	1994	-0.7232	0.8819	0.0013	0.929	28980
12	1995	-0.2319	0.9166	0.0004	0.938	32207
12	1996	-0.1864	0.9556	0.0002	0.937	30016
12	1997	-0.4029	0.8706	0.0015	0.939	31979
12	1998	0.1048	0.7470	0.0027	0.951	32436
12	1999	0.4627	0.6816	0.0035	0.940	31268
14	1997	-1.4670	1.3965	-0.0064	0.935	31695
14	1998	-0.1328	1.2448	-0.0048	0.933	30864
14	1999	-0.2841	1.1501	-0.0029	0.950	33353
14	2000	0.5327	1.0149	-0.0012	0.943	32084
14	2001	-0.0964	1.0062	-0.0009	0.958	32844
14	2002	0.6088	0.8688	0.0009	0.954	31427
14	2003	0.1951	0.9600	-0.0002	0.959	33361
15	2000	0.1642	0.7380	0.0029	0.939	33651
15	2001	-0.4642	0.7937	0.0027	0.956	33059
15	2002	0.1410	0.6751	0.0042	0.960	32359
15	2003	-0.3656	1.1889	-0.0031	0.966	33340
15	2004	0.0403	1.0301	-0.0009	0.976	31080
15	2005	0.0837	0.9788	0.0001	0.970	33509
15	2006	0.0000	1.0000	0.0000	1.000	33877
15	2007	0.5517	0.9891	0.0002	0.966	31159
16	2004	-0.2095	0.9014	0.0007	0.958	31752
16	2005	-0.5565	1.1083	-0.0021	0.970	33618
16	2006	-0.4076	0.8657	0.0020	0.970	31893
16	2007	0.4299	0.6412	0.0046	0.972	32308
16	2008	0.2339	0.7200	0.0033	0.966	32271
16	2009	0.3699	0.7898	0.0022	0.962	28894
18	2010	1.8024	0.2926	0.0092	0.931	31117
18	2011	1.6726	0.4687	0.0062	0.936	31245
18	2012	1.6511	0.3815	0.0078	0.954	32151
18	2013	1.5803	0.4479	0.0064	0.957	32181

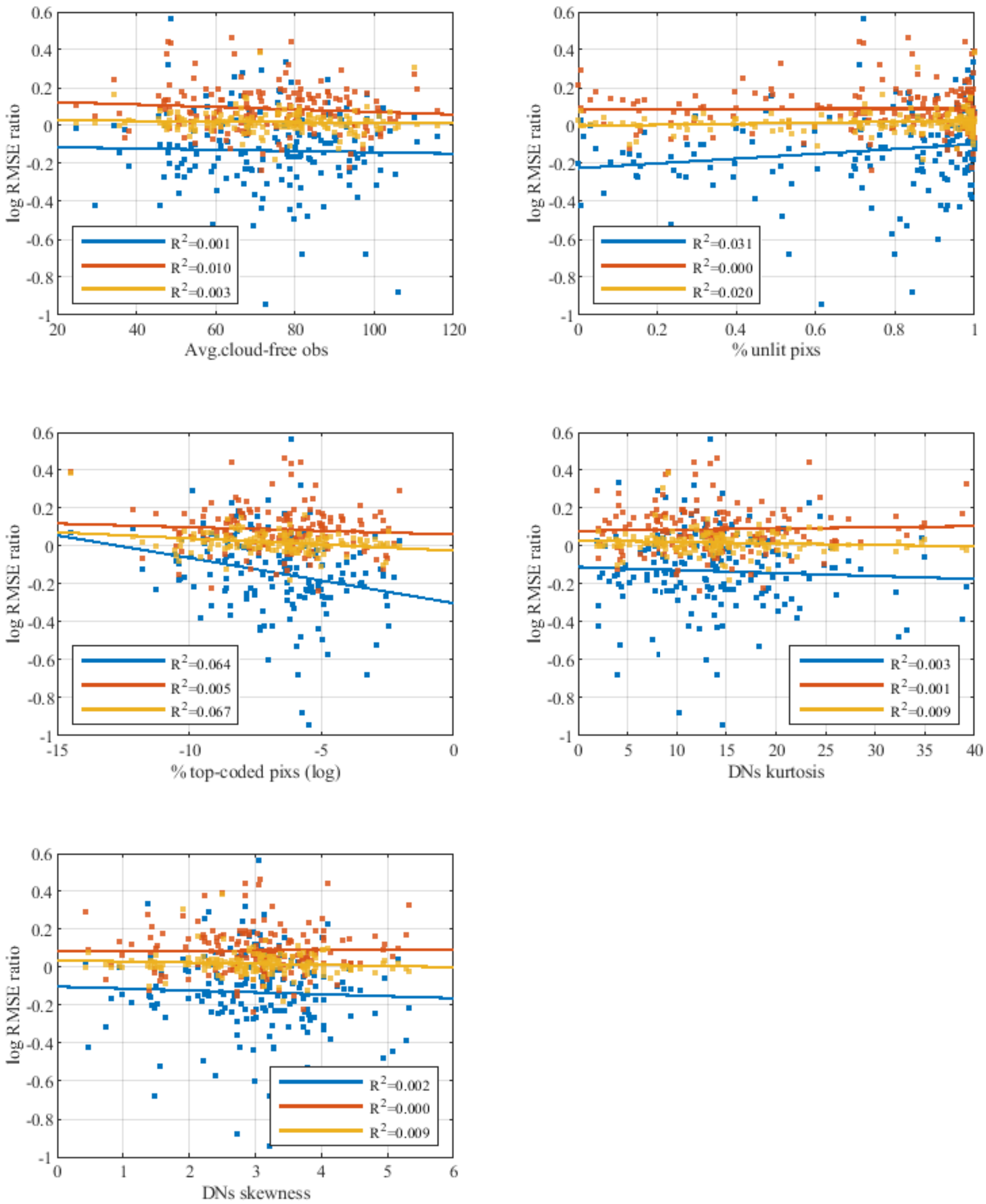
Notes: The estimates refer to equation (A.7) and are based on Sicily's night lights data.

