Forecasting GDP growth from the outer space*

Jaqueson K. Galimberti†
ETH Zurich, Switzerland

February 17, 2017

Abstract

We evaluate the usefulness of satellite-based data on nighttime lights for the prediction of annual GDP growth across a global sample of countries. Going beyond traditional measures of luminosity, such as the sum of lights within a country's borders, we propose several innovative distribution- and location-based indicators attempting to extract new predictive information from the night lights data. Whereas our findings are generally favorable to the use of the night lights data to improve the accuracy of simple autoregressive model-based forecasts, we also find a substantial degree of heterogeneity across countries on the estimated relationships between light emissions and economic activity: individually estimated models tend to outperform pooled specifications, even though the latter provide more efficient estimates for out-of-sample forecasting. The estimation uncertainty affecting the country-specific estimates tends to be more pronounced for low and lower middle income countries. We conduct bootstrapped inference in order to evaluate the statistical significance of our results.

Keywords: night lights, remote sensing, business cycles, leading indicators, panel models.
JEL codes: E37, E01, C82, R12.

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

*A preliminary version of this paper was presented at the 36th International Symposium on Forecasting, held in Santander, 2016, and I thank to the participants for useful comments and discussions. Financial support from the SAS-IIF grant is also gratefully acknowledged.

†Contact e-mail: galimberti@kof.ethz.ch. Personal website: https://sites.google.com/site/jkgeconoeng/. Correspondence address: KOF Swiss Economic Institute, ETH Zurich, LEE G 116, Leonhardstrasse 21, 8092 Zürich, Switzerland. Phone: +41 44 6328529. Fax: +41 44 6321218.

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 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.

For that purpose we develop several night lights-based indicators, broadly classified into three categories: (i) aggregate indicators, including the Sum of Lights (SoL) measure traditionally employed in the past literature, and a balance indicator tracking the number (instead of the intensity) of lights that increased/decreased; (ii) distribution-based indicators, such as the median (central tendency), the kurtosis (fat tails), the skewness (asymmetry), the spatial Gini (inequality), and entropies (diversity) of the measured light intensities; and, (iii) location-based indicators, designed to focus on the SoL emitted from selected areas instead of the entire country's territory, such as areas showing a positive/negative correlation to the country's past GDP growth, and areas located within clusters of low/high degrees of entropy. Whereas the underlying reasoning is the same across all these measures, i.e., exploring the predictive content of the relationship between economic activity and nighttime light emissions, they are aimed to uncover different aspects of such a relationship from the signals obtained from the satellite data. Further details about the night lights data sources and the construction of these indicators are provided in section §2.

We process these measures for a sample of 172 countries at an annual frequency over the period from 1992 to 2013. 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. Importantly, here we distinguish between two alternative specifications

with respect to the estimation of these models, namely, a panel and country-individual specifications. This is in contrast to the previous literature which has usually assumed the existence of a common relationship between the night lights indicators and GDP growth across countries. Our estimation results, presented together with details about the model specifications in section §3, put this assumption into question as we observe wide ranges of parameter estimates under the individual specifications.

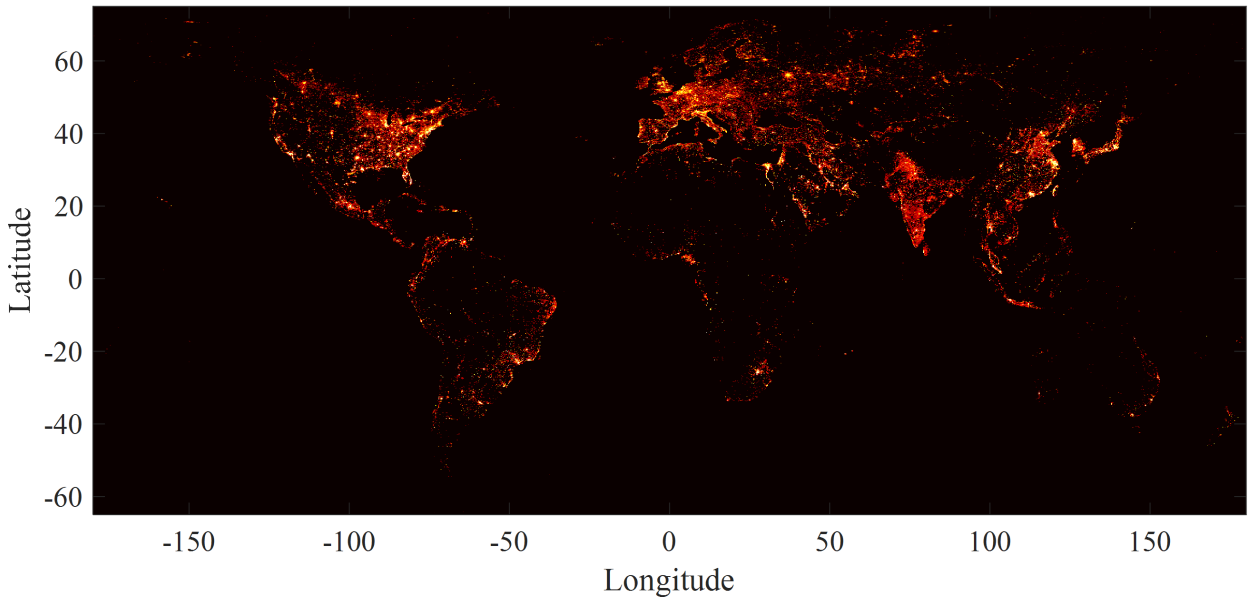
We then proceed with the comparative evaluation of the accuracy of the night lights-based forecasts relative to the benchmark model forecasts. We consider three exercises with respect to the sample used for the estimation of the models' parameters. First, an in-sample evaluation, where the full sample 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. Third, in order to capture the effects of parameter estimation in the recursive exercise, we re-run the out-of-sample evaluation using parameter estimates obtained with the full sample of data. Our main measure of evaluation is the forecasts' root mean square error (RMSE), and the results of these evaluations are presented in section §4.

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 accuracy improvements ranging from 2.9% to 7.2% (cross-country weighted averages) relative to the benchmark model. 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. Interestingly, we find that richer countries tend to be less prone to the effects of such estimation uncertainty. Among the night lights indicators, we find that the distribution-based ones tend to be more strongly affected by the cross-country heterogeneity of model estimates, while those based on the location of lights provide the greatest improvements to the accuracy of the GDP growth forecasts.

Finally, we also evaluate the statistical significance of our results by conducting bootstrapped tests of equal predictive accuracy. Particularly, we adopt the Clark and West (2007) test for comparison of nested models, simulating the empirical distribution of the test statistics according to the bootstrap procedure proposed by Clark and McCracken (2012). The results point to relatively small rejection rates across the countries in our sample, suggesting that this particular sample may have been too restricted to provide informative comparisons between the evaluated models.

In section §5 we conclude with some remarks.

Figure 1: Snapshot of world stable night lights averaged over the year of 2013.



Notes: Based on OLS stable lights data obtained from NOAA-NGDC. The averaged night lights intensity measures are depicted in logarithmic scale after satellite intercalibration (see Appendix A.1) and normalization of grid cell areas to Earth's curvature across the latitudinal dimension.

2 Night lights data and indicators

2.1 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). An illustrative snapshot is presented in figure 1.

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.1.

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[1] observations, the availability of which can vary substantially across countries depending on weather conditions: from a minimum average (across the country's cells) of 3.27 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 2a indicate. Second, sensor saturation, caused by signals exceeding the sensor's detection range, interfere in the measurement of brighter sources of light. These signals are recorded with the highest DN value in the OLS scale (i.e., 63), or "top-coded", and tend to be more frequently observed in the more densely populated countries; see figure 2b. Third, as evidenced in figure 2c, 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 measures based on the the distribution 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.: $0.85km^2$ at equator, $0.37km^2$ at S65º, and $0.22km^2$ at N75º. That is important for the aggregation of night lights at the country level because the closer the detected lights (and their changes) are to equator, the higher 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[2].

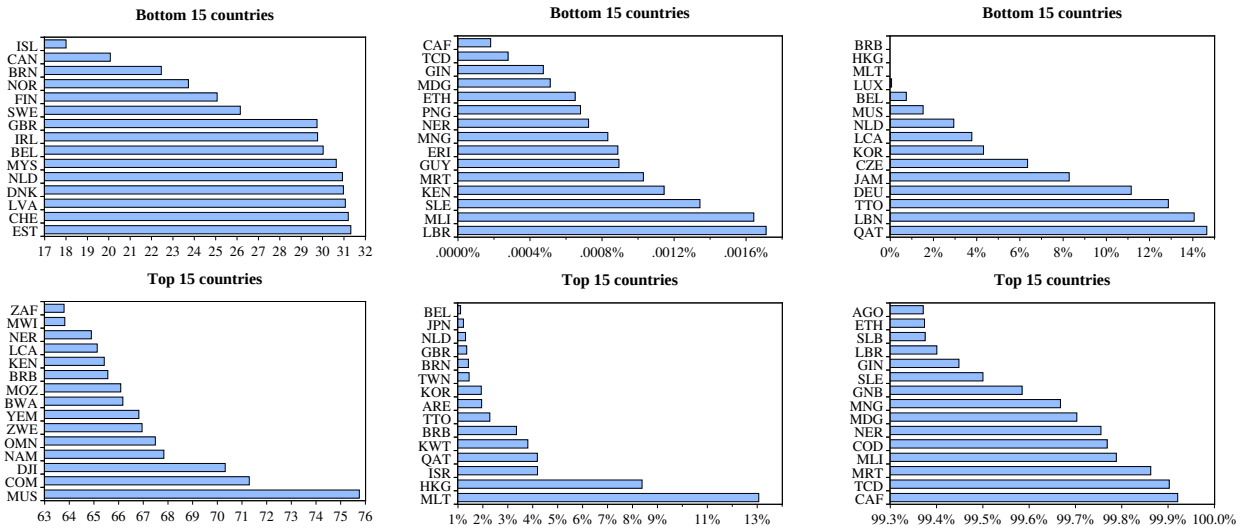
Before the computation of the night lights indicators (detailed below), the global composite images need to be processed for the extraction of light intensity measures within the countries borders. For that purpose we use the Database of Global Administrative Areas, GADM version 2.8 (http://gadm.org/), which contains definitions of 256 countries/territories borders across the globe. This sample reduces to 188 countries after matching the records to those of the International

[1] 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).

[2] 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).

Figure 2: 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.3.

Monetary Fund's (IMF) World Economic Outlook (WEO) database, which is our source of data on countries GDP. Our sample of countries is finally reduced to 172 countries after excluding those with a population smaller than 100,000, a large fraction of top-coded pixels (>40%), 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.3.

2.2 Night lights indicators

The geolocated time series data on night lights provides 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 three major classes: (i) aggregate indicators; (ii) distribution-based indicators; and, (iii) 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). Particularly, a country's **Sum of Lights** (SoL) 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) produc-

tion, 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 SoL indicator is that it does not account for the number of pixels that have actually increased/decreased within the country's borders. I.e., an increase in SoL may be due to a large DN increase in a single location, whereas lights remained constant or even decreased slightly in the rest of the country. To capture this feature we propose the use of a **Balance** indicator, calculated as

$$B_t = \frac{I_t}{I_t + D_t}, \tag{1}$$

where I_t (D_t) stands for the number of pixels within the country's borders showing a DN increase (decrease) in relation to the previous year. Notice $B_t \in [0, 1]$ and $B_t \gtreqless 1/2$ when $I_t \gtreqless D_t$.

Our second class of indicators goes beyond the aggregate to focus on the distribution of light intensity measures[3]. First, based on the country's distribution of DNs in a given year, we calculate some standard descriptive statistics: the **DNs median**, capturing the central tendency in the country's night lights profile; the **DNs kurtosis**, which focus on the tails of the distribution of lights, increasing with the frequency of extreme DN values; and the **DNs skewness**, as a measure of the asymmetry of the distribution of lights within the country, increasing (decreasing) as the mass of the distribution gets more concentrated on the left (right) tail. An illustration of these measures for the case of Spain, 2013, is presented in figure 3.

Another dispersion measure that received attention in the literature is the **Spatial Gini** coefficient (Henderson et al., 2012), which is calculated on the basis of the cumulative shares of the pixels DNs relative to the country's SoL. As with the income Gini, the Spatial Gini decreases (increases) towards zero (one) as the distribution of lights become more (less) equally distributed over the country's territory.

