CREDIT Research Paper

No. 18/04

Blinded by the Light? Heterogeneity in the Luminosity-Growth Nexus and the 'African Growth Miracle'

by

Lionel Roger

Abstract

Night-time light emissions are a popular proxy for growth in circumstances where official data are deemed unreliable. We show that the underlying relationship varies substantially across countries, undermining the imposition of a single slope common in the literature. We propose a two-step method to improve country-specific growth estimates informed by night-light data, making use of a machine-learning algorithm to discern factors driving differences in the luminosity-growth elasticity across countries. The improved performance of this strategy over existing approaches is established in a number of simulation exercises. Applied to African data between 1992 and 2013 we find little evidence of an `African Growth Miracle' undetected by official statistics, as suggested by Young (2012); instead, we observe that countries which recently revised their GDP figures tend to report substantially inflated growth rates over recent years, in line with Jerven (2014)'s hypothesis of purely `statistical growth'.

JEL Classification: E01, N17, O40

Keywords: Nigh lights, Economic Growth, African Growth Miracle

Centre for Research in Economic Development and International Trade, University of Nottingham

CREDIT Research Paper

No. 18/04

Blinded by the Light? Heterogeneity in the Luminosity-Growth Nexus and the 'African Growth Miracle'

by

Lionel Roger

Outline

1. Introduction
2. Literature review
3. Methodology: Addressing heterogeneity in the relationship between GDP and lights
4. Simulation exercise
5. Data
6. Heterogeneity in the relationship between GDP and lights
7. Explaining the relationship between GDP and lights
8. Application: Economic growth in Africa, 1992-2013
9. Conclusion

References

Appendix

The Author

PhD candidate at the University of Nottingham. Email: lionel.roger@nottingham.ac.uk

Acknowledgements

I thank Oliver Morrissey and Markus Eberhardt for their support throughout the research process. This paper is part of the work for my PhD, which was supported by the Economic and Social Research Council [grant number 1511835].

Research Papers at www.nottingham.ac.uk/economics/credit/

1 Introduction

In 2011, Shanta Devarajan, then the World Bank's Chief Economist for Africa, declared 'Africa's statistical tragedy' (Devarajan, 2013): While the continent finally seemed to have overcome its 'growth tragedy' (Easterly and Levine, 1997), the trouble was that no one could actually be quite sure about it. Poor statistical capacity, paired with the possibility of politically motivated misreporting of economic performance led to widespread concerns about the reliability of GDP figures, especially from sub-Saharan African countries (Jerven, 2013). The amplitude of the problem is prominently illustrated by the case of Nigeria, where, after an extensive revision of the statistical office's methodology, GDP figures were revised upwards by 60% overnight. Indeed, since 2000, at least 14 African countries have performed a similar move, with revisions averaging around +20% of the initial GDP estimates (NNBS, 2013).

While improving future estimates of economic performance does come with its own challenges (financial, institutional and political), it is a relatively straightforward task from a technical point of view: compliance with international reporting standards, more comprehensive and regular surveys across the economy, and appropriate processing of the data can be implemented where there exists political will and resources are made available. What is inherently more difficult to rectify, however, is information on past growth performances: If Nigeria is now 60% richer than we used to think, when exactly did this growth come about? Was there indeed an 'African Growth Miracle' in the 1990s and 2000s that escaped traditional measurement methods, as implied by Young (2012)? Or did African countries already start into the 1990s with a substantially higher level of income, and much of the growth that has been attributed to recent periods had actually already happened much earlier in history (e.g., Jerven, 2015)?

The fundamental issue here is that any effort to correct past growth rates using conventional methods will be likely to suffer from at least some of the same deficiencies that also undermined the reliability of the original estimates: Surveys of the economy – provided there is a will for them to be made available – may not have been regularly collected. If beliefs about the structure of the economy at the time of data collection were false, the way in which samples were constructed may not have been representative of the actual structure of the economy. Historical tax records may be inaccurate and by construction only cover the formal sector, which in low-income countries typically accounts only for a small share of the economy. And in case there has been any politically motivated tinkering with the data, data recovered from statistical archives may itself have been subject to manipulation. Statistically speaking, it seems impossible to obtain estimates of past growth rates with errors that are uncorrelated with those of the originally reported growth rates.

Perhaps the most promising and innovative avenue to address this problem is the use of historical satellite imagery, specifically man-made light emissions at night. Several studies have established that there exists a high correlation between economic activity and luminosity, and that lights can therefore serve as a surprisingly powerful proxy of GDP (e.g., Chen and Nordhaus, 2011; Ghosh et al., 2010). Crucially, it can convincingly be argued that the measurement error in light emissions at night is unrelated to the error associated with the figures reported by national statistical offices. Henderson et al. (2012) offer a statistical framework that exploits this property in order to construct estimates of GDP growth with reduced overall measurement error. Combining growth estimates based on luminosity and reported GDP data, they present revised average growth rates for the period between 1992 and 2006 for countries with low statistical capacity. While they cannot discern any systematic pattern of over- or under-reporting in these countries, the suggested revisions are very large for individual economies: In Nigeria and Angola, their growth estimates based on luminosity suggest annual growth rates about half as high as the official ones: 1.92% instead of 4.04% in Nigeria, and 3.88% instead of 6.99% in Angola. On the other hand, some notoriously disappointing growth performances are radically corrected upwards: In Côte d'Ivoire, the Democratic Republic of Congo, and Burundi, the change in luminosity suggests growth rates about 3 percentage points higher than official figures. Based on assumptions about the amplitude of the measurement error in official data, the authors then report weighted averages of the official growth rates and those obtained from luminosity. For countries with supposedly bad data, they suggest an optimal weight of about 50% for the luminosity measure.

The argument put forward in this paper is that the relationship between luminosity and GDP varies across countries, contrary to one of the central assumptions of the literature that exploits this relationship. Under these circumstances, the suggested revisions for individual countries may be highly misleading. To illustrate this, consider Poland and Thailand, two countries that are very different in many respects, but with a similar growth performance over the past two decades and data that is generally considered reliable: While in both countries, GDP approximately increased by 230% between 1992 and 2013, the luminosity emitted by Poland increased by a factor of 2.5, while that of Thailand increased by a factor of almost 5. One possible explanation is of course that at least one of these countries has vastly mis-measured or mis-reported their GDP figures. Perhaps a simpler explanation is that, for some reason, GDP and luminosity interacted differently in these places. For instance, Poland started the period with a much higher level of income, it has a very different location (nearer the pole), a different economic structure, more people living in urban areas, etc.

The aim of this study is to obtain estimates of past GDP growth rates that take into account the heterogeneity in the relationship between luminosity and GDP that can be explained based on such country characteristics. To this end, we introduce an approach that consists in two steps: In a first step, we will estimate country-specific coefficients determining the relationship between GDP and lights[1] – that is, our regression will allow for slopes to vary freely between countries. In a second step, we seek to put some structure in the relationship between luminosity and GDP: we now treat the coefficients obtained in the first exercise as the dependent variable, and seek to identify the part of their variation that can be explained by observable country-characteristics. This allows us to make predictions of GDP based on luminosity values that take into account cross-country heterogeneity in the relationship between these variables, to the extent that this heterogeneity can be traced back to generalisable relationships.

Intuitively speaking, the reasoning behind this two-step procedure is that, while the first step (estimation of country-specific slopes) does indeed allow for full heterogeneity in terms of lights-GDP elasticity across countries, the resulting estimates have no value when it comes to revising the reported growth rates. Instead, any predictions of GDP growth based on lights using the parameters estimated in the first step simply tend to replicate the reported growth rates, whether these were mis-reported / mis-measured or not. The second step serves the purpose of identifying the systematic part of the variation between them. We can then predict *expected* elasticities between GDP and lights conditional on any country's characteristics. These conditional elasticities will then form the basis for our predicted GDP growth rates: By allowing for elasticities to vary between countries based on their characteristics, we acknowledge the fact that luminosity doesn't behave the same in every country of the world, and avoid suggesting large revisions where these are unwarranted. But by restricting the heterogeneity to the systematic component that is generalisable across countries, and based on 'legitimate' factors, we avoid the simple replication of potentially misreported growth rates.

One fundamental difficulty we encounter when attempting to discern structure in the elasticities between lights and GDP is that, unlike for most well-researched economic phenomena (say, economic growth), there does not exist any theoretical framework to offer guidance as to which variables may matter, and in what way.

[1] To clarify: Our study is concerned with GDP *growth*, not levels of GDP. For the latter, lights are a relatively poor predictor, that is, they do not predict cross-country variation well (Henderson et al., 2012). We do, however, predict relative GDP levels *within* countries, from which growth rates are then derived. See section 3.2 for details.

That is, any attempt to specify a model that is parsimonious from the outset would necessarily be based on relatively arbitrary decisions, and risk the omission of important predictors. However, in view of predictive accuracy, a certain degree of parsimony is necessary in order to avoid issues of over-fitting. To overcome this problem, we employ elastic net regularisation, and leave-one-out (LOO) cross-validation as a means of obtaining optimal predictive properties. The intuition behind this is simple: At the core, the elastic net is the ordinary least squares estimator (OLS), augmented with penalty terms that have the role of shrinking parameters (or to drop variables, where parameters are shrunk to zero). The severity of these penalty terms is determined by parameters (α and λ) that can be adjusted (*tuned*), typically with the aim of finding the model that yields the most accurate predictions. The role of cross-validation is then to iteratively divide the existing sample into a training-set (used to estimate the model) and a test-set (used to test its predictions), so as to assess every model's *out-of-sample* predictive power. In other words, it finds the model that is most *generalisable* beyond the narrow sample the estimator is presented with. We exploit this property in order to identify the share of the variation in the estimated lights-GDP coefficients that can credibly be attributed to observable country characteristics, and, crucially, is not a result of mis-reported or mis-measured GDP data.

As the asymptotic and finite sample properties of our procedure are unknown *a priori*, we test and demonstrate its potential on simulated data. We create a setup in which some countries measure and report their GDP accurately, while others systematically misreport it. Luminosity series based on the *true* growth rates are then generated for every country. Crucially, we introduce variation across countries in the elasticity at which GDP induces luminosity. This variation is driven by a range of variables, some of which are simultaneously determinants of GDP growth. The results suggest that, indeed, our suggested methodology has the potential to improve the accuracy of GDP growth estimates based on luminosity. This is mainly because, as opposed to estimators that assume a single elasticity for all countries, our approach is less prone to suggesting revisions where the reported data was correct in the first place. When applying our methodology to the real data, this property appears to hold as well: Our estimator suggests substantially smaller revisions to growth rates where the official data is recognised to be more reliable (as suggested by the Penn World Table's grades of data quality, ranging from A to E). The alternative estimator that imposes a single slope applies larger corrections throughout country grades, suggesting equally large revisions for the countries with the supposedly worst data quality as for those with the best.

We then use our estimates to assess the growth performance of sub-Saharan African countries between 1992 and 2013, and derive estimates of GDP growth rates based on changes in luminosity for each of the African countries in our sample. We find no indications of an 'African Growth Miracle' that would have gone unnoticed by official statistics, as implied by Young (2012): While official growth figures and luminosity based estimates diverge substantially for individual countries, there does not appear to be any particular directionality in this discrepancy. On aggregate across the continent, the official growth rates are quite well aligned with those obtained from the luminosity proxy. On the other hand, it appears as though countries that recently revised their GDP figures after a prolonged period of time (since 2000 and after at least 10 years; this applies to Botswana, Ethiopia, Ghana, Niger and Nigeria) had a tendency to report inflated growth rates for recent years. Between 2003 and 2013, these countries reported annual growth rates of 7.5% on average, while our luminosity based estimates suggest an average growth rate of 5.4%. This pattern is consistent with Jerven (2014)'s hypothesis of 'statistical growth'. According to this narrative, recent improvements in statistical capacity would have revealed a lot of economic activity that had previously gone unnoticed by statistical authorities, and spuriously been accounted for as recent economic growth. Finally, we consider growth rates of individual countries across three sub-periods. The general pattern here is that the most extreme growth performances – stellar growth episodes in individual countries or cataclysmic recessions – tend to appear more moderate seen through the luminosity proxy.

We do consider our results to be an improvement over existing lights-based estimates of growth rates, but emphasise that – like all results in this literature – they are best considered to be of suggestive nature. Lights are inevitably an imperfect proxy for economic activity, and idiosyncratic differences in the relationship between lights and GDP beyond what we are able to discern in this study, paired with randomness due to measurement error, may be driving substantial shares of the remaining discrepancy between our predictions and reported growth rates. The type of conclusion that we wish to draw from this exercise is not that country X should revise its historical growth rates by precisely amount Y. Instead, we consider our results suggestive of broad patterns, and as a valuable complement to other sources of information, such as traditional types of data and historical records. Furthermore, we stress the importance of the observation that a substantial share of the discrepancy between reported growth rates and those inferred from luminosity can be closed once we take into account the factors that drive this relationship: National accounts data may come with deficiencies, but the risk of overstating those deficiencies is as real as the risk of ignoring them.

The remainder of this paper proceeds as follows: section 2 discusses the previous literature on night lights. Section 3 describes our methodology. Section 4 applies our methodology to simulated data in order to test and illustrate its basic properties. The empirical data we employ and any transformations to it are described in section 5. Section 6 substantiates the observation that there exists a sizeable amount of heterogeneity between countries in the lights-GDP relationship. In section 7, we bring the method to the actual data, and in section 8 we apply our estimates to recent growth rates from African countries. Section 9 concludes.

2 Literature review

Night time luminosity as a proxy for economic activity is generally employed for two purposes: First, because of the resolution at which the data is available, it allows researchers to move beyond the usual administrative units (typically countries) when investigating the determinants and dynamics of economic development (e.g., Alesina et al., 2012; Michalopoulos and Papaioannou, 2013; Hodler and Raschky, 2014). Second, as it is inherently independent of the measurement or reporting error by statistical agencies, it has a potential to augment these statistics and to reduce overall measurement error (Chen and Nordhaus, 2011; Henderson et al., 2012). It is the latter type of application this study is concerned with. This section will provide a brief overview of the relevant literature.

To the best of our knowledge, the link between economic activity and luminosity emitted into space was first noted by Croft (1978). However, the data, collected by weather satellites of the National Oceanic and Atmospheric Administration (NOAA), were not systematically stored and made available for research until relatively recently. The first studies that systematically investigated and exploited the lights-GDP relationship therefore only emerged in the early to mid-2000s, with Sutton and Costanza (2002), Ebener et al. (2005), Doll et al. (2006) and Sutton et al. (2007) each aiming to provide estimates of incomes at the sub-national level. Ghosh et al. (2009, 2010), while still devoting much of their analysis to sub-national distributions of economic activity, shift the focus to augmenting reported national accounts data; in particular, they provide luminosity based estimates that are meant to incorporate informal economic activity not reflected in official statistics.

A more formalised statistical treatment followed by Chen and Nordhaus (2011) and Henderson et al. (2012). In both studies, the central aim is to augment official income data at the country level with the information derived from luminosity. In the presence of measurement error in official income (alternatively, output) data, and with luminosity as a strong predictor of incomes with errors uncorrelated to those in official statistics, there must be some linear combination of official figures and those derived from lights emissions that minimises the overall measurement error. In Henderson et al. (2012)'s notation:

$$\hat{y}_i = \lambda z_i + (1 - \lambda)\hat{z}_i$$

where z_i are the reported official growth figures, $\hat{z}_i$ is the luminosity based proxy, and $\hat{y}_i$ is the new synthetic measure of income (output) growth in country (or region) i. The weight on the luminosity based proxy, $(1 - \lambda)$, must then be chosen such that, on expectation, the overall measurement error is minimised.

In their version of the analysis, Chen and Nordhaus (2011) derive such optimal weights both for output levels and long-term growth rates, separately for countries with different levels of statistical quality (A–E as classified in older versions of the Penn World Tables (PWT)). Their findings suggest that luminosity adds value for countries with low statistical quality (ratings D and E), and that its value (as quantified by the optimal weights $(1 - \lambda)$, or θ in their notation) is typically higher when considering long-run growth rates rather than output levels. The optimal weights they derive for the luminosity proxy reach just over 30% for long-run growth rates in countries rated D, and less for other countries. Henderson et al. (2012) focus on growth rates entirely. Instead of using the A–E rating from the PWT, the authors rely on the World Bank's rating of statistical capacity and divide countries into only two categories, *good data* and *bad data* countries (*bad data* being defined as scoring less than 3 out of 10 in the statistical capacity rating). The optimal relative weights of lights and GDP for good and bad data countries crucially rely on assumptions made about the signal to variance ratio in reported GDP growth rates from *good data* countries. The authors' preferred choice, assuming 90% signal to variance in good data countries, implies a weight of about 52% for lights for the 30 countries they qualify as having bad data, and still 15% for those with good data.

Based on these weights, Henderson et al. (2012) compute the optimal GDP growth rates for the 30 *bad data* countries in their sample (optimal in the sense of minimising the expected total measurement error). The first important observation is that they do not find evidence of systematic over- or under-reporting of growth rates across *bad data* countries. Instead, the discrepancies between official growth rates and the lights-proxy average around zero. However, many of the revisions suggested by their results are substantial: In Burundi, where official figures suggest a negative average growth rate of -0.71% annually between 1992 and 2006, the lights proxy suggests positive growth of 2.89% annually; weighted optimally, the suggested average annual growth rate is 1.13%. At the other end of the spectrum (and focussing on the African countries in the sample), Angola's growth rate is revised downwards from 6.99% annually to 5.38% (with lights suggesting 3.88%), and Nigeria's from 4.04% to 2.94% (with lights suggesting 1.92%).

A handful of studies have since been dedicated to gaining a deeper understanding of the relationship between economic activity and lights. Using county-level GDP series from Brazil, India, the United States and Western Europe, Bickenbach et al. (2016) show that at a sub-national level, there exists large variability in the relationship between lights and GDP. Within all of the economies they consider, they find that the estimated elasticity between lights and GDP differs substantially across regions. For instance, when regressing growth in lights on growth in GDP, they report a small but significant positive coefficient on lights of 0.09 in East India, but a *negative* significant one in West India (-0.14). For Brazil, their estimates across five regions range between 0.37 (Centro Oeste) and 0.04 (Sul). Similar discrepancies are observed in the USA and Western Europe, where the reported GDP data are likely to be particularly accurate. The authors explore several avenues to restoring parameter consistency (mainly through the inclusion of a range of control variables, and restricting the sample in a number of ways), but none of these succeed. They conclude that, at the regional level, luminosity does not appear to be a useful proxy of economic growth. However, as the authors note, their findings do not necessarily translate to the national level, where parameters may still be constant.

Wu et al. (2013) explicitly seek to identify factors affecting the relationship between lights and GDP at the country level. Their analysis differs in one important aspect from the bulk of the literature: Instead on regressing lights on GDP, which studies that are concerned with proxying for GDP typically do, they employ lights as the dependent variable and regress a number of potential determinants on it. Their choices of variables derive from a rudimentary theoretical framework that models lights as a normal consumption good. On this basis they hypothesise that, beyond incomes, luminosity may be affected by factors such as the share of agriculture, savings, and latitude. Empirically, their model translates into a standard linear regression with luminosity on the left hand side, and each of its hypothesised determinants (including GDP) on the right hand side. GDP per capita, latitude, the degree of spatial agglomeration and the savings rate are found to be significant determinants of luminosity. Additionally, they separate GDP into its agricultural and non-agricultural component, finding that non-agricultural production is a much stronger driver of luminosity than agricultural production.[2] In a conceptually similar way, Levin and Zhang (2017) investigate the correlates of luminosity in urban areas using the newer and higher resolution VIIRS luminosity

[2]Keola et al. (2015) make a similar observation about agriculture not being reflected as distinctly by luminosity as other sectors, but to a more problematic degree than that observed by Wu et al. (2013). In order to overcome the issue, they suggest combining luminosity data with data on landcover, where growth in agricultural production is approximated based on changes in cultivated land.

data.[3] They identify numerous variables (anthropogenic and physical) that affect the luminosity of urban areas. The strongest and most persistent correlates beyond incomes are found to be snow cover (especially during winter months), rents from fossil fuels, latitude and road density.

Studies of this type deliver useful insights as to what drives luminosity beyond only GDP; it is important to highlight, however, that effectively they investigate the determinants of *luminosity*, as opposed to the determinants of the *relationship* between luminosity and GDP. The study at hand seeks to specifically address the latter. Furthermore, beyond simply exploring factors that may alter the way in which economic growth translates into changes in luminosity, we will offer an empirical framework to improve the predictive power of the lights-based GDP proxy in the presence of this type of heterogeneity.

3 Methodology: Addressing heterogeneity in the relationship between GDP and lights

The literature that exploits luminosity as a proxy for economic activity typically assumes a relationship (implicitly or explicitly) where economic activity generates lights at some constant elasticity (e.g., Chen and Nordhaus, 2011; Henderson et al., 2012; Keola et al., 2015). This elasticity is generally assumed to be the same across all countries, an assumption this study aims to relax.

3.1 Conceptual framework

Our version of the framework therefore describes a basic relationship where economic activity (Y) generates luminosity (L) at some *country-specific* elasticity β_i:

$$L_{it} = Y_{it}^{\beta_i} * exp(\varepsilon_{it}).$$

The relationship is perturbated by an error term ε_{it}, which we note for later has the structure $\varepsilon_{it} = \epsilon_i + \epsilon_t + \epsilon_{it}$, that is, the error term is a composite of time-specific perturbations in year $t = 1, ..., T$ (e.g., because of degrading and replaced lights-sensors over time), country-specific perturbations in country $i = 1, ..., N$ (e.g., different levels of light emission at the baseline), and an idiosyncratic error term ϵ_{it}. Generally, $E[\varepsilon_{it}] = E[\epsilon_i] = E[\epsilon_t] = E[\epsilon_{it}] = 0$. The country-specific elasticity

[3] The recently launched VIIRS series is superior to the DMSP-OLS series employed in this study in a number of respects (Dai et al., 2017). However, data is only available from 2012 onwards, making the investigation of historical growth rates impossible.