Figure A.2: Scatterplots between forecasting performances of night lights indicators and country-specific factors – size-related factors.



Notes: The countries (log) RMSE ratios (depicted in the vertical axis of each plot) are those obtained under the out-of-sample exercise using the individually estimated forecasting models. Yellow, red, and blue points refer to the performances of the forecasting models using the total SoL, real-time correlated pixels SoL, and full-sample correlated pixels SoL growth variables, respectively.

Figure A.3: Scatterplots between forecasting performances of night lights indicators and country-specific factors – night lights data factors.



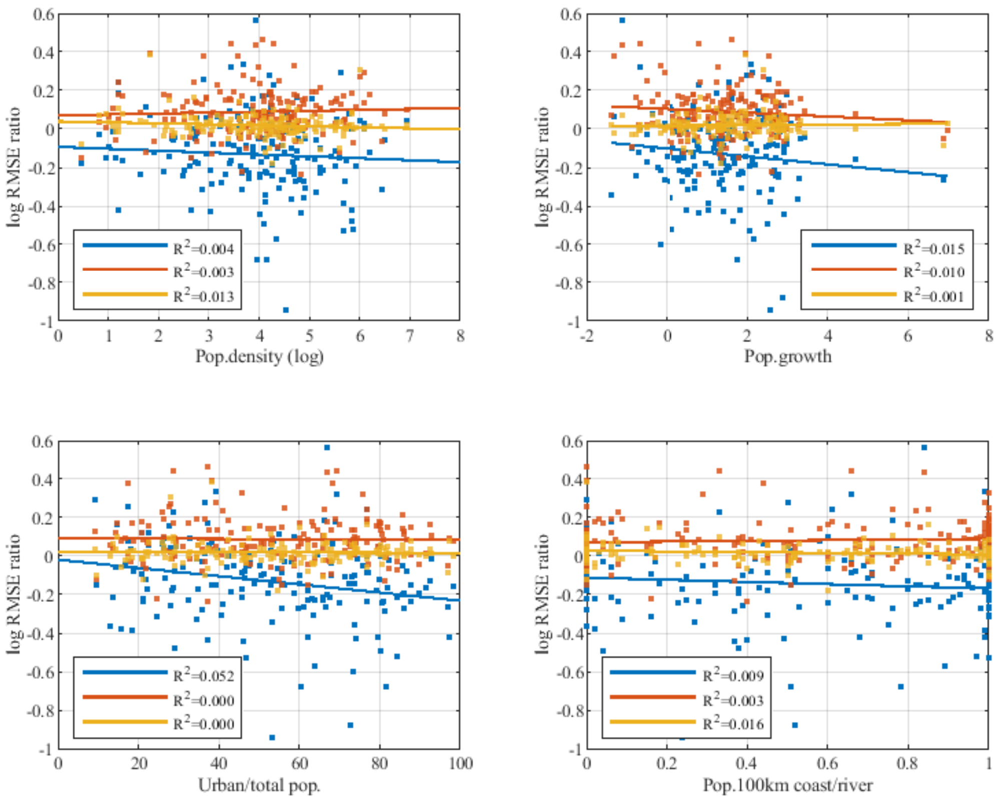
Notes: Same as notes to Figure A.2.

Table A.2: Forecast evaluation statistics against AR(2) benchmark.

Model	Panel RMSE	Panel Ratio	Panel SSR	Individual RMSE	Individual Ratio	Individual SSR	Sample
(i) In-sample evaluation.							
AR(2) benchmark	2.188	-	66%	2.036	-	68%	3437
AR(2) + Lagged SoL indicators:							
+ Total SoL growth	2.184	1.00	66%	1.978	0.97	70%	3307
+ Correlated pixels SoL growth							
Real-time classification	2.205	0.99	67%	1.897	0.90	68%	2843
Full-sample classification	1.792	0.80	74%	1.289	0.63	77%	3290
(ii) Out-of-sample evaluation.							
AR(2) benchmark	2.416	-	66%	2.696	-	62%	2205
AR(2) + Lagged SoL indicators:							
+ Total SoL growth	2.418	1.00	66%	2.736	1.01	63%	2181
+ Correlated pixels SoL growth							
Real-time classification	2.533	1.07	64%	2.847	1.06	59%	2117
Full-sample classification	2.020	0.82	70%	2.244	0.83	69%	2121

Notes: Same as notes to Table 5.1.