We experiment with two further indicators based on the application of the concept of entropy to image processing (see Gonzalez et al., 2004, p. 287). The entropy is a statistical measure of randomness, which in the present context can be used to characterize the diversity of night lights intensity measures over a country's territory, and is calculated as

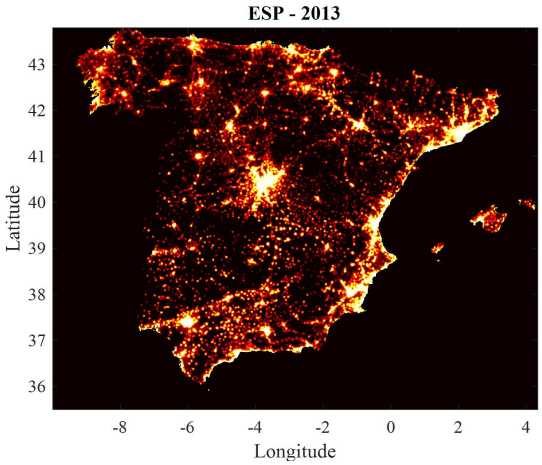
$$E = -\sum_{j=1}^{n} p_j \log_n p_j, \tag{2}$$

where p_j stands for the probability of observing a pixel with DN value in the range of bin j, obtained from the DNs histogram with n ($= 64$) bins. The use of the logarithm in the base n ensures that $E \in [0, 1]$, and the **DNs entropy** is seem to increase with the diversity of observed

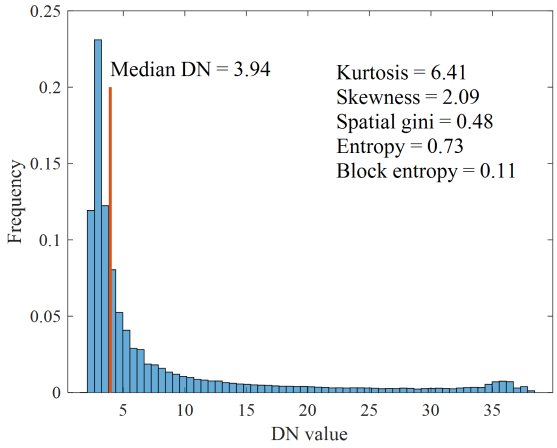
[3]All the distribution-based indicators are calculated after excluding unlit pixels so as to make these measures less sensitive to country-specific geographical conditions (see figure 2c). The exclusion of top-coded pixels was found to have no major effect on the indicators, hence we kept them in.

Figure 3: Night lights distribution-based indicators for Spain-2013.

(a) Snapshot of night lights.



(b) Histogram and distribution-based indicators.



Notes: The snapshot image is produced according to figure 1 notes, though "turning off" the lights outside the country's borders. The histogram excludes unlit pixels. See the text for details on the calculation of the indicators.

night lights[4]. As an alternative, we also calculate the **DNs block entropy**, which is based on the frequencies of 3×3 blocks of pixels[5], hence capturing the diversity of clustering patterns in the distribution of lights.

Finally, in our third class of indicators we develop two location-based measures, both attempting to improve the aggregate SoL indicator by focusing on specific pixels instead of the entire country's territory. Here the idea is to decompose a country's SoL by dividing its pixels according to a given criterion. Our first of such criteria is again based on the concept of block entropy, here adopted to distinguish between pixels with respect to their surrounding diversity of light intensities. For that purpose, local entropies are calculated based on a 9×9 neighborhood around each pixel, and then used to divide the country's pixels in two groups of low and high entropy pixels according to their median. Based on this classification we then obtain a first decomposition of the country's total SoL into the **High/Low entropy SoL** indicators.

A second location-based indicator is similarly constructed by classifying 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 again distinguish between two sets of pixels leading

[4]Mathematically we have that as $p_j \to 1$ (for any j), $E \to 0$, while as $p_j \to 1/n$ (for all j), $E \to 1$.

[5]The number of possible "block states" (n) is given by the number of pixel states to the power of 9 (pixels per block), which would require the infeasible computation of a histogram with 64^9 bins under the original 64-states scale. To make this feasible the DN values are first re-classified into a ternary system (low, medium, high intensity), so that $n = 3^9$.

to two new indicators: **Positive/Negatively correlated pixels SoL**; these can be further specialized to account only for pixels with statistically significant correlations. Importantly, 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 that year and before; other than enforcing the information restrictions required for a proper forecasting evaluation, this recursive approach also propitiates the tracking of regional changes in the lights predictive signals for economic activity. An illustration of these location-based indicators is presented in figure 4 for the case of France.

3 Forecasting approach

3.1 Models

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

$$y_{i,t} = \alpha_i + \beta y_{i,t-1} + \delta_t + \varepsilon_{i,t}, \tag{3}$$

for the pooled specification, and

$$y_{i,t} = \alpha_i' + \beta_i y_{i,t-1} + \epsilon_{i,t}, \tag{4}$$

for the individual specification, where $y_{i,t}$ stands for country i's $(= 1, \ldots, 172)$ GDP growth rate for year t $(= 1992, \ldots, 2013)$, $\alpha_i^{(\prime)}$ and δ_t for country and time fixed effects, respectively.

The night lights-based forecasts are obtained by augmenting the benchmark models with the night lights indicators discussed in the previous section[6]. Letting $\mathbf{x}_{k,i,t}$ stand for a vector k of indicators, the augmented models are given by

$$y_{i,t} = \alpha_{k,i} + \beta_k y_{i,t-1} + \boldsymbol{\theta}_k \mathbf{x}_{k,i,t-1} + \delta_{k,t} + \varepsilon_{k,i,t}, \tag{5}$$

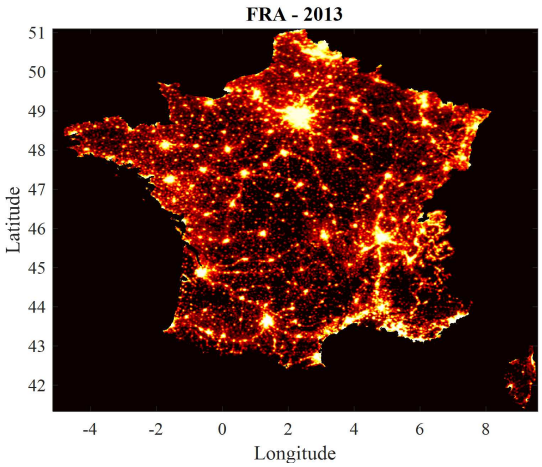
$$y_{i,t} = \alpha_{k,i}' + \beta_{k,i} y_{i,t-1} + \boldsymbol{\theta}_{k,i} \mathbf{x}_{k,i,t-1} + \epsilon_{k,i,t}, \tag{6}$$

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

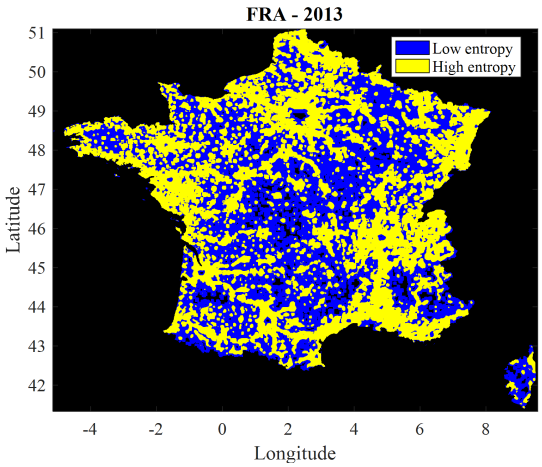
[6]Models including only the night lights indicators, i.e., without the AR(1) term, yield poor forecasting performance relative to the benchmark, which is not surprising considering the relevance of persistence in the GDP series.

Figure 4: Night lights location-based indicators for France.

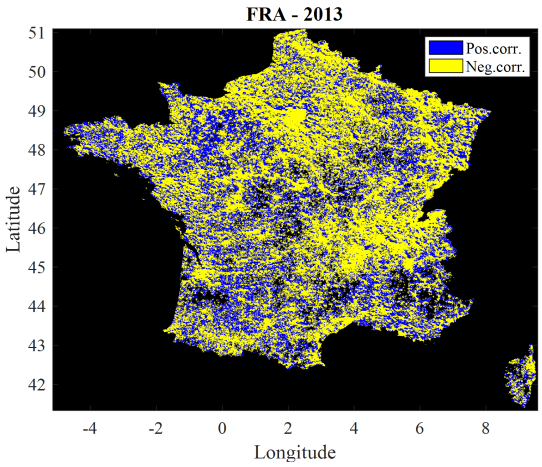
(a) Snapshot of night lights.



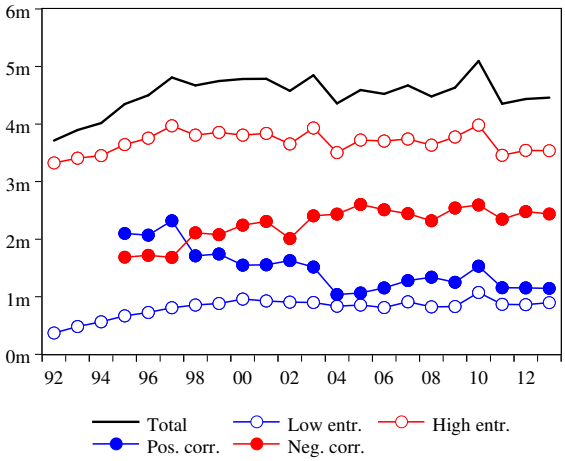
(b) Local entropies.



(c) Correlated pixels.



(d) Sum of Lights time series.



Notes: The snapshot image is produced according to figure 3 notes. See the text for details on the calculation of the indicators.

decomposition of a country's pixels in a given year: high and low entropy pixels; positive and negatively correlated pixels. Figures with the time/cross-country dispersion of all these variables are presented in Appendix A.2.

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 panel model specifications with all available observations and derive their corresponding residuals. Outliers are then detected based on the statistical significance (p-value smaller than 0.1%) of each disturbance. In the second stage we obtain the final estimates of the models, both panel and individual specifications, excluding the detected outliers from the sample.

The estimates of the panel models, (3) and (5), are reported in table 1. Overall, the estimates of the autoregressive parameter are robust to the inclusion of the night lights indicators, remaining statistically significant and close to the benchmark estimates on most cases[7]. In terms of statistical significance, the evidence is not strongly favorable to the night lights indicators: except for some distribution-based measures, there is no statistically significant relationship detected between (lagged) night lights and GDP growth after accounting for the persistence of the latter. However, these results do not preclude the existence of such relationship contemporaneously, where in fact we find results similar to the past literature[8].

The assumption of a common relationship between the night lights indicators and GDP growth is put into question when we look at the estimates for the country-individual model specifications. Namely, we observe a wide range of individual parameter estimates: figure 5 presents the distributions of the deviations of these individual estimates from those obtained under the panel specifications. Although these estimates tend to be centered around the panel ones, it is clear that they are overly dispersed to justify the pooling. Hence, it seems important to also consider this alternative on the evaluation of the predictive performance of the night lights-based forecasts.

3.2 Evaluation exercises

In order to evaluate the quality of the night lights indicators as predictors of annual GDP growth we conduct three exercises, differing mainly with respect to the sample used for the estimation of the

[7] An exception is the case of the correlated pixels indicators, under which the AR(1) estimate is seem to decrease to almost half of the benchmark estimate; further inspection reveals that this effect is due to the reduced sample available for this indicator.

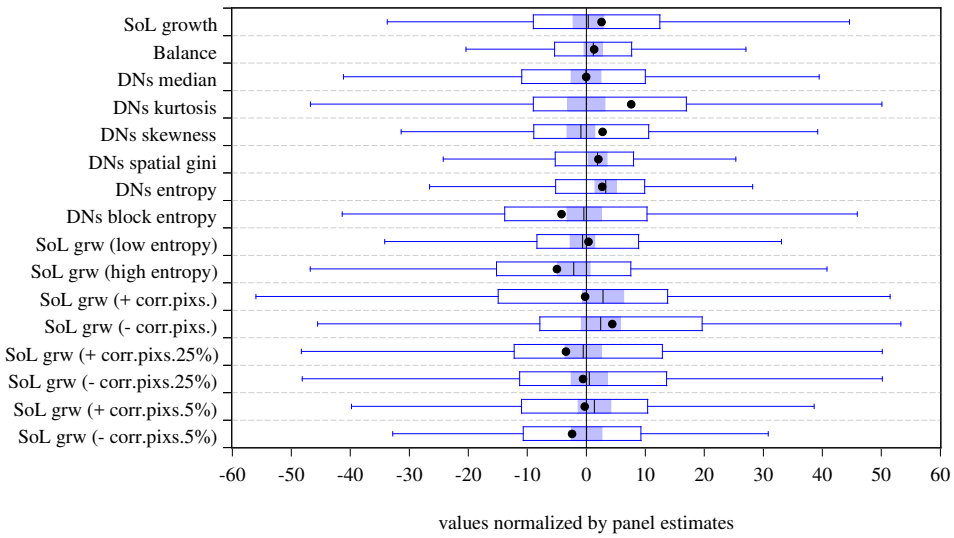
[8] The panel estimates of the contemporaneous relationships between the night lights indicators and GDP growth, both with and without the autoregressive term, are reported in Appendix A.2.

Table 1: Panel estimates of forecasting models.