β_i quantifies the extent to which economic activity translates into luminosity for each country. Taking logarithms, equation 3.1 can be rewritten in linear form:

$$\ell_{it} = \beta_i y_{it} + \varepsilon_{it}$$

with $\ell = ln(L)$ and $y = ln(Y)$. Since we are interested in predicting GDP from luminosity, we rearrange this to

$$y_{it} = \gamma_i \ell_{it} - \gamma_i \varepsilon_{it}, \qquad \gamma_i = \frac{1}{\beta_i}.$$

Since $E[\varepsilon_{it}] = 0$, and ℓ_{it} is fixed, $E[Y_{it}] = exp(\gamma_i \ell_{it})$, for some given inverse elasticity γ_i in country i. This relationship can then be exploited in order to approximate (true) GDP Y in places where there is uncertainty whether the reported data is accurate, or where such data is unavailable. Note that the focus lies in growth rates (i.e., relative changes of Y over time *within* a given country), so the absence of an intercept will have no bearing on the results of interest. We consider the inverse elasticities γ_i to be a function of a set of determinants that vary at the country level:

$$\gamma_i = f(\Phi_i, \Psi_i, \eta_i)$$

where Φ_i are observable determinants, Ψ_i are unobservable determinants, and η_i is a random variable.[4] We deliberately do not attribute a specific functional form to $f(.)$, reflecting the fact that there is little theoretical understanding on how exactly luminosity and GDP interact, and that, from an economic point of view, there is little interest in describing it *per se*. Instead, our interest lies in exploiting the aggregate manifestation of this relationship for predictive purposes. To this end, we will seek to capture the share of the variation in γ that can be attributed to the observables Φ, while remaining widely agnostic about the underlying mechanisms.

3.2 Empirical approach: A procedure in two steps

The broad strategy that we adopt consists in two steps: First, as γ_i are unknown, we obtain estimates $\hat{\gamma}_i$ from a fixed effects regression that allows for country-specific slopes (equation 1 below). Second, we employ the flexible elastic net estimator to discern the part of the variation in the $\hat{\gamma}_i$s that can be attributed to observables (Φ_i) in a way that is generalisable across countries. This will allow

[4] For all intents and purposes, Ψ_i and η_i can be considered the same thing. The differentiation is merely to indicate that we neither believe that we can empirically capture all factors that do affect the relationship between lights and GDP, nor that it is entirely deterministic.

us to predict *expected* inverse elasticities $\tilde{\gamma}_i$, conditional on observable country characteristics: $\tilde{\gamma}_i = E[\gamma_i|\Phi_i]$. These can, in turn, be used to predict Y_{it}, on the basis of which we can derive estimated growth rates.

Step 1: Estimating naive country-specific slopes ($\hat{\gamma}_i$): We obtain our first stage estimates, $\hat{\gamma}_i$ by estimating the following equation using least squares:

$$z_{it} = \sum_{i=1}^{N} \gamma_i(\ell_{it} \times D_i) + \alpha_i + \alpha_t + \varepsilon_{it} \tag{1}$$

where z_{it} and ℓ_{it} are the logs of reported GDP figures and luminosity in country i at time t respectively, D_i are country dummies, and α_i and α_t are country and year fixed effects. The time fixed effects are mainly necessary in order to account for differences in sensor sensitivity across years (see data description in section 5). The country fixed effects serve to absorb differences in levels across countries that are constant over time, such as different baseline levels of brightness and differences in units (GDP will be included in local currency units). In order to estimate the parameters $\hat{\gamma}_i$, we therefore exploit the temporal variation of luminosity and GDP *within* each country.

Step 2: Explaining $\hat{\gamma}_i$ using the elastic net: The main challenge in our second step is to select and weight relevant predictors of $\hat{\gamma}$ from a large set of candidate variables. At the core, this is because of the inherent lack of knowledge about $f(.)$: we have very little guidance when it comes to selecting the variables to include in this predictive exercise (potential elements of Φ, labelled Φ^*). Econometrically, this creates a situation with a large number of variables compared to the number of observations (N countries). In order to avoid over-fitting – that is, in order to identify the determinants that apply in a generalisable manner and beyond our narrow sample – we make use of elastic net regularisation (Zou and Hastie, 2005). The elastic net estimator is suited to situations with large amounts of explanatory variables compared to the number of observations, and has been shown to perform very well in selecting the relevant predictors. It minimises the following criterion:

$$L(\lambda_1, \lambda_2, \boldsymbol{\delta}) = |\hat{\boldsymbol{\gamma}} - \boldsymbol{\Phi^*}\boldsymbol{\delta}|^2 + \lambda_2|\boldsymbol{\delta}|^2 + \lambda_1|\boldsymbol{\delta}|_1 \tag{2}$$

with

$$|\boldsymbol{\delta}|^2 = \sum_{j=1}^{p} \delta_j^2,$$

$$|\boldsymbol{\delta}|_1 = \sum_{j=1}^{p} |\delta_j|$$

.

where $\boldsymbol{\delta}$ is the vector of coefficients (weights) attributed to the elements of $\boldsymbol{\Phi^*}$, which in turn is the set of potential determinants of γ we take into consideration.[5] $\hat{\boldsymbol{\gamma}}$ is the dependent variable, that is, a vector containing the N $\hat{\gamma}_i$s obtained from estimating equation 1 above. Equation 2 is best discussed by considering the terms seperately. The first term, $|\hat{\boldsymbol{\gamma}} - \boldsymbol{\Phi^*}\boldsymbol{\delta}|^2$ is the square of the residuals, reflecting the same minimalisation problem as in OLS. The second term, $\lambda_2|\boldsymbol{\delta}|^2 = \lambda_2 \sum_{j=1}^{p} \delta_j^2$, is a penalty term, penalising for large coefficients. It is in fact the same term as in a ridge regression (Hoerl and Kennard, 1970). The third term, $\lambda_1|\boldsymbol{\delta}|_1 = \lambda_1 \sum_{j=1}^{p} |\delta_j|$, is another penalty term, and corresponds to the one employed in a lasso regression (Tibshirani, 1996). In fact, Zou and Hastie (2005) show that the minimisation of $L(\lambda_1, \lambda_2, \boldsymbol{\delta})$ can be expressed as the optimisation problem

$$\hat{\boldsymbol{\delta}} = arg\ \min_{\boldsymbol{\delta}} |\hat{\boldsymbol{\gamma}} - \boldsymbol{\Phi^*}\boldsymbol{\delta}|^2, \qquad s.t.\ \alpha|\boldsymbol{\delta}|_1 + (1-\alpha)|\boldsymbol{\delta}|^2 < s \text{ for some } s \tag{3}$$

with $(1-\alpha) = \lambda_2/(\lambda_1 + \lambda_2)$. In that sense, the elastic net penalty is a convex combination of the lasso and the ridge penalty, with α determining the weight between the two. In fact, both the lasso and the ridge estimator are nested in the elastic net, as $\alpha = 1$ corresponds to the lasso, while $\alpha = 0$ corresponds to ridge. Both α and λ (or, via algebraic detours, s) will then be calibrated using n-fold cross-validation in order to minimise the out of sample prediction error (as measured by the root mean squared error; see appendix F for a more detailed discussion).

The elastic net estimator is particularly well suited for our application for a number of reasons. First, it performs particularly well in the presence of large numbers of predictors compared to the number of observations; in our empirical application, we will include 57 variables for 129 observations (cf. section 7). Second, as opposed to ridge, it has the lasso's characteristic of selecting variables (setting weights of poor predictors to zero). This facilitates the interpretation of the output.[6] Third, compared to the lasso, the results and predictions from the elastic net tend to be more stable, especially if some of the independent variables are strongly correlated. Fourth, lasso and ridge – the main contenders for the choice of estimator – are nested in it, so if any of them is superior in terms of

[5]Ideally, the weights $\delta = 0$ for elements of Φ^* that are not part of Φ.

[6]Although we do not focus on the interpretation of the output *per se*, as our focus is on predictive power. See Mullainathan and Spiess (2017) for a discussion of the interpretability of output from machine learning procedures in general.

out-of-sample predictive power, it will be selected ($\alpha = 0$ or $\alpha = 1$).[7]

Based on the estimated weights $\boldsymbol{\delta}$ and the known values of $\boldsymbol{\Phi^*}$, we can then obtain our predicted parameters $\tilde{\gamma}_i = E[\gamma_i|\Phi_i^*]$ and fit model 1 using $\tilde{\gamma}_i$ rather than $\hat{\gamma}_i$. The result is, for each country, a full time series of predicted GDP values based on luminosity emissions, where luminosity enters with a weight that is consistent with the country's characteristics ($\tilde{Y}_{it} = E[Y_{it}|\Phi_i^*, \ell_{it}]$). Growth rates are then derived from the annual changes in these predicted values in the usual manner ($\tilde{g}_{it} = \frac{\tilde{Y}_{it} - \tilde{Y}_{it-1}}{\tilde{Y}_{it-1}}$).

3.3 Alternative estimators

While the main focus of our analysis lies on the results obtained using the two-step procedure outlined above, we do use alternative modes of estimation for descriptive purposes, for comparability with the earlier literature, and as a robustness check. We briefly introduce and discuss the respective equations in this subsection, all of which will be estimated using least squares.

In order to explore some patterns of the heterogeneity of γ in section 6, we will divide our sample into G groups that are relatively homogeneous with respect to selected characteristics. We then impose a single slope coefficient within each of these groups, estimated via

$$z_{it} = \sum_{i=1}^{G} \gamma_g(\ell_{it} \times D_g) + \alpha_i + \alpha_t + \varepsilon_{it}, \tag{4}$$

where D_g are dummy variables indicating membership to group g. We also estimate a conventional fixed effects model that imposes a single inverse elasticity γ for all countries:

$$z_{it} = \gamma x_{it} + \alpha_i + \alpha_t + \varepsilon_{it}. \tag{5}$$

This is the model employed by the bulk of the literature, including Henderson et al. (2012) in their baseline estimations. Throughout the paper, equation 5 will be treated as our main benchmark specification. As a robustness check and for comparability, our results in section 8 will also be derived on the basis of this specification.

[7] Another class of variable selection procedures we considered were step-wise *General to Specific* approaches, e.g. Hoover and Perez (2004), which are common in the cross-country growth literature. The general procedure there would be to run conventional OLS regressions, and to iteratively eliminate variables based on some measure of explanatory power, e.g., their t-ratios. We found these to be very sensitive even to small changes (like the order in which the variables are included) in terms of selected variables as well as in terms of predicted values. Overmore, they had a strong tendency to overfit.

Finally, note that when Henderson et al. (2012) proceed to inferring long-run growth rates to augment the GDP data, they estimate a substantially simplified cross-sectional equation. In the context of their study, this maintains tractability of the statistical framework, and enables them to compute optimal weights for the official data and the lights predictions (see section 2). We do not engage in such computations, but rather seek to improve the accuracy of the predicted growth rates directly. Nevertheless, we present basic results (estimates of long term growth rates over the entire period) inferred from their simplified model. To this end, we follow them in averaging our variables over the first and last two years in the sample period (1992/93 and 2012/13), and take the log-difference between these periods. The estimated equation is then

$$z_i^{LR} = \alpha + \ell_i^{LR} + \varepsilon_i, \tag{6}$$

where the superscript LR indicates long run growth rates as described.

4 Simulation exercise

The asymptotic and finite sample properties of the methodology outlined in section 3 are unknown. This is partly due to the econometrically unconventional challenge that we are facing: We effectively seek to reduce measurement error in our target variable (GDP growth) using a proxy that has an independent measurement error (lights); at the same time, we wish to account for heterogeneity in the relationship between this proxy and the measure of interest. Moreover, our methodological strategy partly relies on machine learning techniques that have no analytical solution. In order to assess the basic properties of our suggested methodology, this section will therefore present a simple simulation exercise. Section 4.1 provides a brief description of the data generating process (DGP), which is elaborated in appendix B. Sections 4.2 and 4.3 illustrate different estimation strategies (long difference, fixed effects with common slope, and our two-stage procedure described in section 3) on a single simulated dataset. In section 4.4, we repeat this exercise under 625 different parameter combinations on 31,250 simulated datasets in a simple Monte Carlo type simulation, and discuss the key insights about our estimator's performance.

4.1 Data generating process

Our data generating process (DGP) is designed to reflect the econometric challenges we are facing, while being tractable in its statistical properties. The simulated dataset needs to feature the following series: (i) a true measure Y ('GDP')

(ii) an indicator plagued with measurement error, Z ('reported GDP'), (iii) a proxy for Y that has a measurement error independent of that in Z, ℓ ('lights'), and (iv) determinants (analogous to Φ) and potential determinants (Φ^*) of how Y and L interact in every unit of observation i ('country'); that is, determinants of γ. We design our DGP to exhibit the following features:

1. The resulting series are designed to have econometric properties similar to the observed data; this concerns time series behaviour, orders of magnitudes and distributions of growth rates, and measurement error.

2. We introduce a set of countries with accurately reported data, and a set of countries with poorly reported data. The latter category is modelled to systematically report biased growth rates; the bias in any bad data country's growth rates is determined at random according to a uniform distribution. Depending on the specific parameters, this distribution can imply a tendency for bad data countries to over-report or under-report growth, or for the bias to be centred around zero (no directional bias overall).

3. The elasticity with which y translates into ℓ, γ, is designed to vary between countries based on a set of determinant variables; our estimator will crucially face the challenge of discerning those determinant variables from irrelevant variables.

4. As a potential source of confusion for the estimator, we introduce some overlap between the determinants of y and the determinants of the elasticity γ; for instance, investment levels may both affect GDP growth and the lights-GDP elasticity.

Table 1: Variable parameters in DGP

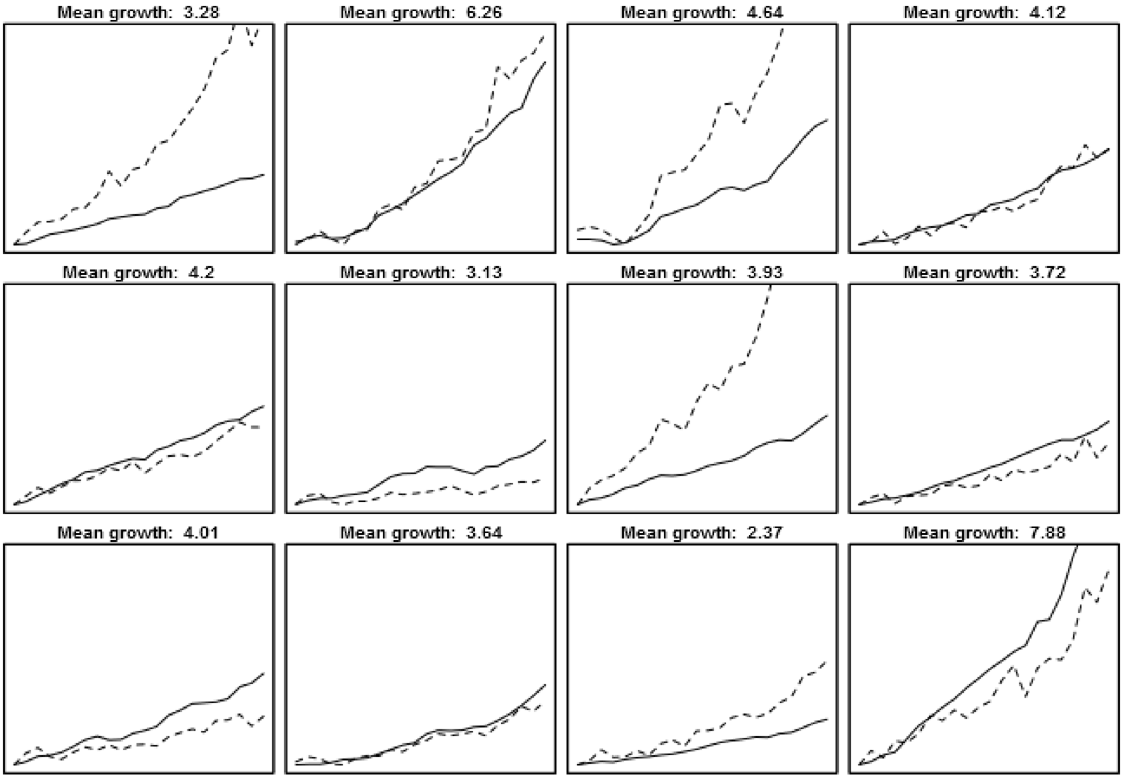
Parameter	Description	Example	MC min	MC max
N_{bad}	Number of bad data countries	20	10	30
SD_ϵ	Error in lights-GDP relationship	μ_g	$\mu_g/0.5$	$\mu_g/2$
Range misrep.	Range of misreporting	$[-0.02; 0.03]$	$[-0.05; 0]$	$[0; 0.05]$
SD_γ	Randomness in inverse elasticities	0.05	0	0.1

Notes: This table summarises the parameters chosen for the DGP both in the illustrative example above, as well as the range iterated though in the Monte Carlo type simulation. All parameters are iterated through over an equally spaced grid with 5 values ranging from MC min to MC max. μ_g is the respective country's mean growth rate.

Appendix B provides a detailed summary of the construction of the series (i)–(iv) designed to fulfil criteria 1–4. The dataset generated for this exercise comprises

$N = 150$ units ('countries') over $T = 22$ periods ('years'); figure 1 illustrates the resulting GDP and luminosity series (solid and dashed lines respectively) by plotting them for 12 randomly selected units of the dataset we use to illustrate our procedure in sections 4.2 and 4.3. The key parameters are reported in table 1, along with the range of parameters that will be explored in section 4.4.

Figure 1: 12 simulated GDP and lights series



Notes: The solid lines depict simulated GDP series, the dashed lines the corresponding proxy L as generated with the DGP described above for 12 randomly selected units.

4.2 Benchmark estimations

We compare the predictive qualities of our two-step estimator to two more conventional estimators that have been employed in the literature, specified in equations 6 and 5 in section 3.3. Table 2 reports the estimated coefficients and prediction errors for the single simulated dataset we analyse in this section for illustrative purposes across different methods.

Long differences Henderson et al. (2012) obtain their estimates of long-run growth rates (average annual growth between 1992/93 and 2005/06) from a simplified, cross-sectional regression (equation 6 in section 3). On the left hand side, they include the log-difference of GDP between the first and the last period, on

Table 2: Predictive performance of estimators (example)

	Long diff.	FE (single slope)	Two-stage
Estimated γ (true: 1.05)	0.24	0.29	0.57
Mean prediction error of average growth rate			
Good data countries	1.01%	1.03%	0.66%
Bad data countries	1.15%	1.18%	1.04%
Total	1.03%	1.05%	0.70%

Notes: The estimated γ in the two-stage method refers to the mean across the N $\hat{\gamma}_i$s. The prediction errors are mean absolute deviations from the true average growth rate in each country across the entire period.

the right hand side an intercept and the log-difference of luminosity between the first and the last period. Based on this regression, they then predict the long-run growth rate for every country based on luminosity. They then report annualised growth rates derived from these values.

This drastic reduction of the dataset is necessary for their statistical framework to be applicable (or remain tractable), where one key ambition is to derive optimal weights for the luminosity based series and official GDP data. In this study, we are not concerned with deriving such optimal weights, but rather with improving the lights-based estimates directly by addressing slope heterogeneity. The framework we suggest to this end crucially relies on the temporal variation *within* countries over time, and we cannot (and need not) base our estimates on a purely cross-sectional dataset. However, for direct comparison, we estimate Henderson et al. (2012)'s specification on our simulated data. The estimated inverse elasticity $\hat{\gamma}^{LR}$ is 0.24 (compared to a true value of 1.05 on average). The predicted average annual growth rates are, on average, off by about 1.03% in good data countries, and by 1.15% in bad data countries. This is an improvement for the bad data countries, where the reported GDP growth deviates from its true value by 1.54% on average in this dataset. However, the prediction error is of a similar amplitude for countries where the data is perfectly accurate, effectively deteriorating growth estimates in these places.

Fixed Effects with single slope Next, we turn to the panel fixed effects estimation with a common slope $\hat{\gamma}$, equation 5. This is most similar to what most of the empirical literature that uses night lights as a proxy of economic activity employs (e.g., Keola et al., 2015; Bickenbach et al., 2016); in Henderson et al. (2012), the baseline estimates are derived from this specification. As it is better comparable to our specification with country-specific slopes, we consider this specification our main benchmark. In order to derive growth rates from the resulting estimates, we then produce predictions in levels, derive annual growth rates ($\frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}$), and average these over all periods for each country (see section 3.2).

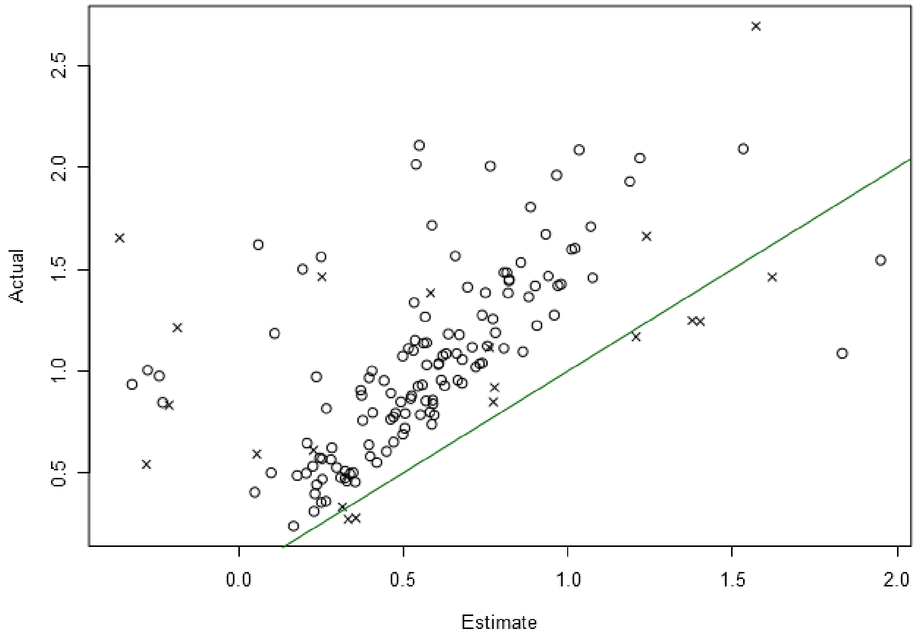
The estimated coefficient $\hat{\gamma}$ is 0.29, again substantially smaller than the true average coefficient of 1.05. In the presence of measurement error, we would indeed expect the measured coefficient to be smaller than the structural one due to attenuation bias. It is worth noting that a large portion of the bias still persists even when we remove every source of measurement error from our DGP. In this particular example, the absence of any measurement error in the DGP (leaving everything else constant) leads to an even *less* accurate estimate of γ (0.25). This is consistent with the finding that – where the heterogeneity in the slope is correlated with the independent variable (as it is in our DGP, and as we suspect it to be the case in the empirical data), a substantial aggregation bias can result from erroneously imposing a single slope across all units of observation (Ul Haque et al., 1999). Note that the direction of the bias depends on the sign of the correlation between the independent variable and the slopes, and may as well be positive.

From a predictive perspective, the fixed effects estimates with a common slope are roughly at par with the long differences: On average, they are off by 1.03% in good data countries, and by 1.18% in bad data countries. As with the long differences method, this is an improvement for bad data countries, but with revisions being almost equally substantial for good data countries.