Figure A.4: Scatterplots between forecasting performances of night lights indicators and country-specific factors – demographic factors.



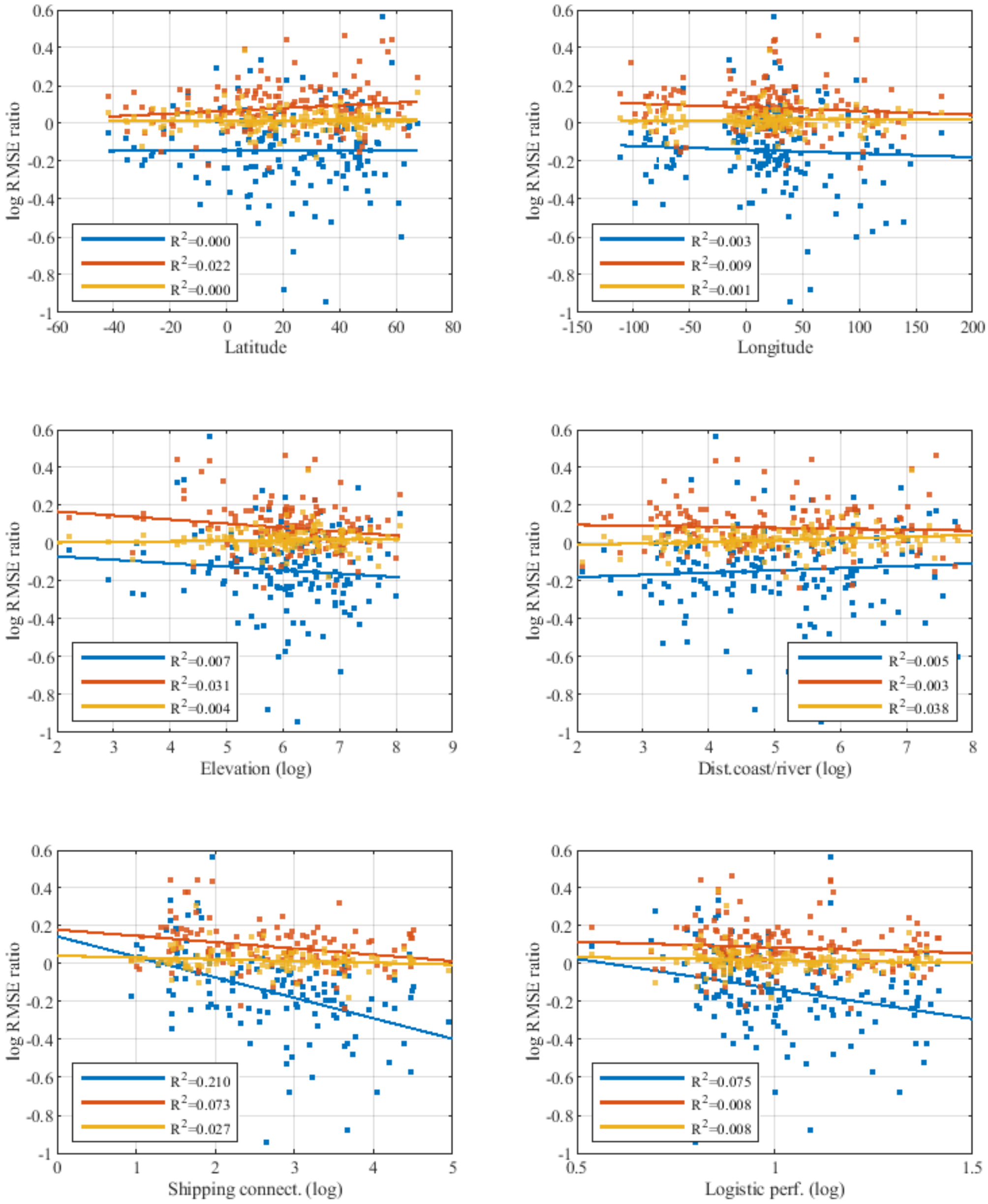
Notes: Same as notes to Figure A.2.

Table A.3: Explanatory factors – variables and data sources.

Control groups	Sources	Variables
Economic and demographic variables	IMF-WEO	GDP per capita (PPP), GDP (PPP), average population count and density.
	WB-DI	Ratios of consumption, agriculture, industry, services, natural resources rents, and trade to GDP, urban population (ratio to total).
Energy	WB-DI	Access to electricity (% of pop.), electric power consumption (per capita), electric power trans./distr. losses (% output), renewable energy consumption (% total), GDP per unit of energy use.
Geographic	WB-DI	Area (km^2), logistic performance, liner shipping connectivity.
	GSM	Latitude and longitude centroids, mean elevation, mean distance to nearest navigable river/coast, pop. within 100km of coast/nav.river.
Informality	WB-DI	% firms competing against informal firms, % firms formally registered when started, n. years operated informal, % firms identifying informal competitors as a major constraint.
	ILO	% informal employment.
Data quality	WB-DI	Assessments of statistical capacity: overall score on periodicity and timeliness, source data, methodology.
	PWT	Quality index.
	OECD-RTRD§	Mean absolute revisions, revisions autocorrelation.

