What Do We See in the Lights?
Lights at Night and Measures of National Growth

Xiaoman Luo* and Ashish Shenoy*

1 Introduction

Measuring national-level economic activities is fundamental and crucial to studying economic growth and poverty. While Gross Domestic Product (GDP) is the most popular metric, the quality by which it is measured has been a serious concern, especially for less-developed countries lacking strong government statistical infrastructure. A nation's ability to collect, analyze, and disseminate high-quality data about its population and economy is referred to as "statistical capacity". [1] According to the Penn World Table (PWT) Version 10.0 (Feenstra et al., 2015; Zeileis, 2021), the statistical capacity has a highly unequal distribution across countries and over time, which leads to the uncertainty in international comparisons of GDP (Deaton and Heston, 2010). Recently, alternative approaches to measure economic activities have been proposed based on large-scale surveys. For example, Young (2012) uses the expenditure measures in the Demographic and Health Survey (DHS) data to estimate the growth rate of real consumption in sub-Saharan countries, which is higher than that indicated by official accounts. Despite the much higher quality, the availability is hindered by the high cost of national-level surveys, especially in developing countries.

To break the trade-off between quality and availability, creative proxy measures have been proposed. Among others, the use of nighttime light data from satellite images has become popular in recent years. A vast literature argues that the nighttime light luminosity and economic activity levels are highly correlated, ranging from the world' largest economy to some least developed countries (e.g. Croft, 1978; Elvidge et al., 1997; Sutton and Costanza, 2002; Ghosh et al., 2009, 2010; Henderson et al., 2018; Hu and Yao, 2021). Unlike GDP, which has high availability but potentially low quality, or the survey-based measures, which have high quality but low availability, the nighttime light-based measures excel on both ends — advances in remote sensing and image processing technologies have increased the quality of nighttime light measures which are publicly available at low cost. On top of that, the nighttime light-based measures are available at relatively

*University of California, Davis
[1]Source: https://datatopics.worldbank.org/statisticalcapacity/

high temporal frequency and spatial resolution — the raw data is collected on a daily basis and further processed by the National Oceanic and Atmospheric Administration (NOAA) into monthly imagery at a spatial resolution of 30 x 30 arc seconds, or approximately 1000 x 1000 meters at the equator.

Although nighttime light-based measures have unique advantages in assessing the level of economic activities, the extremely strong correlation with other measures such as the reported GDP found in the literature is suspicious. For example, Elvidge et al. (1997) found that the correlation between GDP and nighttime light luminosity (after log-transformation) is as high as 0.97 for 21 countries. Similarly high correlations have been found for other sub-populations (e.g. Sutton and Costanza, 2002; Chen and Nordhaus, 2011). Taking a step back, if we believe that reported GDP is not reliable enough, the reliability of any other measure that has $80\% - 90\%$ correlation with GDP should be questionable as well. As a result, an extremely strong correlation with the reported GDP disqualifies the nighttime light measures. On the other hand, it is apparent that the nighttime could not reflect every aspect of economic activities and, thus, cannot be near-perfectly correlated with GDP.

In this paper we demonstrate that these high correlations are indeed unreasonable and the main reason is the failure to adjust for non-stationarity. The correlation between two non-stationary time series tends to be excessively large even if no such relationship is present in the data generating process. This phenomenon is called "spurious correlation" in macro-econometrics (Granger and Newbold, 1974); see also standard textbooks (e.g. Stock et al., 2012; Wooldridge, 2015). After adjusting for non-stationarity properly, the association declines drastically. The substantial drop also corroborates that the issue of spurious correlation cannot be neglected.

Another challenge is the heterogeneity in the light-GDP association. One contribution of this paper is to show that nighttime light is an appropriate measure of GDP in some contexts but not in others. The marginal impact of "true" GDP on nightlight intensity can differ across countries with different geographic or cultural characteristics. In addition, the relationship between "true" GDP and measured GDP varies depending on statistical capacity and other things. While most existing works implicitly assumes the homogeneity in their analyses, we explicitly build the heterogeneity into the model. Under our model, we show that the usual ordinary least squares (OLS) and fixed-effect regression estimators can be misleading in that they weigh countries in a highly imbalanced way. In particular, they put very low weights on smaller or less-developed countries while the weights on other countries are unequal and uninterpretable.

To address the heterogeneity, we carefully define our inferential targets. The first class of targets are the average correlation coefficients (ACC), defined as a weighted average of the individual correlation coefficients (ICC) between nighttime light and GDP measures for each country with user-specified weights. In particular, we are interested in the ACC over all countries and ACC over a certain sub-population, such as the middle-income countries. We propose a weighted least squares (WLS) first-differencing regression estimator that is an unbiased estimator of the ACC without assuming stationarity. The p-value and confidence interval can be computed by standard software. The second class of inferential targets are non-zero ICCs. We cannot expect to estimate all ICCs to a reasonably high accuracy because the panel is short and data is scarce for each country. Fortunately, our preliminary analysis shows that the ICC is close to zero for most countries. We exploit this sparsity and apply the LASSO regression to identify and estimate non-zero ICCs. To control the fraction of false discoveries, we apply the advanced "knockoff" method (Barber and Candès, 2015) to control the false discovery rate (FDR) at level 0.2, meaning that the fraction of true positives is at least 80% on average. This identifies in 10 countries.

1.1 Data

The nighttime light data we analyze in this paper is provided by Proville et al. (2017) and collected by the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) satellites, which provides global daily measurements of nocturnal light. Annual average composite images of nighttime light have been released by NOAA since 1992, including 30 x 30 arc-second grids covering -180 to 180 degrees longitude and -65 to 75 degrees latitude. In particular, we use the stable light measurements that measures persistent lighting, which dismiss fires as ephemeral events and remove background noises. Each pixel in the nighttime light image has a digital number (DN) representing light intensity, ranging from 0 to 63. The measure we use for light intensity is the area lit during 1992 to 2013. To calculate area lit, the first step is to calculate pixel count, which is the sum of DN for all pixels with $DN > 31$ (to remove weak light signals) in grids of 30 x 30 arc seconds. Next, the pixel counts are converted to the equivalent area coverage in square kilometers. This is the area lit measure we use. On the other hand, we use the World Bank nominal GDP at the current US dollar levels.

