Forecasting GDP growth from outer space*

Jaqueson K. Galimberti†
ETH Zurich, Switzerland

Abstract

We evaluate the usefulness of satellite-based data on nighttime 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 the night lights data to improve the accuracy of model-based forecasts. We also find a substantial degree of heterogeneity across countries on 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 income level, logistics infrastructure, and 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:
- Night lights time series indicators: http://bit.ly/2IAn0v4

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

---

*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), and the JRC-CAS Workshop on Big Data and Macroeconomic Forecasting (Ispra, 2018). I thank to the participants of these meetings for comments and suggestions, and to my colleagues at the Chair of Applied Macroeconomics. Financial support from the SAS-IIF grant is gratefully acknowledged; I also thank the corresponding grant committee for the feedback.

†Correspondence address: KOF Swiss Economic Institute, ETH Zurich, LEE G 116, Leonhardstrasse 21, 8092 Zürich, Switzerland. Phone: +41 44 6328529. Fax: +41 44 6321218.
Personal website: https://sites.google.com/site/jkgeconoeng/
Contact e-mail: galimberti@kof.ethz.ch
expectation of the future state of the overall market at which the decision-maker is operating. In practice such a decision-maker is then continuously faced with an intricate forecasting challenge of finding leading indicators to the variables that are relevant to her/his business. In this paper we propose and evaluate the usage of satellite-based data on nighttime 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 to the country’s history of GDP growth rates. As our results reveal, a substantial portion 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 income level, expenditure 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 (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, 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, e.g., informal economic activity or less developed statistical agencies, the night lights data can provide an alternative proxy to economic activity. 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...
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\(^1\). 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. Here we also distinguish between two alternative specifications with respect to the estimation of these models, namely, a panel and country-individual specifications. Importantly, both the benchmark definition and pooling assumptions are later expanded to check for the robustness of the results: (i) forecasts from the International Monetary Fund (IMF) are used as an alternative to the AR(1) benchmark; and, (ii) partial pooling specifications are considered 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 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 the models estimation, starting with forecasts for 2001. Here we also evaluate alternative estimation assumptions in order to improve our understanding on the effects of estimation uncertainty; namely, a partial recursive estimation exercise is proposed to reduce the number of parameters estimated, while still remaining faithful to the informational restrictions of a (simulated) real-time forecasting exercise.

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

\(^1\)One major limitation of the night lights data available and employed here and in 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; here we attempt to compensate for the short time dimension with a broad sample on the cross-section dimension.
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 10% (cross-country 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 of data, the average relative improvement rises to a remarkable 33% rate.

Out-of-sample, the performance of the individual specifications deteriorate 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 41% on average across the indicators, most of which statistically significant at the 20% significance level; focusing on the full-sample correlation-based indicator, statistically significant improvements are obtained for more than 70% of the countries in our sample. The magnitude of these out-of-sample improvements also vary quite substantially across the countries, 27 of which achieve gains on the range between 20-48%, 45 on the range between 10-20%, and 46 on 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 using a meta analysis approach, evaluating several country-specific factors that could be associated with the performance of the night lights. Interestingly, our estimates indicate that the night lights appear more useful for economic forecasting in higher income countries. Going beyond income levels, we also find that countries with lower consumption expenditure (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 other four 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 forecast evaluation results, robustness checks, and an attempt to uncover key factors explanatory 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 indicate the presence of economic activity\[1\]. Clearly, the direction of causality between these variables goes from GDP to lights, i.e., 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

\[ y_t = z_t + u_t, \]

(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, e.g., 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

\[ x_t = \beta z_t + e_t, \]

(2.2)

where \(x_t\) stands for measured lights growth, \(\beta\) for the elasticity of lights with respect to true real GDP, and \(e_t\) for a measurement error in this relationship due to, e.g., satellite sensor’s noise (see Section 3.1) or changes in how production is translated into lights within a 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-pone measures can be combined to recover the true value of the common

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

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

and the lights-based forecasts by

$$\tilde{y}_t = \rho y_{t-1} + \theta x_{t-1},$$

where $\rho$, $\varrho$, and $\theta$ 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 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

$$\frac{\beta^2 \sigma^2_u}{\beta^2 \sigma^2_u + \sigma^2_z (1 + \sigma^2_z / \sigma^2_u)},$$

where $\sigma^2_z$, $\sigma^2_u$, and $\sigma^2_e$ stand for the variances of $z_t$, $u_t$, and $e_t$, respectively. Intuitively, the usefulness of the lights data increase with $|\beta|$, and decrease with the magnitude of the measurement error in the growth vs. lights relationship.

### 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. In this context, the night lights data can provide an alternative predictor, particularly considering that such investments often generate intense lights during construction time. I.e., tracking the variation of geo-located lights can signal
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.

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 of 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, when construction started until the factory opening 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
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 timely captured by the lights data than other sources. That may be particularly relevant for forecasting purposes in countries where businesses tend to start operations informally, but end up entering the formal economy (and GDP statistics) only after succeeding to mature. 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

---

Footnote: 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).
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 of which operating simultaneously; hence, for some years two composite images have been produced, in which cases we adopt their average after intercalibration. The intercalibration is necessary because the OLS has no on-board calibration of the visible band while the sensors performance degrade over time, not to mention the evolution in sensors 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 little changes of light brightness 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.

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, 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

5Other 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).
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.6.

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 (i.e., 63), or “top-coded”, 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 Earth’s curvature, the area covered by each pixel depends on its latitude, e.g.: 0.85 km\(^2\) at equator, 0.37 km\(^2\) at S65°, and 0.22 km\(^2\) at N75°. 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 are their amplitude on the ground; to make these pixels comparable (and aggregable), we re-scale the gridded light intensity measures by multiplying them by their latitude-implied area\(^6\).

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 6. 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.

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, i.e., 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).
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 countries GDP. Notice the GDP data is 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 1,000 km$^2$, 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.6.

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: (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. I.e., 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 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.