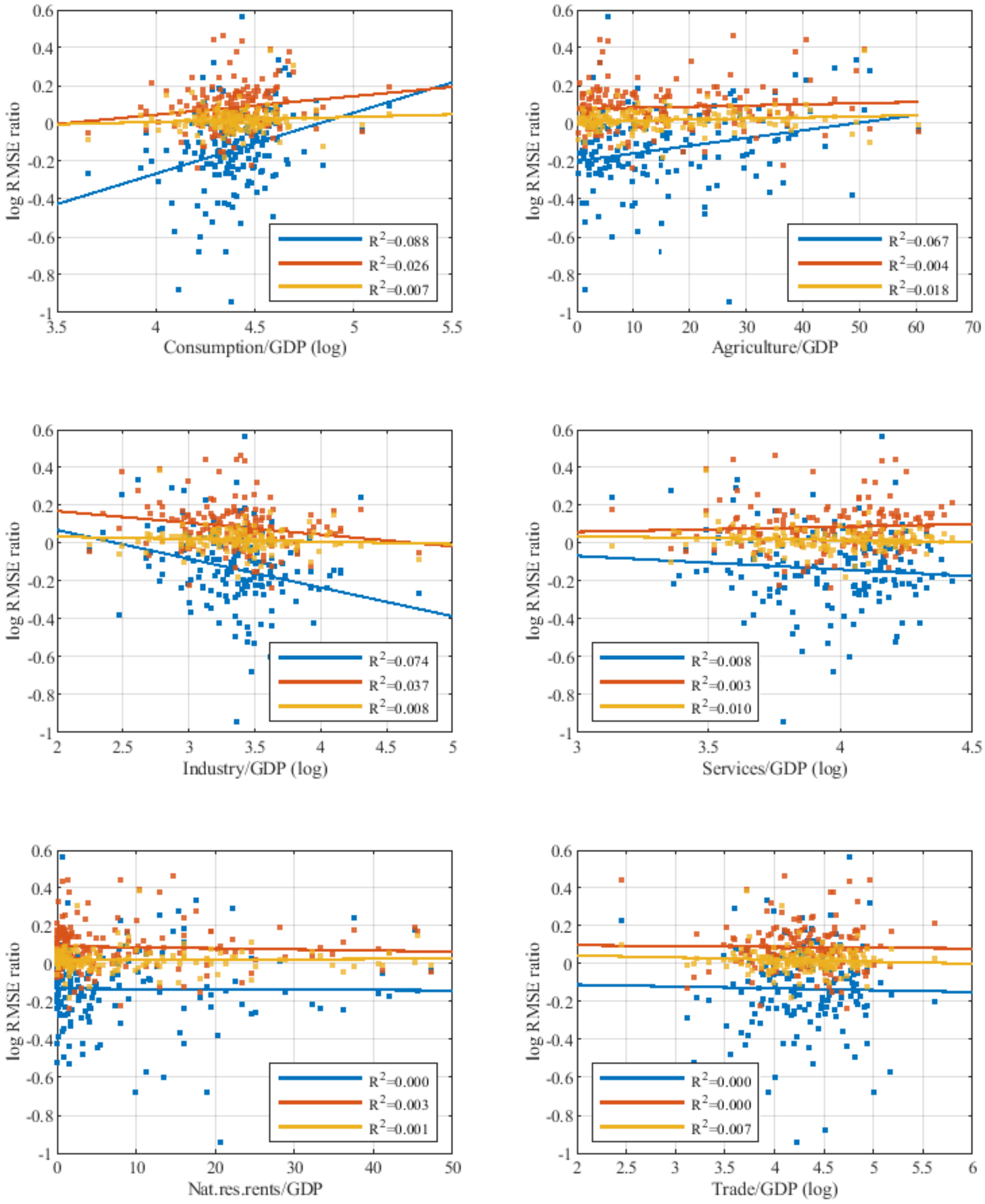
Notes: §Revisions statistics calculated between first published and latest available data.

Figure A.5: Scatterplots between forecasting performances of night lights indicators and country-specific factors – geographic/logistic factors.