1.2 Other related work

The nighttime light measures have been used as a proxy for the level of economic activity in different problems. Ghosh et al. (2009) estimated the light-GDP association in the United States, and then applied the estimated association in Mexico to predict Mexico's economic activity levels, concluding that Mexico's informal economy and remittances are much higher than reported in national accounts. Henderson et al. (2018) studied the association between trade and agricultural factors and the distribution of economic activities proxied by the nighttime light measure. The nighttime light measures are also used to construct better predictors for economic growth. For example, Henderson et al. (2012) developed a measure as a weighted average of the income growth from national accounts and the income growth predicted by the nighttime light measures. The weight reflects the statistical capacity and differs by countries.

Because of the high spatial resolution, the nighttime light measures are also used as proxies for economic activities at the sub-national level (e.g. Sutton and Costanza, 2002; Sutton et al., 2007; Hodler and Raschky, 2014; Harari, 2020), the region level (e.g. Doll et al., 2006), or even abstract areas such as ethnic homelands (Michalopoulos and Papaioannou, 2013).

The rest of this paper is structured as follows. We begin Section 2 by describing our model and comparing the limitation of popular methods in estimating model parameters. We then propose a weighted least square estimator in Section 3, which addresses the limitations of methods mentioned in Section 2. We also present the estimated relationship between nighttime light and GDP — the highest and strongest relationship exists only in middle-income countries. With these results of subgroup average association, we are curious about the individual-level association. In Section 4, we further estimate the correlation coefficients between nighttime light and GDP for each country. Inspired by the fact that many country-level associations observed in the preliminary analysis is close to zero, we apply the LASSO regression to identify and estimate non-zero individual correlation coefficients. We further apply the "knockoff" method to control the false discovery rate among the selected countries with non-zero coefficients, and the results from LASSO regression and "knockoff" method are very consistent.

2 Model

2.1 A panel data model with country-wise heterogeneity and non-stationarity

For country $i \in \{1, \ldots, I\}$ in year $t \in \{1, \ldots, T\}$, let L_{it} denote the logarithmic nighttime light measure, Y_{it} denote the reported measure of GDP, and Y_{it}^* denote the (unobserved) measure of actual level of economic activities. In this paper, we take Y_{it} as logarithm of reported annual GDP, and conceptualize Y_{it}^* as the true annual GDP. Intuitively, Y_{it}^* is associated with both Y_{it} and L_{it}, though not perfectly correlated. Their relationships can be abstracted through the following simplified model:

$$L_{it} = \eta_i^L + \gamma_i^L Y_{it}^* + \epsilon_{it}^L, \quad Y_{it} = \eta_i^Y + \gamma_i^Y Y_{it}^* + \epsilon_{it}^Y, \tag{1}$$

where $(\epsilon_{it}^L, \epsilon_{it}^Y)$ are error terms that capture the effects of other economic variables. Here, we consider a country-specific intercept and slope in both equations in order to capture the country-wise heterogeneity. Indeed, the effect of actual economic activities on the nighttime light measure depends on the level of industrialization, environmental policies, lifestyles of the citizens, and so on, while the mechanisms through which each country misreports their GDP are affected by the statistical capacity, political ideology, technology, etc. It is thus unrealistic to impose a constant parameter for all countries.

Since Y_{it}^* is a latent variable, the parameters are unidentifiable without further structural or distributional assumptions. While some previous works attempted to estimate γ^L and γ_Y and recover Y_{it}^* (e.g. Henderson et al., 2012), this paper is focused on exploring the marginal association between L_{it} and Y_{it}. A reduced-form model yielded by (1) is

$$L_{it} = \alpha_i + \beta_i Y_{it} + \epsilon_{it},$$

where

$$\beta_i = \gamma_i^L/\gamma_i^Y, \quad \alpha_i = \eta_i^L - \eta_i^Y \gamma_i^L/\gamma_i^Y, \quad \epsilon_{it} = \epsilon_{it}^L - \epsilon_{it}^Y \gamma_i^L/\gamma_i^Y. \tag{2}$$

Under the above reduced-form model, the inferential targets are β_is, which measures the country-wise association between the nighttime light measure and the reported GDP.

Since we are analyzing a fairly long panel (with 17 years), we cannot neglect the non-stationarity of the nighttime light measure and reported GDP. It is widely accepted that GDP measures have a unit root (e.g. McCallum, 1993; Libanio, 2005). Unless the nighttime light measure is cointegrated with the reported GDP, ϵ_{it} should be a non-stationary process for each country i. The cointegration is arguably impossible since the nighttime light measure only captures part of economic activities. For this reason, we model the error terms as a unit-root process of order 1.

For reference, we summarize the model below:

$$L_{it} = \alpha_i + \beta_i Y_{it} + \epsilon_{it}, \quad \epsilon_{it} = \epsilon_{i(t-1)} + \nu_{it}, \quad \nu_{it} \text{ is a mean-zero stationary process for each } i. \tag{3}$$

The above model is similar to a non-stationary dynamic panel model because it can be equivalently formulated as

$$L_{it} = L_{i(t-1)} + \beta_i Y_{it} - \beta_i Y_{i(t-1)} + \nu_{it}. \tag{4}$$

Nevertheless, (3) is more complicated due to the country-wise heterogeneity, which introduces the incidental parameter problem.

2.2 Inferential targets: average and individual correlation coefficients

It would be ideal to estimate every individual correlation coefficient (ICC) β_i precisely. However, this can only be achieved with a sufficiently long panel because there are only T observations $(L_{it}, Y_{it})_{t=1}^T$ carrying information on β_i, and there are at least two more nuisance parameter: the intercept α_i and the variance of ν_{it}. In our case, $T = 22$ and thus only 20 degree-of-freedom can be used to estimate β_i. Therefore, it is impossible to estimate all β_i's with a desirable accuracy without further assumptions (i.e, under unrestricted heterogeneity).

In the presence of unrestricted heterogeneity, a routine approach to summarize the effects is to consider an average correlation coefficient (ACC), as an average of ICCs with *user-specified* weights.