---

8In a previous version of this paper we have also considered distribution-based indicators (e.g., 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 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.

Notes: The snapshot image is produced according to figure 1.1 notes, though “turning off” the lights outside the country’s borders. The pixels’ classification is based on the correlations between GDP growth and lagged lights growth using observations up to 2013. The presented classification does not account for the significance of the correlations.

We assess three important dimensions on the design of the correlation-based measures. First, when calculating the pixels correlations one has to decide between using the contemporaneous lights changes or their lagged values in relation to the countries GDP growth rates. As it will become clear from our estimation and forecasting results this choice depends on the timing of the forecasting exercise: when the contemporaneous relationship between lights and growth is used to obtain the forecasts, such as for nowcasting purposes, the contemporaneous correlations provide the most meaningful and more accurate classifications; when the forecasts are based on lagged lights, in contrast, the lagged correlations are more appropriate. Figure 3.3 presents a comparative of these two alternative definitions, again focusing on the case of France for illustrative purposes.

Second, the vintage of data used for the calculation of the pixels’ correlations is another 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 3.3), where the pixels correlations are (re-)calculated recursively according to the data availability, i.e., 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 3.3), 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 enforces 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-
Figure 3.3: Time series of correlation-based sum of lights indicators for France.

(a) Contemporaneous. (b) Lagged.

Notes: The colored lines depict the evolution of the sum of lights over pixels classified according to the past correlation of their contemporaneous (panel a) or lagged (panel b) lights growth and the country’s GDP growth. The solid lines refer to the classification obtained using observations up to 2013, whereas the dashed lines refer to previous vintages. Pixels with correlation significance above 0.25 are discarded. All series are indexed to the value of 100 at 1992.

based indicators can be further specialized to account only for pixels with statistically significant correlations. I.e., before summing up the lights in each classification (positive/negative) we also calculate the associated significances of the pixels 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.

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

$$y_{i,t} = \alpha_i + \rho y_{i,t-1} + \epsilon_{i,t}, \quad (4.1)$$

for the pooled specification, and

$$y_{i,t} = \alpha'_i + \rho_i y_{i,t-1} + \epsilon_{i,t}, \quad (4.2)$$
for the individual specification, where \( y_{i,t} \) stands for country \( i \)'s (\( i = 1, \ldots, 167 \)) GDP growth rate for year \( t \) (\( t = 1993, \ldots, 2014 \)), and \( \alpha_i^{(\ell)} \) for country fixed effects. \(^9\)

The night lights-based forecasts are obtained by augmenting the benchmark models with the night lights indicators discussed in the previous section. \(^10\) Letting \( x_{k,i,t} \) stand for a vector containing indicator \( k \), the augmented models are given by

\[
y_{i,t} = \alpha_k + \varrho_k y_{i,t-1} + \theta_k x_{k,i,t-1} + \varepsilon_{k,i,t},
\]

\[
y_{i,t} = \alpha_k' + \varrho_k y_{i,t-1} + \theta_k x_{k,i,t-1} + \epsilon_{k,i,t},
\]

for the pooled and individual specifications, respectively, where the parameter vectors \( \theta_{k(i)} \) have dimensions conformable to the number of combined indicator measures. E.g., for the standard SoL indicator, \( x_{k,i,t} \) is univariate, i.e., containing only one indicator series at a time; for the cases of the location-based indicators, two SoL growth measures are produced 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 \(^11\) and we report the results of such exercise in the Appendix A.3.

One important issue on the estimation of these models using a global sample of countries is the likely presence of outliers to 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-stages estimation approach for outliers detection. First, we estimate the benchmark panel model specification with all available observations and derive the corresponding residuals. Outliers are then detected based on the statistical significance (\( p \)-value smaller than 1\%) of each disturbance; a total of 64 outliers are detected according to this procedure (these are listed in the Appendix A.6). In the second stage we obtain the final estimates of the models, both panel and individual specifications, excluding the detected outliers from the sample.

\(^9\) 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 (e.g., 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.

\(^10\) Models including only the night lights indicators, i.e., 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.

\(^11\) Although the data we use in this paper is only available for free 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.
4.2 Model estimates

The estimates of the panel models, (4.1) and (4.3), are reported in Table 4.1. For the models augmented with the night lights indicators we also distinguish between the contemporaneous and the lagged estimated relationships. Interestingly, for the estimates based on the standard total SoL indicator only the contemporaneous relationship is found to be statistically significant, a result that anticipates the weakness of this indicator for forecasting purposes.

Table 4.1: Panel estimates of forecasting models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimates</th>
<th>Regression stats.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR(1)</td>
<td></td>
</tr>
<tr>
<td>(i) AR(1) benchmark</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.256***</td>
<td>3601</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>0.274</td>
</tr>
<tr>
<td>(ii) AR(1) + Contemporaneous SoL growth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.238***</td>
<td>3311</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>0.277</td>
</tr>
<tr>
<td>Correlated pixels A (+/-)</td>
<td>0.014***</td>
<td>3294</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>0.374</td>
</tr>
<tr>
<td>Correlated pixels B (+/-)</td>
<td>-0.033***</td>
<td>3306</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>0.276</td>
</tr>
<tr>
<td>(iii) AR(1) + Lagged SoL growth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.247***</td>
<td>3337</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>0.275</td>
</tr>
<tr>
<td>Correlated pixels A (+/-)</td>
<td>-0.012***</td>
<td>3317</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>0.282</td>
</tr>
<tr>
<td>Correlated pixels B (+/-)</td>
<td>0.013***</td>
<td>3332</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>0.366</td>
</tr>
</tbody>
</table>

Notes: Estimates obtained regressing real GDP growth on the corresponding night lights indicator values, plus an autoregressive term, and country fixed effects. The correlated pixels SoL indicators are based on the contemporaneous (A) or lagged (B) correlations of pixel’s lights growth and the country’s GDP growth, calculated using observations up to 2013. Estimation by least squares in two stages: first, with all available observations; second, excluding outliers (64 in total; see text for details). Values inside parentheses are country-clustered robust standard errors. ***, **, and * stand for 1%, 5%, and 10% levels of statistical significance, respectively.