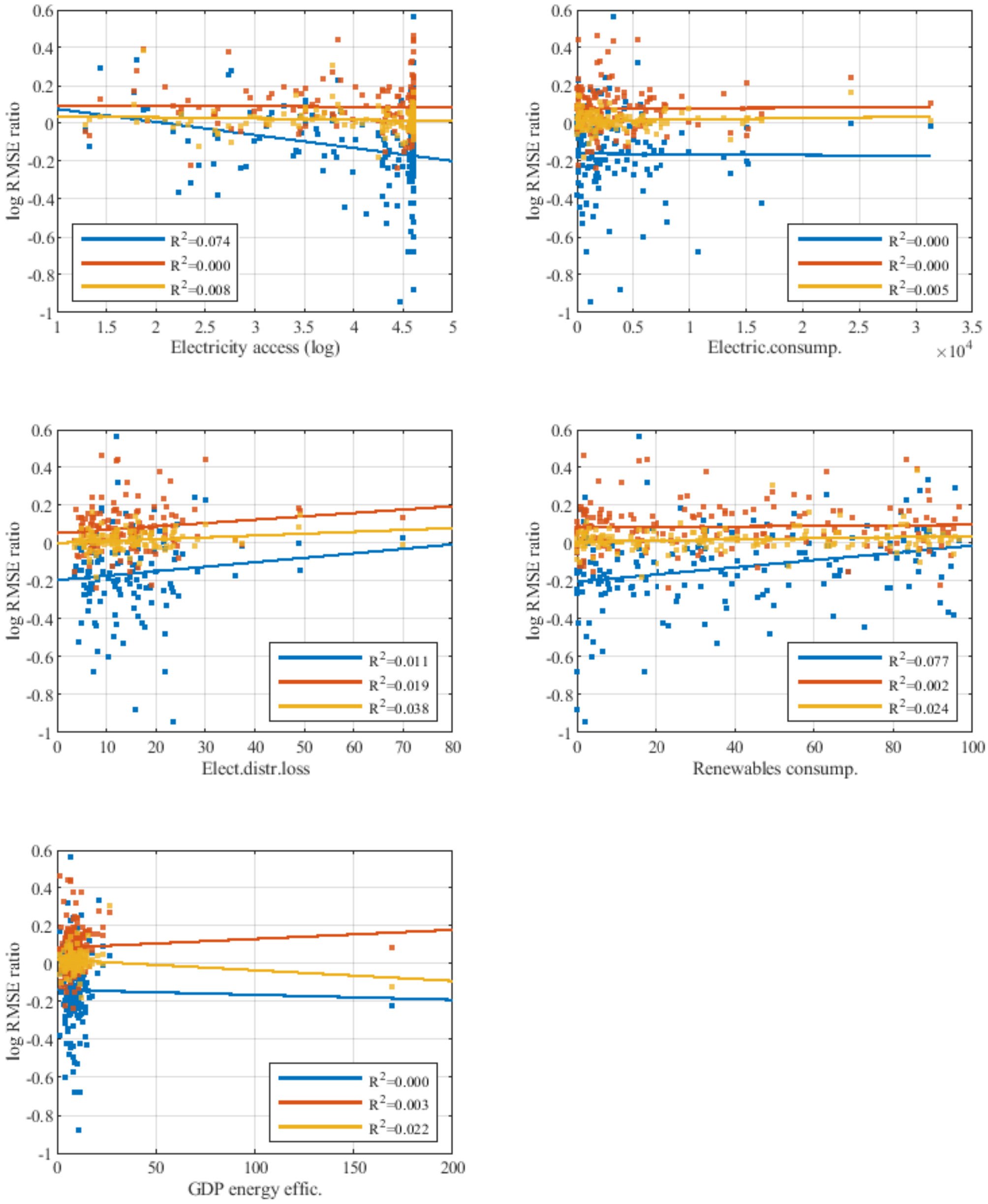
Notes: Same as notes to Figure A.2.

Figure A.6: Scatterplots between forecasting performances of night lights indicators and country-specific factors – economic structure factors.