For example, we can consider the ACC over all countries

$$\beta_{\text{all}} \triangleq \frac{1}{I}\sum_{i=1}^{I}\beta_i. \tag{5}$$

Sometimes it could be more informative to investigate a weighted average of β_i:

$$\beta(v) = \frac{\sum_{i=1}^{I} v_i\beta_i}{\sum_{i=1}^{I} v_i}, \tag{6}$$

where the weights depend on, for example, the population size. Similarly, we can define the ACC over a sub-population, say, the middle-income countries. As we will show in Section 3, the effective sample size for ACC is much larger than T, and hence the variance of the ACC estimator is much lower than that of the ICC estimator.

The ICCs could also be of interest, when the goal is to find the countries with largest ICCs, or to explore how the ICC varies with other factors. As mentioned above, we need to impose additional plausible assumptions in order to ensure inferential reliability. In this paper, we exploit the sparsity of β_i's observed from the preliminary analysis. It may sound implausible because the overall correlation between the nighttime light measure and GDP has been found to be high in the literature (e.g. Elvidge et al., 1997; Sutton and Costanza, 2002; Chen and Nordhaus, 2011). However, we will show in Section 2.3.1 that the high correlation is spurious due to the failure of adjusting for non-stationarity. After the correct adjustment, most ICCs are substantially reduced and close to zero. On the other hand, recalling that β_i is the ratio between dL_{it}/dY_{it}^* and dY_{it}/dY_{it}^*. it is large only when either the nighttime light captures most economic activities or there is substantial amount of under-reporting. Intuitively, neither of them is realistic. Therefore, we argue that it is reasonable to assume sparsity.

2.3 Challenges

2.3.1 Non-stationarity and spurious correlation

It is known in the macro-econometric theory that the correlation coefficient between two non-stationary sequences is typically large and can sometimes be close to ± 1, even if the two sequences are independent (e.g. Phillips, 1998; Ernst et al., 2019). This surprising phenomenon is dubbed "spurious correlation" in the literature. Failure to adjust for spurious correlation properly can yield misleading conclusions. In fact, for the countries under study in this paper, we test for unit root with ADF test (critical ADF p-value > 0.1) (Cheung and Lai, 1995) and KPSS test (critical KPSS p-value < 0.1) (Kwiatkowski et al., 1992). We find that about 70% of the countries have non-stationarity in GDP, and 40% of countries have non-stationarity in nighttime light.

Unfortunately, this issue has not been addressed in the literature on nighttime light measures.

Suppose the non-stationarity is absent and thus the spurious correlation does not exist, the correlation coefficient between L_{it} and Y_{it} (or R^2 equivalently) should be similar to that between $\Delta L_{it} = L_{it} - L_{i(t-1)}$ and $\Delta Y_{it} = Y_{it} - Y_{i(t-1)}$. As a sanity check, I perform the ADF test and KPSS test again after taking the first difference for GDP and nighttime light, and the corresponding share of countries demonstrating non-stationarity drop to about 20% and less than 5% respectively. Note that we use the 10% significance level here, so there is 10% chance that we get reports of non-stationarity when there actually is not, Thus, 20% demonstrating non-stationarity is not quite a large number that we should worry about.

Using our data, if L_{it} and Y_{it} are given by the levels (without any transformation), the first correlation is 0.96 (95% confidence interval $[0.95, 0.97]$), coinciding with the observation by Elvidge et al. (1997), while the second correlation is 0.18 (95% confidence interval $[0.10, 0.26]$); if L_{it} and Y_{it} are given by the log-levels, as in our analyses, so that the first-order difference measures the growth rate, the first correlation is 0.31 (95% confidence interval $[0.23, 0.39]$), while the second correlation is 0.00 (95% confidence interval $[-0.08, 0.08]$). The confidence intervals are non-overlapping with a large gap in both cases, suggesting strong evidence for spurious correlation.

2.3.2 Failure of OLS estimators on the pooled sample

A popular approach is to run the OLS regression on the pooled panel data (e.g. Henderson et al., 2012). However, this approach relies on three assumptions: homogeneity of ICCs ($\beta_i \equiv \beta$), homogeneity of intercepts ($\alpha_i \equiv \alpha$), and stationarity of error terms. Even if the last two assumptions both hold, the heterogeneity in ICCs can make the resulting estimator uninterpretable. To illustrate the failure of the OLS estimator, assume $\alpha_i = 0$ and ϵ_{it} are stationary for a moment. It is not hard to see that

$$\mathbb{E}[\hat{\beta}_{\mathrm{OLS}}] = \frac{\sum_{i=1}^{I} w_i^{\mathrm{OLS}} \beta_i}{\sum_{i=1}^{I} w_i^{\mathrm{OLS}}},$$

where

$$w_i^{\mathrm{OLS}} = \sum_{t=1}^{T} Y_{it}^2.$$

This limit is a weighted average of ICCs, where the weights are data-dependent. One implication is that the countries with larger levels of GDP have larger weights. Given the dramatic imbalance in country-level GDP, the weights for smaller or less developed countries are almost zero. Therefore, the OLS estimator essentially estimates a weighted average of a few large countries in the world, which is clearly misleading.

A more sophisticated approach is to add country fixed-effects into the regression (e.g. Henderson et al., 2012). This relaxes the assumption of homogeneous intercepts. Nevertheless, the heterogeneity in β_i's is still problematic even if the error terms are stationary. Through some tedious calculations, we can show that,

$$\mathbb{E}[\hat{\beta}_{\text{FE}}] = \frac{\sum_{i=1}^{I} w_i^{\text{FE}} \beta_i}{\sum_{i=1}^{I} w_i^{\text{FE}}},$$

where

$$w_i^{\text{FE}} = \sum_{t=1}^{T} (Y_{it} - \bar{Y}_i)^2, \quad \bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it}.$$

Again, the limit of this estimator severely penalizes smaller or less-developed countries.

Another commonly used method in panel data analysis is the first-difference regression estimator, which regresses $\Delta L_{it} = L_{it} - L_{i(t-1)}$ on $\Delta Y_{it} = Y_{it} - Y_{i(t-1)}$, though we are not aware of the application in the literature on nighttime light measures. Under our model (3),

$$\Delta L_{it} = \beta_i \Delta Y_{it} + \nu_{it}.$$

Therefore, the first-difference estimator relaxes the assumptions on both the homogeneity of intercepts and the stationarity of error terms. Despite being more robust than the aforementioned two estimators, it still suffers from the same issue under the heterogeneity of ICCs. As with $\hat{\beta}_{\text{OLS}}$, we can show that

$$\mathbb{E}[\hat{\beta}_{\text{FD}}] = \frac{\sum_{i=1}^{I} w_i^{\text{FD}} \beta_i}{\sum_{i=1}^{I} w_i^{\text{FD}}},$$

where

$$w_i^{\text{FD}} = \sum_{t=1}^{T} (\Delta Y_{it})^2.$$

When Y_{it} is given by the log-GDP, ΔY_{it} approximately measures GDP growth rate. While the variation of GDP growth rate is much smaller than that of GDP, the FD weights are still non-uniform and they are difficult to interpret.