In contrast, the estimates are more favorable to the correlation-based indicators, most of which are found to have statistically significant relationships with GDP growth; that is particularly the case when the pixels correlations are also calculated according to the estimated relationship, i.e., contemporaneous/lagged correlated pixels (denoted by A/B, respectively, in Table 4.1) for the contemporaneous/lagged relationship with GDP growth. Moreover, using the appropriate timing in the calculation of the pixels correlations also yields higher explanatory power, as evidenced by the adjusted $R^2$ statistics, and reasonable signs on the estimated coefficients, i.e., positive (negative) for the indicator reflecting the positively (negatively) correlated pixels.
Figure 4.1: Individual estimates of forecasting models.

(a) Total SoL.

(b) Correlation-based SoL.

Notes: The estimates refer to those obtained under the lagged night lights specifications, and lagged correlations for the correlation-based SoL estimates (i.e., panel b estimates correspond to the type B estimates from Table 4.1-iii).

Nevertheless, the assumption of a common relationship between the night lights indicators and GDP growth across countries, as implied by the panel specifications, is put into question when we look at the 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 overly dispersed to justify the 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
\[
\hat{y}_{k,i,t+1}^S = \hat{\alpha}_{k,i}^S + \hat{\varphi}_{k(i)} y_{i,t} + \hat{\theta}_{k(i)} x_{k,i,t}, \tag{4.5}
\]
where the superscript $S$ 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, i.e., with $t = \{1992, \ldots, 2013\}$ in equation (4.5), using model parameters estimated with our full sample of data, i.e., with $S = \{1992, \ldots, 2014\}$ in (4.5). Naturally, this is not a realistic real-time forecasting exercise considering that data beyond the forecast base period is 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** (OoS) 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 $t = \{2000, \ldots, 2013\}$, $S_t = \{i\}_i=1992$.

Furthermore, in order to assess the robustness of these informational assumptions to the effects of parameter estimation we conduct two additional OoS exercises: (i) partially estimating 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 (OoS-partial recursive); and, (ii) using the in-sample parameter estimates, but focusing only on the forecasts over the 2001 to 2014 evaluation sample (OoS-non-recursive). Whereas both exercises aim to uncover the effects of estimation uncertainty, which is particularly relevant for the evaluation of the forecasts based on the country-individual specifications, only the former exercise remains faithful to the restrictions of real-time forecasting.

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 the countries GDPs (in PPP terms) as weights. A similar weighted averaging is applied to summarize the RMSE ratios, except that these are averaged geometrically.

---

12 Using equation (4.5)’s notation the OoS-partial exercise is equivalent to setting $\hat{\alpha}_{k(i)}^S = \hat{\alpha}_i^S$, $\hat{\varphi}_{k(i)} = \hat{\varphi}_{0(i)}^S$, for all $k$, while $\hat{\theta}_{k(i)}^S$ is obtained by regressing $(y_{i,j+1} - \hat{\varphi}_{k(i)}^S y_{i,t})$ on $x_{k,i,j}$ recursively with $j = S_t$, whereas in the OoS-non-recursive exercise we evaluate the forecasts obtained for $t = \{2000, \ldots, 2013\}$ with $S = \{1992, \ldots, 2013\}$. 

---

18
5 Forecasts 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 25% more accurate than the panel estimated models. Across the night lights indicators, the correlation-based ones stand out, with accuracy improvements reaching up to 33% (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 performances suggest that part of the improvements brought by the night lights are 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 pixels correlations. That is the case even to the point of averting the effects of estimation uncertainty affecting the out-of-sample recursive results, as we will discuss below. 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 venue 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 estimated models, these results seem to be driven mainly by parameter estimation errors, due to the small samples available for the first recursive estimations. E.g., 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 OoS exercise favored the latter.

5.2 Out-of-sample robustness

In order to shed further light on the effects of estimation uncertainty on the individual models’ OoS performance, we compare the forecasts performance under the two additional OoS exercises described above: (i) a partial recursive estimation OoS exercise, where the number of recursively estimated parameters in the night lights-augmented models is reduced to only one; and, (ii) a non-recursive estimation OoS exercise, where the models’ parameters are fixed to those obtained with the full-sample of data.
Table 5.1: Forecast evaluation statistics.

<table>
<thead>
<tr>
<th>Model</th>
<th>Panel</th>
<th>Individual</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) In-sample evaluation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1) benchmark</td>
<td>2.242</td>
<td>2.116</td>
<td>3468</td>
</tr>
<tr>
<td>AR(1) + Lagged SoL indicators:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Total SoL growth</td>
<td>2.239</td>
<td>2.051</td>
<td>3337</td>
</tr>
<tr>
<td>+ Correlated pixels SoL growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-time classification</td>
<td>2.248</td>
<td>1.955</td>
<td>2861</td>
</tr>
<tr>
<td>Full-sample classification</td>
<td>1.911</td>
<td>1.433</td>
<td>3332</td>
</tr>
<tr>
<td>(ii) Out-of-sample evaluation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1) benchmark</td>
<td>2.443</td>
<td>2.564</td>
<td>2233</td>
</tr>
<tr>
<td>AR(1) + Lagged SoL indicators:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Total SoL growth</td>
<td>2.464</td>
<td>2.794</td>
<td>2221</td>
</tr>
<tr>
<td>+ Correlated pixels SoL growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-time classification</td>
<td>2.612</td>
<td>3.008</td>
<td>2195</td>
</tr>
<tr>
<td>Full-sample classification</td>
<td>2.093</td>
<td>2.376</td>
<td>2202</td>
</tr>
</tbody>
</table>

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. All model specifications contain an AR(1) term and country fixed effects. The correlation-based SoL indicators are based on correlations between lagged lights changes and the country’s GDP growth (type B measure). The real-time classification of pixels 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.
Figure 5.1: Out-of-sample estimation exercises.