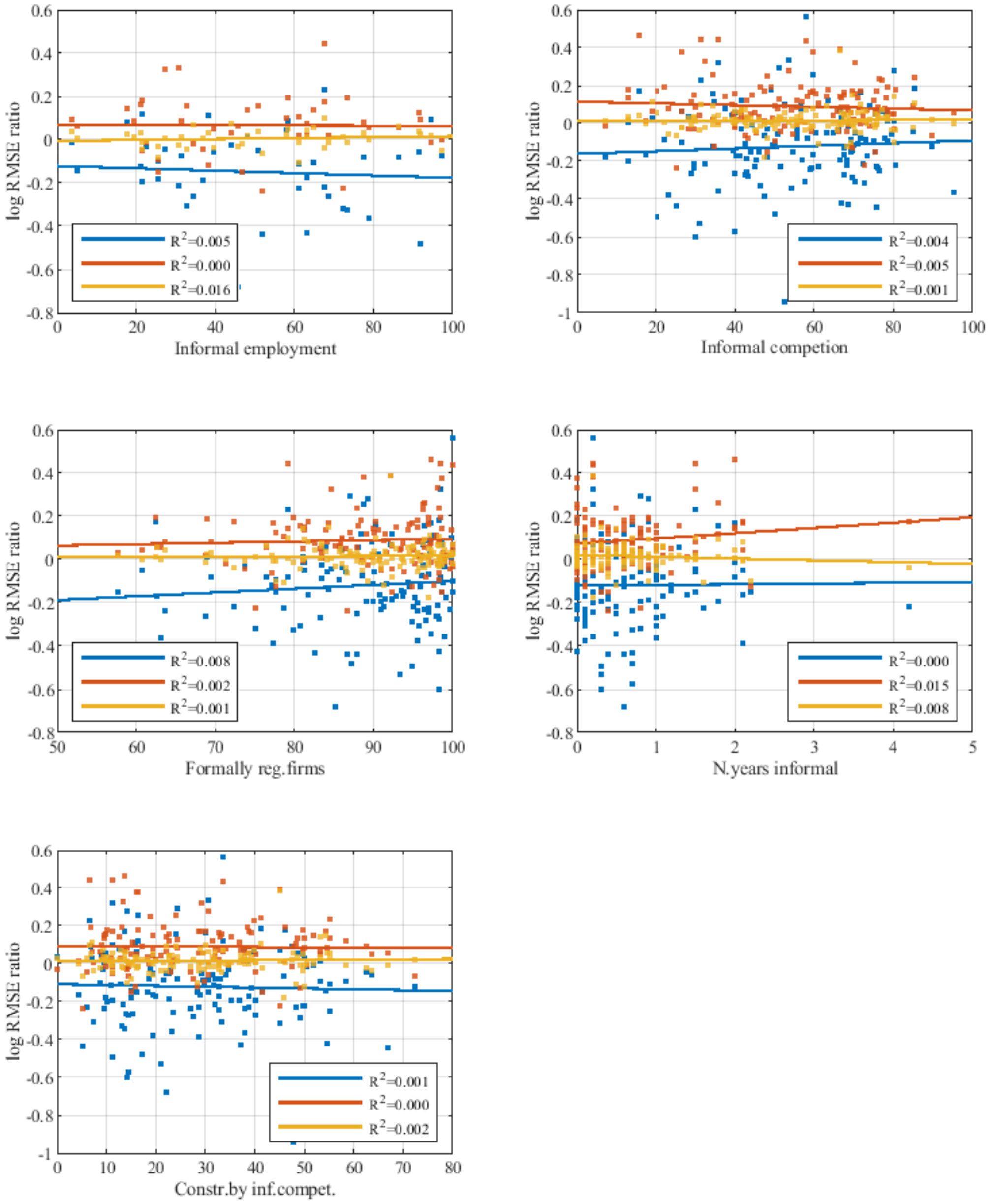
Notes: Same as notes to Figure A.2.

Figure A.7: Scatterplots between forecasting performances of night lights indicators and country-specific factors – energy-related factors.



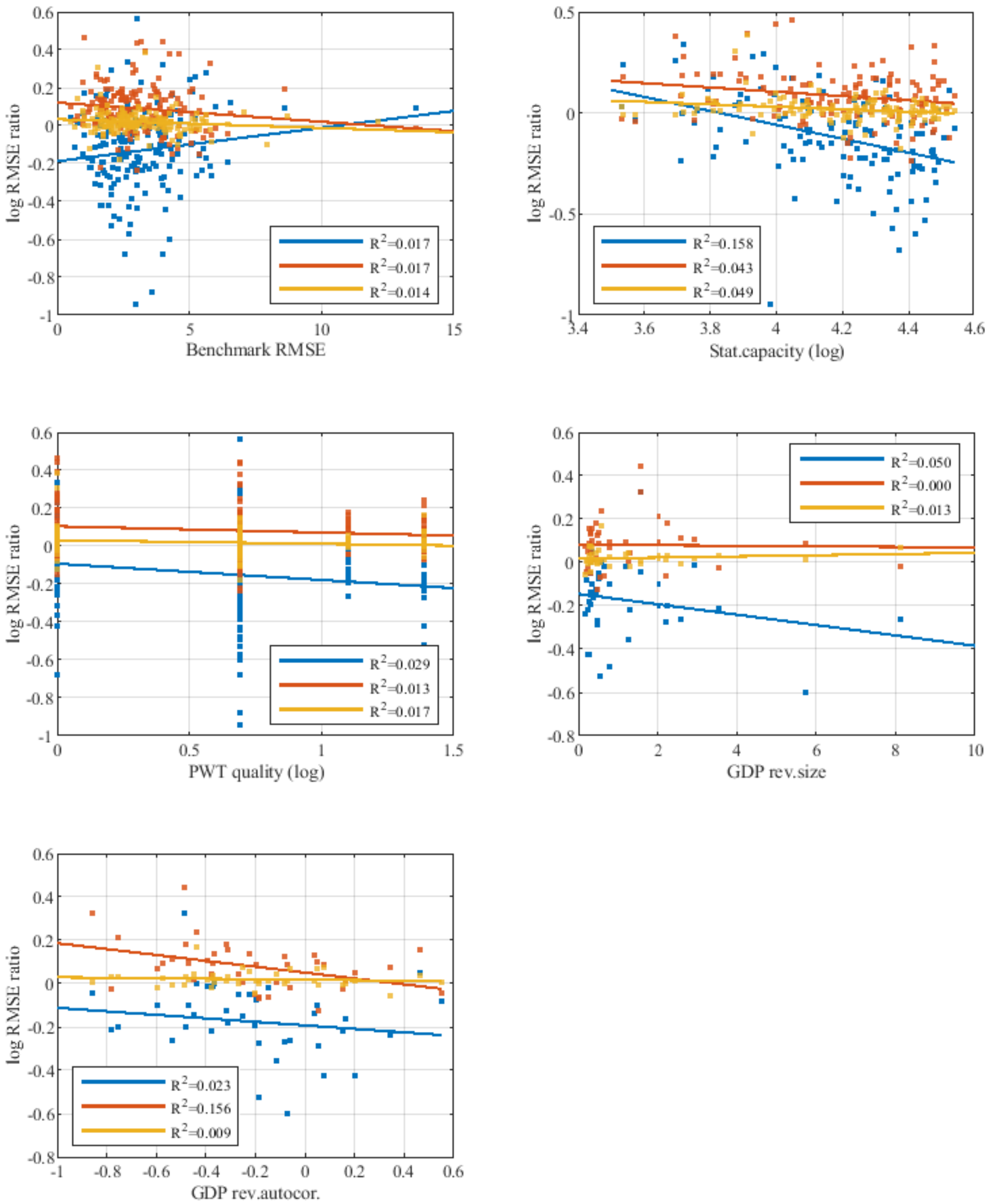
Notes: Same as notes to Figure A.2.