Table 1: Methods for estimating model parameters

Method	Assumptions	Regression (in `R` syntax)
Ordinary Least Squares regression	$\alpha_i \equiv \alpha$, $\beta_i \equiv \beta$, ϵ_{it} is stationary	`lm(`$L_{it} \sim \alpha + \beta \cdot Y_{it}$`)`
(Country) Fixed-Effects regression	$\beta_i \equiv \beta$, ϵ_{it} is stationary	`lm(`$L_{it} \sim \alpha_i + \beta \cdot Y_{it}$`)`
First-Difference regression	$\beta_i \equiv \beta$	`lm(`$\Delta_t L_{it} \sim \beta \cdot \Delta_t Y_{it}$`)`

3 Estimating Average Correlation Coefficients

3.1 Weighted first-difference regression estimator

As discussed in Section 2.3.2, the first-difference regression estimator requires neither homogeneity of intercepts nor stationarity of errors. In this section, we will propose an adjustment for the first-difference estimator to handle the heterogeneity of ICCs. To set the stage, we start by reformulating the model (3) based on the first-differenced quantities:

$$\Delta L_{it} = \beta_i \cdot \Delta Y_{it} + \nu_{it}. \tag{7}$$

A natural estimator for $\beta(v)$ is the aggregated OLS estimator, defined as

$$\hat{\beta}(v) = \frac{\sum_{i=1}^{I} v_i \hat{\beta}_{i,\text{OLS}}}{\sum_{i=1}^{I} v_i},$$

where $\hat{\beta}_{i,\text{OLS}}$ is obtained by the OLS regression for unit i (i.e., on $(L_{it}, Y_{it})_{t=1}^{T}$). Clearly, it is an unbiased estimator of $\beta(v)$ and

$$\text{Var}\left[\hat{\beta}(v)\right] = \frac{\sum_{i=1}^{I} v_i^2 \text{Var}[\hat{\beta}_{i,\text{OLS}}]}{\left(\sum_{i=1}^{I} v_i\right)^2}.$$

Since $\hat{\beta}_{i,\text{OLS}}$ is the best linear unbiased estimator (BLUE) for β_i, $\hat{\beta}(v)$ is BLUE for $\beta(v)$. For example, for β_{all}, the variance of the above estimator is

$$\frac{1}{I^2} \sum_{i=1}^{I} \text{Var}[\hat{\beta}_{i,\text{OLS}}].$$

When the variances of all individual OLS estimators are similar, the above variance can be I times lower than each, implying that the efficiency to estimate ACC is much higher than the efficiency to estimate an ICC.

In Section 2.3.2, we have seen that the OLS estimator, whose objective function treats every country equally, does not treats each β_i equally. Therefore, the estimator $\hat{\beta}(v)$, which treats every β_i equally, must treat each country differently. It turns out that $\hat{\beta}(v)$ can be equivalently formulated as a weighted least squares (WLS) regression estimator with properly chosen weights:

$$\hat{\beta}(v) = \text{argmin}_{\beta} \frac{1}{IT} \sum_{i=1}^{I} \gamma_i \sum_{t=1}^{T} (\Delta L_{it} - \beta \Delta Y_{it})^2, \quad \text{where } \gamma_i = \frac{v_i}{\sum_{t=1}^{T} (\Delta Y_{it})^2}. \tag{8}$$

To see the equivalence, via the standard computation,

$$\hat{\beta}(v) = \frac{\sum_i \gamma_i \sum_t \Delta Y_{it} \Delta L_{it}}{\sum_i \gamma_i \sum_t (\Delta Y_{it})^2} = \frac{\sum_i \gamma_i \hat{\beta}_{i,\mathrm{OLS}} \sum_t (\Delta Y_{it})^2}{\sum_i \gamma_i \sum_t (\Delta Y_{it})^2} = \frac{\sum_{i=1}^I v_i \hat{\beta}_{i,\mathrm{OLS}}}{\sum_{i=1}^I v_i}.$$

Consider the ACC over all countries as an example, in which case γ_i is inversely proportional to $\sum_{t=1}^T (\Delta Y_{it})^2$. Then $\hat{\beta}(v)$ down-weights countries with larger growth rates while up-weights countries with smaller growth rates. This is unsurprising because the unweighted OLS estimator tends to up-weight the former.

Another advantage of the WLS formulation (8) is that the variance of $\hat{\beta}(v)$, and thus the p-value and confidence interval, can be computed by standard software directly. Comparisons of these methods are summarized in Table 1.

3.2 Results

First, we estimate β_{all}, the ACC over all countries defined in (5). The point estimate is 0.176 with the 95% confidence interval $[0.094, 0.258]$. Though it is significant at the 1% level, the magnitude is substantially smaller than what is found in previous works which did not adjust for spurious correlation (e.g. Chen and Nordhaus, 2011; Henderson et al., 2012). The huge gap between our estimates and the previous ones suggests that the non-stationarity is non-negligible in studying the relationship between nighttime light measures and measures of wealth such as GDP. Put another way, both variables should be measured by growth rates instead of levels.

Next, we estimate the ACC over subpopulations. The association between the nighttime light measure and GDP is affected by the industrialization levels. As a result, we expect the ACCs to vary with the income levels. We stratify the countries by income groups defined by World Bank. The results are summarized in Table 2. Again, the point estimates are not large in general. The estimates are significant at the 1% level for middle-income countries. This is partly attributed to the "capping" effect: for low-income countries, the luminosity is too low to be detected by the satellites, while for high-income countries, the luminosity is beyond the maximal detectable level. The latter is called the "saturation effect" in the remote sensing literature. Therefore, there is sufficient variation in the nighttime light measure only for middle income countries.