(a) Panel. (b) Individual.

Notes: The plotted statistics correspond to the weighted cross-country geometric averages of the RMSE ratios relative to the AR(1) benchmark model. The in-sample results are based on forecasts over the period from 1993 to 2014, whereas the out-of-sample results cover the period from 2001 to 2014 under different estimation assumptions: under the Recursive exercises the models are estimated using only past observations available at the forecast base period, while under the Non-recursive exercise the forecasts are based on the in-sample parameter estimates; under the Partial recursive exercise only the relationships with the night lights indicators are re-estimated, whereas the AR(1) and intercept coefficients are fixed to the recursive benchmark estimates; see also the explanation in footnote (12).

The results obtained under these exercises, presented in Figure 5.1, corroborate the explanation that the poorer out-of-sample results of the individual models are due to parameter instability. Particularly, notice that as we attenuate the effects of estimation uncertainty, i.e., moving from the recursive to the partial recursive and to the non-recursive exercises, the out-of-sample performances of the night lights-based forecasts tend to converge to those obtained in-sample.

The in-sample superior performance of the individually estimated models is consistent with our earlier findings with respect to the heterogeneity of the parameter estimates on the relationships between GDP growth and the night lights indicators. Although the out-of-sample performance of this approach can be substantially affected by small sample biases, we note that the correlation between in-sample and OoS performances of the individual models is highly positive, with an average value of 0.70. Hence, our in-sample assessments can be considered to provide a reliable guidance on the OoS potential of the night lights indicators to improve GDP growth forecasts for the countries in our sample.
Table 5.2: Frequencies of night lights improvements over benchmark model across countries.

<table>
<thead>
<tr>
<th>Statistics / specifications</th>
<th>In-sample</th>
<th>Out-of-sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Panel</td>
<td>Individual</td>
</tr>
<tr>
<td>(i) Indicator-specific improvement rates:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a) Total SoL growth</td>
<td>67.1%</td>
<td>50.3%</td>
</tr>
<tr>
<td></td>
<td>(10.2%)</td>
<td>(9.6%)</td>
</tr>
<tr>
<td>Correlated pixels SoL growth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(b) Real-time classification</td>
<td>74.9%</td>
<td>23.4%</td>
</tr>
<tr>
<td></td>
<td>(12.6%)</td>
<td>(10.2%)</td>
</tr>
<tr>
<td>(c) Full-sample classification</td>
<td>82.0%</td>
<td>84.4%</td>
</tr>
<tr>
<td></td>
<td>(79.6%)</td>
<td>(81.4%)</td>
</tr>
<tr>
<td>Average: (a) + (b) + (c)</td>
<td>74.7%</td>
<td>52.7%</td>
</tr>
<tr>
<td></td>
<td>(34.1%)</td>
<td>(33.7%)</td>
</tr>
<tr>
<td>(ii) Max-t frequency of rejections at 20% significance:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Models set: (a) + (b)</td>
<td>11.4%</td>
<td>11.4%</td>
</tr>
<tr>
<td></td>
<td>24.6%</td>
<td>22.8%</td>
</tr>
<tr>
<td>Models set: (a) + (b) + (c)</td>
<td>68.3%</td>
<td>72.5%</td>
</tr>
<tr>
<td></td>
<td>58.7%</td>
<td>67.1%</td>
</tr>
</tbody>
</table>

Notes: The frequencies measure the percentage of countries for which the night lights augmented model outperformed the benchmark. The out-of-sample results refer to those obtained under the partial recursive estimation exercise. 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.

5.3 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 (the statistical tests results will be discussed below).

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 OoS evaluation. Notwithstanding, these results also reveal that even in OoS forecasting the night lights data can still be useful for a considerable portion of countries, providing improvements to about 53%, using the panel specification, and 41%, 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 two thirds 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 5.2 we present the countries RMSE ratios, focusing on the forecasts obtained with the individual model augmented with the full-sample correlated pixels indicators, and grouped according to the World Bank income classification. We note that the improvement rates reported above are found to be slightly skewed against the low income countries, with OoS improvements for only 53% of the countries in that group, compared to 82%, 86%, and 73%, for the high, upper middle, and lower middle income groups of countries, respectively.

5.4 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, i.e., 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 models, also adapting the inference to finite samples by simulating the empirical distribution of the test statistics according to the bootstrap procedure proposed by Clark and McCracken (2012).

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. As expected, these hit-rates tend to be smaller than the improvement rates observed above, based solely on the sample RMSEs. Noteworthy exceptions include the OoS results with individual

---

13Country-specific evaluation reports are also available in a supplementary file.

14One 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, i.e., 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).

15In 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.

16The individual tests underlying these rates are provided in a separate worksheet, 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.
Notes: The RMSE ratios refer to forecasts obtained with the correlated pixels (full-sample classification) night lights model estimated individually for each country. The out-of-sample RMSE ratios refer to those obtained under the partial recursive estimation exercise. 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 of the Clark and West (2007) one-sided test for equal predictive accuracy in nested models, following the Clark and McCracken (2012) resampling approach for nested model reality checks. The countries are grouped according to their World Bank income classification and sorted within each group in ascending order according to their corresponding out-of-sample RMSE ratios.
correlation-based SoL forecasts, where the CW rejections are even 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 of similar out-of-sample evidence is likely to be spurious, showing that such 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 depicted in figure 5.2.

5.5 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 so as 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 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 countries individual estimates (see, e.g., Vahtid, 1999; Sarafidis and Weber, 2015, for previous applications of clustering methods in panel data); more details about this approach are provided in the Appendix A.4. The results of these exercises are presented in Table 5.3, which also reports the previous panel and individual results for comparative purposes.