Model	Estimates AR(1)	Night lights ($t-1$)		N.Obs. [outliers]	Adj. R^2
AR(1) benchmark	0.288***	-		3598	0.335
	(0.015)	-		[36]	
Aggregate night lights indicators:					
+ SoL growth	0.273***	0.581		3321	0.340
	(0.016)	(0.386)		[34]	
+ Balance	0.274***	0.288		3321	0.339
	(0.016)	(0.402)		[34]	
Distribution-based night lights indicators:					
+ DNs median	0.278***	-0.194**		3457	0.343
	(0.015)	(0.090)		[38]	
+ DNs kurtosis	0.276***	0.043**		3456	0.342
	(0.015)	(0.018)		[39]	
+ DNs skewness	0.275***	0.446***		3456	0.342
	(0.015)	(0.166)		[39]	
+ DNs spatial gini	0.278***	-0.072***		3458	0.341
	(0.015)	(0.027)		[37]	
+ DNs entropy	0.275***	-0.046***		3456	0.343
	(0.015)	(0.016)		[39]	
+ DNs block entropy	0.279***	-0.012		3457	0.341
	(0.015)	(0.274)		[38]	
Location-based night lights indicators:					
+ SoL grw. (low/high entropy)	0.259***	0.295	-0.237	3150	0.354
	(0.016)	(0.192)	(0.472)	[33]	
+ SoL grw. (+/- corr. pixs.)	0.157***	0.046	-0.302	2787	0.358
	(0.018)	(0.262)	(0.247)	[24]	
+ SoL grw. (+/- corr. pixs.25%)	0.165***	-0.065	-0.116	2621	0.367
	(0.019)	(0.106)	(0.100)	[21]	
+ SoL grw. (+/- corr. pixs.5%)	0.164***	0.028	-0.099	2179	0.383
	(0.020)	(0.080)	(0.078)	[14]	

Notes: The estimates are obtained by regressing real GDP growth on the corresponding night lights indicator lagged values, plus a constant, an autoregressive term, and country and year fixed effects. Models are estimated by least squares in two stages: first, with all available observations; second, excluding outliers, as defined in the text. The values inside parentheses are country-clustered robust standard errors. ***, **, and * stand for 1%, 5%, and 10% levels of statistical significance, respectively. The last two correlated pixels indicators are restricted to pixels with correlations statistically significant at 25% and 5% levels of significance.

Figure 5: Distributions of individual estimates.



Notes: Boxplots are based on the normalized individual countries estimates of the coefficient associated to the corresponding night lights indicator (lagged) in the model specification augmented by an AR(1) term, i.e., parameter(s) $\boldsymbol{\theta}_{k,i}$ in (6). The individual coefficient estimates ($\hat{\theta}_{k,i}$) are normalized by the panel estimates ($\hat{\theta}_k$) according to $\left(\hat{\theta}_{k,i} - \hat{\theta}_k\right)/\hat{\sigma}$, where $\hat{\sigma}$ is the estimated standard error of the panel estimate. The filled circles represent the mean estimates. Shaded areas represent approximate 95% confidence intervals for the medians, which are pictured as vertical lines inside the interquartile boxes. Outliers, defined as estimates outside the inner fences (1st/3rd quartile -/+ 1.5 multiplied by the interquartile range), are not presented.

model parameters. 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[9]

$$\hat{y}^{\mathcal{S}}_{k,i,t+1} = \hat{\alpha}^{\mathcal{S}}_{k,i} + \hat{\beta}^{\mathcal{S}}_{k(,i)} y_{i,t} + \hat{\boldsymbol{\theta}}^{\mathcal{S}}_{k(,i)} \mathbf{x}_{k,i,t}, \tag{7}$$

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

First, we evaluate the models' **in-sample** predictive performance. To that purpose we construct GDP growth forecasts for every year in the period from 1993 to 2014, i.e., with $t = \{1992, \ldots, 2013\}$ in (7), using model parameters estimated with our full sample of data, i.e., with $\mathcal{S} = \{1992, \ldots, 2013\}$ in (7). Naturally, this is not a realistic 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 (7), while $t = \{2000, \ldots, 2013\}$, $\mathcal{S}_t = \{i\}^t_{i=1992}$. Finally, in order to assess the effects of parameter estimation in the comparison between the in-sample and the OoS assumptions, we propose a third exercise where forecasts for the same OoS period are constructed on the basis of the full-sample model estimates; i.e., the **full-sample OoS** forecasts are obtained for $t = \{2000, \ldots, 2013\}$ with $\mathcal{S} = \{1992, \ldots, 2013\}$.

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 172 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.

4 Forecast evaluation

4.1 Averaged statistics

The averaged results for the in-sample evaluation are presented in table 2, where we observe that the usefulness of the night lights data depends on whether the forecasting model was estimated

[9]Notice the time fixed effects from the panel model specifications, (3) and (5), are not used for the computation of conditional forecasts but only included during model estimation in order to improve identification.

Table 2: In-sample evaluation statistics.

Model	Panel		Individual		Sample	
	RMSE	Ratio	RMSE	Ratio	Countries	Forecs.
AR(1) benchmark	2.508	-	2.376	-	172	3666
Aggregate night lights indicators:						
+ SoL growth	2.502	0.995	2.310	0.971	172	3526
+ Balance	2.500	0.995	2.310	0.971	172	3526
Distribution-based night lights indicators:						
+ DNs median	2.516	1.006	2.305	0.963	172	3666
+ DNs kurtosis	2.502	0.999	2.292	0.958	172	3666
+ DNs skewness	2.507	1.002	2.292	0.958	172	3666
+ DNs spatial gini	2.505	1.000	2.306	0.966	172	3666
+ DNs entropy	2.528	1.016	2.306	0.967	172	3666
+ DNs block entropy	2.505	0.999	2.267	0.950	172	3666
Location-based night lights indicators:						
+ SoL grw. (low/high entropy)	2.500	0.994	2.234	0.928	172	3508
+ SoL grw. (+/- corr. pixs.)	2.513	1.024	2.237	0.946	170	2975
+ SoL grw. (+/- corr. pixs.25%)	2.503	1.018	2.233	0.948	163	2781
+ SoL grw. (+/- corr. pixs.5%)	2.491	1.017	2.234	0.950	147	2284

Notes: Statistics are based on forecasts over the period from 1993 to 2014. The reported RMSEs are weighted cross-country averages of the individual RMSEs using the countries GDPs (in PPP terms) as weights. The ratios are also first computed individually, relative to the corresponding benchmark specification, and then geometrically averaged using countries GDPs as weights. All model specifications contain an AR(1) term. Panel specifications are estimated with country and year fixed effects, where the latter are not used to compute forecasts.

pooling all the countries together or individually. Here the evidence is strongly in favor of the individual estimation, on average yielding in-sample predictions about 9% more accurate than the panel estimated models. Across the night lights indicators, the location-based ones stand out, with accuracy improvements reaching up to 7.2% (low/high entropy indicator) in relation to the benchmark model.

One criticism to the in-sample evaluation of forecasts is that it may not provide reliable assessments of how the forecasts will perform under a more realistic context, i.e., before the actual value for the forecasted variable becomes available. Because the model estimates will carry information about the forecasting targets, the use of the same sample for model estimation and forecast evaluation can distort the quality assessment in favor of the model-based forecasts. In order to validate the robustness of our results to this issue, we now focus on the results obtained under our out-of-sample evaluation exercises, presented in table 3.

Table 3: Out-of-sample evaluation statistics.

Model	Recursive estimates				Full-sample estimates					
	Panel		Individual		Panel		Individual		Sample	
	RMSE	Ratio	RMSE	Ratio	RMSE	Ratio	RMSE	Ratio	Countries	Forecs.
AR(1) benchmark	2.632	-	2.715	-	2.426	-	2.329	-	172	2371
Aggregate night lights indicators:										
+ SoL growth	2.532	0.994	2.823	1.072	2.318	0.985	2.168	0.957	172	2360
+ Balance	2.530	0.992	2.832	1.075	2.317	0.985	2.163	0.957	172	2360
Distribution-based night lights indicators:										
+ DNs median	2.539	0.999	2.857	1.082	2.323	0.988	2.180	0.961	172	2360
+ DNs kurtosis	2.505	0.980	2.859	1.072	2.313	0.985	2.174	0.960	172	2360
+ DNs skewness	2.504	0.980	2.845	1.070	2.316	0.986	2.178	0.961	172	2360
+ DNs spatial gini	2.512	0.983	2.909	1.093	2.334	0.998	2.176	0.960	172	2360
+ DNs entropy	2.548	1.004	2.863	1.078	2.344	1.001	2.191	0.967	172	2360
+ DNs block entropy	2.535	0.998	2.887	1.079	2.317	0.986	2.159	0.947	172	2360
Location-based night lights indicators:										
+ SoL grw. (low/high entropy)	2.528	0.993	2.951	1.113	2.311	0.982	2.083	0.911	172	2336
+ SoL grw. (+/- corr. pixs.)	2.614	1.039	4.522	1.510	2.334	0.996	2.081	0.903	170	2113
+ SoL grw. (+/- corr. pixs.25%)	2.590	1.028	3.801	1.395	2.318	0.989	2.082	0.915	162	1957
+ SoL grw. (+/- corr. pixs.5%)	2.559	1.014	3.809	1.386	2.308	0.986	2.102	0.928	147	1541

Notes: Statistics are based on forecasts over the period from 2001 to 2014. Recursive forecasts are computed with models estimated using only past observations available at the forecast base period; full-sample forecasts, in contrast, use model estimates based on the whole sample of available observations. Further details on the construction of the evaluation statistics are provided in the notes of table 2.

One first observation from the results in table 3 is that the recursive estimation of the individual models lead to a substantial deterioration of these models' 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 9 data points, from 1992 to 2000, for each individual country; compare that to the (about) 1,500 observations used under the panel estimation and it is not surprising that the recursive OoS exercise strongly favored the latter. Estimation uncertainty seems to be behind the performance deterioration of the location-based indicators too, considering that these models require estimation of one additional parameter in relation to the others.

In order to shed further light on this issue, we also compare the forecasts performance under a full-sample OoS exercise, i.e., with forecasts computed from model estimates based on the full sample of data, but keeping the focus on the OoS evaluation sample from 2001 to 2014. As expected, these results are in line with our in-sample conclusions. Namely, the individually estimated models are generally favored against the panel estimation approach, and the location-based indicators provide the best performing forecasts.

The 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, our results suggest that the benefits of allowing for country-individual forecasting models outweighs the accuracy losses due to estimation biases.

4.2 Cross-country distribution

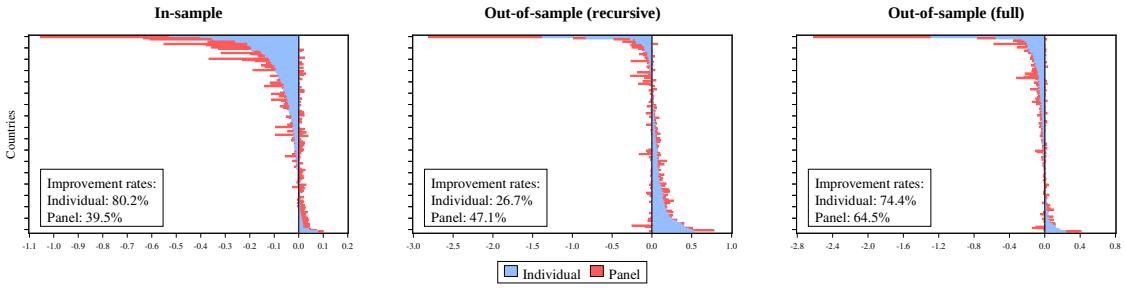
The averaged statistics can conceal the cross-country heterogeneity of performances too. Particularly, the averages can be affected by large outliers that push the evaluation measures towards a direction that does not reflect the majority of the results. To approach this issue we now focus on the distribution of results across countries. In figure 6 we present some selected figures on the (log) RMSE ratios for the models augmented with the SoL aggregate and location-based indicators[10].

Overall, the distribution of forecasting performances across countries reported in figure 6 confirm our findings with the averaged statistics. Namely, the individual models achieve higher improvement rates than the panel ones under the in-sample and the full-sample OoS exercises, whereas the opposite is observed under the recursive OoS evaluation. We also find that the SoL

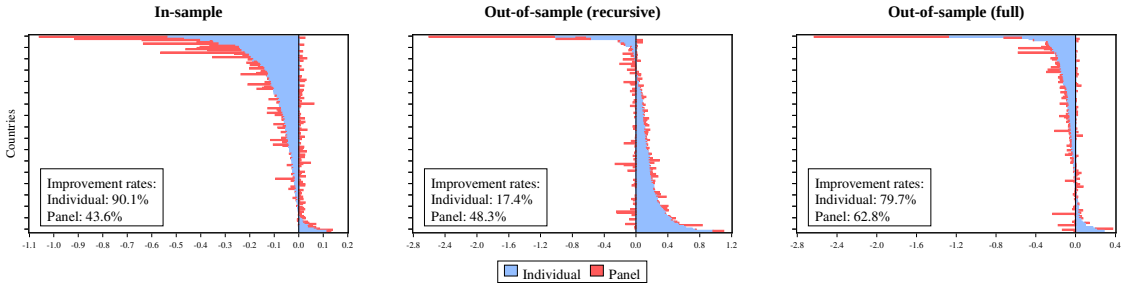
[10]Similar results for the remaining model specifications are reported in Appendix A.2.