4.3 Two-step procedure

Step 1: Estimating naive country-specific slopes ($\hat{\gamma}$): In the first step, we aim to derive country-specific slopes for each of the countries in the dataset. These estimates are naive in the sense that they take the reported data at face value, irrespective of whether these estimates are biased by any mis-measurement or misreporting of GDP. We obtain our (naive) country-specific slopes $\hat{\gamma}_i$ from estimating equation 1 – the fixed effects regression from above, augmented by $\ell_{it} \times D_i$, that is, N interaction terms between lights and the country dummies. Figure 2 plots the estimated coefficients against the true values from the simulated data. The circles indicate good data countries, the crosses indicate bad data countries as defined above. The green diagonal line indicates $\hat{\gamma}_i = \gamma_i$, that is, the closer an estimate is to the line, the more precise it is. Attenuation bias due to measurement error biases the estimates down compared to their true values, that is, they are plotted left of the green line in figure 2. At least in this example however, this bias is weaker than the aggregation bias occuring where a single slope is erroneously assumed, like in the estimators above. Note also that, due to measurement error,

Figure 2: Actual versus estimated γ (simulated data)



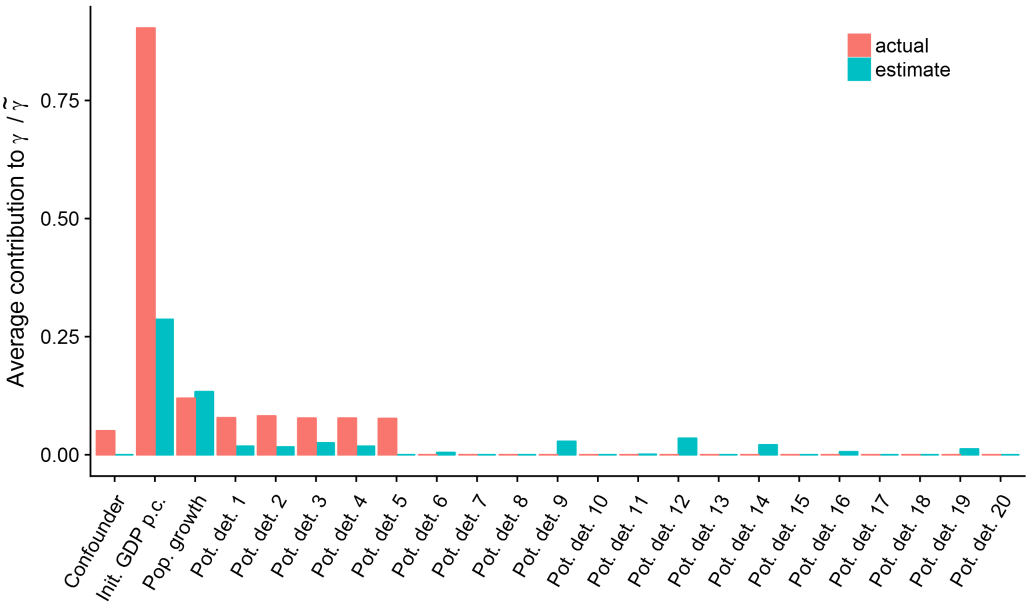
Notes: Coefficients obtained from estamating equation 1 on the simulated data plotted against the true coefficients. $\circ$ indicate good data countries, $\times$ are bad data countries. The green line indicates $\hat{\gamma}_i = \gamma_i$. Attenuation bias from measurement error means the estimated values are systematically below the actual ones; in the absence of measurement error, the green line and the datapoints would entirely coincide.

a few $\hat{\gamma}_i$ are negative, even though there are no true negative values of γ.[8]

On average, the luminosity coefficients obtained from estimating equation 1 are 0.57, much closer to the true average of 1.05 than both methods that impose a common slope (setting all measurement error to zero in our DGP would yield $E[\hat{\gamma}] = E[\gamma] = 1.05$ in this example, with all estimated slopes being accurate; there is no aggregation bias). Using the country-specific $\hat{\gamma}$ as estimated to predict GDP growth from luminosity does, however, by construction not have much potential to improve estimates from bad data countries. Rather, it essentially reproduces reported growth rates: For the countries with accurately reported GDP, the predicted growth rates are only 0.26% off compared to the true (and reported) growth rates. For countries with systematically misreported GDPs (bad data countries) it is off by 1.22% on average, which is worse than the previously discussed estima-

[8]It should be noted that the possibility of negative coefficients points at a limitation of the conceptual framework adopted by most of the literature, including the study at hand: If the coefficients obtained are strictly interpreted as the inverse elasticity $\gamma = \frac{1}{\beta}$ with $L = Y^{\beta}$, then a positive $\hat{\gamma}$ near zero would imply a structural elasticity β near *positive* infinity. A negative $\hat{\gamma}$ that approaches zero, however, would imply a structural elasticity β that approaches *negative* infinity. We note this as a caveat from the narrative point of view, and where this is central to the analysis it would be desirable to introduce a functional form that does not have such a drastic discontinuity. In the study at hand, however, the purpose is purely predictive, with the relevant parameter being γ.

Figure 3: Actual and estimated determinants of γ



Notes: The bars depict the absolute contributions of each of the determinants of γ. The red bars indicate the actual contributions, known from our DGP, the blue bars are the contributions to the predicted $\tilde{\gamma}$, based on the elastic net estimation. Note that the typical importance of initial GDP appears inflated in this representation, as the mean is driven by some outliers due to the exponential distribution we chose for the variable.

tors. The aim of of the next section will be to discern the systematic component in the variation in γ and improve the predictions for countries where data is poor.

Step 2: Explaining $\hat{\gamma}_i$ using the elastic net In order to avoid simple replication of the reported growth rates by taking the country-specific slopes at face-value, we will now employ the elastic net estimator in order to discern the systematic component in the variation in slopes between countries. In the DGP described in section 4.1 and specified in appendix B, γ had 8 determinants in total: 3 of them are simultaneously determinants of economic growth, and 5 others are unrelated. We also generated another 15 random variables that do not actually affect γ, but that we will consider *potential* determinants. All of these variables are included as independent variables in the elastic net estimator, with $\hat{\gamma}$ as the dependent variable. As discussed in section 3 and further in appendix F, the parameters of the elastic net estimator are then tuned to attribute weights to the variables that minimise the out-of-sample prediction error. The aim of this second step estimation is therefore to discern the relevant predictors Φ from the pool of potential predictors Φ^*, and to attribute optimal weights to each of them. These can then be employed to derive estimates of γ_i *conditional on country-characteristics* Φ_i, which we label $\tilde{\gamma}_i$.

Figure 3 depicts the average composition of the γs in red, and the estimated contributions in blue. This is quantified by multiplying the estimated weights $\hat{\delta}$ with the values of Φ_i^* for each country, and then averaging these across countries. If the elastic net was doing a perfect job in identifying the predictors and their respective weights, the bars would therefore be identical. However, as we introduced measurement error in our DGP, and the estimates of γ_i, $\hat{\gamma}_i$, are only approximations of the true values (figure 2) the elastic net will naturally only capture parts of the structure. In this example, 6 out of 8 elements of Φ are correctly identified, and 3 variables are falsely attributed non-negligible weights. Note that the accuracy of these predictions strongly hinges on the amplitude of the diverse sources of measurement error in our DGP: Setting them to zero completely aligns true and estimated determinants; with increasing error, the elastic net attributes smaller and smaller weights to the determinants, and – as a tendency – converges to the unconditional mean. The R^2 of the second stage in this illustrative example is 50%, slightly lower than the 55% we will obtain in our empirical application.

Discerning the determinants of γ, however, is mainly an auxiliary goal in our strategy. Ultimately, we aim to improve the predicted GDP growth rates from changes in luminosity. Indeed, it appears as though we achieve this goal: In the present example, the predicted GDP growth rates for good data countries are, on average, 0.70% off when compared to the true growth rates; for good data countries, the mean error is about 0.66%, a substantial improvement over the other estimators. For countries with poor data the improvement is still present but less substantial, with predictions now deviating from actual growth rates by 1.04% on average.

4.4 A simple Monte-Carlo simulation

The example presented above refers to a single iteration using a particular combination of parameters. While any artificial DGP will only provide a very stylised view on the matter, we will now repeat the exercise above using several parameter combinations for the generated data. In order to keep computing time at reasonable levels and the exercise tractable, we focus our attention on four key parameters.[9] For each parameter, we iterate through a grid of 5 equally spaced values, which leads to a total of 625 parameter combinations. Each of these we repeat 50 times, leading to 31,250 iterations in total. The parameters we vary are:

[9] The elastic net is a computationally demanding procedure, and one iteration of the procedure outlined above (including data generation, estimation and prediction) takes about 1.4 seconds to carry out on the computer that we have at our disposal (including λ-tuning, but excluding α-tuning). In order to obtain reliable indicators of the estimators' expected performance given any parameter combination, we run 50 repetitions per parameter combination. In total, the simulation takes approximately 12 hours to run in this specification.

- Number of countries with flawed data (N_{bad}): When deriving the determinants of γ, our setup treats all countries equally, which avoids any *a priori* assumptions about the quality of their data. The larger the number of bad data countries, the more influential these observations become and potentially undermine the precision of our estimates.

- Error in lights-GDP relationship (SD_ϵ): The precision of any estimates will obviously be influenced by the signal-to-noise ratio in our proxy variable (luminosity). We therefore vary the standard deviation of the error term ϵ_{it} in the relationship between lights and GDP.

- Bias in bad data countries (*Range misrep.*): Misreporting is modelled as the systematic misreporting of growth figures within a country, where the amplitude of the misreporting is a (uniformly distributed) random number. We shift the limits of this distribution, so as to cover the hypothetical case where bad data countries generally *under*report growth rates (e.g, their statistical offices fail to pick it up), the case where they *over*report (e.g., they tend to manipulate figures upwards), and the case where misreporting has no direction in particular.

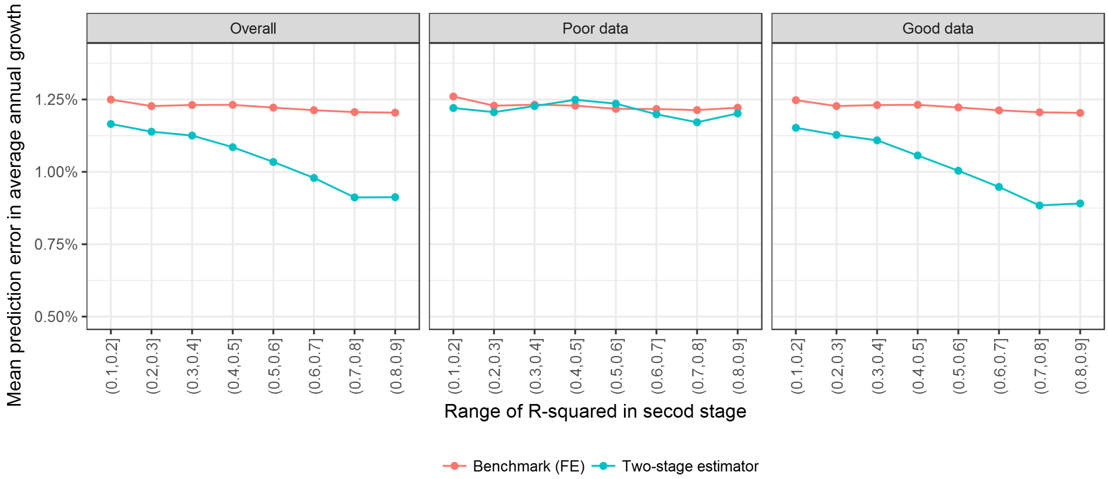
- Randomness in γ (SD_γ): The (inverse) elasticity of the relationship between GDP and lights is unlikely to be fully deterministic, and not all of the determinants are necessarily observable. Our method crucially relies on the predictability of γ, so we will assess its performance under different values of it.

The general pattern that emerges is that, in countries where the reported data are accurate, our two-stage procedure is less prone to suggest false corrections. Where the reported data are poor, the results are more ambiguous, and overall, the estimators appear to be more or less on par in these cases. Especially as the bias in mis-reporting becomes more systematic (that is, on expectation more different from 0), its performance deteriorates, while the benchmark estimator assuming a single slope γ remains relatively unaffected by this: In these cases, our estimator has a tendency of being too 'lenient', and increasing parts of the bias from misreporting are reflected in the $\tilde{\gamma}_i$s.[10] Appendix C compares the performance of the estimators across the range of parameters specified in table 1.

Irrespective of the underlying parameters, we observe that the quality of our predictions from the two-step procedure strongly hinges on the predictability of

[10] One promising avenue for future research would be to explore weighting country observations by their statistical capacity or a proxy thereof, that is, giving reliable data more weight when deriving the systematic component of the variation in $\hat{\gamma}$. This would come at the cost of making *a priori* assumptions about the quality of the data, but potentially enhance predictive accuracy.

Figure 4: Mean absolute prediction error by R^2 of second stage



Notes: The plotted values are obtained from sorting iterations into bins by their R^2 in the second stage (explanatory power of the elastic net), and then taking the mean absolute prediction error within each of these bins.

our parameter γ_i, or more precisely of its estimate $\hat{\gamma}_i$: The larger the proportion of the inverse elasticity that we can explain based on Φ^*, the more precise the subsequent predictions of growth rates, conditional on luminosity. Figure 4 plots the average prediction error in annual growth rates by the range of R^2 in the second stage, that is, how much of the variation in $\hat{\gamma}_i$ is captured by the elastic net. The values are obtained by grouping iterations of the simulation into bins, according to the R^2 of the elastic net in the second stage. The prediction errors from the two-stage procedure are plotted in blue. For comparison, the average prediction errors from the corresponding iterations obtained from the fixed effects estimator assuming a common slope γ is plotted in red. For bad data countries, the two-stage procedure does not substantially outperform the benchmark; instead the measures are more or less on par. For good data countries, however, it does outperform the fixed effects estimation even where very little (10-20%) of the variation in the $\hat{\gamma}_i$ is captured; fewer false corrections are applied to countries countries with accurate data. As the explanatory power of the second stage goes up, so does the relative advantage of the two-stage procedure.

Overall, this exercise suggests that there is a potential for the two-stage estimator to improve the predictions based on luminosity by capturing some of the heterogeneity, in particular by avoiding false corrections. Note also that, compared to alternative approaches in the literature, we do not need to make any *a priori* assumptions about countries' data quality in order to obtain that pattern (unlike approaches that compute optimal weights depending on the assumed data quality). In the empirical application that follows, we will report both the estimates resulting from our procedure, as well as more conventional estimates based on the assumption of a single slope. Reassuringly, the big picture of the results remains similar irrespective of the estimator employed.

5 Data

The data we employ in our empirical application can be divided into two datasets: The primary dataset consists in the series on GDP (Y) and luminosity (L), our main variables of interest. The second dataset contains the potential determinants of this relationship (Φ^*). This encompasses a wide range of geographical and socio-economic variables at the country level.

5.0.1 Main dataset: Lights and GDP

Lights: The series on luminosity we employ is derived from the version 4 of the DMSP-OLS nighttime lights time series provided by the US National Oceanic and Atmospheric Administration (NOAA). The data span from 1992 to 2013, and come in 30 arc second grids, corresponding to a spatial resolution of about 1km^2 at the equator. It covers the surface of the earth between -65 and 75 degrees latitude, which corresponds to almost the entirety of the earth's inhabited land. For each year, each of the ca. 700 million grid cells is attributed a luminosity value labelled Digital Number (DN), ranging from 0 (very dark) to 63 (very bright). These values are the result of pre-processing by NOAA staff, who combine cloud-free imagery collected throughout the year, and remove glare from solar light, moonlit data, as well as features from the aurora (northern lights).

One major issue with the use of luminosity data is the contamination with gas flaring. In areas where residual gas from the production of petroleum-production is burned, large flames almost constantly illuminate the night sky; in the DMSP-OLS data, even moderately sized plants can lead to the top-coding (DN values of 63) of several kilometres square. In line with most of the literature, we therefore remove gas flaring areas as identified by Elvidge et al. (2009). However, these areas tend to be large and sometimes go far beyond the spots where the flares do induce excess luminosity. In some areas (Nigeria in particular), this leads to a loss of substantial parts of the country's surface. In order to mitigate the problem, we use MODIS data on landcover (version 5.1) to identify agglomerations (defined as adjacent built-up areas larger than 10 km^2) in the gas-flaring areas, and include them with a buffer of 50km.

Top-coding is generally an important issue with the DMSP-OLS data: the scale is capped at 63, and a large share of the lit cells take on this value. This can potentially come with a substantial loss of information at the top end of the distribution (cells that are in reality brighter than what is required for a DN value of 63). A number of issues follow from this: First, urban centres often only consist of lit cells. Any increase in luminosity in these places will not be reflected in the cells' DN, and small decreases will go unnoticed where luminosity is high (larger by some margin than the threshold for being top-coded). Second, particularly densely lit countries (e.g., rich and densely populated ones) may be disproportionately affected by top-coding, altering the relationship between GDP and measured luminosity in these places. Our methodology aims at capturing such effects, by deriving country-specific elasticities based on countries' properties such as population density and income per capita. However, this effect is taken to the extreme in areas that are very small, like city-states and small island countries. For this reason, and in line with most of the literature, we exclude countries with a surface area smaller than 5,000 km^2. This affects 29 countries, all of them small island states and city-states.

Especially where the interest lies in growth rates, we must be concerned with the consistency of luminosity measurements over time. Because the sensitivity of the sensors declines over time, and satellites are replaced, the reported luminosity data is indeed not immediately comparable across years. Broadly speaking, the literature offers two ways of addressing this issue: The most common way, employed for instance by Henderson et al. (2012), is to include year fixed effects in the regressions, absorbing global fluctuations in luminosity. Indeed, most of the time–inconsistencies appear to be global, and mostly consist in shifts in levels. Another option is to calibrate the luminosity data beforehand, especially where the nature of the estimation does not lend itself to the inclusion of year fixed effects. For instance, Tanaka and Keola (2017) propose a methodology for such adjustments using presumably reliable GDP data from OECD countries as a benchmark to calibrate luminosity values. Elvidge et al. (2009) seek to identify spots on earth that arguably emit constant levels of luminosity as a benchmark for calibration (and choose Sicily). In our application, the inclusion of fixed effects appears to be the econometrically more tractable option. Moreover, we can thereby avoid making any strong assumptions about certain places' constant emissions of luminosity, or the accuracy of their GDP figures.

In 12 out of 22 years of the DMSP-OLS data, there do actually exist two series, as satellites were in orbit at the same time and registered luminosity in parallel. In line with Henderson et al. (2012) and most of the other literature, we average across satellites in these years. The luminosity values we employ in our series are expressed in $DN/cell$ and computed for every country and year. To this end, we take the sum of the luminosity values (DN) on the territory of each country, and divide it by the overall number of cells contained in it.[11] Country borders are obtained from the GADM Global Administrative Areas database version 2.8. Where countries' territories have changed over time, we follow the World Development Indicators in how we treat them. Where country GDP series have retrospectively been split (e.g., Yugoslavia), we use these series and compute luminosity using current borders. Where the GDP series is incomplete, we include those years for which it is available. Depending on the exercise, countries with incomplete data are excluded from the analysis (this is the case for the estimation of country-specific slopes using equation 1).

Finally, when discussing luminosity data, it is worth noting that the DMSP-OLS series has effectively been superseded by the VIIRS series. As shown by Dai et al. (2017), the latter is superior in many respects: It has higher (effective) resolution, is more consistent across years, and less plagued by issues of top-coding. However, the purpose of our investigation is to recover historical growth rates, and VIIRS data is available only from 2012 onwards.

Gross Domestic Product: The GDP series are obtained from the World Bank's World Development Indicators 2015 (WDI). As we are interested in real GDP growth rates, we use the series in constant local currency units (labelled *NY.GDP.MKTP.KN* in the WDI), that is, adjusted for inflation but without any transformations aiming at rendering income values internationally comparable. This is to minimise any distortions to growth rates stemming from adjustments for purchasing power or exchange rates (see Deaton and Heston, 2010); furthermore, it is the series with the widest coverage. The lack of international comparability in levels is no concern, as our analysis does not touch on cross-sectional GDP levels, and differences in units will be fully absorbed by country fixed effects.

Both luminosity and GDP values are included as logarithms in order for the coefficients to be interpretable in terms of elasticities and to conform to the methodological framework outlined in section 3.

5.0.2 Auxiliary dataset: Potential determinants of γ

Our secondary dataset consists in factors that are potential determinants of the (inverse) elasticity between GDP and lights, labelled Φ^* in the methodological framework. Its purpose is to derive the key determinants of γ, and then to predict

[11] This follows a convention in the literature of considering the density rather than the absolute level of luminosity emitted by each country. While this may be somewhat sensible from an intuitive perspective, it has no bearings on our empirical results as the area of a country is fixed over time, and the scaling will be fully absorbed by country fixed effects.

expected values of $\tilde{\gamma}$ that are conditional on these factors for each country. As in the present setup, γ varies across countries but is assumed to be constant over time, this dataset is cross-sectional. Time-varying variables will therefore need to enter in some aggregated form. Depending on the variable, we take averages over the period 1992-2013, or initial values. Furthermore, for the lack of theoretical knowledge of the lights-GDP nexus, it is unclear from the outset which transformation of any given variable is the most sensible one, and has the strongest predictive power for $\hat{\gamma}$. We therefore include the variables in up to three transformations: As square roots (where this is numerically possible), in levels, and squared. Similarly, the choice of the variables themselves cannot follow any rigid theoretical understanding of the relationship of interest, and a parsimonious specification cannot be motivated on strong theoretical grounds (see section 3). The included variables are therefore partly motivated from earlier studies concerned with the relationship between GDP and lights (especially Wu et al., 2013; Levin and Zhang, 2017; Keola et al., 2015; Bickenbach et al., 2016), and partly on speculative grounds. We then leave it to the elastic net estimator to select the variables with a strong predictive power for γ across and beyond the sample, and to attribute appropriate weights.

Table 3 lists all included factors and the included transformations. Most variables are obtained from WDI, with the following exceptions: Snow cover is obtained from the National Snow and Ice Data Center; it corresponds to yearly average snow cover (as a share of the land surface area), aggregated from monthly data and within the borders provided by the GADM database. The absolute latitude is the latitude of the centroid of a country, computed by the author based on GADM borders. Furthermore, the sectoral growth variables (agricultural, industrial, services and manufacturing growth) are relative contributions – they express the share of total growth that is attributable to the respective macro-sector, that is, they are not growth rates *per se*.