Table 2: WLS estimates by income group

	Income Group	Obs.	WLS estimate	p-value	95% C.I.
1	Low income	638	0.130	0.121	(-0.034, 0.295)
2	Upper middle income	1100	0.416^{***}	<0.001	(0.312, 0.521)
3	High income: non-OECD	924	0.004	0.330	(-0.004, 0.011)
4	Lower middle income	968	0.297^{***}	<0.001	(0.177, 0.416)
5	High income: OECD	704	0.111	0.126	(-0.031, 0.253)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: WLS estimates by continent

	Continent	Obs.	WLS estimate	p-value	95% C.I.
1	South Asia	154	0.465**	0.023	(0.058, 0.872)
2	Europe & Central Asia	1188	0.017	0.664	(-0.058, 0.092)
3	Middle East & North Africa	440	0.301***	<0.001	(0.196, 0.407)
4	East Asia & Pacific	616	0.003	0.382	(-0.004, 0.01)
5	Sub-Saharan Africa	1012	0.112*	0.058	(-0.004, 0.227)
6	Latin America & Caribbean	858	0.459***	<0.001	(0.34, 0.578)
7	North America	66	0.154	0.617	(-0.446, 0.754)

Note: *p<0.1; **p<0.05; ***p<0.01

Lastly, we stratify the countries by geographical locations because the human lifestyles heavily depend on the geography and climate. In particular, we estimate the ACC over seven continents. The results are presented in Table 3. Interestingly, we found significant ACCs (at at least the 5% level) at South Asia, Middle East & North Africa, and Latin America & Caribbean, where most developing countries reside in. This corroborates the findings in Table 2.

4 Estimating Individual Correlation Coefficients

4.1 Naive estimates: country-wise regressions

As shown in Section 2.2, the ICCs can be estimated by running a first-difference regression for each country separately. Though this naive approach is inefficient due to data scarcity, we present the result here as a benchmark.

For visualization, we plot the estimated ICCs against initial income levels, measured by GDP or per capita GDP in 1992, the initial year when the nighttime light data is available. Figure 1(a) and 1(b) present the scatter-plots of ICC estimates versus GDP and per capita GDP, respectively. Both figures show the inverse-U shapes, motivating us to fit a quadratic regression:

$$\hat{\beta}_i = \alpha_0 + \alpha_1 Z_{i,1992} + \alpha_2 (Z_{i,1992})^2 + \zeta_i, \tag{9}$$

where $\hat{\beta}_i$ denotes the estimated ICC and $Z_{i,1992}$ denotes GDP or per capita GDP. Table 4 summarizes the regression results. As expected, the coefficient on $Z_{i,1992}^2$ is negative and statistically significant, and the coefficient on $Z_{i,1992}$ is positive and statistically significant. This suggests that countries distributed in the middle range of income tend to have higher positive ICCs, while countries distributed at the lowest or highest end tend to have near-zero ICCs.

Next, we compute the p-values for each ICC estimate. Figure 2 plots the histogram of -log(p-values) for ICCs. The vertical blue, red, and green lines correspond to the $p = 0.1, 0.05, 0.01$ thresholds. Only 35 out of the 179 countries or regions have a significant ICC estimate at the 10%

level. This could be either due to the fact that the actual ICC is low or that the variance is too large. Furthermore, Figure 1(c) and 1(d) show the scatter-plots of -log(p-values) for each country versus GDP and per capita GDP in the initial year.

Table 4: Relationship between β_i and initial wealth

	Dependent variable:	
	β_i	
	(1)	(2)
(Initial GDP)^2	-0.009*	
	(0.005)	
Initial GDP	0.056*	
	(0.033)	
(Initial GDP per capita)^2		-3.657**
		(1.431)
Initial GDP per capita		1.119**
		(0.519)
Constant	0.200***	0.208***
	(0.048)	(0.046)
R^2	0.023	0.040
Adjusted R^2	0.010	0.028
Residual Std. Error (df = 160)	0.350	0.347
F Statistic (df = 2; 160)	1.842	3.328**
Note:	*p<0.1; **p<0.05; ***p<0.01	

The p-values displayed above are only marginally valid for each country. Since there are 179 p-values, we must adjust for multiplicity to control the number of false discoveries, i.e., the countries with zero ICCs that are claimed to have significant ICCs. Simply thresholding the p-values at level 10% does not guarantee the fraction of false rejections to be controlled. In particular, we apply the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995) on these p-values to control the false discovery rate (FDR). Given a target FDR level α, the BH procedure sorts the p-values in ascending order, denoted as $p_{(1)} \le \ldots \le p_{(I)}$, and rejects all p-values less than or equal to $p_{(R)}$ where

$$R = \max\left\{r : p_{(r)} \le \frac{r\alpha}{n}\right\}.$$

Unfortunately, the BH procedure cannot reject any p-value even if the target FDR level $\alpha = 1$.

This suggests that the naive estimates are so inefficient that the evidence is too weak to support any multiplicity adjustment.

Next, we apply the BH procedure in each income group separately. This is easier than the previous task because each group involves far fewer countries. Again, no p-value is rejected from any group with $\alpha = 20\%$, which is a commonly considered level. If the target level is raised to 50%, only two countries are rejected: Lao People's Democratic Republic from the lower middle income group and Qatar from the high income non-OECD group. This further confirms the inefficiency of the naive estimates.