Figure 6: Log RMSE ratios across countries.

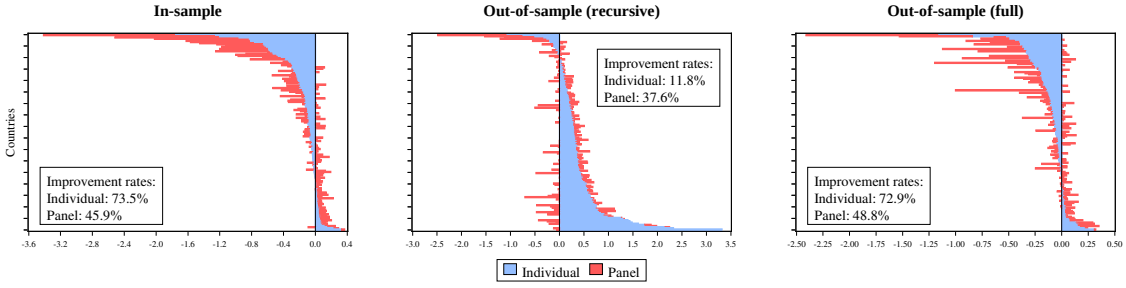
(a) Forecasts based on AR(1) plus SoL growth model.



(b) Forecasts based on AR(1) plus SoL growth in high/low entropy pixels model.



(c) Forecasts based on AR(1) plus SoL growth in pos./neg. correlated pixels model.



Notes: Each bar represents the logarithm of the ratio between the RMSE of the night lights based forecasts and that for the forecasts obtained with benchmark AR(1) model for each country. Hence, negative (positive) values indicate countries for which the use of night lights data was (dis)advantageous. The countries are presented in the vertical axis and sorted within each panel in ascending order according to their corresponding log RMSE ratios under the individually estimated specifications (in blue). The evaluations obtained under the panel specifications are presented (in red) stacked on top of the individual ones.

indicator distinguishing between pixels with low/high entropy achieves the best performance, improving the forecasts of up to 90% of the countries in the sample. Another inference from these results is that the country-specific individual and panel performances are, generally, positively correlated, e.g.: the average cross-country correlation between individual and panel log RMSE ratios for the aggregate and location-based indicators amounts to approximately 0.78, 0.46, and 0.70, for the in-sample, the recursive OoS, and the full-sample OoS evaluations, respectively; interestingly, the in-sample average correlation decreases to 0.19 for the distribution-based indicators, suggesting a greater impact of heterogeneity for this type of indicators.

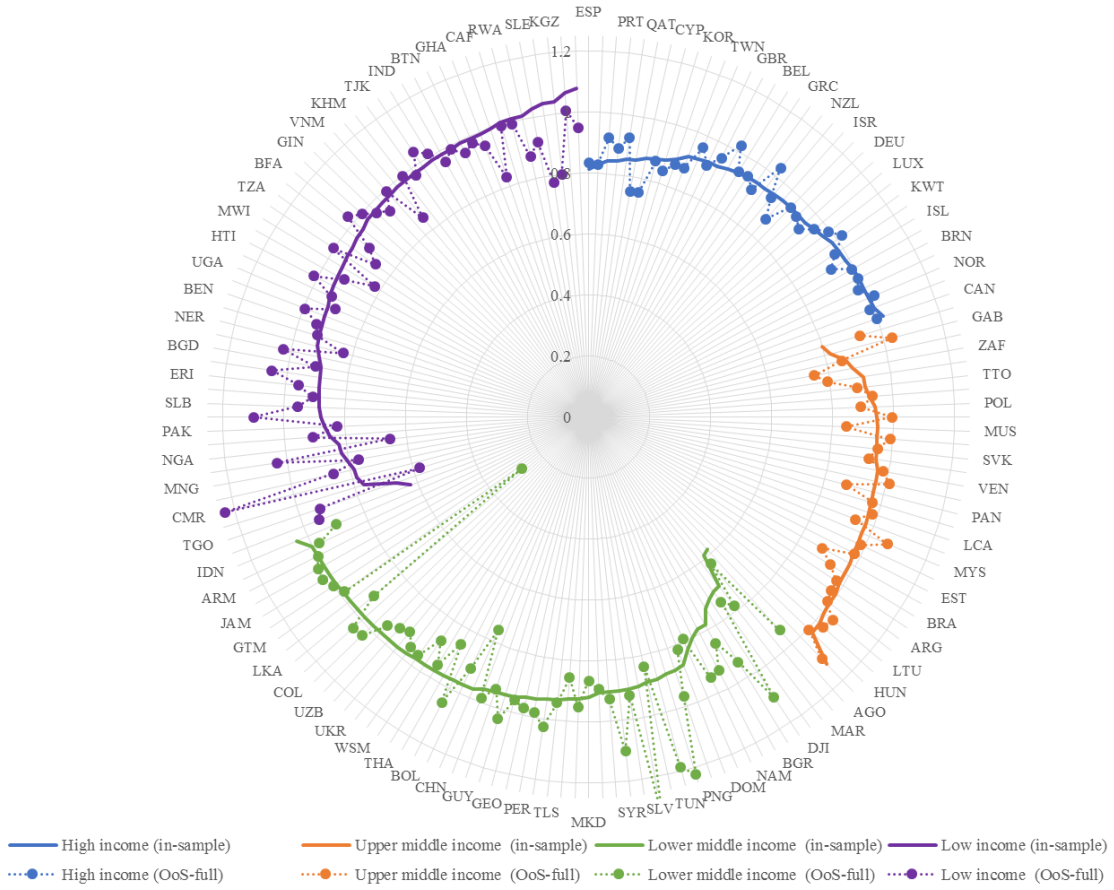
Finally, it is interesting to look at how the usefulness of the night lights data may depend on the country's level of income. In figure 7 we present the countries RMSE ratios, for the forecasts based on the individual model with the SoL low/high entropy indicator, grouped according to the World Bank income classification. Whereas the improvement rates reported above are found to be fairly equally distributed across the income groups, the correlation between in-sample and OoS performances looks stronger for the higher income countries. Hence, our in-sample assessments may be considered to provide a more reliable guidance on the OoS potential of the night lights indicators for countries with higher income than for those classified under the lower middle and low income groups.

4.3 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 now conduct statistical tests comparing the predictive accuracy of the night lights-augmented forecasts to those obtained under the AR(1) benchmark model.

One traditional test for the hypothesis of equal predictive accuracy is the Diebold and Mariano (1995, DM) test, which is based on the mean difference in sample average losses (here assumed to be the squared forecast error). However, one important disadvantage of the DM test is that it does not account for parameter estimation uncertainty, an issue that is of great relevance in our application. Besides, our comparative evaluation involves so-called nested models, i.e., the night lights-augmented specifications, eqs. (5) and (6), would converge to the benchmark specifications, eqs. (3) and (4), 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). To account for these features we follow the approach suggested by Clark and West (2007, CW)

Figure 7: Countries RMSE ratios grouped by income level.



Notes: The RMSE ratios refer to forecasts obtained with the model augmented by the SoL growth in high/low entropy pixels, estimated individually for each country. Values below (above) one indicate countries for which the use of the night lights indicator was (dis)advantageous. The countries are grouped according to their World Bank income classification and sorted within each group in ascending order according to their corresponding in-sample RMSE ratios.

for comparison of nested models, also adapting the inference to finite sample biases by simulating the empirical distribution of the test statistics according to the bootstrap procedure proposed by Clark and McCracken (2012)[11,12].

The statistical tests for predictive improvement are again conducted separately for each country, model, pooling assumption, and evaluation sample. In order to synthesize these results we adopt two approaches. First, we look at the hit-rates, defined as the cross-country frequency of rejections of the null hypothesis of equal predictive accuracy at the 10% level of significance; hence, the higher the hit-rate of a model's forecasts, the stronger the statistical evidence is in favor to that model's improvement against the benchmark.

The hit-rates for the bootstrapped CW test are presented in table 4, also including the Max-t reality check proposed in Clark and McCracken (2012) as an extension for multiple-model comparisons. As expected, these hit-rates are much smaller than the improvement rates observed above, based solely on the sample RMSEs. Whereas this indicates caution should be taken in extrapolating our assessments beyond our sample, some of our previous inferences are corroborated. Namely, the underperformance of the panel specifications in forecasting in-sample is observed for most cases, particularly across the distribution-based indicators. Also, the OoS performance of the individual specifications is again found to be affected by parameter estimation uncertainties, as indicated by the mostly higher hit-rates under the full-sample OoS setup against the recursive one.

As a second approach to synthesize our cross-country results we combine the bootstrapped p-values associated to the CW tests according to the Fisher's method from meta-analysis (see Hedges and Olkin, 1985, Ch.3). In short, this method is used to combine the significance of several tests, assumed to be independent[13], for the same hypothesis. The resulting combined p-values are presented in table 5, and provide a clearer picture on the statistical significance of the improvements achieved with the use of the night lights data to forecast GDP growth. First, we notice that the hypothesis of equal predictive performance can be generally rejected for the individually estimated models under the in-sample and the full-sample OoS exercises. Second, whereas most of the distribution-based indicators show robustness to the estimation issue affecting the recursive esti-

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

[12] Results based on the theoretical distribution of the test statistics are reported in Appendix A.2, and tend to show higher rejection rates, on average, than those obtained under the bootstrapped tests.

[13] Although the independence assumption may be questionable for our cross-country application, for practical purposes we adopt the method in its simplest form. Besides, the random resampling adopted in the bootstrapping procedure (see footnote (11)) is expected to account for the cross-country dependence between the test statistics.

Table 4: Hit-rates on bootstrapped CW tests for predictive improvement.

Predictor	In-sample		OoS recursive		OoS full-sample	
	Panel	Indiv.	Panel	Indiv.	Panel	Indiv.
Aggregate night lights indicators:						
+ SoL growth	14.0%	15.1%	18.0%	13.4%	17.4%	21.5%
+ Balance	11.6%	15.7%	18.0%	12.8%	20.9%	19.2%
Distribution-based night lights indicators:						
+ DNs median	9.3%	15.7%	14.0%	14.0%	14.0%	13.4%
+ DNs kurtosis	10.5%	19.2%	13.4%	17.4%	16.3%	21.5%
+ DNs skewness	11.0%	22.7%	11.6%	17.4%	12.8%	20.3%
+ DNs spatial gini	15.1%	19.2%	19.2%	12.2%	19.2%	10.5%
+ DNs entropy	12.8%	19.2%	15.7%	12.8%	14.5%	14.5%
+ DNs block entropy	3.5%	16.3%	15.1%	15.1%	19.8%	18.0%
Location-based night lights indicators:						
+ SoL grw. (low/high entropy)	14.5%	15.1%	17.4%	9.3%	17.4%	16.3%
+ SoL grw. (+/- corr. pixs.)	25.3%	10.6%	15.3%	10.6%	24.7%	15.3%
Averages	12.8%	16.9%	15.8%	13.5%	17.7%	17.1%
Max-t reality check	11.6%	13.4%	14.5%	9.9%	14.5%	12.2%

Notes: Hit-rates 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 10% 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. Samples and exercises are equivalent to those reported in tables 2 and 3.

Table 5: Fisher's combined p-values of bootstrapped CW tests.

Predictor	In-sample		OoS recursive		OoS full-sample	
	Panel	Indiv.	Panel	Indiv.	Panel	Indiv.
Aggregate night lights indicators:						
+ SoL growth	3.1%	0.4%	0.0%	8.9%	0.0%	0.0%
+ Balance	26.5%	1.3%	0.0%	3.6%	0.0%	0.0%
Distribution-based night lights indicators:						
+ DNs median	41.9%	0.0%	0.0%	1.4%	1.3%	1.2%
+ DNs kurtosis	13.8%	0.0%	0.1%	0.0%	0.0%	0.0%
+ DNs skewness	7.4%	0.0%	0.4%	0.5%	0.1%	0.0%
+ DNs spatial gini	0.0%	0.0%	0.0%	8.0%	0.0%	1.1%
+ DNs entropy	17.5%	0.0%	1.3%	2.5%	1.0%	0.2%
+ DNs block entropy	83.6%	0.2%	6.1%	1.0%	0.0%	0.1%
Location-based night lights indicators:						
+ SoL grw. (low/high entropy)	2.5%	3.7%	0.0%	52.6%	0.0%	0.2%
+ SoL grw. (+/- corr. pixs.)	0.0%	1.8%	0.0%	84.3%	0.0%	0.9%
Averages	19.6%	0.7%	0.8%	16.3%	0.2%	0.4%
Max-t reality check	43.6%	1.8%	11.2%	46.1%	5.5%	1.6%

Notes: The reported p-values are drawn from a χ^2 distribution based on the Fisher's statistic, $-2\sum \ln p_i$, combining the bootstrapped p-values for the CW tests across the countries (same tests behind hit-rates reported in table 4).

mation of the individual specifications, these indicators show a poor in-sample performance with the panel specifications; this finding is consistent with our earlier observation that the relationship between the distribution-based night lights indicators and GDP growth is subject to a greater degree of heterogeneity across countries.