Figure A.8: Scatterplots between forecasting performances of night lights indicators and country-specific factors – informality-related factors.



Notes: Same as notes to Figure A.2.

Figure A.9: Scatterplots between forecasting performances of night lights indicators and country-specific factors – data quality factors.



Notes: Same as notes to Figure A.2.

A.5 List of countries

ISO	COUNTRY	GDP SAMPLE	ISO	COUNTRY	GDP SAMPLE
HIGH INCOME COUNTRIES:			UPPER MIDDLE INCOME COUNTRIES:		
ARE	United Arab Emirates	1992 - 2014	ARG	Argentina	1992 - 2014*
AUS	Australia	1992 - 2014	BRA	Brazil	1992 - 2014
AUT	Austria	1992 - 2014	BWA	Botswana	1992 - 2014
BEL	Belgium	1992 - 2014	CHL	Chile	1992 - 2014
BHS	The Bahamas	1992 - 2014	CRI	Costa Rica	1992 - 2014
BRN	Brunei Darussalam	1992 - 2014	CZE	Czech Republic	1996 - 2014
CAN	Canada	1992 - 2014	EST	Estonia	1994 - 2014*
CHE	Switzerland	1992 - 2014	GAB	Gabon	1992 - 2014
CYP	Cyprus	1992 - 2014	HRV	Croatia	1993 - 2014
DEU	Germany	1992 - 2014	HUN	Hungary	1992 - 2014
DNK	Denmark	1992 - 2014	LBN	Lebanon	1992 - 2013
ESP	Spain	1992 - 2014	LBY	Libya	1992 - 2014*
FIN	Finland	1992 - 2014	LTU	Lithuania	1996 - 2014*
FRA	France	1992 - 2014	LVA	Latvia	1993 - 2014*
GBR	United Kingdom	1992 - 2014	MEX	Mexico	1992 - 2014
GRC	Greece	1992 - 2014	MNE	Montenegro	2001 - 2014
IRL	Ireland	1992 - 2014	MUS	Mauritius	1992 - 2014
ISL	Iceland	1992 - 2014	MYS	Malaysia	1992 - 2014
ISR	Israel	1992 - 2014	OMN	Oman	1992 - 2014
ITA	Italy	1992 - 2014	PAN	Panama	1992 - 2014
JPN	Japan	1992 - 2014	POL	Poland	1992 - 2014
KOR	Korea	1992 - 2014	SAU	Saudi Arabia	1992 - 2014
KWT	Kuwait	1992 - 2014*	SRB	Serbia	1999 - 2014
LUX	Luxembourg	1992 - 2014	SVK	Slovak Republic	1994 - 2014
NLD	Netherlands	1992 - 2014	TTO	Trinidad and Tobago	1992 - 2014
NOR	Norway	1992 - 2014	TUR	Turkey	1992 - 2014
NZL	New Zealand	1992 - 2014	URY	Uruguay	1992 - 2014
PRI	Puerto Rico	1992 - 2014	VEN	Venezuela	1992 - 2014*
PRT	Portugal	1992 - 2014	ZAF	South Africa	1992 - 2014
QAT	Qatar	1992 - 2014*			
SVN	Slovenia	1993 - 2014			
SWE	Sweden	1992 - 2014			
TWN	Taiwan Province of China	1992 - 2014			
USA	United States	1992 - 2014	LOW INCOME COUNTRIES:		
			AFG	Afghanistan	2003 - 2014
LOWER MIDDLE INCOME COUNTRIES:			AZE	Azerbaijan	1993 - 2014*
AGO	Angola	1992 - 2014*	BDI	Burundi	1992 - 2014
ALB	Albania	1992 - 2014*	BEN	Benin	1992 - 2014
ARM	Armenia	1993 - 2014*	BFA	Burkina Faso	1992 - 2014
BGR	Bulgaria	1992 - 2014	BGD	Bangladesh	1992 - 2014
BIH	Bosnia and Herzegovina	1997 - 2014	BTN	Bhutan	1992 - 2014
BLR	Belarus	1993 - 2014*	CAF	Central African Republic	1992 - 2012
BLZ	Belize	1992 - 2014	CIV	Côte d'Ivoire	1992 - 2014
BOL	Bolivia	1992 - 2014	CMR	Cameroon	1992 - 2014