6 Heterogeneity in the relationship between GDP and lights

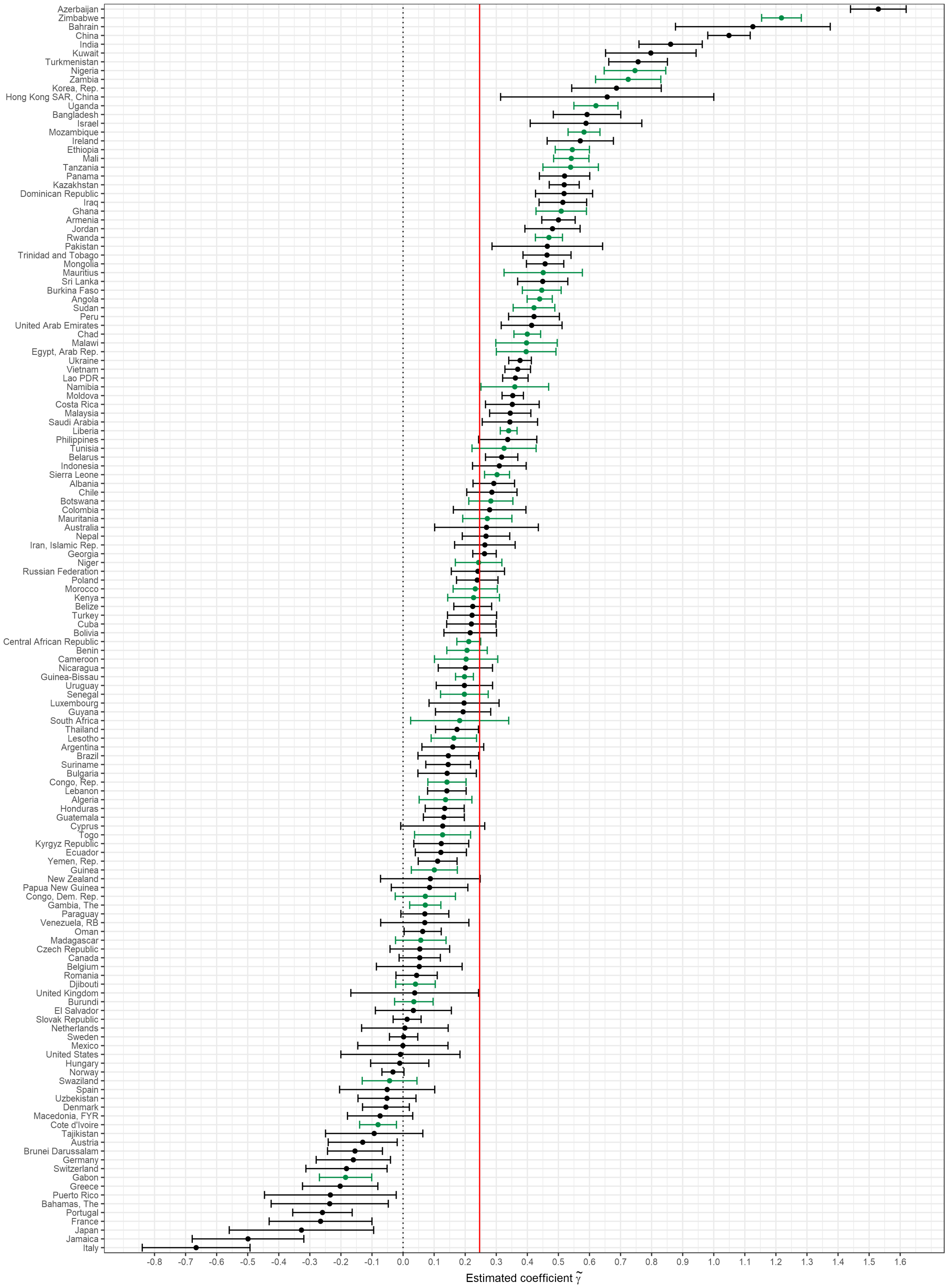
Henderson et al. (2012), as well as the bulk of the studies using a methodology similar to theirs, estimate a model of the type specified in equation 5 or 6. On this basis, they estimate an inverse elasticity $\hat{\gamma}$ (see section 3), finding a value of about 0.28. Throughout their analysis, this coefficient is assumed to be constant across countries, and serves as a basis for assessing the accuracy of reported growth rates in countries with poor statistical capacity.[12]

[12] In order to derive optimal weights and compute lights-adjusted growth measures, Henderson

Table 3: Potential determinants of γ (Φ^*)

Variable	Transformations: Level	Square	Root	Aggregation
Geographic (non-anthropogenic)				
Snow cover	✓	✓		Mean
Absolute latitude (centroid)	✓	✓	✓	First
Surface area	✓	✓	✓	First
Geographic (anthropogenic)				
Population density	✓	✓	✓	Mean
Population growth	✓	✓		Mean
Urban population (share of total)	✓	✓	✓	Mean
Forest area (share of total)	✓	✓	✓	Mean
Agricultural land (share of total)	✓	✓	✓	Mean
Economic				
GDP level	✓	✓	✓	First
GDP per capita	✓	✓	✓	First
Investment (% GDP)	✓	✓	✓	Mean
Consumption (% GDP)	✓	✓	✓	Mean
Agriculture (% GDP)	✓	✓	✓	Mean
Industry (% GDP)	✓	✓	✓	Mean
Services (% GDP)	✓	✓	✓	Mean
Manufacturing (% GDP)	✓	✓	✓	Mean
Agricultural growth (relative)	✓	✓		Mean
Industrial growth (relative)	✓	✓		Mean
Services growth (relative)	✓	✓		Mean
Manufacturing growth (relative)	✓	✓		Mean
Fossil fuel revenues (% GDP)	✓	✓		First

Figure 5: Estimated $\hat{\gamma}$ in 142 countries



The plotted estimates are obtained from a two-ways fixed effects regression with countryspecific slopes as specified in equations 1. Mean point estimate of $\hat{\gamma}$: 0.25 (red line). Mean SE: 0.05. Standard Errors are clustered at the country-level following MacKinnon and White (1985). The errorbars indicate 95% confidence intervals. Bars highlighted in green correspond to African countries.

Figure 5 plots the coefficients obtained from estimating equation 1, that is, a twoway fixed effects regression augmented by N $\ell_{it} \times D_i$ interaction terms. This allows us to obtain country-specific estimates of the slope coefficient γ_i. While each of these estimated slopes is based on a small sample of 22 observations (1992-2013), the strength of the relationship is such that most of the obtained coefficients are statistically significant from zero at the 5% level (114 out of 142).[13] The figure plots the country-specific coefficients in descending order of the point estimate; the precise values are reported in tabular form in appendix D. The errorbars around the point estimates indicate 95% confidence intervals based on standard errors clustered at the country level computed following MacKinnon and White (1985). In view of our later application, the bars associated with coefficients for African countries are highlighted in green. On average, the estimated $\hat{\gamma}$ is 0.25, which is marginally lower than the 0.28 obtained by Henderson et al. (2012), and lower than what we obtain when estimating equation 5 on the sample at hand (0.30, see section 7). But, crucially, few of the coefficients do actually correspond to that mean: in 109 out of 142 countries, the coefficient is statistically significantly different from the mean at the 5% level. At the extremes, Azerbaijan has the highest $\hat{\gamma_i}$ at 1.53, and Italy has the lowest at -0.67.

Naturally, the dispersion of the coefficients is the result of a conflation of factors. First, the OLS coefficient has its own variance by construction, and even under perfect parameter homogeneity one would expect some variation across coefficients obtained from different random samples, even more so in the presence of measurement error and where there is an element of randomness in the relationship between the variables. However, as suggested by the fairly narrow confidence intervals, this can only be a small part of the story. Second, there may be structural differences between countries in the relationship between luminosity and economic activity that are driven by *unobservable* factors (Ψ); say, a culturally driven preference for bright streets. Third, structural differences across countries that are driven by *observable* factors (Φ) may lead to differences in the elasticities. And finally, systematic mis-reporting or mis-measurement of GDP figures will be reflected in different estimated elasticities across countries. The aim of this study is then to identify the observable factors Φ that drive the relationship between GDP and lights, compute *expected* true coefficients for each country conditional on these observed factors, and derive luminosity-based growth rates based on those conditional, country-specific coefficients.

et al. (2012) rely on equation 6 and exclude high-income countries specifically because their γ appears to differ.

[13] As the luminosity data only starts in 1992, earlier studies had to rely on sample sizes that were prohibitively short in order to rely on the time dimension to estimate elasticities based merely on within country variation.

Visually inspecting figure 5 already appears to be suggestive of some patterns. Looking at the very top of the distribution, there appears to be a high number countries where petroleum production accounts for a large share of the economy: with Azerbaijan, Bahrain, Kuwait, Turkmenistan and Nigeria, 5 of the 8 countries with the highest estimated inverse elasticities $\hat{\gamma}$ are major oil producers. Beyond misreporting and measurement error, this is compatible with a number of explanations. For one thing, while we do remove areas that have been identified as contaminated by gas-flaring by Elvidge et al. (2009), it is likely that some areas with gas flares still remain (see section 5). Any growth-induced increase in luminosity may then be dwarfed by the brightness emitted by gas flares irrespective of other economic activity, resulting in a low elasticity of lights with respect to GDP, and therefore a high inverse elasticity γ_i. Alternatively, the relative cheapness of electricity may imply that even at lower stages of development, luminosity is already high, and further increases in GDP do not induce much additional luminosity. At the bottom of the distribution of $\hat{\gamma}_i$s, on the other hand, there appears to be a bunching of highly developed economies: 10 out of 34 OECD members are among the 20 countries with the lowest $\hat{\gamma}$. The average reported per capita income in these places in the year 2000 was 24,865$ US, compared to 9,048$ US in the remaining 122 countries (WDI 2015).

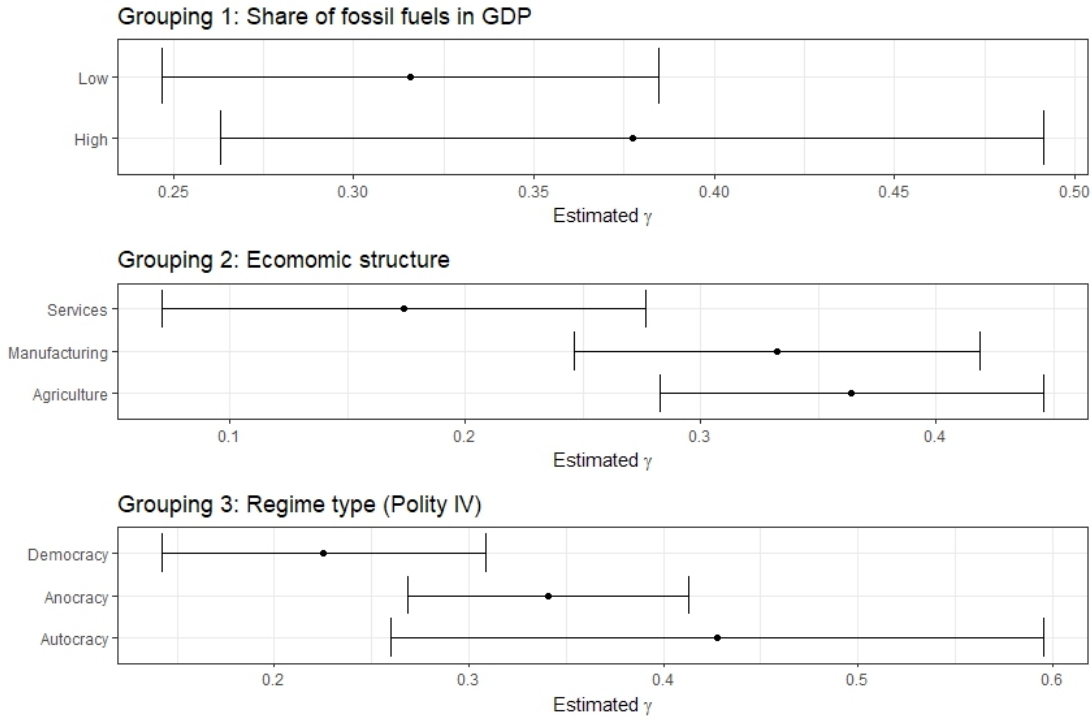
Table 4: Composition of the groups (all values are group means)

Grouping 1: Share of fossil fuels in GDP				
Group	Fossil fuels (%GDP)			N
1	37.6			25
2	1.6			138
Grouping 2: Economic structure				
Group	Agric. (%GDP)	Manuf. (%GDP)	Services (%GDP)	N
1	26.3	6.8	38.4	33
2	15.6	17.4	49.9	65
3	2.8	13.5	63.6	58
Grouping 3: Regime type (Polity IV)				
Group	PolityIV			N
1	-8.0			16
2	0.3			65
3	8.5			75

Notes: Groupings 1 and 2 are obtained from a k-means clustering algorithm, grouping 3 follows the Polity IV classification whereby any country with a rating below -6 is classified as an autocracy, any above 6 as a democracy, and those in between are labelled anocracies. See appendix A for details.

Figure 6 depicts the results from a grouped regression of the form specified in equation 4. The plotted coefficients $\hat{\gamma}_g$ are those on the $\ell_{it} \times D_g$ interactions, and

Figure 6: Estimated $\hat{\gamma}$ by groups of countries



Notes: The plotted estimates are obtained from a two-ways fixed effects regression augmented by $\ell_{it} \times D_g$ interaction dummies, as specified in equation 4. Standard Errors are clustered at the country-level following MacKinnon and White (1985). The errorbars indicate 95% confidence intervals. Details on the construction of the groups can be found in appendix A.

indicate the estimated inverse elasticity as obtained for each of the sub-groups of countries. The compositions of the groups are summarised in table 4, with details in appendix A. Grouping 1 uses a k-means sorting algorithm to divide the countries into two distinct groups (see appendix A), based on the share of their GDP that that is attributed to exports of fossil fuels (oil, coal, or natural gas). The resulting groups are unequal in size, with a group of 25 countries that have an average share of fossil fuels in GDP of 37.6% across the sample period, and a group of 135 countries where fossil fuels make a very small contribution to GDP, on average 1.6%. As it would be expected from what we observed in figure 5, the point estimate of $\hat{\gamma}_g$ is larger for the fossil fuel producing countries (0.38) than for those with little reliance on fossil fuels (0.32). However, the precision of these estimates is too low in order for these differences to reach statistical significance.

Grouping 2 is again based on k-means clustering, where countries are sorted into clusters that are similar in terms of their economic structure, as measured by the relative shares in GDP of the main sectors (agriculture, manufacturing, and services, averaged over the sample period). The resulting groups are (1) a group of 33 countries with a particularly high agricultural share in GDP (26.3% on average, compared to 15.6% and 2.8% in groups 2 and 3), (2) a group of 65 countries with a relatively even sectoral composition, and the largest average share of manufacturing (17.4%), and (3) a group of 58 countries dominated by services (63.6% of GDP, on average); the latter group comprises all 32 OECD members. In line with what is suggested by figure 5, group 3 exhibits a much lower $\hat{\gamma}_g$ (0.17) than the two groups of countries that rely more on agriculture (0.36) and manufacturing (0.33).

Grouping 3 divides the countries according to their regime type, as suggested by the Polity IV index (averaged over the sample period; Marshall and Jaggers, 2002, updated version from 2014). Polity IV rates countries' political regimes on a scale that ranges from -10 (fully autocratic) to 10 (fully democratic). In order to form groups, we apply the same threshold values employed by the Polity IV project, whereby any country with an overall rating below -6 counts as an autocracy, any country with a rating above 6 counts as a democracy, and any country with a rating between these values counts as an 'anocracy'. Again, the estimated standard errors are too large in order for statistically significant differences to emerge at the 95% level. But the point estimate for democracies (0.23) is, in economic terms, substantially lower than that for anocracies (0.34) or autocracies (0.43). This highlights another factor that may be contributing to the dispersion of (inverse) elasticities: systematic misreporting. The observed pattern suggests that autocracies emit less luminosity for every increase in reported GDP. This is consistent with a situation where autocracies have a tendency to over-report their growth rates. In fact, Magee and Doces (2015) draw precisely this conclusion based on luminosity data.

Of course, there is some overlap between the groups of countries discerned above and we have thus far not made any attempts at identifying the underlying factors that drive differences in the $\hat{\gamma}_g$: For instance, rich economies tend to rely more on services, and be more democratic (see appendix A). The analysis that follows will therefore seek to isolate the relevant drivers of the elasticity between GDP and lights, and to provide estimates of actual growth rates that take these factors into account. Note that, in order to avoid replication of systematically misreported growth rates, e.g. by undemocratic regimes, only factors that are arguably 'legitimate' drivers of the lights-GDP relationship (that is, not suggestive of misreporting, like regime type) will be considered in the predictive exercise in section 8.

7 Explaining the relationship between GDP and lights

7.1 Estimating the relationship between GDP and lights

In what follows, we will establish the relationship between GDP and lights that our predictions later on will be based on. In line with the discussion above, we do so in three different ways: First, for reference, we estimate the long difference as specified in equation 6; as in Henderson et al., we conflate the first and last two observations in the sample period (1992/93 and 2012/13). Second, we estimate the panel FE model as specified in equation 5. And third, we estimate country-specific slopes in a fixed effects regression augmented by $N\ \ell_{it} \times D_i$ interaction terms, as in equation 1. Table 5 summarises the results obtained from these regressions. Columns 1 to 3 summarise the cross-sectional long difference equation (equation 6), columns 4-6 the panel regressions (equation 5), and column 7 refers to the panel regression with country-specific slopes. As the latter is only estimated for countries where observations are available for all 22 years, we report the results for the other models using the same sub-sample for reference (columns labelled 'Compl.'). For the models with a single coefficient (all except column 7), we also report the results obtained when excluding OECD countries (columns labelled 'No OECD'). This is to create comparability to Henderson et al. (2012) who exclude rich economies from their benchmark estimations, as they find their $\hat{\gamma}$ to be smaller.

Table 5: Regression output for different models

	Long difference			Panel fixed effects			Individual
Sample	Max	No OECD	Compl.	Max	No OECD	Compl.	Compl.
ln(lights)	0.45***	0.43***	0.42***	0.30***	0.31***	0.30***	0.25†
	(0.05)	(0.06)	(0.05)	(0.04)	(0.04)	(0.03)	(0.05)†
R^2	0.36	0.32	0.36	0.20	0.20	0.22	0.53
Adj. R^2	0.36	0.32	0.36	0.15	0.15	0.17	0.48
N	145	113	142	166	132	142	142
T (avg.)				21.16	21	22	22
Obs.	145	113	142	3512	2773	3124	3124

Notes: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$. †Average of country-specific coefficients (standard errors). The results for individual countries are depicted in figure 5 and listed in appendix D. The standard errors of all panel estimations (columns 4-7) are clustered at the country level. R^2 and adjusted R^2 are within-R^2 based on a FE model where the FE have been differenced out, which explains why they are much smaller than the values typically reported in the literature (see main text).

Focussing on the long differences reported in columns 1-3, the sample is naturally limited to those countries where luminosity and GDP values are available in the years 1992/93 and 2012/13. Using the maximally available sample of 145 countries, the estimated coefficient is 0.45. Restricting the sample to the 113 non-OECD countries does not substantially reduce this estimate (0.43). In fact, the impact of restricting the sample by only three countries (Finland, Iceland, Myanmar) in order to match those employed for the estimation of country-specific coefficients has a larger impact, reducing the estimate to 0.42. Either way, the differences between the three estimates are small and within about half a standard error of either estimate, therefore statistically negligible. Note that the values of the estimated coefficients lie above the 0.32 obtained by Henderson et al. (2012). This is likely to be due to differences in the sensitivity of satellite sensors across time, which in the cross-sectional setup cannot be absorbed through time fixed effects. The magnitude of the coefficient may then be sensitive to the choice of sample period, without however affecting the predictive qualities within the respective period.

Turning to the estimates obtained from panel fixed effects regressions (equation 5), reported in columns 4 to 6, these estimates are very much in line with the bulk of the literature, where γ is typically estimated at around 0.3 for the global sample (Henderson et al., 2012; Chen and Nordhaus, 2011; Keola et al., 2015). The reported standard errors are clustered at the country level following MacKinnon and White (1985). Across the sub-samples, there is very little variation in the coefficient, with the exception that it slightly increases to 0.31 as OECD countries are excluded from the estimation (column 5). Indeed, our inspection of figure 5 in section 6 had suggested a bunching of OECD members among the countries with the lowest $\hat{\gamma}$, and we also found a lower value for economies that rely heavily on services (figure 6).

Compared to most of the literature, consider also the substantially smaller R^2 and adjusted R^2. This is due to computational differences: The within-R^2 typically reported is based on the model with time fixed effects included as dummy variables, and accounts also for the predictive power of those terms. We report the R^2 based on the same model but differencing out the time fixed effects; this is computationally identical, but the resulting R^2 does not encompass the fixed effects' explanatory power, which we deem more informative.[14]

Column 7 reports the aggregate results from the estimation of equation 1. The reported coefficient of 0.25 corresponds to the mean point estimate across all countries in the sample. The standard error, 0.05, is the mean estimated standard error for each of these coefficients. The point estimates of the coefficients themselves

[14]This is related to differences in the software packages employed. Most studies rely on Stata, which reports a within- and a between R^2 that includes the fixed effects' contribution. We use the `plm` package for R for our computations, which differences the fixed effects out.