5 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 172 countries over the period from 1993 to 2014. For that purpose we have developed several night lights-based indicators, classified into three broad categories: (i) aggregate indicators; (ii) distribution-based indicators; and, (iii) location-based indicators. In order to evaluate the predictive content of such measures, we have constructed forecasts based on an AR(1) GDP growth model augmented by the lagged values of the night lights indicators. We have also considered two alternatives with respect to the estimation of the relationship between these night lights indicators and GDP growth, namely, a panel and country-individual 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 the individually estimated models tend to outperform the pooled specifications, even though the former are subject to stronger estimation biases due to the use of smaller samples. These biases have been particularly harmful under an out-of-sample forecast evaluation exercise. Across the countries, estimation uncertainty was found to be more relevant in out-of-sample forecasting for low and lower middle income countries, potentially due to a less stable relationship between night lights and economic activity in these countries.

Another major contribution of this paper was the development and evaluation of new measures attempting to explore the richness of information provided by the night lights dataset. Interestingly, we have found that the forecasts based on the distribution-based night lights indicators are more strongly affected by country-wise heterogeneity. Finally, our location-based indicators, designed to focus on the night lights emitted from selected areas instead of the entire country's territory, were found to provide the greatest improvements to the accuracy of the GDP growth forecasts. We hope these measures can uncover further useful relationships in future research.

References

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

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

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

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

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

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

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

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

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

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

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

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

Gonzalez, R. C., R. E. Woods, and S. L. Eddins (2004). *Digital Image processing using MATLAB*. Pearson/Prentice Hall.

Hedges, L. V. and I. Olkin (1985). *Statistical Methods for Meta-Analysis*. San Diego: Academic Press.

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

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

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

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

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

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

A Appendix

A.1 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

$$DN_{s,p} = \phi_0 + \phi_1 DN_{r,p} + \phi_2 DN_{r,p}^2, \tag{8}$$

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 8 and table 6, respectively.

A.2 Supplementary results

Results supplementary to those in the main text are presented here.

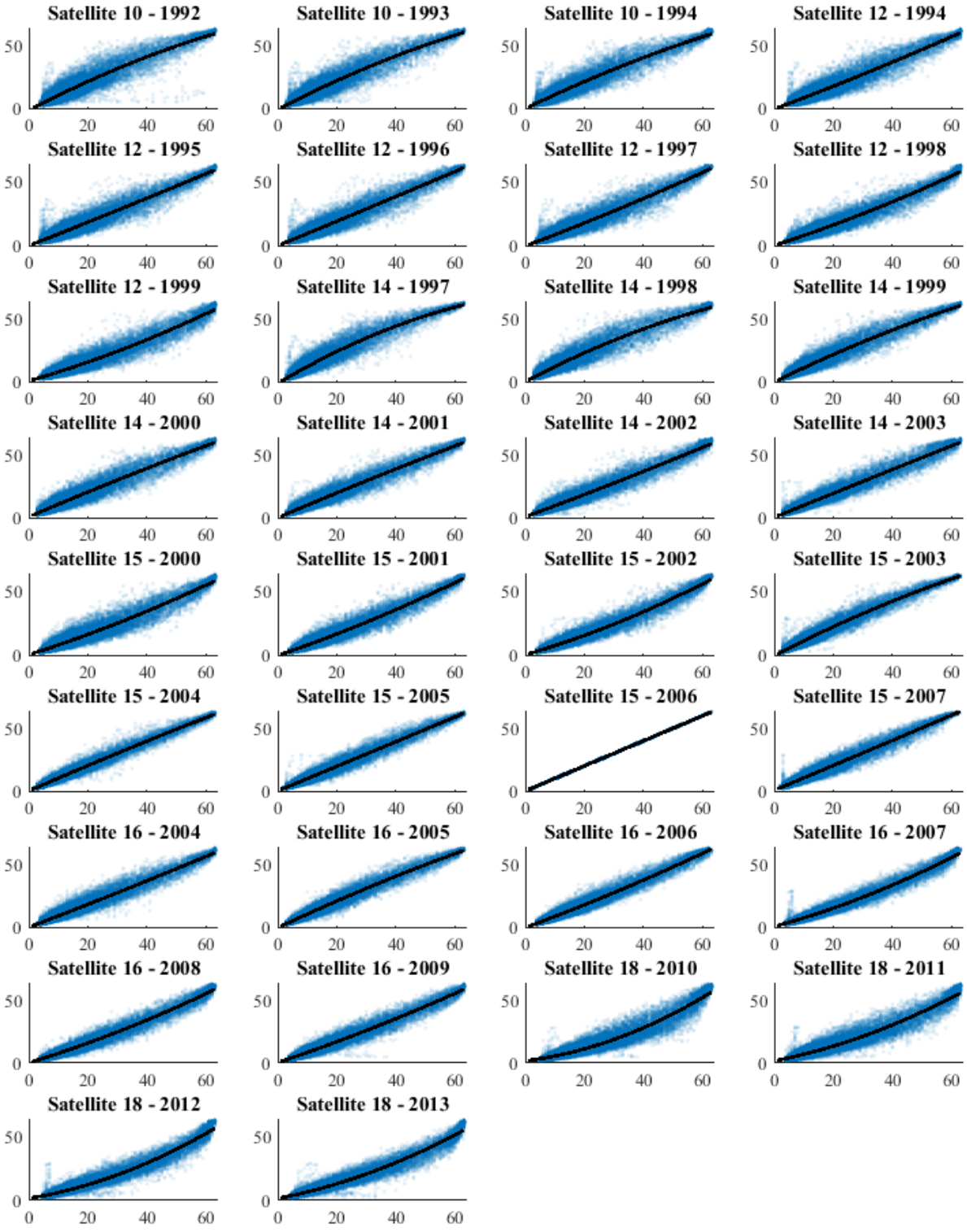
The time series and cross-section dispersion of the night lights variables used in this study are presented in Figures 9-12.

A summary of panel estimates on the relationship between the night lights indicators and GDP growth is presented in table 7, while the distributions of the corresponding individual estimates are depicted in figure 13.

The distributions of log RMSE ratios between the night lights-based forecasts and the benchmark forecasts are presented in Figures 14, 15, 16, for the in-sample, recursive OoS, and full-sample OoS exercises, respectively.

Hit-rates for the DM tests of equal accuracy are presented in table 8, while hit-rates for the CW test based on theoretical p-values (instead of bootstrapped as in the main text) are presented in table 9.

Figure 8: Scatter plots of Sicily's pixels DNs used for intercalibration of satellites.



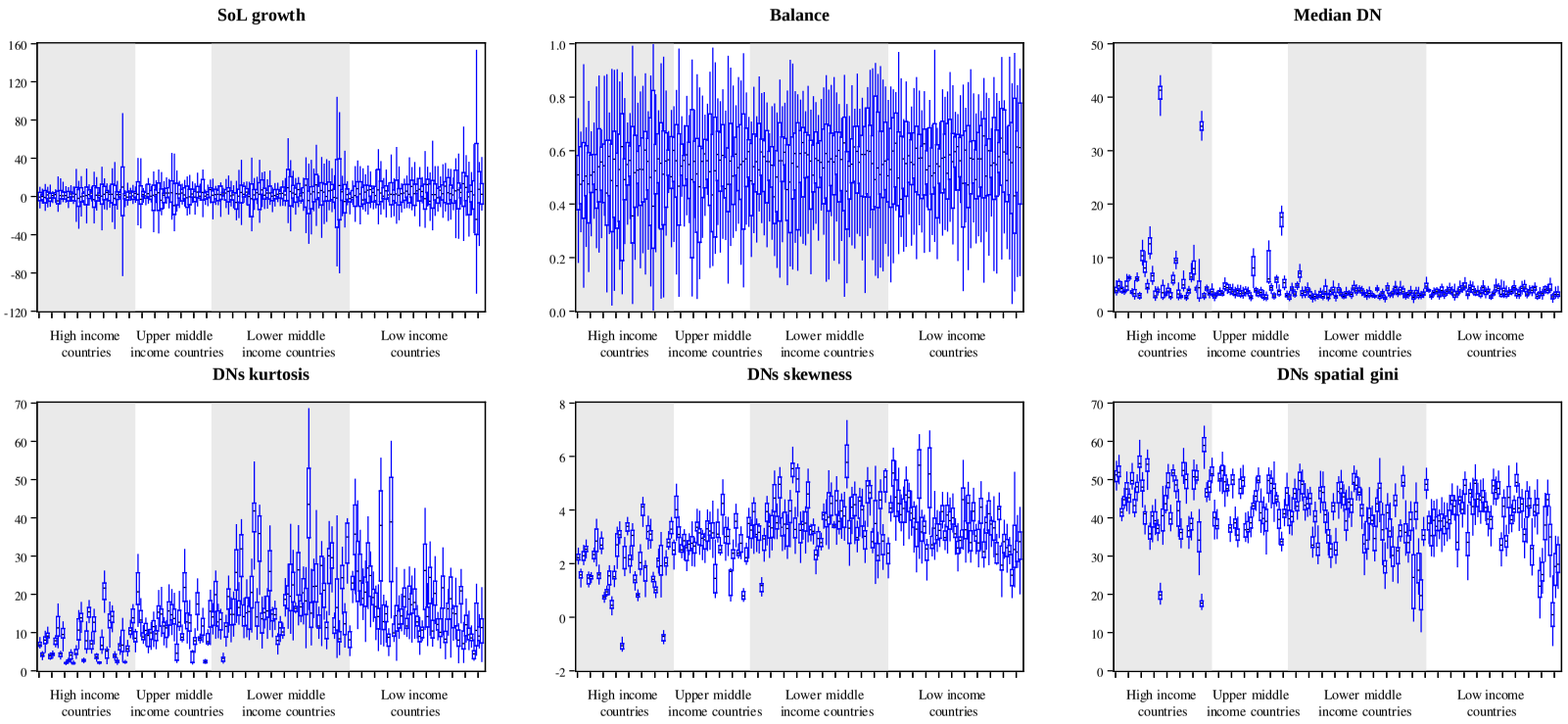
Notes: The DNs observed over Sicily from each satellite-year composite are plotted (y-axis) against their corresponding DN values for the reference satellite F152006 (x-axis). The black line depicts the fitted values according to equation (8).

Table 6: Intercalibration regression estimates.

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

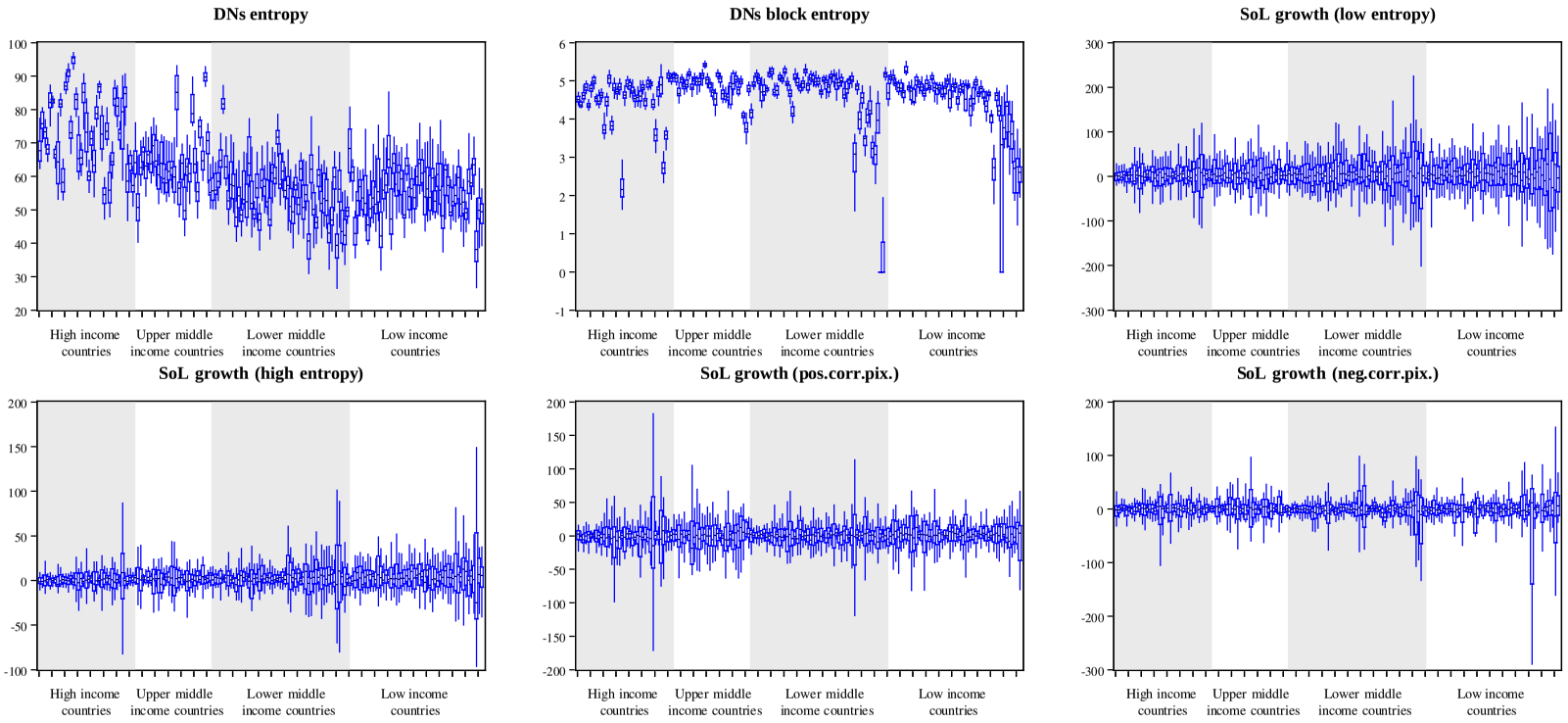
Notes: The estimates refer to equation (8) and are based on Sicily's night lights data.