As the RMSE ratios in Table 5.3 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 most of the exercises, with the exception of the OoS exercise with the full-sample correlation-based night lights indicators, where the clustering method achieves the best performance relative to the AR(1) benchmark forecasts. Hence, whereas there is some space for improvement with the partial pooling alternative, more elaborated methods may be required to single out the optimal partitioning of countries with respect
Table 5.3: Partial pooling RMSE ratios.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Panel</th>
<th>Partially pooled by</th>
<th>Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Income</td>
<td>Region</td>
</tr>
<tr>
<td>(i) In-sample evaluation.</td>
<td>Total SoL growth</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Correlated pixels SoL growth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Real-time</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Full-sample</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>(ii) Out-of-sample evaluation.</td>
<td>Total SoL growth</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Correlated pixels SoL growth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Real-time</td>
<td>1.07</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>Full-sample</td>
<td>0.85</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Notes: See the notes to Table [5.1] for details about the construction of the RMSE ratio statistics. The out-of-sample results refer to those obtained under the partial recursive estimation exercise. 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; the number of clusters is also determined separately by maximizing the average silhouettes, yielding selections of 13, 9, and 4 clusters for the total SoL, the real-time and the full-sample correlated pixels SoL specifications, respectively; see the Appendix [A.4] for more details.
to the relationship between the night lights and GDP growth. We leave this analysis for future research.

5.6 Alternative benchmark

Although the AR(1) model provides a parsimonious and well-known benchmark for the relative evaluation of the night lights-based forecasts, its practical value for annual macroeconomic forecasting is limited. Indeed, a variety of other sources of data, such as business tendency surveys and financial data, become available at a higher frequency and can be informative about the recent and future evolution of economic activity. In order to provide a more realistic evaluation of the night lights-based forecasts we reassess our results using the IMF WEO forecasts as a benchmark.

The IMF WEO forecasts are published twice a year, in April and September, and its historical database (October 2017 vintage) consists of forecasts 6-years ahead and covering up to 193 countries since 1990. These forecasts have been extensively evaluated in the literature and, in spite of some evidence of persistent biases, their performance have been found to be comparable to that of private sector consensus forecasts (Timmermann, 2007). Here we compare the performance of the current year and the 1-year-ahead forecasts to those obtained using the night lights data for the same target year over our sample of countries\(^\text{17}\). The results of these comparisons are summarized in Table 5.4 in the form of relative RMSE ratios, where values below one indicate the night lights outperformed the IMF forecasts, and vice versa.

As with our previous evaluation the results show some favorable evidence for the night lights data in forecasting in-sample, but more challenging in forecasting out-of-sample. Considering that our previous results indicate that the out-of-sample performance is affected by estimation errors we focus on the in-sample results. In that cases, both the real-time and the full-sample versions of the correlation-based SoL indicators perform better than the 1-year-ahead IMF forecasts. Naturally, the advantage of the night lights tends to shrink, and eventually reverse for current-year forecasts, as the timing gap between the target year and the base period of the forecast decreases, reflecting the incorporation of new information into the IMF forecasts. On the other hand, it is also important to mention that the night lights-based forecasts enjoy some informational advantage in relation to the previous year IMF forecasts considering that the previous year GDP and night lights data points were not available at the time the latter forecasts were published.

5.7 Potential explanatory factors

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

\(^{17}\)There is only one country for which the IMF has not produced GDP growth forecasts over our evaluation sample: Puerto Rico.
Table 5.4: Evaluation using IMF forecasts as benchmark.

<table>
<thead>
<tr>
<th>Models</th>
<th>RMSE ratios vs. WEO forecasts from</th>
<th>Improv. rates vs. WEO forecasts from</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Previous year</td>
<td>Current year</td>
</tr>
<tr>
<td></td>
<td>Spring</td>
<td>Fall</td>
</tr>
<tr>
<td>(i) In-sample evaluation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlated pixels SoL growth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-time</td>
<td>0.74</td>
<td>0.83</td>
</tr>
<tr>
<td>Full-sample</td>
<td>0.54</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>98%</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>(ii) Out-of-sample evaluation.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlated pixels SoL growth:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-time</td>
<td>1.08</td>
<td>1.28</td>
</tr>
<tr>
<td>Full-sample</td>
<td>0.83</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>35%</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>67%</td>
<td>54%</td>
</tr>
</tbody>
</table>

Notes: The benchmark forecasts come from the International Monetary Fund (IMF) World Economic Outlook (WEO) Spring (April) and Fall (September) issues. See the notes to Tables 5.1 and 5.2 for details about the construction of the RMSE ratio statistics and the improvement rates, respectively.

forecasting out-of-sample. Whereas the countries income level is found to be positively associated with the night lights performance (recall the analysis of Figure 5.2), we now attempt to uncover other key factors underlying the usefulness of the night lights data. For that purpose we conduct a meta analysis by examining 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 5.3, focusing on the four most important 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 the Appendix A.5.

The first factor presented in Figure 5.3 is the share of consumption in the country’s total expenditure, which is 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 usefulness of geo-located signals is also evidenced by the results for the second factor, which indicate that the night lights tend to achieve better performance in countries with better logistics infrastructure.

Finally, the third and fourth factors presented in Figure 5.3 are related to the measurement error hypothesis outline in Section 2.1. Intriguingly, panel (c) indicates that 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
Figure 5.3: Scatterplots between night lights performances and key country-specific factors.

Notes: The countries (log) RMSE ratios (depicted in the vertical axis of each plot) are the same as reported in Figure 5.2. Fitted regression lines, all statistically significant at the 1% significance level, are depicted as solid lines. The night lights coefficient estimates reported in panel (d) are normalized by their corresponding estimated standard errors.

lights to improve GDP measures in countries with less reliable statistics\textsuperscript{18}. 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 to better estimates propitiated by the higher quality of GDP data for those countries. The results presented in panel (d) seem to corroborate this explanation, where we find that the night lights performance tends to improve as the estimates of the relationship between GDP growth and (lagged) lights growth in the forecasting models increase.