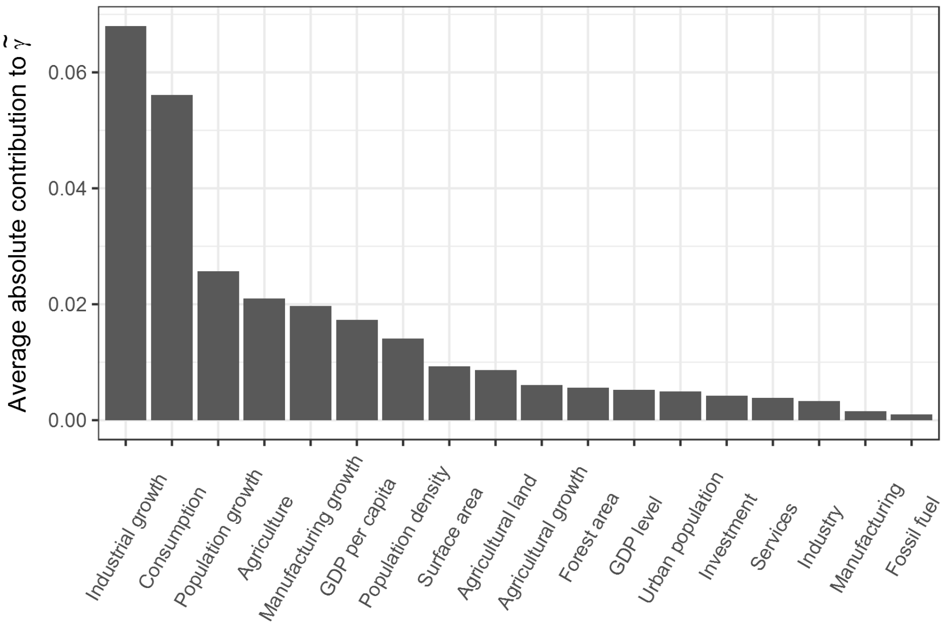
range between -0.67 (Italy) and 1.52 (Azerbaijan) with a standard deviation of 0.31 between them. Figure 5 in section 6 plotted the complete distribution of the measured coefficients including their 95% confidence intervals, and appendix D reports them in tabular form. The R^2 is greatly increased when compared to the other models, even when accounting for the inclusion of additional explanatory variables (adjusted R^2), which once more underlines the importance of slope heterogeneity in this context. The estimated average $\hat{\gamma}_i$ of 0.25 is smaller than the one estimated under the assumption of a common γ; the most direct comparison is with column 6, the fixed effects estimate with a single slope (0.30). As discussed in section 4, this discrepancy can essentially stem from two sources. First, measurement error biases OLS estimates towards zero, more so in smaller samples. As each of the coefficients we estimate is essentially based on 22 observations, this may induce a slight decrease in the estimated γs. Second, an aggregation bias can occur in a fixed effects estimation where slope heterogeneity is falsely assumed, as we argue is the case (Ul Haque et al., 1999). The direction of this bias depends on the sign of the correlation between the coefficient (γ_i) and the explanatory variable (ℓ), and can therefore not be determined *ex ante*.

7.2 Predicting country-specific $\tilde{\gamma}_i$ and growth rates

As noted in sections 3 and 4, taking the country-specific $\hat{\gamma}$s as obtained from estimating equation 1 defies the purpose of our investigation. Instead of providing an indication of potential mis-measurement or mis-reporting of GDP, the resulting predictions would merely replicate the reported long-run growth rates, along with some noise. In order to circumvent this issue, we will now – analogous to section 4.3 with the simulated data – seek to identify the systematic component of the variation in $\hat{\gamma}$. To this end, we employ the elastic net estimator as described in section 3.2, using the variables described in section 5 as potential determinants of $\hat{\gamma}$, Φ^*. Figure 7 plots the average absolute contributions of the variables to the predicted coefficients $\tilde{\gamma}_i$, for those variables that have been attributed non-zero weights by the elastic net. Note that, for the ease of representation, the graph aggregates the contributions of the variables in levels and their non-linear transformations (squares and square roots). Similar to figure 2 in section 4.3, the depicted values are average absolute contributions: for each country, we multiply the value of a variable with the respective estimated coefficient ($\hat{\delta}$) – the variable's contribution to $\tilde{\gamma}_i$ – and then average these values across all countries.

When considering the determinants as elicited by the elastic net, it is important to bear in mind the limitations that are inherent to most methods of machine learning. While in predictive terms, these methods can often dramatically outper-

Figure 7: Contribution of determinants to $\tilde{\gamma}$

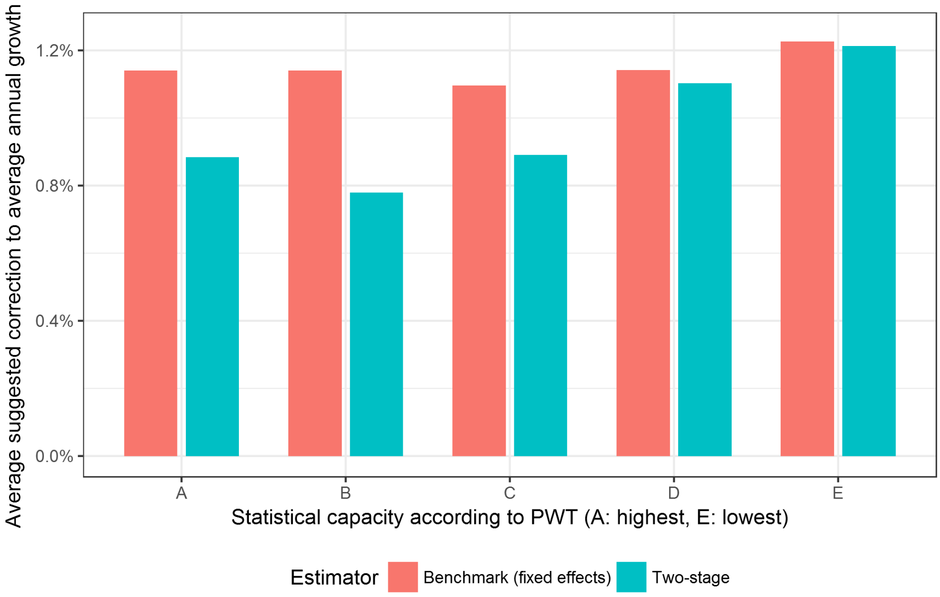


Notes: The bars depict the absolute contributions of each of the determinants of $\tilde{\gamma}_i$, on average across all countries. Non-linear transformations and the variables in levels are conflated here for readability. 'Industrial growth' etc. correspond to *relative* contributions of the sectors to total GDP growth, that is, they do not correspond to actual growth rates.

form traditional statistical methods, great care must be exerted when interpreting their output. As shown by Mullainathan and Spiess (2017), the resulting selection of variables and weights can be quite unstable, even across models that have similar predictive power. For the interested reader, we note however that among the largest six contributors, (i) the share industrial growth, population growth and the share of agricultural in GDP all *increase* $\tilde{\gamma}$, while (ii) the share of consumption and levels of GDP per capita *decrease* $\tilde{\gamma}$. For the share of manufacturing growth in total growth, the contribution is ambiguous, with the variable in levels entering with a negative sign, and the squared variable entering positively. The estimated coefficients are reported in appendix G. We are interested primarily in the predictive properties; in this respect, our variables can explain 55% of the variation in $\hat{\gamma}$ (as measured by the R^2), and the out-of-sample mean squared error at the optimal combination of α and λ is 0.07 (see appendix F for the cross-validation of α and λ).

We can now use the *expected* elasticities $\tilde{\gamma}_i$ conditional on country-characteristics Φ_i^* and luminosity values from each country to predict GDP levels, and then derive growth rates from the relative changes in these levels (see section 3.2). The discrepancy between the figures thereby obtained and the officially reported growth rates (according to WDI) is then what we call the 'suggested revision' (as in, suggested by the lights proxy). Figure 8 plots the amplitude of the revisions that

Figure 8: Average corrections by statistical capacity



Notes: The bars describe the average correction suggested by lights based on the respective estimator. While FE is indiscriminate as it makes similar corrections across countries of different statistical capacity, the twostage estimator discriminates in the sense that growth rates of poor data countries are corrected more.

are suggested by our two-stage method, compared to those obtained from a fixed effects estimation that assumes a single γ. It disaggregates this effect by the statistical quality of the countries as rated in the Penn World Table (see Summers and Heston, 1991), with grades ranging from A (best, e.g., USA) to E (worst, e.g., Chad).[15] Two things can be noted: First, our estimator generally applies smaller corrections to the countries' average annual growth rates than the fixed effects estimates (0.91 pp vs. 1.15 pp across the entire sample). Second, its estimates suggest substantially smaller revisions on average for countries with supposedly better data quality, and larger corrections to places with supposedly less reliable data. This pattern is of course desirable, but almost entirely absent in the alternative estimates based on a single slope: there, the suggested revisions are equally large across levels of statistical quality. We consider this observation indicative of the idea that our estimates are less prone to suggest undue revisions in places where the reported data is in fact accurate, a property that is also suggested by our simulation exercise in section 4.

[15]Recent vintages of the PWT do not report such country grades anymore. An alternative measure of statistical capacity is provided in the WDI, but this measure is not provided for most high-income countries. Furthermore, the PWT estimates date from the beginning of our sample period, and should therefore give a good representation of the initial quality of statistics.

8 Application: Economic growth in Africa, 1992-2013

We now turn to the assessment of recent growth performance in sub-Saharan Africa. First, we will briefly review the academic debate, and then present our own results. Note that in our own analysis, we do not engage in any formal hypothesis testing in the statistical sense of the word. Instead, we seek to provide well-grounded indicative evidence of growth patterns in SSA and whether these are reflected in the luminosity data; our focus is therefore on the point estimates.[16]

8.1 An 'African Growth Miracle'?

Perhaps the most optimistic view on recent economic developments in SSA comes from Young (2012), who summarises his findings as the 'African Growth Miracle'. Based on increases in assets owned by households observed in Demographic and Health Surveys (DHS) since 1987 and using education as a proxy for income, he estimates that real household consumption per capita in sub-Saharan Africa has been growing at an annual rate of 3.4% to 3.7% between 1991 and 2004. That is, up to four times the rate reported in national accounts figures. From a methodologically very different perspective, Pinkovskiy and Sala-i Martin (2014a,b) arrive at similarly optimistic conclusions. Combining growth rates from national accounts with distributional data derived from household surveys, and assuming a log-normal distribution of incomes, they find that poverty in Africa had been falling at a much higher rate then previously thought. While optimistic with regards to poverty, unlike Young (2012), they do not challenge aggregate figures on GDP or consumption.[17]

Not all research shares this optimism. Harttgen et al. (2013), for instance, strongly question the empirical foundations of Young's 'African Growth Miracle'. They highlight that Young's conclusions rest on a number of assumptions that are unlikely to be met in reality. Crucially, Young assumes constant income elasticities for assets, ignoring changes in preferences or relative prices. Using educational attainment as a proxy for income, his results also rest on the assumption of constant

[16]The correct estimation of standard errors for coefficient resulting from estimators with regularisation terms, such as the elastic net we employ, is an open debate. For the elastic net, there exists no analytical solution to derive standard errors. One possible avenue would be to bootstrap the errors, but this does not overcome some fundamental conceptual issues. See, e.g., Dezeure et al. (2015) for a more comprehensive discussion of the issue.

[17]In fact, Pinkovskiy and Sala-i Martin (2014b) explicitly compare the accuracy of national accounts data versus survey data as a measure of true income using light emissions at night as a third and independent measure of income. They conclude that national accounts data consistently outperform aggregate income estimates derived from household surveys.

returns to education, ignoring the existence of a substantial body of literature indicating that this is not the case. Beyond these methodological objections, Harttgen et al. (2013) question the validity of asset based measures of consumption in general: They show that, empirically, there is almost no correlation between the two measures. Growth in asset ownership would therefore be more likely to reflect lower disposal rates (new items last longer), or shifts in preferences and relative prices, than unobserved growth in consumption. They conclude that the traditional view, whereby African economies stagnated until 1995 and then started to grow, cannot convincingly be challenged on these grounds. In contrast to what Young finds using changes in asset ownership, those income or consumption measures that are derived immediately from survey means typically indicate lower growth than those reported in national accounts data. Mediating between Young (2012) and Harttgen et al. (2013), Johnston and Abreu (2016) note that asset based indexes are best used as a proxy for changes in wealth, rather than for changes in income. While Young's calculations may not be a fair challenge of national accounts data, they would however reveal substantial increases in household wealth in wide parts of the continent.

Jerven (2014) reviews the evidence on the growth performance across Africa over the past two decades, and reaches the conclusion that recent growth rates are likely to have been over-reported. This is because GDP levels as such are typically *under*estimated, as witnessed by recent revisions of GDP figures in many countries across SSA, e.g., 63% in Ghana or 60% in Nigeria. As statistical capacity is increasing, economic activity is being registered that previously went unnoticed. This 'statistical growth', as Jerven labels it, appears to be particularly prevalent in the years that precede a GDP rebasing. Because rebasing is a process that spans over several years, the direction of future GDP revisions can become clear long before the actual rebasing is declared; it may then be in the best interest of statistical offices and incumbent governments to attribute much of the newly measured economic activity to recent growth. Jerven emphasises that there is likely to be substantial heterogeneity in the misreporting of recent growth figures, and that countries with recent revisions in GDP levels are more likely to have inflated recent growth figures than those where no such revisions took place.

Although they do not explicitly focus on African countries, Henderson et al. (2012)'s examination of growth patterns based on changes in luminosity still contributes to this debate. 24 out of 30 countries for which they present revised growth rates (the countries with the lowest statistical capacity rating by the World Bank) are African. The suggested revisions are negative for 13 of these countries, and positive for 11; on average, their results suggest a moderate upwards correction of about 0.12 pp annually between 1992 and 2006 for the 24 African countries with the lowest statistical capacity. The aim of the statistical exercise in this paper is to re-assess and refine these estimates, accounting for the fact that the relationship between luminosity and GDP growth may indeed differ across countries. Furthermore, we refine the analysis by considering growth patterns within sub-periods, and across relevant groupings of countries.

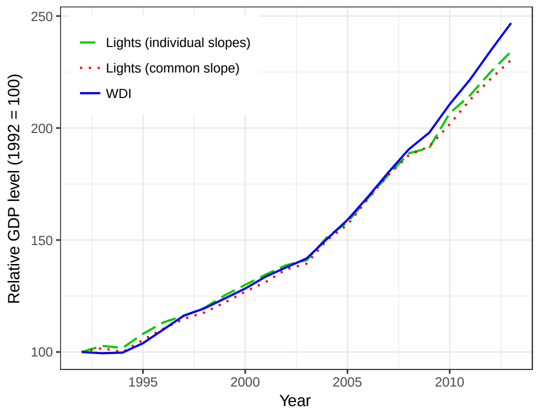
8.2 The aggregate picture

We now assess the sub-Saharan African growth experience from 1992 to 2013 in the light of our estimates from an aggregate persepective. Besides looking at the continent as a whole, we will disaggregate it into relevant groups in two ways. First, we consider growth experiences of coastal, landlocked and resource-rich countries separately; this typology by 'opportunity group' was introduced by (Collier and O'Connell, 2009), and has arguably dominated and structured much of the debate about African growth in recent years (see Thorbecke, 2015, for a comprehensive overview of typologies). If there was evidence that across these groups, there has been a substantial bias in reported GDP figures, this could be considered an indication that at least parts the recent discourse have been flawed. Second, we sort countries into groups according to whether they recently updated their base years after a prolonged period of time, in order to examine whether there is evidence supporting Jerven (2014)'s hypothesis of 'statistical growth'.

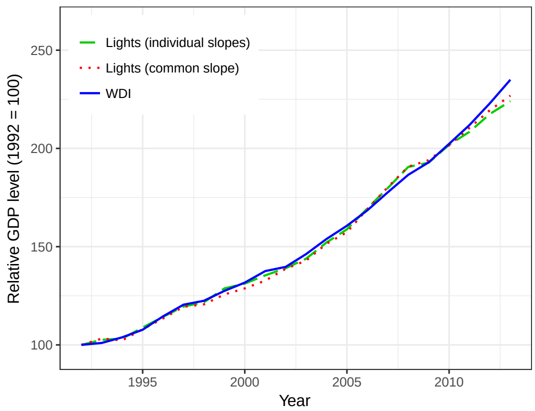
We start by looking at the African continent as a whole, to assess whether there is any support for an 'African Growth Miracle' as declared by Young (2012) in the lights data. Figure 9 plots the trajectory of real GDP across three sub-groups of African countries in our final sample, as well as across all these countries as a whole. The reported values are the unweighted averages of real GDP levels, normalised to their 1992 values. The solid blue line is derived directly from the official values, as reported in the WDI 2015. The dotted red line reports the predictions based on luminosity under the assumption of a common elasticity between lights and GDP across all countries (the predictions obtained from equation 5, or the model reported in column 6, table 5). The dashed green line plots the predictions obtained from our two-step method with heterogeneous slopes as described in section 3 and 4, and therefore accounts for differences in the relationship between luminosity and GDP across countries.

Panel (a) presents the most aggregate picture of the evolution of real GDP across the African countries in our sample. As suggested by the trajectory of the blue line, the WDI report consistently high growth rates starting from 1994. Across the entire period, the reported growth rates in the sample average at 4.28%. A noticeable acceleration in growth rates occurs around 2003: the average growth rate between 1993 and 2003 is 3.34%, compared to 5.32% between 2004 and 2013. Throughout the two decades we consider here, there appears to be, on average across countries, no substantial discrepancy between officially reported growth rates and those suggested by the luminosity series, whether we allow for heterogeneity in the γs or not. The series follow each other very closely, and only from 2007 onwards there is a slight discrepancy. However, as opposed to the narrative of an 'African Growth Miracle' that would have gone unrecorded in the official national accounts data, this discrepancy is indeed negative, with the luminosity measure suggesting slightly less economic growth than the official data. Note however that the discrepancy is moderate: Over the entire period, the average growth rate suggested by luminosity (with heterogeneous slopes) is 4.14%, 0.14 pp below what the official data suggests. For the period after 2007, these figures are 5.32% and 0.19 pp respectively. Overall, our aggregate results confirm Henderson et al. (2012)'s finding, whereby there does not appear to be a systematic directional bias in the mis-reporting or mis-measurement of GDP figures in SSA over the past decades.

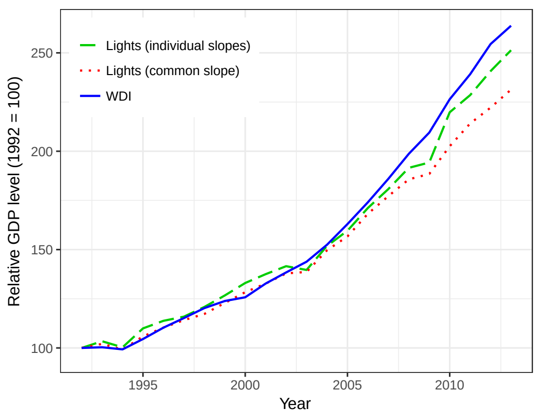
Figure 9: Growth performance by opportunity groups (Collier and O'Connell, 2009)



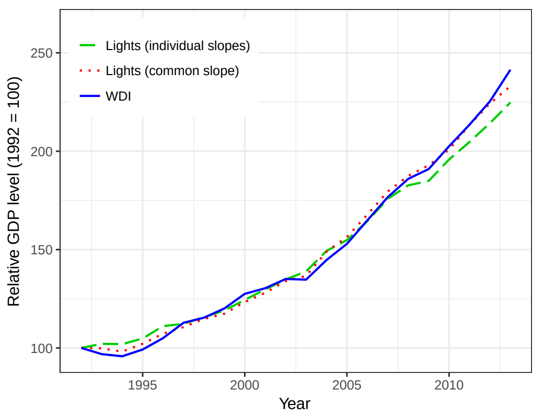
(a) Africa (all)



(b) Coastal countries



(c) Landlocked countries



(d) Resource-rich countries

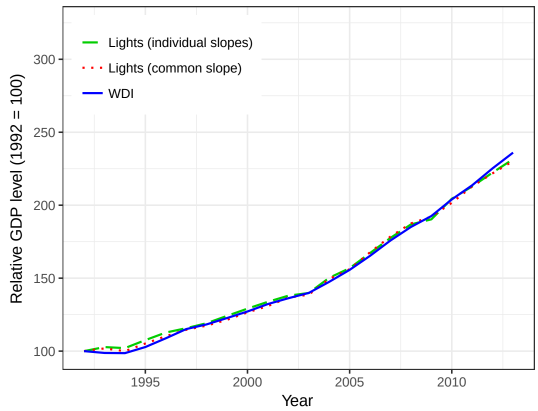
Coastal, landlocked and resource-rich countries: Arguably the most influential economic typology of African countries is that by Collier and O'Connell (2009), who divide them into (resource-scarce) coastal, (resource-scarce) landlocked, and resource-rich countries (Collier and O'Connell, 2009, p. 126–127). For the period they consider, 1960–2000, they find that these are central defining features of the growth performance of sub-Saharan African countries and developing countries in general; while growth in SSA was disappointing during that period across all three categories, it was worst for the category of landlocked countries, a feature that is commonly considered a major impediment to the participation in international trade and economic growth. Implicitly or explicitly, much of the discourse about the economic performance of African countries has evolved along these lines; if growth rates had been systematically mis-reported across these categories over the past two decades, this may therefore likely have flawed considerable parts of the debate.

Panels (b)-(d) disaggregate the African sample into the categories suggested by Collier and O'Connell (2009).[18] All of the series we report agree on the stylised fact that countries of all three categories have, on average, experienced substantial growth over the past two decades. The most substantial discrepancy between official growth rates and those suggested by changes in luminosity is for landlocked countries: The WDI suggest some convergence here, with the historically stagnating landlocked countries growing at 4.42% annually, as opposed to 3.89% for coastal and 4.29% for resource-rich countries respectively. Under the assumption of a common slope across all countries, the luminosity data is somewhat more pessimistic in that respect. For the later part of the sample period (2007 onwards), it indicates an average growth rate of 4.11% as opposed to the 4.42% suggested by WDI for landlocked economies. However, once we account for the characteristics that determine the relationship between GDP and lights, most of this discrepancy is bridged; in fact, the arithmetic average of annual growth rates suggested by the two-stage estimator is slightly higher than what WDI suggests, at 4.50%.

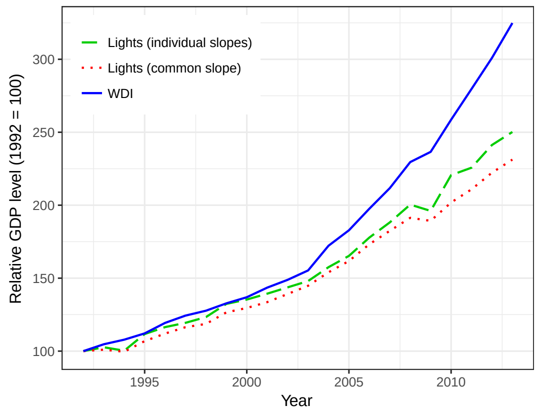
Countries that recently rebased GDP: Next, we seek to examine the hypothesis that those African economies that have recently performed major rebasings of their GDPs tended to report inflated recent growth rates (Jerven, 2014). As discussed in section 8.1, this could be the case for a number of resasons: First, the process of rebasing GDP estimates takes several years, during which the statistical capacity typically increases. Some previously unrecorded activity may therefore be

[18]Note that Collier and O'Connell (2009, p. 77) report per capita figures and weight countries by their population numbers when computing growth rates, as they seek to best describe the 'experience of the typical African'. We discuss total GDP and focus on the accuracy of the reported figures, and countries are therefore given equal weight.