Figure 9: Time dispersion of variables across countries (1 of 2).



Countries (left to right): USA, JPN, DEU, FRA, GBR, ITA, ESP, CAN, KOR, AUS, TWN, NLD, ARE, BEL, CHE, SWE, AUT, HKG, NOR, GRC, PRT, DNK, KWT, ISR, FIN, IRL, QAT, NZL, SVN, LUX, BRN, CYP, ISL, MLT, BHS, BRA, MEX, SAU, TUR, POL, ARG, ZAF, MYS, VEN, CZE, CHL, HUN, SVK, LBY, OMN, SRB, HRV, LTU, LBN, PAN, URY, CRI, LVA, EST, TTO, GAB, BWA, MUS, MNE, BRB, LCA, CHN, RUS, IDN, IRN, THA, EGY, IRQ, PHL, COL, DZA, UKR, KAZ, ROU, PER, MAR, UZB, BLR, LKA, ECU, BGR, AGO, DOM, TUN, GTM, SYR, TKM, JOR, BOL, PRY, SLV, BIH, GEO, HND, MKD, ALB, JAM, NAM, ARM, MDA, PNG, SWZ, TLS, SUR, FJI, GUY, DJI, CPV, BLZ, VCT, WSM, VUT, TON, FSM, IND, NGA, PAK, BGD, VNM, MMR, AZE, KEN, YEM, ETH, TZA, GHA, AFG, CIV, CMR, UGA, COD, NPL, ZMB, KHM, ZWE, MDG, SEN, MLI, NIC, LAO, MNG, MOZ, COG, BFA, TJK, TCD, KGZ, HTI, BEN, MWI, MRT, NER, RWA, GIN, ERI, TGO, SLE, BDI, LSO, LBR, BTN, CAF, GMB, GNB, COM, SLB, STP.

Figure 10: Time dispersion of variables across countries (2 of 2).



Countries (left to right): USA, JPN, DEU, FRA, GBR, ITA, ESP, CAN, KOR, AUS, TWN, NLD, ARE, BEL, CHE, SWE, AUT, HKG, NOR, GRC, PRT, DNK, KWT, ISR, FIN, IRL, QAT, NZL, SVN, LUX, BRN, CYP, ISL, MLT, BHS, BRA, MEX, SAU, TUR, POL, ARG, ZAF, MYS, VEN, CZE, CHL, HUN, SVK, LBY, OMN, SRB, HRV, LTU, LBN, PAN, URY, CRI, LVA, EST, TTO, GAB, BWA, MUS, MNE, BRB, LCA, CHN, RUS, IDN, IRN, THA, EGY, IRQ, PHL, COL, DZA, UKR, KAZ, ROU, PER, MAR, UZB, BLR, LKA, ECU, BGR, AGO, DOM, TUN, GTM, SYR, TKM, JOR, BOL, PRY, SLV, BIH, GEO, HND, MKD, ALB, JAM, NAM, ARM, MDA, PNG, SWZ, TLS, SUR, FJI, GUY, DJI, CPV, BLZ, VCT, WSM, VUT, TON, FSM, IND, NGA, PAK, BGD, VNM, MMR, AZE, KEN, YEM, ETH, TZA, GHA, AFG, CIV, CMR, UGA, COD, NPL, ZMB, KHM, ZWE, MDG, SEN, MLI, NIC, LAO, MNG, MOZ, COG, BFA, TJK, TCD, KGZ, HTI, BEN, MWI, MRT, NER, RWA, GIN, ERI, TGO, SLE, BDI, LSO, LBR, BTN, CAF, GMB, GNB, COM, SLB, STP.

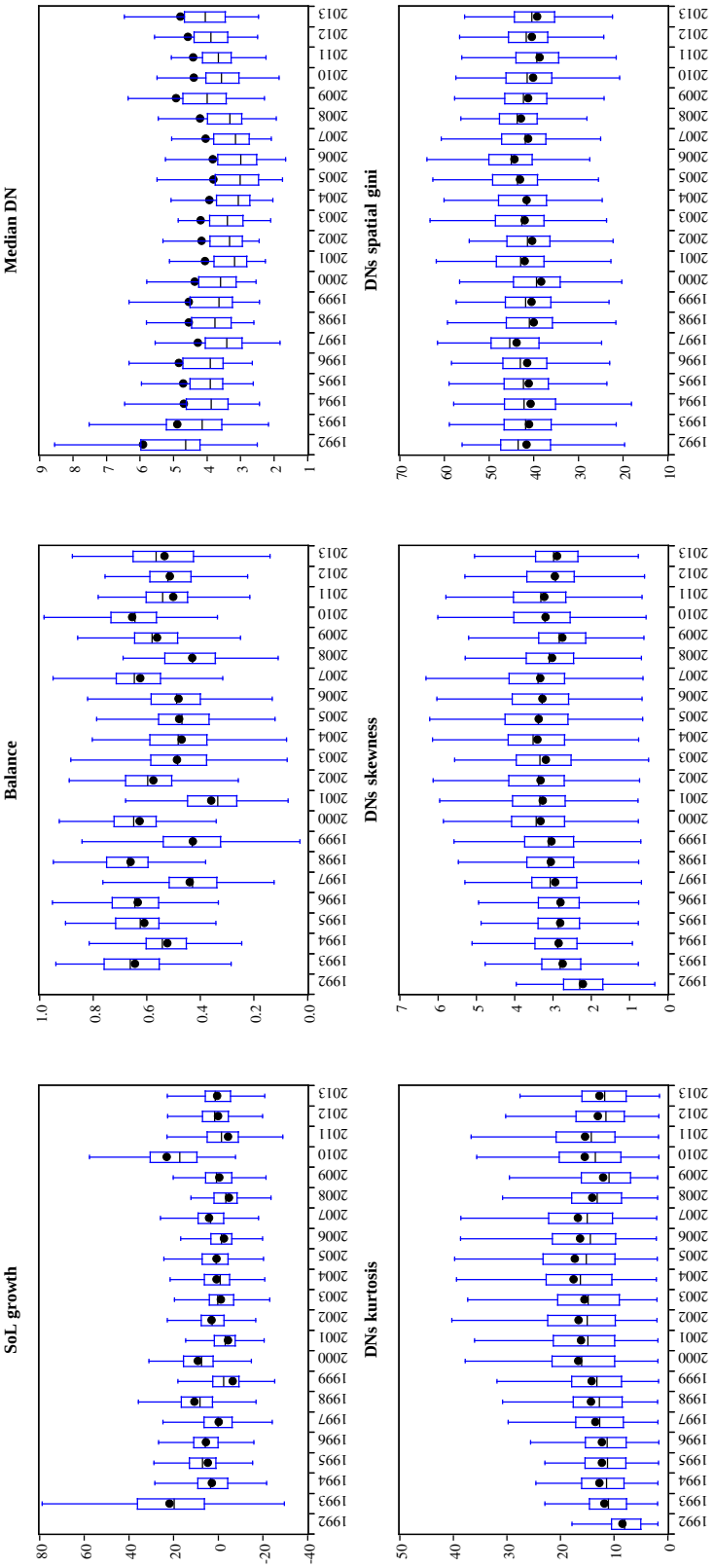


Figure 11: Cross-country dispersion of variables over time (1 of 2).

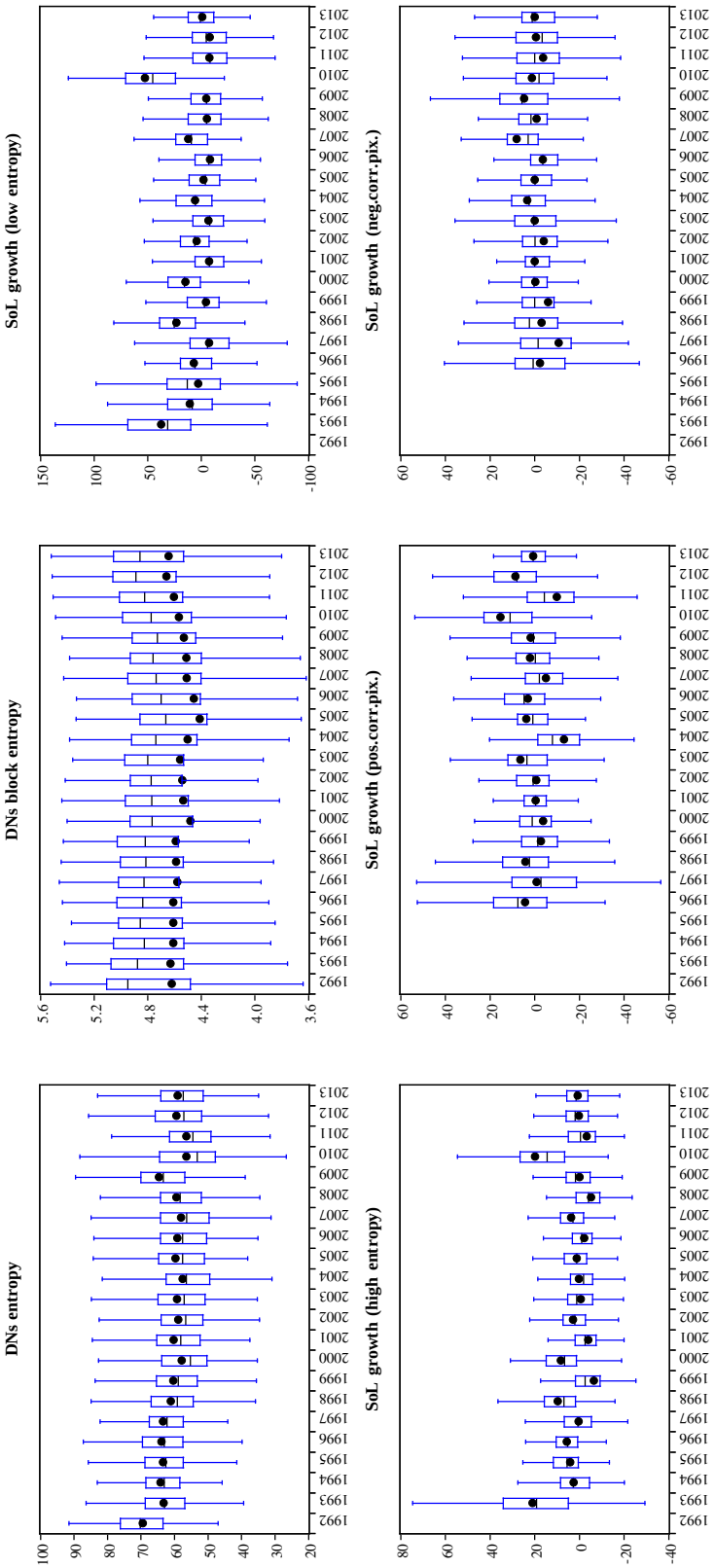


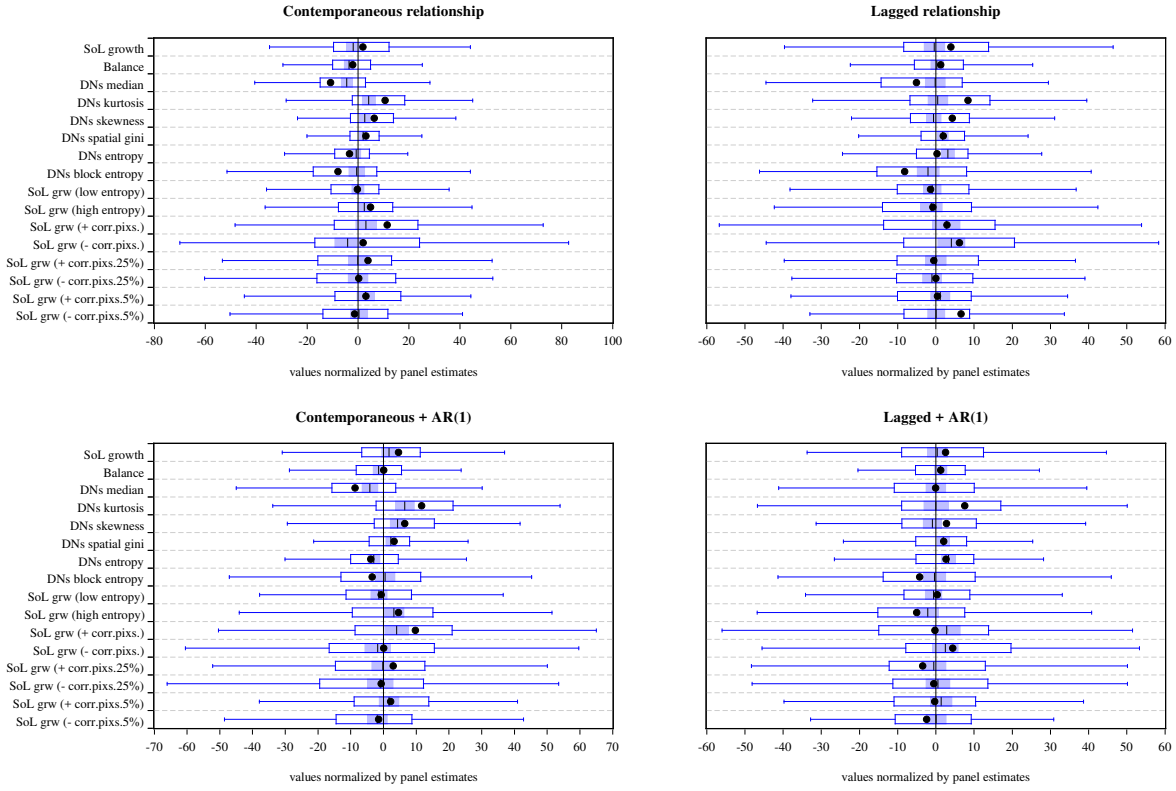
Figure 12: Cross-country dispersion of variables over time (2 of 2).