6 Concluding remarks

In this paper we evaluated the usefulness of satellite-based data on nighttime 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 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 countries 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 stronger estimation biases due to

\textsuperscript{18}Notice equation (2.5) also predicts greater night lights improvements for countries with greater GDP measurement errors, $\sigma^2$. 

29
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 evidenced promising venues for future applications of the night lights data for economic measurement and forecasting.

References


A Appendix

A.1 Theoretical derivation of lights relative accuracy

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

\begin{align*}
\hat{\rho} &= \frac{\text{Cov}(z_t, z_{t-1})}{\sigma_z^2 + \sigma_u^2}, \quad (A.1) \\
\hat{\varrho} &= \frac{\text{Cov}(z_t, z_{t-1}) \sigma_z^2}{\sigma_z^2 \sigma_e^2 + \sigma_u^2 \sigma_z^2}, \quad (A.2) \\
\hat{\theta} &= \frac{\beta \text{Cov}(z_t, z_{t-1}) \sigma_u^2}{\sigma_z^2 \sigma_e^2 + \sigma_u^2 \sigma_x^2}, \quad (A.3)
\end{align*}

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

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

and

\begin{align*}
\tilde{\Delta} &= \sigma_y^2 - \text{Cov}(z_t, z_{t-1})^2 \left( \frac{\beta^2 \sigma_u^2 + \sigma_e^2}{\sigma_z^2 \sigma_e^2 + \sigma_u^2 \sigma_x^2} \right), \\
&= \sigma_y^2 - \text{Cov}(z_t, z_{t-1})^2 \left( \frac{\beta^2 \sigma_u^2}{\sigma_z^2 \sigma_e^2 + \sigma_u^2 \sigma_x^2} \right) \left( \frac{\beta^2 \sigma_u^2}{\sigma_z^2 \sigma_e^2 + \sigma_u^2 \sigma_x^2} \right), \quad (A.5)
\end{align*}

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

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

\[ DN_{r,p} = \phi_0 + \phi_1 DN_{s,p} + \phi_2 DN_{s,p}^2, \]  

(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 Contemporaneous forecasting results

The night lights can be used for contemporaneous forecasting too. Naturally, such forecasts can only be computed after the lights have been observed; whether the timing of these observations represent an advantage for macroeconomic forecasting will depend on the application. Here we report the results obtained using the same annual data adopted for the lagged specifications in the main text. Namely, to obtain the contemporaneous night lights-based forecasts we adjust the augmented models of equations (4.3) and (4.4) to use the contemporaneous values of the night lights indicators, i.e.,

\[ y_{i,t} = \alpha_k + \varrho_k y_{i,t-1} + \theta_k x_{k,i,t} + \epsilon_{k,i,t}, \]  

(A.8)

\[ y_{i,t} = \alpha'_{k,i} + \varrho_{k,i} y_{i,t-1} + \theta_{k,i} x_{k,i,t} + \epsilon_{k,i,t}, \]  

(A.9)

from which the contemporaneous forecasts are obtained as

\[ \hat{y}_{k,i,t+1} = \hat{\alpha}_{k,i} + \hat{\varrho}_{k,i} y_{i,t} + \hat{\theta}_{k,i} x_{k,i,t+1}. \]  

(A.10)

The averaged results for the main evaluation exercises are presented in Table A.2. Overall, the contemporaneous forecasting results are similar to those obtained with the lagged specification, namely: (i) the individual specifications are favored in-sample, and but are subject to a strong effect of estimation uncertainty in forecasting out-of-sample; and, (ii) the location-based measures tend to perform better than the traditional total SoL, whereas the correlation-based measure using the full-sample for the pixels classification obtains remarkable average improvements. Finally, as expected, the night lights improvements are generally higher than those obtained under the lagged specification.

A.4 k-Means partial pooling

We adopt the k-means clustering method in order to group countries according to their proximity in terms of their forecasting models’ parameter estimates. The classifications are obtained
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.