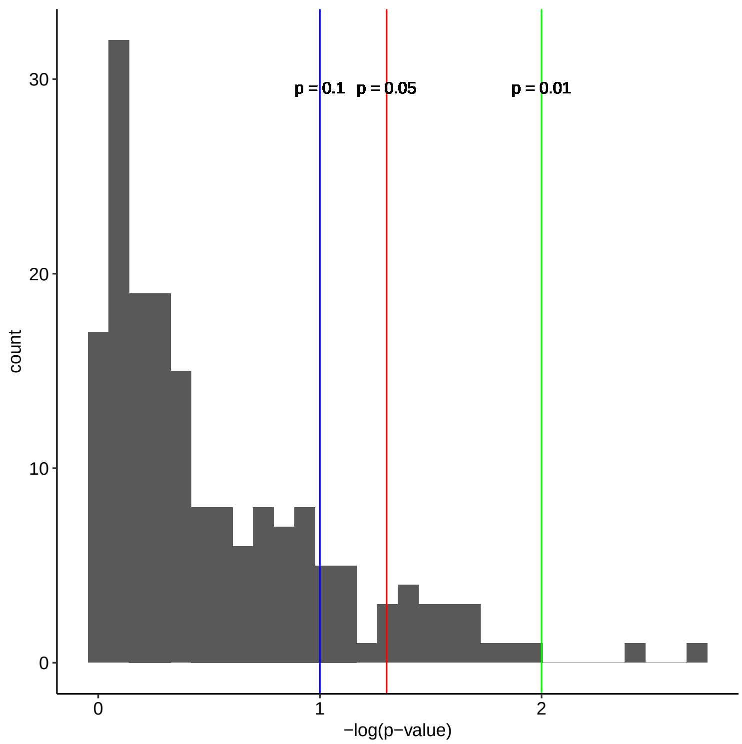


Figure 2: Histogram of -log(p-values) for 179 countries or regions

4.2 Exploiting the sparsity of ICCs via LASSO regression

In Section 4.1, we have seen that the naive estimates of β_i's are not only statistically insignificant but also have small sizes. Inspired by this observation, it is reasonable to assume the sparsity of ICCs, i.e., most β_i's are zero. The sparsity enables reliable estimation even when the number of observations used to estimate each the parameter is small (e.g. Bühlmann and Van De Geer, 2011).

Note that our model (7) can be reformulated as a generic linear model with coefficients $(\beta_1, \ldots, \beta_I)$:

$$\underbrace{\begin{bmatrix} \Delta\mathbf{L}_1 \\ \Delta\mathbf{L}_2 \\ \vdots \\ \Delta\mathbf{L}_I \end{bmatrix}}_{\Delta\mathbf{L}} = \underbrace{\begin{bmatrix} \Delta\mathbf{Y}_1 & & & \\ & \Delta\mathbf{Y}_2 & & \\ & & \ddots & \\ & & & \Delta\mathbf{Y}_I \end{bmatrix}}_{\Delta\mathbf{Y}} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \boldsymbol{\nu}_I \end{bmatrix}, \tag{10}$$

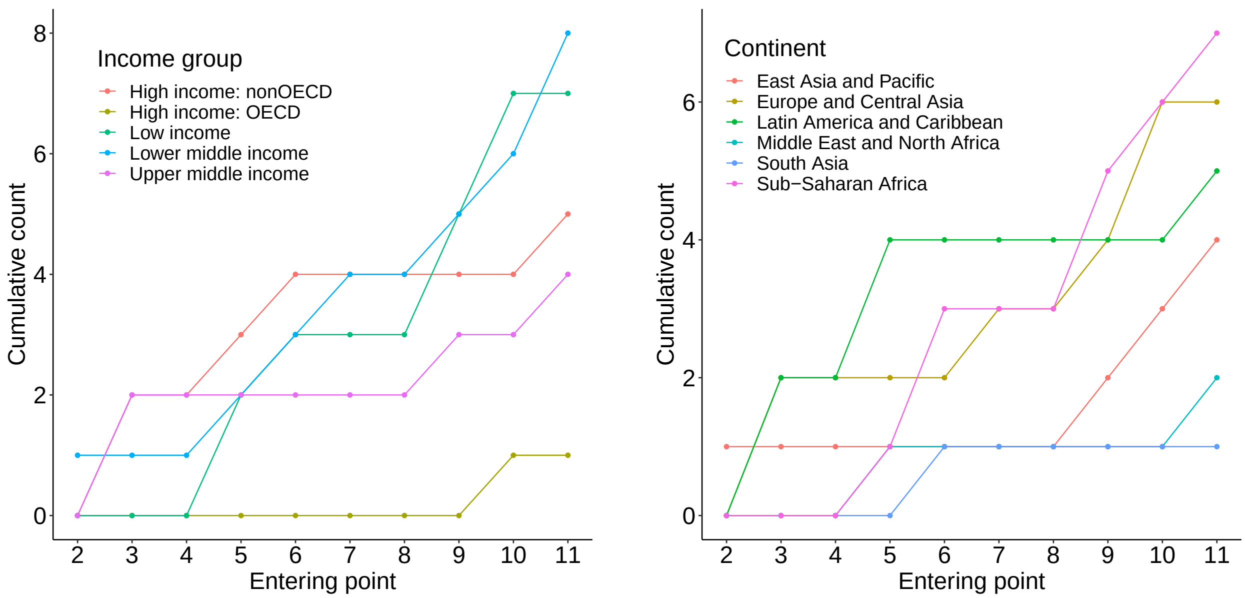
where

$$\Delta\mathbf{L}_i = \begin{bmatrix} L_{i1} \\ L_{i2} \\ \vdots \\ L_{iT} \end{bmatrix}, \quad \Delta\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix}, \quad \boldsymbol{\nu}_i = \begin{bmatrix} \nu_{i1} \\ \nu_{i2} \\ \vdots \\ \nu_{iT} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_I \end{bmatrix}$$

As a result, we can apply the LASSO regression to estimate the sparse $\boldsymbol{\beta}$. The LASSO regression estimator is defined as

$$\hat{\boldsymbol{\beta}}_{\text{LASSO}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \ \frac{1}{IT} \sum_{j=1}^{IT} (\Delta\mathbf{L}_j - \Delta\mathbf{Y}_j \cdot \boldsymbol{\beta})^2 + \lambda \sum_{i=1}^{I} |\beta_i|, \tag{11}$$

where λ is the penalty level. When $\lambda = 0$, the LASSO estimator reduces to the OLS estimator. As λ grows, the regularization term has a greater effect and thus fewer variables will enter the model. One way to sort the ICCs is based on the largest λ at which each coefficient turns non-zero. Specifically, we choose a grid of penalty levels $\lambda_1 > \lambda_2 > \ldots > \lambda_N$, which are chosen by the `cv.glmnet` function in `R`, and then solve (11) to find the set of active countries with non-zero coefficients for each λ_k. Figure 3 plots the number of active countries in each strata as λ decreases until 25 countries or regions enter the model, where the x-axis represents the entering point, namely the minimal k such that the $\hat{\beta}_i$ turns non-zero with $\lambda = \lambda_k$.