Table 7: Summary of panel estimates on relationship between night lights and GDP growth.

Model	Estimates on night lights indicators					
	Lagged (no AR)		Contemp. (no AR)		Contemp. (with AR)	
Aggregate night lights indicators:						
+ SoL growth	1.432***		2.658***		1.766***	
	(0.390)		(0.413)		(0.389)	
+ Balance	1.235***		2.991***		2.049***	
	(0.404)		(0.412)		(0.398)	
Distribution-based night lights indicators:						
+ DNs median	-0.256**		-0.116		-0.016	
	(0.122)		(0.121)		(0.097)	
+ DNs kurtosis	0.070***		0.049*		0.014	
	(0.026)		(0.025)		(0.019)	
+ DNs skewness	0.743***		0.559**		0.204	
	(0.231)		(0.225)		(0.172)	
+ DNs spatial gini	-0.064*		-0.013		-0.045*	
	(0.037)		(0.036)		(0.027)	
+ DNs entropy	-0.080***		-0.066***		-0.034**	
	(0.021)		(0.021)		(0.016)	
+ DNs block entropy	-0.164		0.051		-0.253	
	(0.343)		(0.359)		(0.284)	
Location-based night lights indicators:						
+ SoL grw. (low/high entropy)	0.237	0.198	-0.219	1.903***	-0.067	0.799
	(0.179)	(0.460)	(0.193)	(0.491)	(0.202)	(0.487)
+ SoL grw. (+/- corr. pixs.)	0.159	-0.129	0.597**	0.979***	0.609**	0.852***
	(0.262)	(0.249)	(0.263)	(0.248)	(0.262)	(0.246)
+ SoL grw. (+/- corr. pixs.25%)	-0.059	-0.135	0.142	-0.173*	0.151	-0.195*
	(0.104)	(0.102)	(0.104)	(0.104)	(0.105)	(0.101)
+ SoL grw. (+/- corr. pixs.5%)	0.033	-0.113	0.090	-0.138*	0.105	-0.152*
	(0.079)	(0.079)	(0.081)	(0.081)	(0.080)	(0.079)

Notes: The estimates are obtained by regressing real GDP growth on the corresponding night lights indicator lagged/contemporaneous values, plus a constant, an autoregressive term (when indicated in the column header), and country and year fixed effects. For more estimation details see the notes of table 1, where estimates for the case of lagged indicators with the AR(1) term are presented. The values inside parentheses are country-clustered robust standard errors. ***, **, and * stand for 1%, 5%, and 10% levels of statistical significance, respectively.

Figure 13: Distributions of individual estimates.



Notes: See the notes of figure 5, which corresponds to the fourth panel in this figure.

Figure 14: In-sample log RMSE ratios across countries.

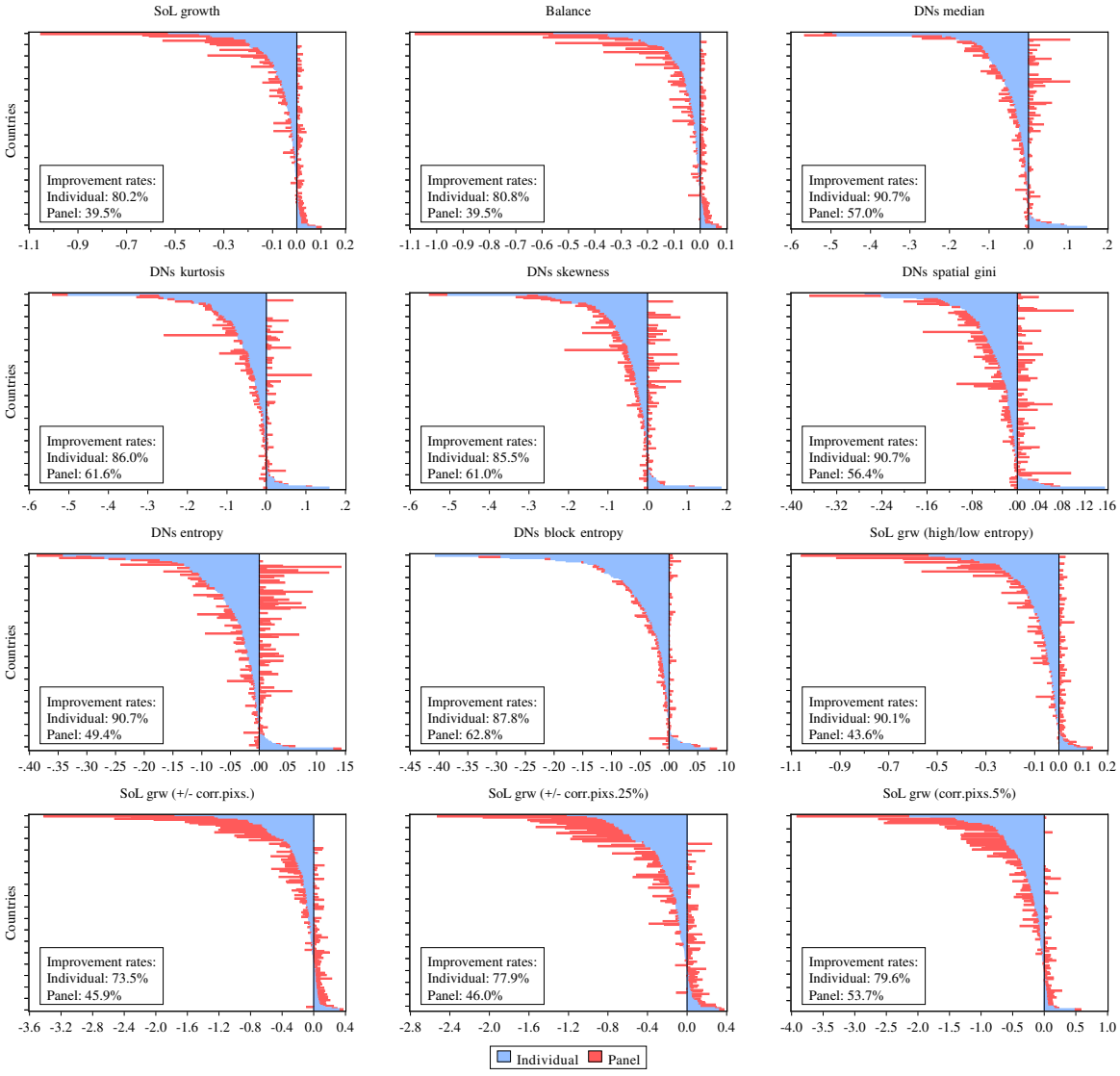


Figure 15: Out-of-sample log RMSE ratios from recursive model estimates across countries.

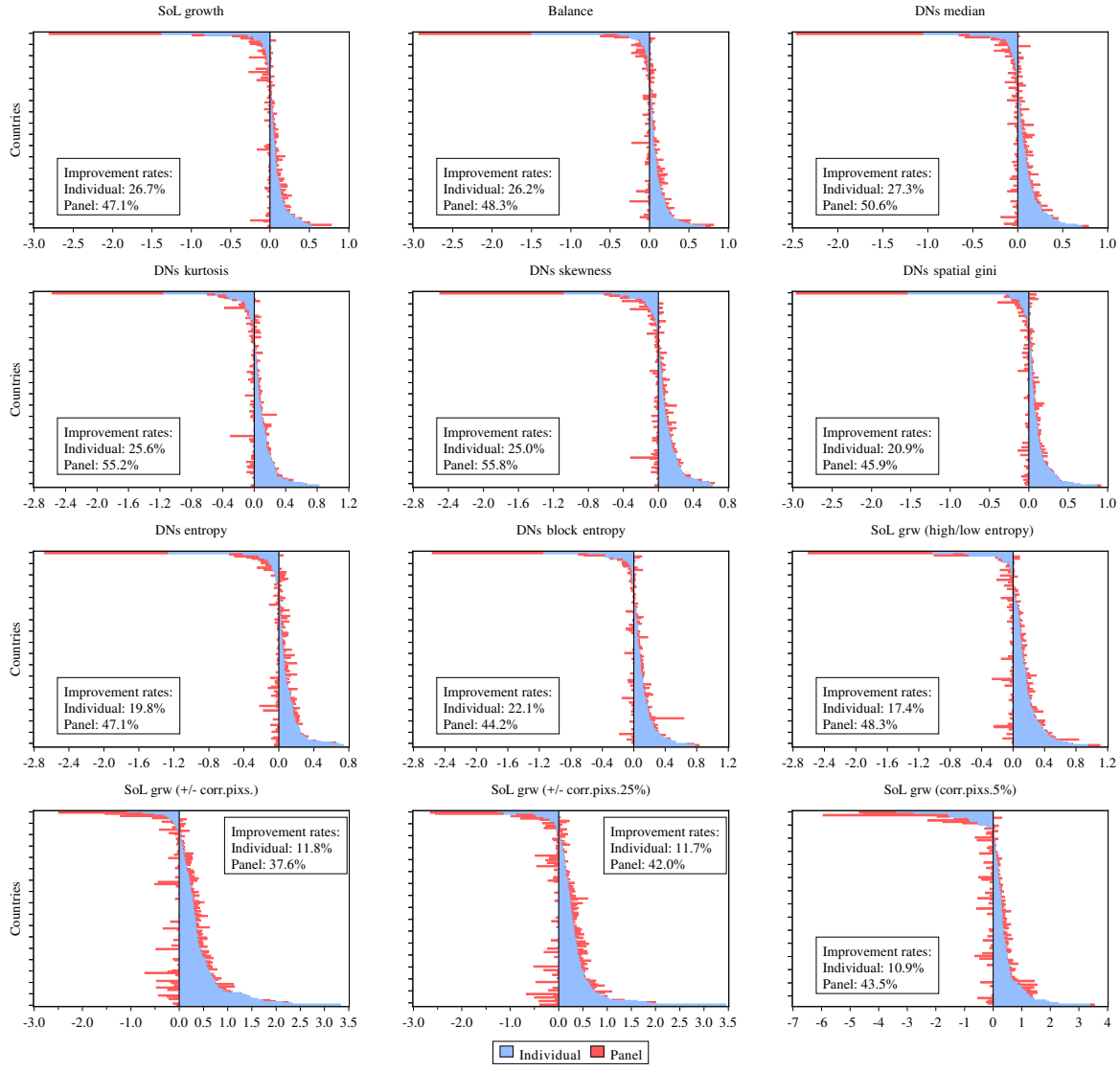


Figure 16: Out-of-sample log RMSE ratios from full-sample model estimates across countries.