<table>
<thead>
<tr>
<th>Sat.</th>
<th>Year</th>
<th>$\phi_0$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$R^2$</th>
<th>N.Pixs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1992</td>
<td>-1.8913</td>
<td>1.2378</td>
<td>-0.0042</td>
<td>0.904</td>
<td>29796</td>
</tr>
<tr>
<td>10</td>
<td>1993</td>
<td>-1.0148</td>
<td>1.2210</td>
<td>-0.0039</td>
<td>0.926</td>
<td>33413</td>
</tr>
<tr>
<td>10</td>
<td>1994</td>
<td>-0.4652</td>
<td>1.1417</td>
<td>-0.0031</td>
<td>0.932</td>
<td>30561</td>
</tr>
<tr>
<td>12</td>
<td>1994</td>
<td>-0.7232</td>
<td>0.8819</td>
<td>0.0013</td>
<td>0.929</td>
<td>28980</td>
</tr>
<tr>
<td>12</td>
<td>1995</td>
<td>-0.2319</td>
<td>0.9166</td>
<td>0.0004</td>
<td>0.938</td>
<td>32207</td>
</tr>
<tr>
<td>12</td>
<td>1996</td>
<td>-0.1864</td>
<td>0.9556</td>
<td>0.0002</td>
<td>0.937</td>
<td>30016</td>
</tr>
<tr>
<td>12</td>
<td>1997</td>
<td>-0.4029</td>
<td>0.8706</td>
<td>0.0015</td>
<td>0.939</td>
<td>31979</td>
</tr>
<tr>
<td>12</td>
<td>1998</td>
<td>0.1048</td>
<td>0.7470</td>
<td>0.0027</td>
<td>0.951</td>
<td>32436</td>
</tr>
<tr>
<td>12</td>
<td>1999</td>
<td>0.4627</td>
<td>0.6816</td>
<td>0.0035</td>
<td>0.940</td>
<td>31268</td>
</tr>
<tr>
<td>14</td>
<td>1997</td>
<td>-1.4670</td>
<td>1.3965</td>
<td>-0.0064</td>
<td>0.935</td>
<td>31695</td>
</tr>
<tr>
<td>14</td>
<td>1998</td>
<td>-0.1328</td>
<td>1.2448</td>
<td>-0.0048</td>
<td>0.933</td>
<td>30864</td>
</tr>
<tr>
<td>14</td>
<td>1999</td>
<td>-0.2841</td>
<td>1.1501</td>
<td>-0.0029</td>
<td>0.950</td>
<td>33353</td>
</tr>
<tr>
<td>14</td>
<td>2000</td>
<td>0.5327</td>
<td>1.0149</td>
<td>-0.0012</td>
<td>0.943</td>
<td>32084</td>
</tr>
<tr>
<td>14</td>
<td>2001</td>
<td>-0.0964</td>
<td>1.0062</td>
<td>-0.0009</td>
<td>0.958</td>
<td>32844</td>
</tr>
<tr>
<td>14</td>
<td>2002</td>
<td>0.6088</td>
<td>0.8688</td>
<td>0.0009</td>
<td>0.954</td>
<td>31427</td>
</tr>
<tr>
<td>14</td>
<td>2003</td>
<td>0.1951</td>
<td>0.9600</td>
<td>-0.0002</td>
<td>0.959</td>
<td>33361</td>
</tr>
<tr>
<td>15</td>
<td>2000</td>
<td>0.1642</td>
<td>0.7380</td>
<td>0.0029</td>
<td>0.939</td>
<td>33651</td>
</tr>
<tr>
<td>15</td>
<td>2001</td>
<td>-0.4642</td>
<td>0.7937</td>
<td>0.0027</td>
<td>0.956</td>
<td>33059</td>
</tr>
<tr>
<td>15</td>
<td>2002</td>
<td>0.1410</td>
<td>0.6751</td>
<td>0.0042</td>
<td>0.960</td>
<td>32359</td>
</tr>
<tr>
<td>15</td>
<td>2003</td>
<td>-0.3656</td>
<td>1.1889</td>
<td>-0.0031</td>
<td>0.966</td>
<td>33340</td>
</tr>
<tr>
<td>15</td>
<td>2004</td>
<td>0.0403</td>
<td>1.0301</td>
<td>-0.0009</td>
<td>0.976</td>
<td>31080</td>
</tr>
<tr>
<td>15</td>
<td>2005</td>
<td>0.0837</td>
<td>0.9788</td>
<td>0.0001</td>
<td>0.970</td>
<td>33509</td>
</tr>
<tr>
<td>15</td>
<td>2006</td>
<td>0.0000</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.000</td>
<td>33877</td>
</tr>
<tr>
<td>15</td>
<td>2007</td>
<td>0.5517</td>
<td>0.9891</td>
<td>0.0002</td>
<td>0.966</td>
<td>31159</td>
</tr>
<tr>
<td>16</td>
<td>2004</td>
<td>-0.2095</td>
<td>0.9014</td>
<td>0.0007</td>
<td>0.958</td>
<td>31752</td>
</tr>
<tr>
<td>16</td>
<td>2005</td>
<td>-0.5565</td>
<td>1.1083</td>
<td>-0.0021</td>
<td>0.970</td>
<td>33618</td>
</tr>
<tr>
<td>16</td>
<td>2006</td>
<td>-0.4076</td>
<td>0.8657</td>
<td>0.0020</td>
<td>0.970</td>
<td>31893</td>
</tr>
<tr>
<td>16</td>
<td>2007</td>
<td>0.4299</td>
<td>0.6412</td>
<td>0.0046</td>
<td>0.972</td>
<td>32308</td>
</tr>
<tr>
<td>16</td>
<td>2008</td>
<td>0.2339</td>
<td>0.7200</td>
<td>0.0033</td>
<td>0.966</td>
<td>32271</td>
</tr>
<tr>
<td>16</td>
<td>2009</td>
<td>0.3699</td>
<td>0.7898</td>
<td>0.0022</td>
<td>0.962</td>
<td>28894</td>
</tr>
<tr>
<td>18</td>
<td>2010</td>
<td>1.8024</td>
<td>0.2926</td>
<td>0.0092</td>
<td>0.931</td>
<td>31117</td>
</tr>
<tr>
<td>18</td>
<td>2011</td>
<td>1.6726</td>
<td>0.4687</td>
<td>0.0062</td>
<td>0.936</td>
<td>31245</td>
</tr>
<tr>
<td>18</td>
<td>2012</td>
<td>1.6511</td>
<td>0.3815</td>
<td>0.0078</td>
<td>0.954</td>
<td>32151</td>
</tr>
<tr>
<td>18</td>
<td>2013</td>
<td>1.5803</td>
<td>0.4479</td>
<td>0.0064</td>
<td>0.957</td>
<td>32181</td>
</tr>
</tbody>
</table>

Notes: The estimates refer to equation (A.7) and are based on Sicily’s night lights data.
Table A.2: Contemporaneous forecast evaluation statistics.