CHN	China	1992 - 2014	COD	Democratic Republic of the Congo	1992 - 2014*
COL	Colombia	1992 - 2014	COG	Republic of Congo	1992 - 2014
CPV	Cabo Verde	1992 - 2014	COM	Comoros	1992 - 2014
DJI	Djibouti	1992 - 2014	ERI	Eritrea	1993 - 2006*
DOM	Dominican Republic	1992 - 2014	ETH	Ethiopia	1992 - 2014*
DZA	Algeria	1992 - 2014	GHA	Ghana	1992 - 2014
ECU	Ecuador	1992 - 2014	GIN	Guinea	1992 - 2011
EGY	Egypt	1992 - 2014	GMB	The Gambia	1992 - 2014
FJI	Fiji	1992 - 2014	GNB	Guinea-Bissau	1992 - 2014*
GEO	Georgia	1995 - 2014	HTI	Haiti	1992 - 2014
GTM	Guatemala	1992 - 2014	IND	India	1992 - 2014*
GUY	Guyana	1992 - 2014	KEN	Kenya	1992 - 2014
HND	Honduras	1992 - 2014	KGZ	Kyrgyz Republic	1993 - 2014*
IDN	Indonesia	1992 - 2014	KHM	Cambodia	1992 - 2014
IRN	Islamic Republic of Iran	1992 - 2014	LAO	Lao P.D.R.	1992 - 2014
IRQ	Iraq	1999 - 2014*	LBR	Liberia	2001 - 2014*
JAM	Jamaica	1992 - 2014	LSO	Lesotho	1992 - 2014
JOR	Jordan	1992 - 2014	MDG	Madagascar	1992 - 2014*
KAZ	Kazakhstan	1993 - 2014*	MLI	Mali	1992 - 2014
LKA	Sri Lanka	1992 - 2014	MMR	Myanmar	1998 - 2014
MAR	Morocco	1992 - 2014	MNG	Mongolia	1992 - 2014
MDA	Moldova	1993 - 2014*	MOZ	Mozambique	1992 - 2014*
MKD	FYR Macedonia	1993 - 2014	MRT	Mauritania	1992 - 2014
NAM	Namibia	1992 - 2014	MWI	Malawi	1992 - 2011*
PER	Peru	1992 - 2014	NER	Niger	1992 - 2014
PHL	Philippines	1992 - 2014	NGA	Nigeria	1992 - 2014
PNG	Papua New Guinea	1992 - 2013	NIC	Nicaragua	1992 - 2014
PRY	Paraguay	1992 - 2014	NPL	Nepal	1992 - 2014
ROU	Romania	1992 - 2014	PAK	Pakistan	1992 - 2014
RUS	Russia	1993 - 2014*	RWA	Rwanda	1992 - 2014*
SLV	El Salvador	1992 - 2014	SDN	Sudan	1992 - 2010
SUR	Suriname	1992 - 2014	SEN	Senegal	1992 - 2014
SWZ	Swaziland	1992 - 2014	SLB	Solomon Islands	1992 - 2014*
SYR	Syria	1992 - 2010	SLE	Sierra Leone	1992 - 2014*
THA	Thailand	1992 - 2014	TCD	Chad	1992 - 2014*
TKM	Turkmenistan	1993 - 2014*	TGO	Togo	1992 - 2013*
TLS	Timor-Leste	2001 - 2014	TJK	Tajikistan	1993 - 2014*
TUN	Tunisia	1992 - 2014	TZA	Tanzania	1992 - 2014
UKR	Ukraine	1993 - 2014*	UGA	Uganda	1992 - 2014
UVK	Kosovo	2001 - 2014	VNM	Vietnam	1992 - 2014
UZB	Uzbekistan	1993 - 2014	YEM	Yemen	1992 - 2008
VUT	Vanuatu	1992 - 2014	ZMB	Zambia	1992 - 2014*
WSM	Samoa	1992 - 2014	ZWE	Zimbabwe	1999 - 2013*

*Outliers: AGO (93), ALB (97), ARG (02), ARM (09), AZE (94, 95, 05, 06), BLR (94, 95), COD (93), ERI (95,01), EST (09), ETH (92), GNB (98), IDN (98), IRQ (00-04), KAZ (94), KGZ (94), KWT (92, 93, 09), LBR (03), LBY (03, 05, 11-14), LTU (09), LVA (09), MDA (94), MDG (02), MOZ (96), MWI (94), QAT (97, 06), RUS (94), RWA (93-95), SLB (00), SLE (96, 97, 01, 02, 13), TCD (04), TGO (93, 94, 95), TJK (94), TKM (94, 97), UKR (94, 09), VEN (04), ZMB (94), ZWE (03, 08, 10, 11).