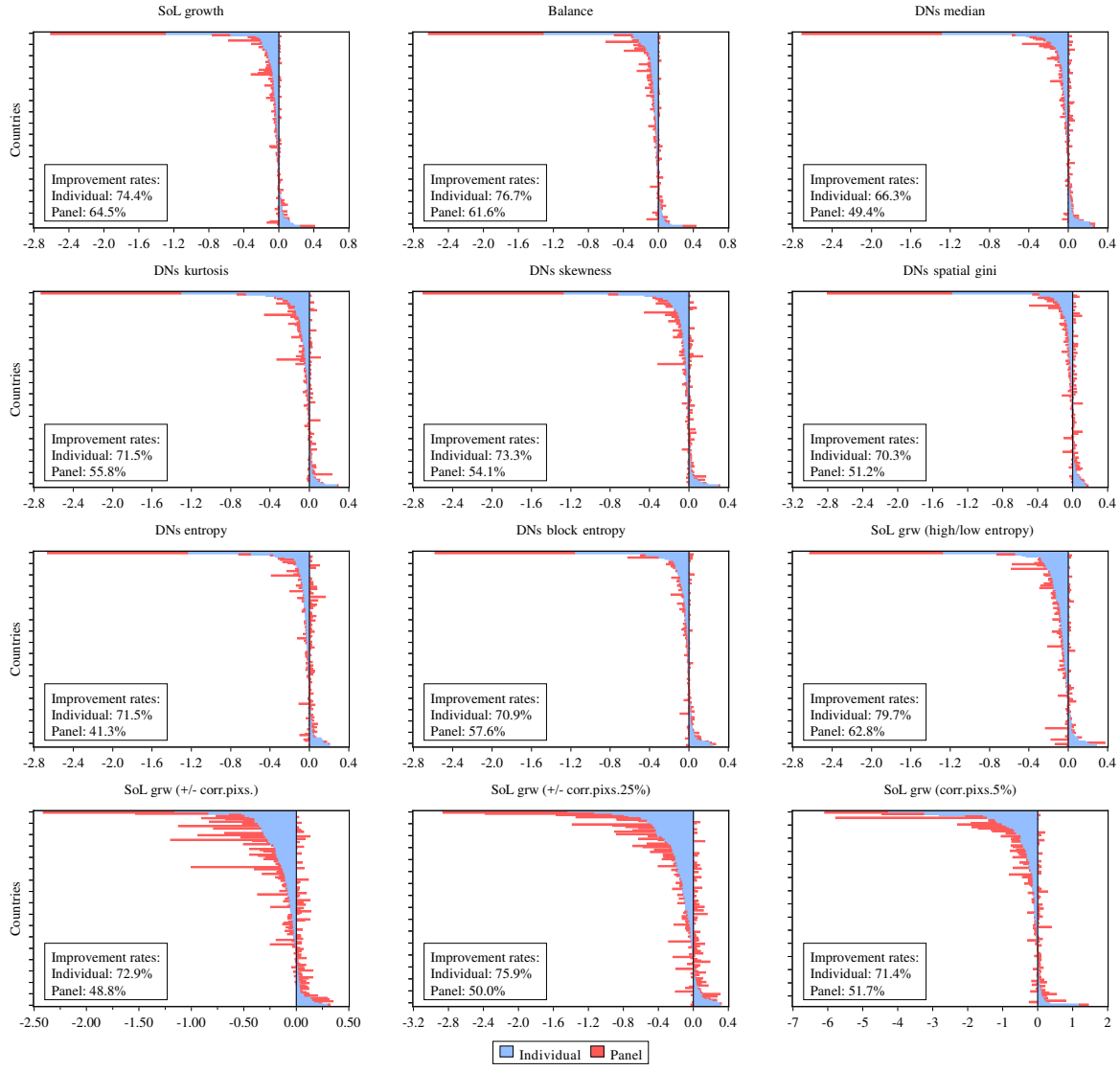


Table 8: Hit-rates on DM tests for predictive improvement.

Predictor	In-sample		OoS recursive		OoS full-sample	
	Panel	Indiv.	Panel	Indiv.	Panel	Indiv.
Aggregate night lights indicators:						
+ SoL growth	12.2%	18.6%	17.4%	14.5%	15.7%	19.8%
	[5.8%]	[7.0%]	[19.2%]	[5.8%]	[16.3%]	[17.4%]
+ Balance	13.4%	16.3%	16.9%	14.5%	18.6%	21.5%
	[4.7%]	[5.8%]	[18.6%]	[4.1%]	[19.2%]	[20.9%]
Distribution-based night lights indicators:						
+ DNs median	7.0%	15.1%	11.0%	14.0%	11.0%	13.4%
	[12.2%]	[9.3%]	[15.7%]	[1.7%]	[15.1%]	[11.6%]
+ DNs kurtosis	7.6%	23.8%	11.6%	17.4%	14.5%	17.4%
	[17.4%]	[14.5%]	[15.1%]	[4.1%]	[22.7%]	[13.4%]
+ DNs skewness	8.7%	24.4%	8.7%	16.9%	10.5%	15.7%
	[16.9%]	[16.3%]	[14.5%]	[4.7%]	[19.2%]	[14.0%]
+ DNs spatial gini	11.6%	19.2%	15.7%	13.4%	16.9%	12.2%
	[9.9%]	[4.1%]	[11.6%]	[3.5%]	[11.0%]	[7.6%]
+ DNs entropy	8.7%	18.0%	11.0%	13.4%	11.0%	14.5%
	[12.2%]	[7.6%]	[11.6%]	[4.7%]	[11.6%]	[14.5%]
+ DNs block entropy	2.3%	18.0%	12.2%	14.5%	18.0%	18.6%
	[1.7%]	[7.0%]	[15.7%]	[4.1%]	[18.6%]	[15.7%]
Location-based night lights indicators:						
+ SoL grw. (low/high entropy)	15.1%	18.6%	16.9%	9.9%	16.3%	18.6%
	[8.7%]	[14.0%]	[19.2%]	[1.7%]	[20.3%]	[20.3%]
+ SoL grw. (+/- corr. pixs.)	18.8%	11.2%	15.3%	10.0%	21.2%	12.9%
	[9.4%]	[18.8%]	[14.7%]	[1.2%]	[15.9%]	[21.2%]
Averages	10.5%	18.3%	13.7%	13.9%	15.4%	16.5%
	[9.9%]	[10.4%]	[15.6%]	[3.6%]	[17.0%]	[15.7%]
Max-t reality check	8.1%	17.4%	14.0%	12.2%	13.4%	14.0%

Notes: Same as in table 4 except that here the hit-rates are based on the Diebold and Mariano (1995) (DM) one-sided test for the null hypothesis of equal predictive accuracy. The hit-rates inside brackets are obtained from tests based on the Student's t-distribution of the statistic, instead of the bootstrapped version (without brackets).

Table 9: Hit-rates on theoretical CW tests for predictive improvement.

Predictor	In-sample		OoS recursive		OoS full-sample	
	Panel	Indiv.	Panel	Indiv.	Panel	Indiv.
Aggregate night lights indicators:						
+ SoL growth	9.3%	35.5%	23.8%	15.1%	21.5%	37.8%
+ Balance	8.7%	34.9%	22.7%	15.1%	23.3%	39.0%
Distribution-based night lights indicators:						
+ DNs median	15.7%	37.2%	19.2%	19.2%	18.6%	30.8%
+ DNs kurtosis	20.9%	39.5%	22.1%	17.4%	25.6%	33.7%
+ DNs skewness	22.7%	41.3%	20.3%	16.9%	22.7%	33.1%
+ DNs spatial gini	15.1%	32.6%	16.9%	13.4%	15.1%	24.4%
+ DNs entropy	15.7%	36.6%	17.4%	13.4%	15.1%	32.0%
+ DNs block entropy	2.3%	36.0%	18.0%	15.7%	19.2%	31.4%
Location-based night lights indicators:						
+ SoL grw. (low/high entropy)	13.4%	57.0%	22.1%	9.9%	23.8%	47.1%
+ SoL grw. (+/- corr. pixs.)	24.7%	64.7%	22.4%	13.5%	26.5%	54.7%
Averages	14.9%	41.5%	20.5%	15.0%	21.1%	36.4%

Notes: See notes of table 4. Here the significance levels are obtained from the Student's t-distribution.

A.3 Countries ISO codes

ISO	COUNTRY
HIGH INCOME COUNTRIES:	
ARE	UNITED ARAB EMIRATES
AUS	AUSTRALIA
AUT	AUSTRIA
BEL	BELGIUM
BHS	BAHAMAS
BRN	BRUNEI DARUSSALAM
CAN	CANADA
CHE	SWITZERLAND
CYP	CYPRUS
DEU	GERMANY
DNK	DENMARK
ESP	SPAIN
FIN	FINLAND
FRA	FRANCE
GBR	UNITED KINGDOM
GRC	GREECE
HKG	HONG KONG
IRL	IRELAND
ISL	ICELAND
ISR	ISRAEL
ITA	ITALY
JPN	JAPAN
KOR	KOREA, REPUBLIC OF
KWT	KUWAIT
LUX	LUXEMBOURG
MLT	MALTA
NLD	NETHERLANDS
NOR	NORWAY
NZL	NEW ZEALAND
PRT	PORTUGAL
QAT	QATAR
SVN	SLOVENIA
SWE	SWEDEN
TWN	TAIWAN, PROVINCE OF CHINA
USA	UNITED STATES
LOWER MIDDLE INCOME COUNTRIES:	
AGO	ANGOLA
ALB	ALBANIA
ARM	ARMENIA
BGR	BULGARIA
BIH	BOSNIA AND HERZEGOVINA
BLR	BELARUS
BLZ	BELIZE
BOL	BOLIVIA
CHN	CHINA
COL	COLOMBIA
CPV	CAPE VERDE
DJI	DJIBOUTI
DOM	DOMINICAN REPUBLIC
DZA	ALGERIA
ECU	ECUADOR
EGY	EGYPT
FJI	FIJI
FSM	MICRONESIA, FEDERATED STATES OF
GEO	GEORGIA
GTM	GUATEMALA
GUY	GUYANA
HND	HONDURAS
IDN	INDONESIA
IRN	IRAN, ISLAMIC REPUBLIC OF
IRQ	IRAQ
JAM	JAMAICA
JOR	JORDAN
KAZ	KAZAKHSTAN
LKA	SRI LANKA
MAR	MOROCCO
MDA	MOLDOVA, REPUBLIC OF
MKD	MACEDONIA, THE FORMER YUGOSLAV REPUBLIC OF
NAM	NAMIBIA
PER	PERU
PHL	PHILIPPINES
PNG	PAPUA NEW GUINEA
PRY	PARAGUAY
ROU	ROMANIA
RUS	RUSSIAN FEDERATION
SLV	EL SALVADOR
SUR	SURINAME
SWZ	SWAZILAND
SYR	SYRIAN ARAB REPUBLIC
THA	THAILAND
TKM	TURKMENISTAN
TLS	TIMOR-LESTE
TON	TONGA
TUN	TUNISIA
UKR	UKRAINE
UZB	UZBEKISTAN
VCT	SAINT VINCENT AND THE GRENADINES
VUT	VANUATU
WSM	SAMOA

ISO	COUNTRY
UPPER MIDDLE INCOME COUNTRIES:	
ARG	ARGENTINA
BRA	BRAZIL
BRB	BARBADOS
BWA	BOTSWANA
CHL	CHILE
CRI	COSTA RICA
CZE	CZECH REPUBLIC
EST	ESTONIA
GAB	GABON
HRV	CROATIA
HUN	HUNGARY
LBN	LEBANON
LBY	LIBYAN ARAB JAMAHIRIYA
LCA	SAINT LUCIA
LTU	LITHUANIA
LVA	LATVIA
MEX	MEXICO
MNE	MONTENEGRO
MUS	MAURITIUS
MYS	MALAYSIA
OMN	OMAN
PAN	PANAMA
POL	POLAND
SAU	SAUDI ARABIA
SRB	SERBIA
SVK	SLOVAKIA
TTO	TRINIDAD AND TOBAGO
TUR	TURKEY
URY	URUGUAY
VEN	VENEZUELA
ZAF	SOUTH AFRICA
LOW INCOME COUNTRIES:	
AFG	AFGHANISTAN
AZE	AZERBAIJAN
BDI	BURUNDI
BEN	BENIN
BFA	BURKINA FASO
BGD	BANGLADESH
BTN	BHUTAN
CAF	CENTRAL AFRICAN REPUBLIC
CIV	COTE D'IVOIRE
CMR	CAMEROON
COD	CONGO, THE DEMOCRATIC REPUBLIC OF THE
COG	CONGO
COM	COMOROS
ERI	ERITREA
ETH	ETHIOPIA
GHA	GHANA
GIN	GUINEA
GMB	GAMBIA
GNB	GUINEA-BISSAU
HTI	HAITI
IND	INDIA
KEN	KENYA
KGZ	KYRGYZSTAN
KHM	CAMBODIA
LAO	LAO PEOPLE'S DEMOCRATIC REPUBLIC
LBR	LIBERIA
LSO	LESOTHO
MDG	MADAGASCAR
MLI	MALI
MMR	MYANMAR
MNG	MONGOLIA
MOZ	MOZAMBIQUE
MRT	MAURITANIA
MWI	MALAWI
NER	NIGER
NGA	NIGERIA
NIC	NICARAGUA
NPL	NEPAL
PAK	PAKISTAN
RWA	RWANDA
SEN	SENEGAL
SLB	SOLOMON ISLANDS
SLE	SIERRA LEONE
STP	SAO TOME AND PRINCIPE
TCD	CHAD
TGO	TOGO
TJK	TAJIKISTAN
TZA	TANZANIA, UNITED REPUBLIC OF
UGA	UGANDA
VNM	VIET NAM
YEM	YEMEN
ZMB	ZAMBIA
ZWE	ZIMBABWE