<table>
<thead>
<tr>
<th>Model</th>
<th>Panel RMSE</th>
<th>Panel Ratio</th>
<th>Panel SSR</th>
<th>Individual RMSE</th>
<th>Individual Ratio</th>
<th>Individual SSR</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) In-sample evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1) benchmark</td>
<td>2.270</td>
<td>-</td>
<td>65%</td>
<td>2.139</td>
<td>-</td>
<td>65%</td>
<td>3311</td>
</tr>
<tr>
<td>AR(1) + Contemporaneous SoL indicators:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Total SoL growth</td>
<td>2.255</td>
<td>0.99</td>
<td>66%</td>
<td>2.064</td>
<td>0.96</td>
<td>69%</td>
<td>3311</td>
</tr>
<tr>
<td>+ Correlated pixels SoL growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-time</td>
<td>2.253</td>
<td>0.98</td>
<td>65%</td>
<td>1.998</td>
<td>0.92</td>
<td>70%</td>
<td>2844</td>
</tr>
<tr>
<td>Full-sample</td>
<td>1.848</td>
<td>0.78</td>
<td>73%</td>
<td>1.366</td>
<td>0.63</td>
<td>79%</td>
<td>3294</td>
</tr>
<tr>
<td>(ii) Out-of-sample evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR(1) benchmark</td>
<td>2.499</td>
<td>-</td>
<td>64%</td>
<td>2.621</td>
<td>-</td>
<td>62%</td>
<td>2078</td>
</tr>
<tr>
<td>AR(1) + Contemporaneous SoL indicators:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Total SoL growth</td>
<td>2.486</td>
<td>0.99</td>
<td>65%</td>
<td>2.818</td>
<td>1.07</td>
<td>57%</td>
<td>2066</td>
</tr>
<tr>
<td>+ Correlated pixels SoL growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real-time</td>
<td>2.666</td>
<td>1.08</td>
<td>65%</td>
<td>2.921</td>
<td>1.10</td>
<td>61%</td>
<td>2024</td>
</tr>
<tr>
<td>Full-sample</td>
<td>2.095</td>
<td>0.82</td>
<td>69%</td>
<td>2.063</td>
<td>0.75</td>
<td>71%</td>
<td>2036</td>
</tr>
</tbody>
</table>

Notes: Same as notes to Table 5.1 except that here the evaluation samples range from 1993 to 2013, in-sample, and from 2001 to 2013, out-of-sample, and that the correlation-based SoL indicators are based on correlations between contemporaneous lights changes and the country’s GDP growth (type A measure).

separately for each model specification based on the Euclidean distance between the countries individual in-sample estimates of the AR(1) and the night lights indicator(s) coefficients. All coefficients are normalized in terms of standard deviations around their cross-country mean estimate. Besides, the number of clusters is also determined separately for each model by maximizing the average silhouettes under classifications of up to 30 clusters; the silhouette value for each point is a measure that ranges from -1 to +1, and captures how similar that point is to points in its own cluster relative to points in other clusters. The silhouettes for the selected cluster sizes are presented in Figure A.2, with 13, 9, and 4 clusters for the total SoL, the real-time and the full-sample correlated pixels SoL specifications, respectively. The corresponding clustering groups are presented in Figure A.3.

A.5 Meta analysis supplementary results and data sources

The data used for the meta analysis 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). A complete list of the variables considered is presented in Table A.3. The scatterplots relating all these variables to the performance of the difference night lights-based forecasts are presented in Figures A.4-A.6.
Figure A.2: k-Means silhouette values.

(a) Total SoL.  
(b) Real-time corr. pixels.  
(c) Full-sample corr. pixels.

Notes: The number of clusters is selected individually for each specification by maximizing the average silhouette value.

Table A.3: Meta analysis variables and data sources.

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

Notes: §Revisions statistics calculated between first published and latest available data.
Figure A.3: k-Means clusters.

(a) Total SoL.

(b) Real-time correlated pixels SoL (+).

(c) Real-time correlated pixels SoL (-).

(d) Full-sample correlated pixels SoL (+).

(e) Full-sample correlated pixels SoL (-).
Figure A.4: Cross-country scatterplots on performance of total SoL indicator.

Notes: The countries (log) RMSE ratios (depicted in the vertical axis of each plot) are the same as reported in Figure 5, i.e., using forecasts obtained from partially recursive estimates of individual models with the full-sample correlated pixels indicators. Statistically significant relationships, at 10% significance level, and based on heteroskedasticity robust standard errors, are depicted as solid lines; dashed lines otherwise.
Figure A.5: Cross-country scatterplots on performance of real-time correlated pixels indicator.

Notes: see notes of Figure A.4
Figure A.6: Cross-country scatterplots on performance of full-sample correlated pixels indicator.

Notes: see notes of Figure A.4
### A.6 List of countries

<table>
<thead>
<tr>
<th>ISO</th>
<th>Country</th>
<th>GDP Sample</th>
<th>ISO</th>
<th>Country</th>
<th>GDP Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIGH INCOME COUNTRIES:</td>
<td></td>
<td></td>
<td>UPPER MIDDLE INCOME COUNTRIES:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESP</td>
<td>Spain</td>
<td>1992 - 2014</td>
<td>EST</td>
<td>Estonia</td>
<td>1994 - 2014*</td>
</tr>
<tr>
<td>IRL</td>
<td>Ireland</td>
<td>1992 - 2014</td>
<td>IND</td>
<td>India</td>
<td>1992 - 2014*</td>
</tr>
<tr>
<td>ITA</td>
<td>Italy</td>
<td>1992 - 2014</td>
<td>KWT</td>
<td>Kuwait</td>
<td>1992 - 2014*</td>
</tr>
<tr>
<td>NGR</td>
<td>Nigeria</td>
<td>1992 - 2014*</td>
<td>MOL</td>
<td>Moldova</td>
<td>1993 - 2014*</td>
</tr>
<tr>
<td>QAT</td>
<td>Qatar</td>
<td>1992 - 2014*</td>
<td>QAT</td>
<td>Qatar</td>
<td>1992 - 2014*</td>
</tr>
<tr>
<td>KOS</td>
<td>Kosovo</td>
<td>2001 - 2014</td>
<td>KOS</td>
<td>Kosovo</td>
<td>2001 - 2014</td>
</tr>
</tbody>
</table>

*Outliers: AGO, ALB, ARM, ARI, AZE, BRA, BEL, BGR, BLM, BOL, CAN, CHE, CZE, CRO, CUB, CYM, CZE, DNK, ESP, FRA, GBR, GRC, HKG, HUN, IRL, ISR, ITA, JPN, KOR, KWT, LUX, MCO, MEX, MLY, MUS, NLD, NGR, NOR, NZL, NRU, PNG, NRU, PER, PNG, PHL, PNG, PSE, PSE, QAT, KOS, TWN, UGU, UZB.*