Figure 10: Growth performance of countries with recent GDP revisions



(a) No major rebase since 2000

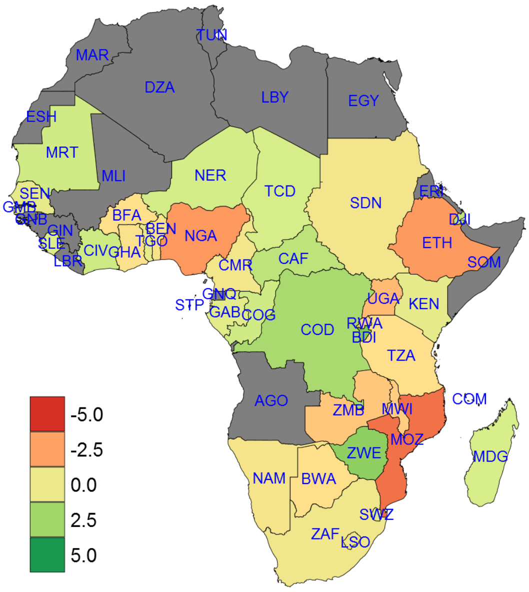


(b) Major rebase since 2000

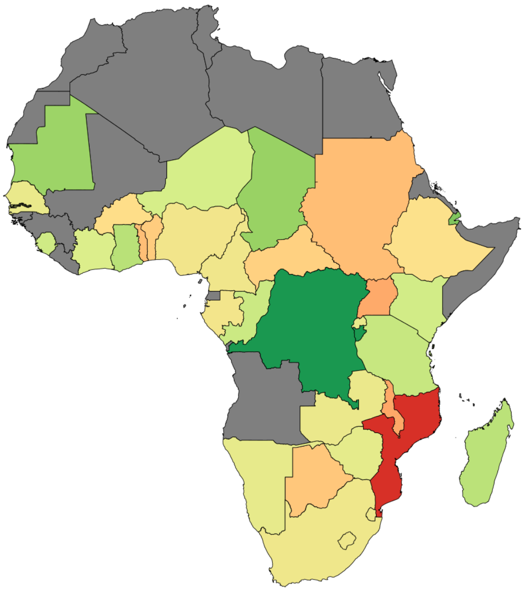
recorded, showing up as economic growth in official figures ('statistical growth'). Second, a rebasing will increase the accuracy of current GDP levels, but it does not as such give any indication as to when the discrepancy between old and new estimates came about - that is, when economic growth occured (or did not occur, in the rare case of downward revisions). It can then simply be easier to attribute it to recent years. Third, it is politically advantageous for current leaders to be associated with good economic performance, and governments may opportunistically push for growth to be statistically attributed to the present.

Figure 10 is identical to figure 9 in its interpretation, but splits the African sample into those countries that recently had a major revision to their base year (panel (b)), and those who didn't (panel (a)). We define those as the countries where (i) GDP has been rebased after the year 2000, and (ii) the previous base year was at least 10 years before the new base year (the IMF recommends an update to the base year every 5 years). This is the case for 5 countries, namely Botswana, Ethiopia, Ghana, Niger and Nigeria. Indeed, the discrepancy is large, in particular after 2003: While the average growth rate according to WDI in these countries in this period is 7.49% (compared to 4.92% in the countries that didn't revise), their performance hardly exceeds the continent's average when proxied by lights. This is true both under the assumption of slope heterogeneity, suggesting 5.44% on average between 2003 and 2013 and when imposing a single slope for all countries (4.83%). The exception among these countries is Niger, where the two-stage estimates suggest a growth rate 1.10 pp higher than WDI (0.06 pp under the assumption of a common slope).

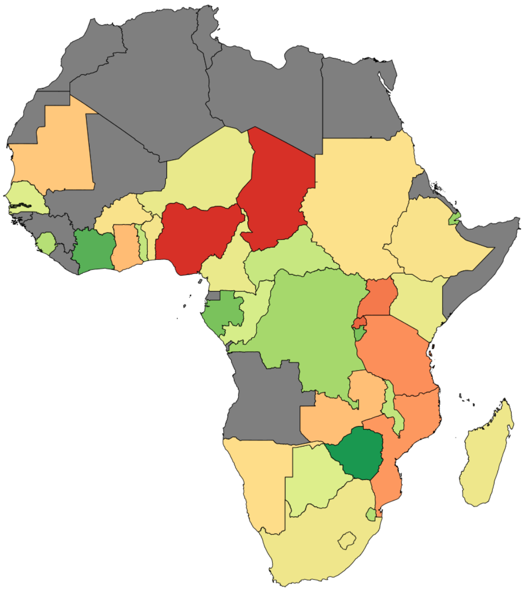
Figure 11: Growth according to official vs. lights data (individual slopes)



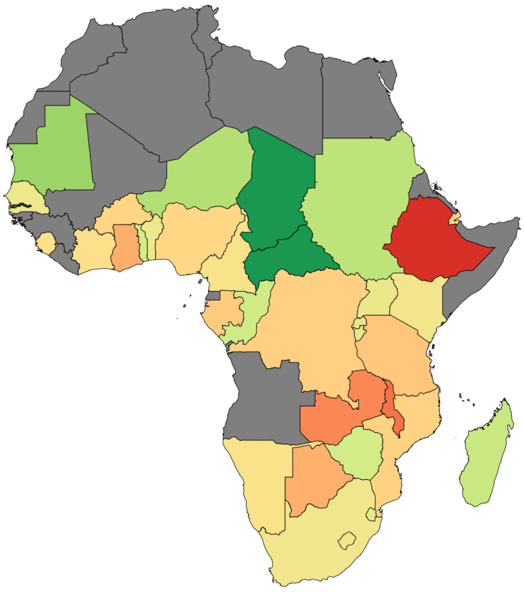
(a) Whole period 1993-2013



(b) 1993-1999



(c) 2000-2006



(d) 2007-2013

Notes: Colours indicate the discrepancy between average WDI growth rates and those obtained from the luminosity proxy, with positive numbers indicating that the luminosity proxy indicates higher growth. The colour coding is capped at -5 pp and 5 pp, meaning that individual discrepancies can be larger. This is the case for MOZ, COD and BDI in panel (b), NGA, TCD and ZWE in panel (c), and ETH, TCD and CAF in panel (d) (see appendix H). Borders are obtained from maplibrary.org for illustrative purposes only, and the authors do not imply the expression of any opinion concerning the legal status of any country, area or territory.

8.3 Country-wise estimates

While a comprehensive discussion of individual countries' growth records and their conformance with the luminosity proxy is beyond the scope of this paper, we will discuss some general patterns and particularly noteworthy cases. The maps in figure 11 provide an overview over the discrepancies between the growth rates reported in WDI and those predicted with the luminosity data, allowing for heterogeneity in γ. Equivalent figures under the assumption of a homogeneous γ as well as all underlying values in tabular form are provided in appendix H. Negative deviations (red) indicate that the lights measure suggests lower growth rates than WDI, positive deviations (green) mean that lights suggest higher growth rates than WDI. Note that the colour coding is capped at 5 pp and -5 pp deviation in average annual growth rates, which is exceeded in a few cases (see notes of the figure). The figure is divided into four panels: panel (a) depicts the average growth rates over the entire period, 1993-2013 (the first period now being 1993, as no growth rate can be computed for 1992 which is where our series in levels starts). Panels (b)-(d) divide the period into three sub-periods of 7 years each.

As a general pattern, the luminosity proxy tends to relativise some of the more extreme growth experiences, both high and low. For instance, a number of stellar growth records do not find their equivalent in terms of luminosity emissions. This is most pronounced for Mozambique, where growth rates average at 8.5% in WDI. Our luminosity estimates suggest a figure that is 3.6 pp lower – a substantial discrepancy, although 4.9% growth annually is still far above the average growth rate on the continent. Note also that the discrepancy is to the largest part concentrated on the first sub-period between 1993 and 1999. While this coincides with the country's transition to democracy and a substantial wave of repatriation (Sheldon and Penvenne, 2018), reported average growth rates of 10.6% (26.8% in 1996) are not corroborated by light emissions (suggesting instead only 3.8% of growth annually on average over that period). A similarly large discrepancy emerges in the case of Nigeria, one of the countries that recently revised its GDP estimates in a major way: WDI suggest an average growth rate of 6.0% per annum over the entire period, luminosity only 3.2%. Again, the disagreement between the luminosity series and official figures are highly concentrated in time. Between 2000 and 2006, WDI report an average annual growth rate of 9.9%, with a maximum of 33.7% in 2004. This spike is not reflected in the luminosity series, which suggests a much more moderate 3.3%. In Ethiopia, the discrepancy between the luminosity based growth rates and the reported ones is concentrated in the latest period between 2007 and 2013. Where WDI report an average growth rate of 11.6%, our luminosity measure suggests 4.11%.

At the other end of the spectrum, a few notoriously under-performing countries have increased their levels of luminosity far beyond what their official growth record suggests. Interestingly, among the countries with the greatest positive divergence between the luminosity proxy and WDI growth rates, there appears to be a bunching of particularly conflict torn countries: The greatest positive discrepancies occur in Burundi (3.3 pp), Zimbabwe (3.0 pp), and the Democratic Republic of Congo (2.31 pp). In Burundi, the episode where the WDI and luminosity disagree most coincides with a prolonged period of extreme ethnic violence and civil war. While between 1993 and 1999, the WDI register a decline of, on average, 3.4% of GDP, luminosity increases to an extent that would suggest economic growth of about 3.1% annually – a difference of 6.5 pp. In the Democratic Republic of Congo, the worst official growth record is also in the first period from 1993 to 1999, a period encompassing the First Congo War (1996–1997). WDI report an average decline of 4.17% for this period (–13.5% in 1993 alone). The evolution of the luminosity data suggests a slow increase in economic activity of 1.72% per year. For Zimbabwe, WDI report an average decline of about 6.1% between 2000 and 2006, again a period marked by severe political tensions, including wide-spread government violence. Luminosity values, however, suggest that the economy would still have grown at 1.5% annually.

Evidently, many regularities collapse in periods of armed conflict, and the relationship between luminosity and GDP may well be one of them. For instance, the destruction of a power plant can abruptly stop entire cities from emitting artificial light, while much of the regular economic activity still takes place. On the other hand, warfare in itself is an activity that may generate substantial amounts of luminosity; at the extreme, a process of destruction may then be misinterpreted as constructive economic activity. Once more, we highlight that we consider our estimations as merely indicative of potential mis-measurement or mis-reporting. While our methodology aims at incorporating some of the factors that alter the relationship between lights and GDP, there are natural limits to this ambition. Especially where the interest lies in individual countries or episodes, interested scholars should carefully weigh the full body of evidence. Historical and institutional knowledge, as well as official statistics, should in our opinion form the main pillars of this evidence, and the luminosity can be a valuable complement where it is weak otherwise.

9 Conclusion

The close relationship between economic activity and light emissions can offer a unique perspective on historical growth records, especially where the data is weak.

In this study, we raised the issue of heterogeneity, in the sense that economic activity may translate into luminosity at different rates in different countries. We offer a method that allows for the relationship to differ, to the extent where this can be explained by observable country characteristics. We then apply this methodology to economic growth in Africa over the past two decades.

At the core, our methodology seeks to disaggregate the elasticity between lights and GDP, allowing for it to differ across countries. To this end, we split the estimation of the relevant coefficients into two parts: First, we run a conventional fixed effects regression, but including a full set of $country \times lights$ interaction terms in order to allow for the slope to differ across countries. The resulting estimates reveal a strong variation in the relationship across countries, but the results have little value where the goal is to discern potential mis-measurement or mis-reporting: Any predictions from this model would merely tend to replicate each country's reported GDP series. This is why, in a second step, we apply the elastic net estimator to discern the part of the variation in the elasticity-coefficients that can be attributed to observable country characteristics. On this basis, we can then infer *expected* elasticities, conditional on countries' economic and geographic characteristics. The resulting coefficients are then used to derive GDP growth rates based on luminosity emissions.

We explore some basic properties of this approach in a simple simulation exercise. The results suggest that, compared to methods that assume a single coefficient across all countries, our method can offer substantial improvements in the accuracy of the inferred growth rates. This is mainly due to the the fact that our estimator suggests less false revisions where the data is indeed accurate. Our simulations show that the relative success of the method strongly hinges on the predictability of the elasticity coefficients based on observable characteristics in a way that is generalisable beyond the narrow sample. With about 55% of the variation in the parameter of interest ($\hat{\gamma}$) explained in our empirical exercise, we believe that our estimates constitute an improvement over those obtained from more conventional methods that assume a common coefficient for the relationship between GDP and lights across all countries. Indeed, we find that the revisions suggested by our estimator are substantially smaller for countries with good statistical capacity than for those with poor statistical capacity, a desirable pattern that does not occur when falsely assuming a common slope.

We apply our estimates to the debate about the recent growth performance of African countries. Our results do not lend support to claims of an 'African Growth Miracle' (Young, 2012), according to which African countries would have grown at a rate several times higher than those recorded in the national accounts data. Instead, our estimates suggest that, overall, the luminosity emissions of African countries appear to be in line with what official estimates suggest. When looking at countries that recently rebased their GDPs after a long period of time (Botswana, Ethiopia, Ghana, Niger and Nigeria), we find that these countries' reported growth rates tend to exceed the growth rates suggested by the luminosity proxy quite substantially. This is in line with Jerven (2014)'s hypothesis of 'statistical growth', according to which already existing economic activity is discovered with increased statistical efforts, and in parts spuriously attributed to recent years.

When looking at individual countries' growth performances, the main pattern is that the most extreme growth records according to the national accounts data – both negative and positive – are typically relativised by the luminosity proxy. For instance, the stellar growth performances of Mozambique and Ethiopia over the past two decades do not find confirmation in the luminosity data. While the proxy still suggests that these countries grew substantially above average, it does not corroborate official series that in some years suggest GDP growth rates of more than 25%. On the other hand, some extremely negative growth records are corrected upwards, and the most substantial discrepancies on this end of the spectrum are for Burundi, Zimbabwe and the Central African Republic. Interestingly, the largest upwards revisions appear to be concentrated around periods of armed conflict.

This latter observations illustrates our case that growth estimates based on luminosity must be generally considered with care, and evaluated within the specific context: For instance, in periods of armed conflict, the mechanisms that are normally at work in a country cease to function. This is true for the collection of economic data by national statistical offices, but likely also for the interaction between GDP and lights. A damaged power plant may cause the remaining economic activity of an entire city to go unregistered by the proxy, while luminosity caused by purely destructive warfare activities may falsely be picked up as economic growth.

The study sought to incorporate in its methodological framework some of the country-specific context, by allowing the elasticity between GDP and lights to vary with a country's observable characteristics. This does not, of course, cover every aspect of how luminosity is linked to GDP. This relationship may vary over time, and be undermined by certain events, observable and unobservable ones. When assessing any country's historical growth performance, researchers have a plethora of evidence to consider. Besides historical and institutional knowledge, national accounts data, and individual and household level surveys, night time luminosity should be treated as a valuable complement.

References

Alesina, A. F., Michalopoulos, S., and Papaioannou, E. (2012). Ethnic inequality. *NBER Working Paper*, (18512).

Bickenbach, F., Bode, E., Nunnenkamp, P., and Söder, M. (2016). Night lights and regional GDP. *Review of World Economics*, 152(2):425–447.

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.

Collier, P. and O'Connell, S. (2009). Opportunities and choices. In Ndulu, B. J., O'Connell, S. A., Bates, R. H., Collier, P., and Soludo, C. C., editors, *The Political Economy of Economic Growth in Africa, 1960–2000*, volume 1, pages 76–136. Cambridge University Press.

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

Dai, Z., Hu, Y., and Zhao, G. (2017). The suitability of different nighttime light data for GDP estimation at different spatial scales and regional levels. *Sustainability*, 9(2):305.

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

Devarajan, S. (2013). Africa's statistical tragedy. *Review of Income and Wealth*, 59(S1):S9–S15.

Dezeure, R., Bühlmann, P., Meier, L., Meinshausen, N., et al. (2015). High-dimensional inference: Confidence intervals, p-values and r-software hdi. *Statistical science*, 30(4):533–558.

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.

Easterly, W. and Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics*, pages 1203–1250.

Ebener, S., Murray, C., Tandon, A., and Elvidge, C. C. (2005). From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics*, 4(1):5.

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

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).

Harttgen, K., Klasen, S., and Vollmer, S. (2013). An African growth miracle? Or: what do asset indices tell us about trends in economic performance? *Review of Income and Wealth*, 59(S1):S37–S61.

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*, 129:995–1033.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hoover, K. D. and Perez, S. J. (2004). Truth and robustness in cross-country growth regressions. *Oxford Bulletin of Economics and Statistics*, 66(5):765–798.

Jerven, M. (2013). *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It.* Cornell University Press.

Jerven, M. (2014). African growth miracle or statistical tragedy? Interpreting trends in the data over the past two decades. *WIDER Working Paper 2014/114*.

Jerven, M. (2015). *Africa: why economists get it wrong.* Zed Books, London.

Johnston, D. and Abreu, A. (2016). The asset debates: How (not) to use asset indices to measure well-being and the middle class in Africa. *African Affairs*, 115(460):399–418.

Keola, S., Andersson, M., and Hall, O. (2015). Monitoring economic development from space: using nighttime light and land cover data to measure economic growth. *World Development*, 66:322–334.

Levin, N. and Zhang, Q. (2017). A global analysis of factors controlling VIIRS nighttime light levels from densely populated areas. *Remote Sensing of Environment*, 190:366–382.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.

Magee, C. S. and Doces, J. A. (2015). Reconsidering regime type and growth: lies, dictatorships, and statistics. *International Studies Quarterly*, 59(2):223–237.

Marshall, M. G. and Jaggers, K. (2002). Polity IV project: Political regime characteristics and transitions, 1800-2002. *Center for Systemic Peace*.

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

Mullainathan, S. and Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106.

NNBS (2013). FAQ on the rebasing / re-benchmarking of Nigeria's gross domestic product (GDP). *Nigerian National Bureau of Statistics, https://tinyurl.com/yb2jukar*.

Pinkovskiy, M. and Sala-i Martin, X. (2014a). Africa is on time. *Journal of Economic Growth*, 19(3):311–338.

Pinkovskiy, M. and Sala-i Martin, X. (2014b). Lights, camera,... income!: Estimating poverty using national accounts, survey means, and lights. *NBER Working Paper*, (19831).

Sheldon, K. E. and Penvenne, J. M. (2018). Mozambique. In *Encyclopædia Britannica*. Encyclopædia Britannica, inc.

Summers, R. and Heston, A. (1991). The Penn World Table (mark 5): an expanded set of international comparisons, 1950–1988. *The Quarterly Journal of Economics*, 106(2):327–368.

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 sub-national scales using nighttime satellite imagery. *International Journal of Ecological Economics & Statistics*, 8(S07):5–21.

Tanaka, K. and Keola, S. (2017). Shedding light on the shadow economy: a nighttime light approach. *The Journal of Development Studies*, 53(1):32–48.

Thorbecke, E. (2015). The anatomy of growth and development in sub saharan africa. In McKay, A. and Thorbecke, E., editors, *Economic Growth and Poverty Reduction in Sub-Saharan Africa*, chapter 2, pages 20–50. Oxford University Press, Oxford, UK.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Ul Haque, N., Pesaran, M. H., and Sharma, S. (1999). Neglected heterogeneity and dynamics in cross-country savings regressions. *IMF Working Paper*, 99/128.

Wu, J., Wang, Z., Li, W., and Peng, J. (2013). Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery. *Remote Sensing of Environment*, 134:111–119.

Young, A. (2012). The African growth miracle. *Journal of Political Economy*, 120(4).