(a) (b)

Figure 3: Coefficients from LASSO

From Figure 3, to order the countries, we apply the LARS algorithm (Efron et al., 2004) which computes the exact λ at which each coefficient turns non-zero. The first 25 countries or regions that enter the model are listed in Table 5.

4.3 Controlling the fraction of false positives via the knockoff method

While the LASSO regression yields the ordering of the countries or regions that enter the model, it does not provide any statistically meaningful control of false positives. As mentioned in Section 4.1, it would be ideal to control the FDR for selected ones. Unfortunately, the BH procedure fails to reject any country when applied to the p-values obtained from the country-wise regressions because they are inefficient. Here, we apply a more advanced procedure called the "knockoff" method which can control FDR for linear models based on the more efficient LASSO estimates (Barber and Candès, 2015).

The knockoff method has gained tremendous popularity in statistics and biology since introduced. Roughly speaking, it constructs a "knockoff" variable as a negative control for each variable of interest and selects variables whose entry to the LASSO path is significantly earlier than its knockoff counterpart. The actual procedure is very complicated and we refer to the readers to Barber and Candès (2015) for a detailed description of the method. Theoretically, the knockoff method controls FDR in finite samples whenever the errors are homoscedastic and Gaussian. The performance is also shown to be robust to heteroscedasticity and non-normality.

Here, we set the target FDR level as 0.2, which is a commonly used level in multiple testing, and

Table 5: The first 25 countries or regions entering the model

Order	Country/Region	Continent	Income group
1	Rwanda	Sub-Saharan Africa	Low income
2	Grenada	Latin America & Caribbean	Upper middle income
3	Albania	Europe & Central Asia	Upper middle income
4	Greenland	Europe & Central Asia	High income: non-OECD
5	Bosnia and Herzegovina	Europe & Central Asia	Upper middle income
6	Vanuatu	East Asia & Pacific	Lower middle income
7	Fiji	East Asia & Pacific	Upper middle income
8	Gambia	Sub-Saharan Africa	Low income
9	Belize	Latin America & Caribbean	Upper middle income
10	Iceland	Europe & Central Asia	High income: OECD
11	Afghanistan	South Asia	Low income
12	Liechtenstein	Europe & Central Asia	High income: non-OECD
13	Equatorial Guinea	Sub-Saharan Africa	High income: non-OECD
14	Sierra Leone	Sub-Saharan Africa	Low income
15	Timor-Leste	East Asia & Pacific	Lower middle income
16	Cape Verde	Sub-Saharan Africa	Lower middle income
17	Slovenia	Europe & Central Asia	High income: OECD
18	Haiti	Latin America & Caribbean	Low income
19	Serbia	Europe & Central Asia	Upper middle income
20	Georgia	Europe & Central Asia	Lower middle income
21	Libyan Arab Jamahiriya	Middle East & North Africa	Upper middle income
22	Guyana	Latin America & Caribbean	Lower middle income
23	Viet Nam	East Asia & Pacific	Lower middle income
24	Madagascar	Sub-Saharan Africa	Low income
25	Armenia	Europe & Central Asia	Lower middle income

apply the `knockoff.filter` function from the `knockoff` package in `R` (Patterson and Sesia, 2018). The knockoff method rejects the first 10 countries or regions in Table 5. An informal interpretation is that at least 80% of the selected ones are true discoveries on average. Table 6 reports the selected countries or regions where we think night lights are the most accurate measure of GDP. Among the ten that demonstrate selected by the knockoff method, six belong to the middle-income group (Grenada, Albania, Bosnia and Herzegovina, Vanuatu, Fiji, and Belize), two belong to the high income group (Greenland, and Iceland), and only two belong to the low income group (Rwanda, and Gambia). This is a warning that we should be very careful when using the nighttime light data as a proxy for GDP for low-income countries or regions.

Table 6: The 10 countries or regions selected by the knockoff method with FDR control at level 0.2

Order	Country/Regiondisc	Continent	Income group
1	Rwanda	Sub-Saharan Africa	Low income
2	Grenada	Latin America & Caribbean	Upper middle income
3	Albania	Europe & Central Asia	Upper middle income
4	Greenland	Europe & Central Asia	High income: non-OECD
5	Bosnia and Herzegovina	Europe & Central Asia	Upper middle income
6	Vanuatu	East Asia & Pacific	Lower middle income
7	Fiji	East Asia & Pacific	Upper middle income
8	Gambia	Sub-Saharan Africa	Low income
9	Belize	Latin America & Caribbean	Upper middle income
10	Iceland	Europe & Central Asia	High income: OECD

5 Conclusion

We investigate the association between the nighttime light measures and reported GDP for 179 countries or regions. Our analyses overcome major limitations in previous works by including non-stationarity and heterogeneity explicitly in the model. To adjust for non-stationarity, we apply the first differencing technique and find a substantially smaller overall correlation than that found in the literature, suggesting that the latter might be attributed to ”spurious correlation”, a well-understood phenomenon in macro-econometrics. To deal with heterogeneity, we propose a weighted least square estimator for the average correlation coefficient by properly re-weighting each country, which resolves the issue of unequal weighting for the standard OLS and fixed-effect regression estimators. We find positive and significant average correlation among middle-income countries. Moving beyond the average association, we apply the LASSO regression to identify and estimate non-zero individual correlation coefficients. This is inspired by the sparsity of country-level associations observed in the preliminary analysis. We further apply the ”knockoff” method to control the false discovery rate among the selected countries. We find that the majority of countries or regions that demonstrate a strong and significant association between nighttime light and GDP belongs to the middle-income group. This is a warning that the light-GDP association is not universally high, and we should be very careful when using the nighttime light data as a proxy for GDP for low-income countries or regions.

References

Barber, R. F. and Candès, E. J. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5):2055–2085.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

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

Cheung, Y.-W. and Lai, K. S. (1995). Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics*, 13(3):277–280.

Croft, T. A. (1978). Nighttime images of the earth from space. *Scientific American*, 239(1):86–101.

Deaton, A. and Heston, A. (2010). Understanding ppps and ppp-based national accounts. *American Economic Journal: Macroeconomics*, 2(4):1–35.

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

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, 32(2):407–499.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., and Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379.

Ernst, P. A., Rogers, L., and Zhou, Q. (2019). The distribution of yule's" nonsense correlation". *arXiv preprint arXiv:1909.02546*.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the penn world table. *American Economic Review*, 105(10):3150–3182.

Ghosh, T., Anderson, S., Powell, R. L., Sutton, P. C., and Elvidge, C. D. (2009). Estimation of mexico?s informal economy and remittances using nighttime imagery. *Remote Sensing*, 1(3):418–444.

Ghosh, T., L Powell, R., D Elvidge, C., E Baugh, K., C Sutton, P., and Anderson, S. (2010). Shedding light on the global distribution of economic activity. *The Open Geography Journal*, 3(1).

Granger, C. W. and Newbold, P. (1974). Spurious regressions in econometrics. *Journal of econometrics*, 2(2):111–120.

Harari, M. (2020). Cities in bad shape: Urban geometry in india. *American Economic Review*, 110(8):2377–2421.

Henderson, J. V., Squires, T., Storeygard, A., and Weil, D. (2018). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics*, 133(1):357–406.

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

Hodler, R. and Raschky, P. A. (2014). Regional favoritism. *The Quarterly Journal of Economics*, 129(2):995–1033.

Hu, Y. and Yao, J. (2021). Illuminating economic growth. *Journal of Econometrics*.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Libanio, G. A. (2005). Unit roots in macroeconomic time series: theory, implications, and evidence. *Nova Economia*, 15(3):145–176.

McCallum, B. T. (1993). Unit roots in macroeconomic time series: Some critical issues.

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

Patterson, E. and Sesia, M. (2018). *knockoff: The Knockoff Filter for Controlled Variable Selection*. R package version 0.3.2.

Phillips, P. C. (1998). New tools for understanding spurious regressions. *Econometrica*, pages 1299–1325.

Proville, J., Zavala-Araiza, D., and Wagner, G. (2017). Night-time lights: A global, long term look at links to socio-economic trends. *PloS one*, 12(3):e0174610.

Stock, J. H., Watson, M. W., et al. (2012). *Introduction to econometrics*, volume 3. Pearson New York.

Sutton, P. C. and Costanza, R. (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.

Sutton, P. C., Elvidge, C. D., Ghosh, T., et al. (2007). Estimation of gross domestic product at subnational scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, 8(S07):5–21.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Young, A. (2012). The african growth miracle. *Journal of Political Economy*, 120(4):696–739.

Zeileis, A. (2021). *pwt10: Penn World Table (Version 10.x)*. R package version 10.0-0.

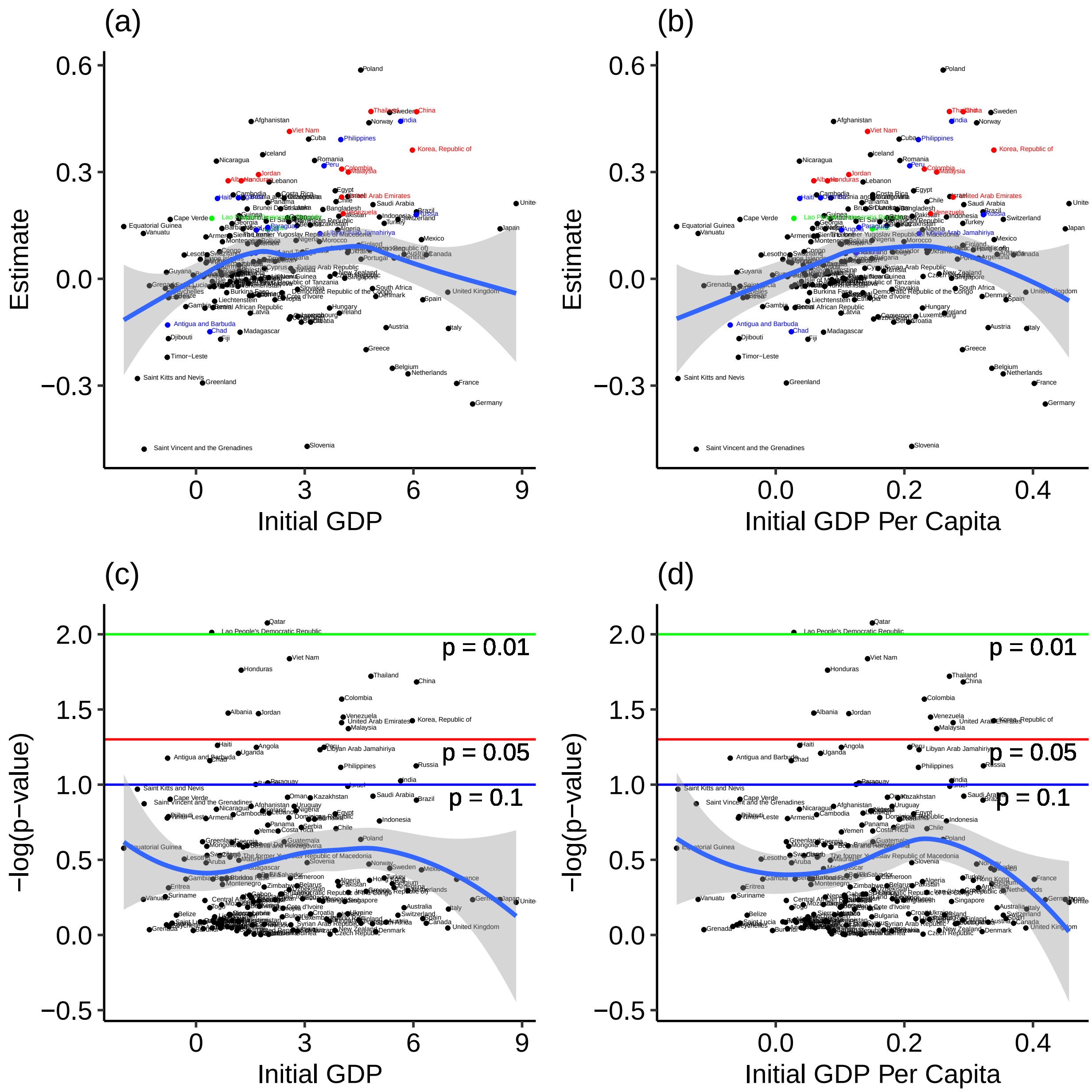


Figure 1: Estimated ICC over initial income levels and corresponding p-values