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Appendix Chapter 2

A Sorting of countries in to groups/clusters

Table 6: Sorting of countries into groups

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Africa				Swaziland	2	2	1
Angola	1	1	2	Chad	1	1	2
Burundi	2	1	2	Togo	2	1	2
Benin	2	2	3	Tunisia	2	2	2
Burkina Faso	2	2	2	Tanzania	2	1	2
Botswana	2	3	3	Uganda	2	1	2
CAF	2	1	2	South Africa	2	3	3
Côte D'Ivoire	2	2	2	Zambia	2	2	2
Cameroon	2	2	2	Zimbabwe	2	2	2
Congo, DR	2	2	2	**Americas**			
Congo, Republic	1	1	2	Argentina	2	2	3
Djibouti	2	3	2	Bahamas	2	3	-
Algeria	1	1	2	Belize	2	2	-
Egypt	2	2	2	Bolivia	2	2	3
Eritrea	2	2	1	Brazil	2	3	3
Ethiopia	2	1	2	Canada	2	3	3
Gabon	1	1	2	Chile	2	3	3
Ghana	2	1	2	Colombia	2	2	3
Guinea	2	1	2	Costa Rica	2	2	3
Gambia	2	1	2	Cuba	2	3	1
Guinea Bissau	2	1	2	Dom. Rep.	2	2	3
Kenya	2	2	2	Ecuador	2	2	3
Liberia	2	-	2	Guatemala	2	2	3
Libya	1	1	2	Guyana	2	1	3
Lesotho	2	2	3	Honduras	2	2	3
Morocco	2	2	2	Haiti	2	-	2
Madagascar	2	2	3	Jamaica	2	3	3
Mali	2	-	3	Mexico	2	3	3
Mozambique	2	2	2	Nicaragua	2	2	3
Mauritania	2	1	2	Panama	2	3	3
Mauritius	2	3	3	Peru	2	2	3
Malawi	2	2	2	Puerto Rico	2	2	-
Namibia	2	2	3	Paraguay	2	2	3
Niger	2	1	2	El Salvador	2	2	3

Continued on next page

Table 6: Sorting of countries into groups (continued)

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Nigeria	1	1	2	Suriname	2	2	2
Rwanda	2	1	2	Trinidad & Tob.	1	3	3
Sudan	2	1	2	Uruguay	2	2	3
Senegal	2	2	2	USA	2	3	3
Sierra Leone	2	1	2	Venezuela	1	2	2
South Sudan	-	-	2				

Continued on next page

Table 6: Sorting of countries into groups (continued)

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Asia				**Europe**			
Afghanistan	2	2	2				
UAE	1	-	1	Albania	2	1	3
Armenia	2	2	2	Austria	2	3	3
Azerbaijan	1	1	1	Belgium	2	3	3
Bangladesh	2	2	2	Bulgaria	2	-	3
Bahrain	1	3	1	Bosnia and Herz.	2	2	2
Brunei	1	3	-	Belarus	2	2	2
China	2	2	1	Switzerland	2	3	3
Cyprus	2	3	3	Czech Republic	2	3	3
Georgia	2	2	2	Germany	2	3	3
Indonesia	2	2	2	Denmark	2	3	3
India	2	2	3	Spain	2	3	3
Iran	1	2	2	Estonia	2	3	3
Iraq	1	-	2	Finland	2	3	3
Israel	2	-	3	France	2	3	3
Jordan	2	3	2	United Kingdom	2	3	3
Japan	2	3	3	Greece	2	3	3
Kazakhstan	1	2	2	Croatia	2	3	2
Kyrgyzstan	2	2	2	Hungary	2	3	3
Cambodia	2	2	2	Ireland	2	3	3
Korea	2	3	3	Iceland	2	3	-
Kuwait	1	3	1	Italy	2	3	3
Laos	2	1	1	Lithuania	2	3	3
Lebanon	2	3	3	Luxembourg	2	3	3
Sri Lanka	2	2	2	Latvia	2	3	3
Myanmar	2	1	2	Moldova	2	2	3
Mongolia	2	1	3	Macedonia	2	2	3
Malaysia	2	2	2	Montenegro	2	3	3
Nepal	2	1	2	Netherlands	2	3	3
Oman	1	3	1	Norway	2	3	3
Pakistan	2	2	2	Poland	2	3	3
Philippines	2	2	3	Portugal	2	3	3
Palestine	2	3	-	Romania	2	2	3
Qatar	1	3	1	Russia	1	3	2
Saudi Arabia	1	3	1	Serbia	2	2	3

Continued on next page

Table 6: Sorting of countries into groups (continued)

	Fossil	Econ.	Regime		Fossil	Econ.	Regime
Syria	1	2	1	Slovakia	2	3	3
Thailand	2	2	2	Slovenia	2	3	3
Tajikistan	2	2	2	Sweden	2	3	3
Turkmenistan	1	2	1	Ukraine	2	2	3
Timor-Leste	2	1	-	**Oceania**			
Turkey	2	2	3	Australia	2	3	3
Uzbekistan	1	1	1	New Caledonia	2	3	-
Viet Nam	2	2	1	New Zealand	2	3	3
Yemen	1	1	2	Papua New G.	2	1	2

B Data Generating Process (detailed)

B.0.1 Income data (Y)

Initial GDP per capita: In order reflect the highly right-skewed distribution of national incomes, we choose an exponential distribution to generate initial values of GDP per capita. We scale up the series to values of a similar order of magnitude of empirically observed income levels as measured in US-Dollars:

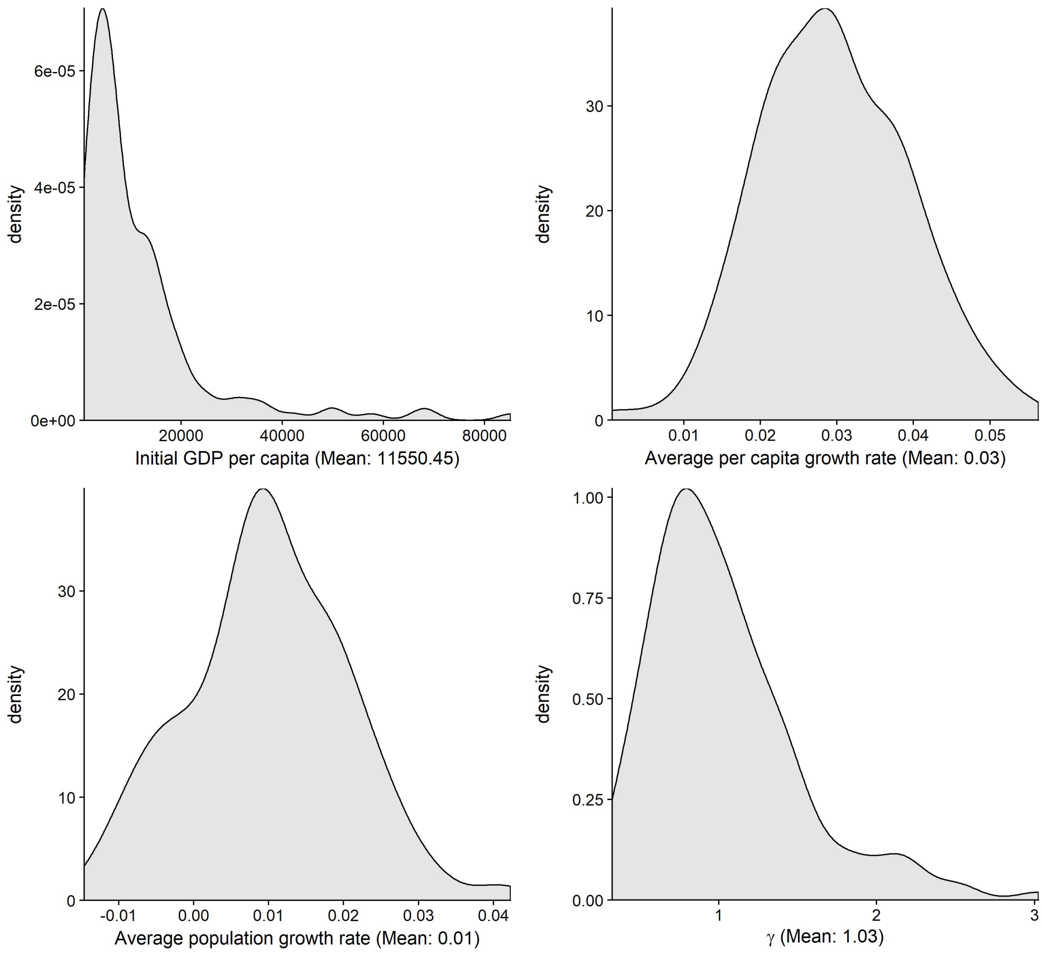
$$y_{init} \sim Exp(1) * 10.000 + 500 \tag{7}$$

The resulting distribution is depicted in the top left panel of figure 12.

Co-determinants of growth and γ: In order to satisfy requirement 4, we introduce three variables that simultaneously determine GDP growth rates of Y and the inverse elasticity between GDP and lights γ for each of the countries. First, we introduce one arbitrary uniformely distributed variable $v_1 \sim U(0.2, 0.8)$ that will both increase GDP growth rates and affect each country's γ (e.g., levels of investment). Second, we introduce some beta-convergence, with countries that have smaller initial GDP per capita values tending to have larger growth rates; GDP per capita will also affect γ. And third, population growth will naturally enter total GDP growth rates, and we include it as another determinant of γ (see below).

Per capita GDP growth: In order to approximately match the mean and variance of empirical per capita growth rates, we choose an unconditional mean of per capita growth rates of 0.03; the expected annual growth rate per capita for

Figure 12: Distribution of key series of the simulated data



Notes: Kernel density plots of the distribution of key values in the simulated data. The plotted distributions refer to the simulated data discussed in sections 4.1 to 4.3.

any given country is then $g_{i,det} = 0.03 + (v_1 - 0.5) + ((\mu_{y_{init}} - y_{init})/10 * \mu_{y_{init}})$, where $\mu_{y_{init}}$ is the mean initial GDP per capita across countries; the last term therefore introduces convergence. Beyond this deterministic component, we add some stochasticity, and the average growth rates in each country are are distributed $g_i \sim N(g_{i,det}, 0.01)$. Panel 2 in figure 12 plots the resulting distribution of average per capita growth rates.

Population growth: To obtain figures similar to empirical figures in their amplitude and distribution, we generate initial population numbers $pop_{init} \sim Exp(1) * 10^6$; these then grow at country-specific rates $pop_{growth} \sim N(0.1, 0.1)$ that remain constant across time. Figure 12, panel 3 plots the distribution of population growth rates.

The final GDP series is then simply $Y_t = y_t * pop_t$.

B.0.2 Generating an imperfect proxy ℓ (lights)

Next, we generate a proxy ℓ of y, where y induces ℓ with (inverse) elasticity γ and some measurement error. Importantly, we want to allow for γ to vary across units of observation (countries) based on a number of determinants. We therefore first introduce a basic DGP for γ.

DGP for γ: The country-specific component of the inverse elasticities γ are partly determined by the confounding factors (v_1, y_{init} and pop_{growth}) that affect both the elasticities as well as GDP growth, and partly by factors that only affect the elasticities. The latter are each generated as $det_j \sim N(0, SD_{det})$. We also generate a number of identically defined variables that *do not* enter the γs, in order to assess the variable selection properties of our proposed estimator. In the present set-up, we generate 20 of these normally distributed variables, 5 of which enter as determinants into the GDP predictions. The deterministic component of the γs is defined as $\gamma_{det} = c_\gamma + \sum_{j=1}^{5} det_j + f(v_1) + f(y_{init}) + f(pop_{growth})$, with $c_\gamma = 0.5$. We add an element of randomness by generating the final γs as $\gamma_i \sim N(\gamma_{det,i}, SD_\gamma)$. Panel 4 in figure 12 depicts the resulting distribution.

Proxy series ℓ: Based on the generated unit-specific γ_is, we can now generate luminosity measures from the true income measure Y. In line with section 3, we determine $L_{it} = Y_{it}^{\beta_i} * exp(\varepsilon_{it})$, where $\beta_i = \frac{1}{\gamma_i}$. Moreover, $\varepsilon_i t = \epsilon_t + \epsilon_i + \epsilon_{it}$, such that the measurement error in the lights-GDP relationship is a composite of a time specific component (e.g., variability in the sensor sensitivity) that will be accounted for with time fixed effects ($\epsilon_t \sim N(1, SD_{time})$ and an idiosyncratic error ($\epsilon_t \sim N(1, SD_{idio})$).

Figure 1 plots Y and L for 12 randomly selected units (countries). The solid line represents the simulated GDP series, and the dashed line represents L; as intended, L follows GDP with some error, and with different elasticities in different countries.

'Bad data' countries In order to simulate misreporting, we create another series of *reported* GDP (z). We therefore divide our sample into two groups: Good data countries, where the reported GDP data is accurate, that is, equal to the true series. And bad data countries, where the reported GDP growth rates are systematically biased. The direction and amplitude of the bias contains an element of randomness. We opt for the systematic error to be distributed uniformly in a range specified by *Range misrep.*. The number of bad data countries is determined by N_{bad}

C Results of simulation exercise

Figure 13: Predictive power across the parameter range

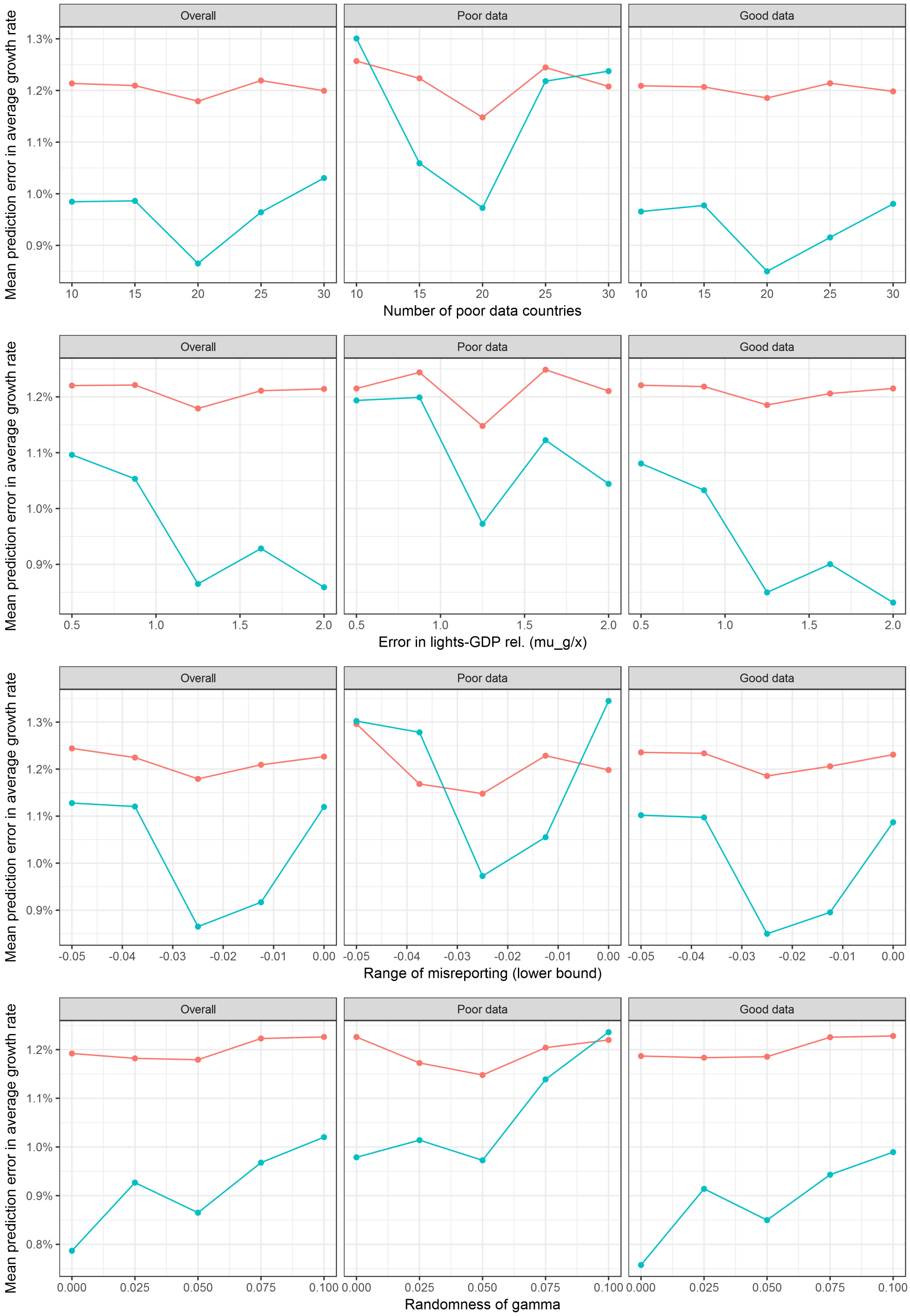


Figure 13 plots the mean error in the predicted average annual growth rates across the (simulated) sample for a range of parameters. Each row shows how this measure changes as the respective parameter (see label of the x-axis) is changed, keeping all other parameters constant. The values at which the (respectively) other parameters are kept constant are $N_{bad} = 20$, $SD_{\epsilon} = \mu_g/1.25$, $SD_{\gamma} = 0.05$ and a range of misreporting of $U[-0.25; 0.25]$ (almost identical to the illustratory example in the main body, with minor deviations as the intervals do not exactly coincide with those values). The blue lines refer to the two-stage estimator, the red line to the single-slope fixed effects.

D Estimated country-specific coefficients $\hat{\gamma}_i$

Table 7: Country-specific $\hat{\gamma}$

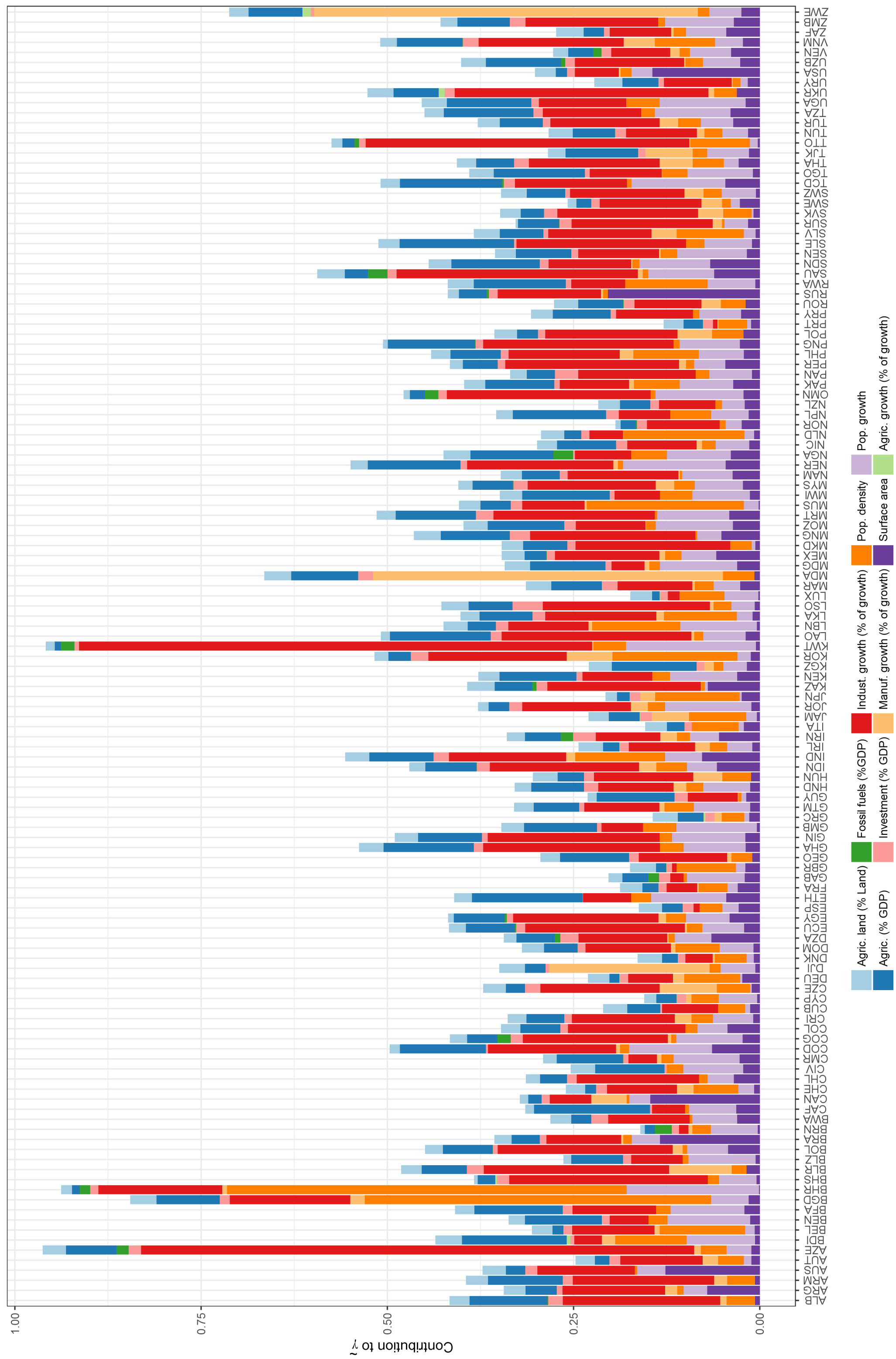
	$\hat{\gamma}$	SE		$\hat{\gamma}$	SE		$\hat{\gamma}$	SE		$\hat{\gamma}$	SE
AGO	0.44	0.02	DEU	-0.16	0.06	KEN	0.23	0.04	POL	0.24	0.03
ALB	0.29	0.03	DJI	0.04	0.03	KGZ	0.12	0.04	PRI	-0.24	0.11
ARE	0.41	0.05	DNK	-0.06	0.04	KOR	0.68	0.07	PRT	-0.26	0.05
ARG	0.16	0.05	DOM	0.52	0.05	KWT	0.79	0.07	PRY	0.07	0.04
ARM	0.50	0.03	DZA	0.14	0.04	LAO	0.36	0.02	ROU	0.04	0.03
AUS	0.27	0.08	ECU	0.12	0.04	LBN	0.14	0.03	RUS	0.24	0.04
AUT	-0.13	0.05	EGY	0.39	0.05	LBR	0.34	0.01	RWA	0.47	0.02
AZE	1.53	0.04	ESP	-0.05	0.08	LKA	0.45	0.04	SAU	0.34	0.04
BDI	0.03	0.03	ETH	0.54	0.03	LSO	0.16	0.04	SDN	0.42	0.03
BEL	0.05	0.07	FIN	-0.02	0.05	LUX	0.19	0.06	SEN	0.20	0.04
BEN	0.20	0.03	FRA	-0.27	0.08	MAR	0.23	0.04	SLE	0.30	0.02
BFA	0.45	0.03	GAB	-0.19	0.04	MDA	0.35	0.02	SLV	0.03	0.06
BGD	0.59	0.05	GBR	0.04	0.11	MDG	0.06	0.04	SUR	0.14	0.04
BGR	0.14	0.05	GEO	0.26	0.02	MEX	0.00	0.07	SVK	0.01	0.02
BHR	1.12	0.12	GHA	0.51	0.04	MKD	-0.08	0.05	SWE	0.00	0.02
BHS	-0.24	0.09	GIN	0.10	0.04	MLI	0.54	0.03	SWZ	-0.04	0.04
BLR	0.32	0.03	GMB	0.07	0.02	MNG	0.46	0.03	TCD	0.40	0.02
BLZ	0.22	0.03	GNB	0.20	0.01	MOZ	0.58	0.03	TGO	0.13	0.04
BOL	0.21	0.04	GRC	-0.20	0.06	MRT	0.27	0.04	THA	0.17	0.03
BRA	0.14	0.05	GTM	0.13	0.03	MUS	0.45	0.06	TJK	-0.09	0.08
BRN	-0.16	0.04	GUY	0.19	0.04	MWI	0.40	0.05	TKM	0.76	0.05
BWA	0.28	0.04	HKG	0.65	0.17	MYS	0.34	0.03	TTO	0.46	0.04
CAF	0.21	0.02	HND	0.13	0.03	NAM	0.36	0.05	TUN	0.32	0.05
CAN	0.05	0.03	HUN	-0.01	0.05	NER	0.24	0.04	TUR	0.22	0.04
CHE	-0.18	0.06	IDN	0.31	0.04	NGA	0.75	0.05	TZA	0.54	0.04
CHL	0.28	0.04	IND	0.86	0.05	NIC	0.20	0.04	UGA	0.62	0.04
CHN	1.05	0.03	IRL	0.57	0.05	NLD	0.00	0.07	UKR	0.38	0.02
CIV	-0.08	0.03	IRN	0.26	0.05	NOR	-0.03	0.02	URY	0.20	0.04
CMR	0.20	0.05	IRQ	0.51	0.04	NPL	0.26	0.04	USA	-0.01	0.09
COD	0.07	0.05	ISL	0.03	0.02	NZL	0.09	0.08	UZB	-0.05	0.05
COG	0.14	0.03	ISR	0.59	0.09	OMN	0.06	0.03	VEN	0.07	0.07
COL	0.28	0.06	ITA	-0.67	0.09	PAK	0.46	0.09	VNM	0.37	0.02
CRI	0.35	0.04	JAM	-0.50	0.09	PAN	0.52	0.04	YEM	0.11	0.03
CUB	0.22	0.04	JOR	0.48	0.04	PER	0.42	0.04	ZAF	0.18	0.08
CYP	0.13	0.07	JPN	-0.33	0.12	PHL	0.33	0.05	ZMB	0.72	0.05
CZE	0.05	0.05	KAZ	0.52	0.02	PNG	0.08	0.06	ZWE	1.22	0.03

E Value and composition of $\tilde{\gamma}_i$ by country

Table 8: Country-specific coefficients $\tilde{\gamma}$

ALB	0.28	COG	0.32	HUN	0.17	MRT	0.43	SEN	0.24
ARG	0.19	COL	0.20	IDN	0.35	MUS	0.30	SLE	0.36
ARM	0.24	CRI	0.20	IND	0.47	MWI	0.21	SLV	0.24
AUS	0.14	CUB	0.08	IRL	0.09	MYS	0.28	SUR	0.16
AUT	0.01	CYP	-0.04	IRN	0.26	NAM	0.25	SVK	0.22
AZE	0.89	CZE	0.23	ITA	-0.14	NER	0.43	SWE	0.00
BDI	0.35	DEU	-0.07	JAM	0.06	NGA	0.33	SWZ	0.20
BEL	0.08	DJI	0.20	JOR	0.25	NIC	0.17	TCD	0.44
BEN	0.21	DNK	-0.07	JPN	-0.18	NLD	0.07	TGO	0.29
BFA	0.30	DOM	0.19	KAZ	0.31	NOR	-0.07	THA	0.30
BGD	0.77	DZA	0.26	KEN	0.28	NPL	0.25	TJK	0.07
BHR	0.81	ECU	0.29	KGZ	0.04	NZL	0.02	TTO	0.48
BHS	0.20	EGY	0.32	KOR	0.34	OMN	0.36	TUN	0.19
BLR	0.36	ESP	-0.07	KWT	0.76	PAK	0.30	TUR	0.24
BLZ	0.12	ETH	0.32	LAO	0.40	PAN	0.21	TZA	0.34
BOL	0.33	FRA	-0.09	LBN	0.26	PER	0.28	UGA	0.37
BRA	0.16	GAB	0.07	LKA	0.31	PHL	0.33	UKR	0.40
BRN	-0.04	GBR	-0.09	LSO	0.23	PNG	0.29	URY	0.09
BWA	0.21	GEO	0.15	LUX	-0.07	POL	0.22	USA	-0.11
CAF	0.15	GHA	0.43	MAR	0.21	PRT	-0.06	UZB	0.32
CAN	0.06	GIN	0.38	MDA	0.52	PRY	0.17	VEN	0.11
CHE	-0.02	GMB	0.22	MDG	0.22	ROU	0.14	VNM	0.43
CHL	0.19	GRC	-0.08	MEX	0.18	RUS	0.26	ZAF	0.14
CIV	0.16	GTM	0.19	MKD	0.21	RWA	0.29	ZMB	0.35
CMR	0.16	GUY	0.10	MNG	0.40	SAU	0.46	ZWE	0.55
COD	0.37	HND	0.20	MOZ	0.26	SDN	0.36		

Figure 14: Contribution of determinants to $\tilde{\gamma}$ by country



Notes: The bars depict the absolute contributions of each of the determinants of $\tilde{\gamma}_i$ for each country individualy. Non-linear transformations and the variables in levels are conflated here for readability. 'Industrial growth' etc. correspond to *relative* contributions of the sectors to total GDP growth, that is, they are not actual growth rates.

F Cross-validation

The goal of cross-validation is to find the parameters that lead to the model with the smallest out-of-sample prediction error. In other words, it serves to avoid over-fitting, which may be a serious issue where the number of variables is large compared to the number of observations: Instead of finding a model of general validity, a regression may then simply end up 'memorising' the data in the sample, that is, as in-sample predictions become better, out of sample predictions become worse. As in most empirical applications, testing the out-of-sample predictions is complicated by the fact that we only observe the sample itself.

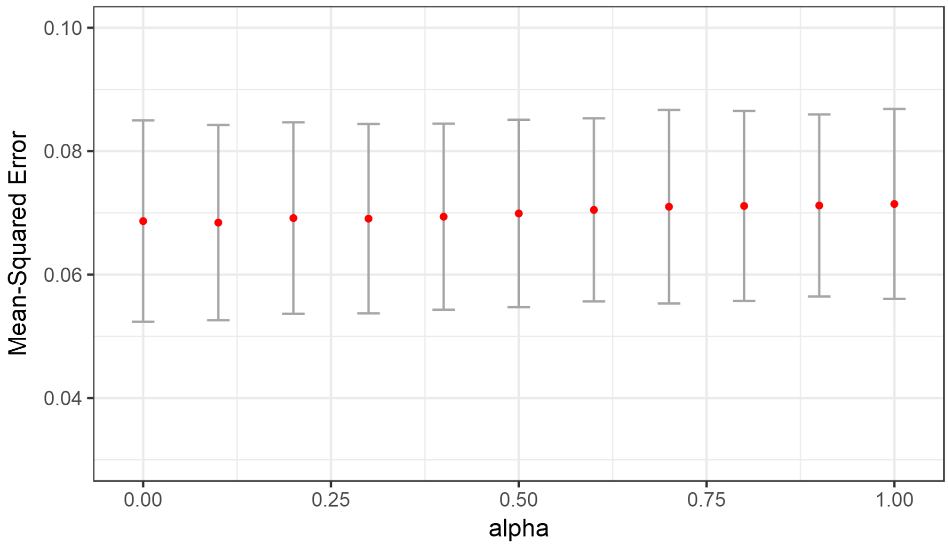
Cross-validation tackles this problem by iteratively dividing the existing sample into a training set and a test set. The training set is the one that is used to estimate the model, while the test set is 'hidden'. Once the model has been estimated, it is then used to predict the dependent variable based on the independent variables (here: $\hat{\gamma}$ based on $\mathbf{\Phi^*}$). The (squared) prediction error – 'out-of-sample', as the test set was not in the sample underlying the estimated model – of this prediction is to be minimised. The key choice to make at this point is about the division of the sample into training and test set. We use 'leave-one-out' cross-validation (LOO), which is particularly rigorous as it uses every country as a test sample once, using the remaining countries as a training sample. For every combination of the parameters to be tuned, the elastic net estimation is therefore carried out N times; thanks to the limited size of our dataset ($N = 129$ in this exercise), this is feasible within reasonable computation times.

We perform this procedure for both *alpha* and λ; note that this notation (with a single λ) refers to a slightly re-parametrised version of equation 2:

$$L(\lambda, \alpha, \boldsymbol{\delta}) = |\hat{\boldsymbol{\gamma}} - \mathbf{\Phi^*}\boldsymbol{\delta}|^2 + \lambda[(1-\alpha)|\boldsymbol{\delta}|^2 + \alpha|\boldsymbol{\delta}|_1] \tag{8}$$

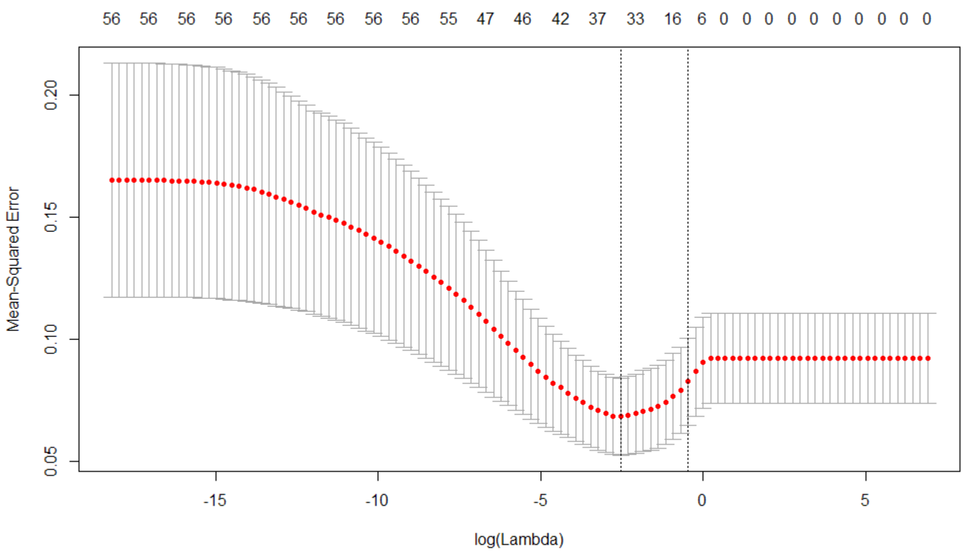
so α still refers to the parameter that dominates the relative importance of the lasso and the ridge penalty, and λ is the overall strength of the penalty. Note that as we tune α, for every value of α we iterate through the entire range of λs. The mean-squared error reported in figure 15 is then the one resulting from the specification with the optimal λ given that α. Figure 16 then reports the mean-squared error obtained from different values of λ given the optimal α. As a general observation, relative importance of the lasso and the ridge penalty hardly matters for the predictive properties of the estimated model: The out-of-sample mean-squared error is almost invariable to the choice of α (figure 15). On the other hand, the choice of λ is very influential, highlighting the importance of the regularisation procedure: With the (virtually) unrestricted model (very small λ, all 56 candidate variables included), the out of sample prediction error is about

Figure 15: Cross-validation of α



Notes: The points indicate the mean squared error at different values of alpha (see text for details). The grey bars indicate the 95% confidence intervals for these point estimates. The value of α appears to have negligible impact on the quality of the predictions, i.e., the lasso and the ridge penalies have very similar impact.

Figure 16: Cross-validation of λ



Notes: The points indicate the mean squared error at different values of λ (see text for details). The grey bars indicate the 95% confidence intervals for these point estimates. The numbers in the top row along the x-axis indicate the degrees of freedom, that is, the number of variables with non-zero coefficient. Compared to the case with (virtually) no regularisation, we can halve the out-of-sample prediction error by tuning lambda.

0.16, compared to a mean of the dependent variable, $\hat{\gamma}$, of 0.25. At the optimum, the MSE is 0.07.

G Coefficients $\hat{\hat{\delta}}$ from elastic net

Table 9: Estimated coefficients, elastic net

Variable	Coeff.	Variable	Coeff.	Variable	Coeff.
Investment	-	Investment (Sq.)	0.000	Investment (Sqrt.)	-
Consumption	-0.000	Consumption (Sq.)	- 0.000	Consumption (Sqrt.)	-0.012
Agriculture	0.000	Agriculture (Sq.)	-	Agriculture (Sqrt.)	0.007
Industry	-	Industry (Sq.)	- 0.000	Industry (Sqrt.)	-
Sevices	-	Sevices (Sq.)	- 0.000	Sevices (Sqrt.)	-
Pop. Density	0.000	Pop. Density (Sq.)	0.000	Pop. Density (Sqrt.)	0.000
Pop. growth	0.021	Pop. growth (Sq.)	0.001		
Urban pop.	-0.000	Urban pop. (Sq.)	-	Urban pop. (Sqrt.)	-0.001
Forest land	-0.000	Forest land (Sq.)	- 0.000	Forest land (Sqrt.)	-0.000
Agric. land	-	Agric. land (Sq.)	-	Agric. land (Sqrt.)	0.002
Manufacturing	-	Manufacturing (Sq.)	- 0.000	Manufacturing (Sqrt.)	-
Agric. growth	-0.633	Agric. growth (Sq.)	- 2.216		
Indust. growth	2.356	Indust. growth (Sq.)	10.459		
Serv. growth	-	Serv. growth (Sq.)	-		
Manuf. Growth	-0.718	Manuf. Growth (Sq.)	21.902		
GDP level	-0.000	GDP level (Sq.)	-	GDP level (Sqrt.)	-0.000
GDP per capita	-0.000	GDP per capita (Sq.)	-	GDP per capita (Sqrt.)	-0.000
Snow cover	-	Snow cover (Sq.)	-		
Latitude	-	Latitude (Sq.)	-	Latitude (Sqrt.)	-
Surface area	0.000	Surface area (Sq.)	-	Surface area (Sqrt.)	0.000
Fossil fuels	-	Fossil fuels (Sq.)	0.000		

Notes: The reported coefficients are normalised by the respective variables standard deviation. Coefficients reported as $(-)0.000$ are in fact rounded very small non-zero coefficients, coefficients that have been eliminated (attributed weights that are actually zero) are reported as a dash (-).

H Discrepancy between lights and GDP series

H.1 Suggested corrections to growth rates (tabular)

Table 10: WDI vs. luminosity estimates of growth (1993-2013, 1993-1999)

	Full period (1993-2013)							1993-1999				
	WDI	EN	Δ EN	FE	Δ FE	LD	Δ LD	WDI	EN	Δ EN	FE	Δ FE
AGO	6.28			5.16	-1.12	5.71	-0.57	2.62			2.28	-0.34
BDI	1.03	4.33	3.30	3.87	2.84	3.36	2.34	-3.41	3.08	6.49	2.51	5.92
BEN	4.40	3.75	-0.65	4.19	-0.21	4.29	-0.11	4.75	3.27	-1.48	3.61	-1.14
BFA	5.88	4.87	-1.01	4.69	-1.20	4.86	-1.03	6.08	5.01	-1.06	4.56	-1.51
BWA	4.82	3.82	-1.00	4.29	-0.53	4.46	-0.36	5.26	3.65	-1.61	4.11	-1.15
CAF	0.64	2.39	1.75	2.27	1.63	1.83	1.19	3.15	1.76	-1.38	1.46	-1.68
CIV	2.40	3.72	1.32	4.69	2.29	4.69	2.29	3.68	4.51	0.83	6.50	2.82
CMR	3.29	3.04	-0.24	3.45	0.16	3.36	0.07	2.49	2.29	-0.20	2.45	-0.04
COD	1.59	3.90	2.31	3.50	1.90	3.42	1.83	-4.17	1.72	5.90	1.43	5.60
COG	3.21	4.47	1.25	4.21	1.00	4.24	1.03	0.33	1.72	1.38	1.38	1.05
DJI	1.91	3.50	1.59	3.89	1.98	3.54	1.63	-1.95	1.38	3.32	0.93	2.88
ETH	7.79	5.10	-2.70	4.74	-3.05	4.97	-2.83	5.67	4.76	-0.91	4.13	-1.54
GAB	2.25	2.79	0.53	3.72	1.46	3.69	1.44	2.36	1.97	-0.40	2.30	-0.06
GHA	5.83	5.10	-0.73	4.11	-1.72	4.14	-1.69	4.31	6.21	1.90	4.41	0.10
GMB	3.42	4.19	0.77	4.69	1.26	4.59	1.17	3.01	3.69	0.68	4.01	1.00
GNB	0.92			1.53	0.61	2.26	1.34	0.05			-0.59	-0.65
KEN	3.72	4.03	0.31	4.01	0.29	3.93	0.21	2.52	3.71	1.19	3.47	0.95
LBR	8.83			6.64	-2.19	7.87	-0.96	15.91			6.06	-9.85
LSO	3.87	3.86	-0.01	4.14	0.27	4.03	0.16	3.16	3.70	0.54	3.91	0.75
MDG	2.73	3.78	1.05	4.15	1.42	3.71	0.98	2.60	4.42	1.83	5.01	2.42
MLI	7.22			4.87	-2.36	4.91	-2.32	4.72			4.74	0.02
MOZ	8.54	4.90	-3.65	5.12	-3.42	5.62	-2.92	10.65	3.78	-6.87	3.73	-6.92
MRT	4.33	5.59	1.26	4.45	0.12	4.43	0.10	3.56	6.24	2.68	4.43	0.87
MUS	4.32	3.55	-0.76	3.44	-0.87	3.29	-1.02	4.78	3.99	-0.79	3.62	-1.16
MWI	4.57	3.13	-1.44	3.34	-1.23	3.21	-1.36	4.89	2.52	-2.37	2.56	-2.33
NAM	4.05	3.53	-0.52	3.66	-0.39	3.54	-0.51	2.59	2.89	0.30	2.80	0.21
NER	4.20	5.31	1.10	4.26	0.06	4.18	-0.03	3.44	4.53	1.09	3.28	-0.16
NGA	5.97	3.23	-2.74	3.06	-2.91	2.65	-3.32	1.95	1.63	-0.33	1.33	-0.62
RWA	6.13	4.88	-1.25	4.79	-1.34	4.28	-1.84	2.84	2.99	0.15	2.65	-0.19
SDN	5.19	4.88	-0.32	4.29	-0.90	4.31	-0.89	5.07	3.22	-1.85	2.62	-2.45
SEN	3.71	3.99	0.27	4.28	0.56	4.14	0.42	3.43	3.55	0.11	3.68	0.25
SLE	4.94	5.69	0.74	4.91	-0.03	4.10	-0.84	-1.84	-0.57	1.27	-0.54	1.30
SWZ	2.60	3.40	0.80	3.78	1.18	3.82	1.22	3.26	3.20	-0.06	3.60	0.34
TCD	6.39	7.49	1.10	5.60	-0.79	6.01	-0.38	1.40	4.19	2.79	2.96	1.56
TGO	3.13	3.45	0.32	3.40	0.26	3.30	0.16	4.45	2.61	-1.83	2.37	-2.08
TZA	5.46	4.45	-1.02	4.07	-1.39	3.78	-1.68	3.28	4.79	1.51	4.05	0.77
UGA	7.04	5.11	-1.92	4.41	-2.63	4.23	-2.80	7.63	5.31	-2.32	4.19	-3.44
ZAF	3.05	2.86	-0.19	3.16	0.12	3.01	-0.04	2.48	2.34	-0.14	2.60	0.13
ZMB	5.41	3.79	-1.62	3.48	-1.93	3.42	-1.99	2.20	2.28	0.09	1.91	-0.29
ZWE	0.29	3.32	3.03	2.75	2.46	2.27	1.98	3.65	4.04	0.39	2.65	-1.00

Notes: EN are the estimates obtained from the two-step procedure involving the elastic net estimator (and equation 1), FE those obtained from the Fixed Effects estimation with a common slope (equation 5), LD those from the long-difference specification (equation 6). LD are only obtained for the full period. Δ indicates the difference between the respective estimate and the growth rates implied by WDI.

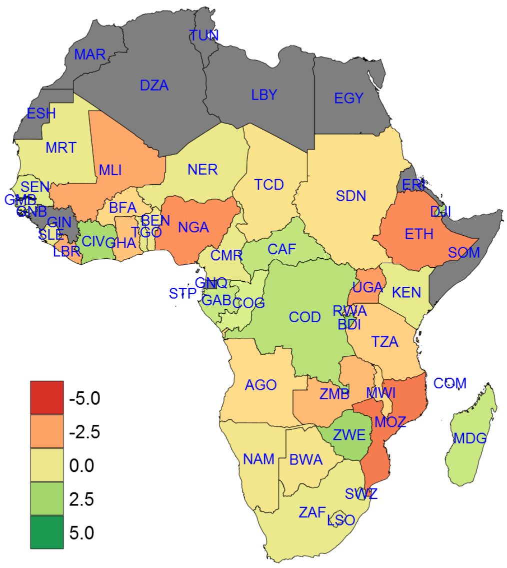
Table 11: WDI vs. luminosity estimates of growth (2000-2013)

	2000-2006					2007-2013				
	WDI	EN	Δ EN	FE	Δ FE	WDI	EN	Δ EN	FE	Δ FE
AGO	7.92			6.15	-1.76	8.30			7.04	-1.26
BDI	2.22	5.68	3.46	5.45	3.23	4.27	4.21	-0.06	3.65	-0.62
BEN	4.20	3.75	-0.45	3.81	-0.38	4.26	4.23	-0.03	5.16	0.90
BFA	5.71	4.99	-0.72	5.09	-0.62	5.85	4.61	-1.24	4.40	-1.45
BWA	4.08	4.88	0.80	5.40	1.32	5.11	2.93	-2.19	3.35	-1.76
CAF	1.53	3.03	1.49	2.34	0.81	-2.77	2.37	5.14	3.00	5.77
CIV	-0.07	4.05	4.13	4.30	4.37	3.59	2.60	-0.99	3.26	-0.33
CMR	3.71	3.81	0.10	3.86	0.15	3.66	3.03	-0.64	4.05	0.39
COD	2.53	4.96	2.43	4.87	2.34	6.42	5.02	-1.40	4.19	-2.23
COG	4.89	5.99	1.10	5.96	1.07	4.41	5.69	1.28	5.28	0.87
DJI	2.87	5.64	2.77	6.54	3.67	4.82	3.48	-1.34	4.20	-0.62
ETH	7.14	6.42	-0.71	6.33	-0.81	10.57	4.11	-6.46	3.77	-6.80
GAB	0.46	3.94	3.48	4.03	3.57	3.94	2.45	-1.48	4.82	0.88
GHA	5.04	3.16	-1.88	3.55	-1.50	8.13	5.92	-2.20	4.37	-3.76
GMB	3.16	4.60	1.43	4.97	1.81	4.09	4.29	0.20	5.07	0.98
GNB	1.18			3.02	1.84	1.61			2.28	0.67
KEN	3.62	3.72	0.10	3.87	0.25	5.02	4.66	-0.36	4.70	-0.32
LBR	3.01			8.17	5.16	7.57			5.69	-1.88
LSO	3.41	4.09	0.68	4.27	0.86	5.03	3.79	-1.24	4.25	-0.78
MDG	3.25	3.14	-0.12	3.03	-0.22	2.34	3.79	1.45	4.40	2.06
MLI	7.46			6.06	-1.40	9.49			3.79	-5.70
MOZ	8.01	5.26	-2.75	5.57	-2.44	6.97	5.65	-1.32	6.07	-0.90
MRT	5.97	4.39	-1.58	4.38	-1.59	3.46	6.15	2.69	4.55	1.09
MUS	4.04	3.93	-0.12	4.07	0.03	4.12	2.73	-1.39	2.63	-1.49
MWI	2.49	4.14	1.65	4.37	1.88	6.34	2.73	-3.61	3.10	-3.24
NAM	5.08	4.00	-1.08	4.16	-0.92	4.49	3.71	-0.78	4.01	-0.48
NER	3.48	3.68	0.19	3.89	0.40	5.68	7.72	2.03	5.61	-0.07
NGA	9.89	3.25	-6.65	3.46	-6.43	6.06	4.83	-1.24	4.38	-1.68
RWA	7.86	3.91	-3.96	4.04	-3.82	7.67	7.73	0.06	7.66	-0.01
SDN	6.92	6.04	-0.88	5.76	-1.16	3.59	5.37	1.78	4.50	0.90
SEN	4.15	4.88	0.73	5.25	1.10	3.55	3.54	-0.02	3.90	0.35
SLE	7.35	9.34	1.99	8.36	1.02	9.32	8.29	-1.03	6.92	-2.40
SWZ	2.23	4.20	1.97	4.46	2.23	2.30	2.79	0.49	3.27	0.97
TCD	12.23	6.13	-6.10	5.46	-6.76	5.54	12.13	6.60	8.36	2.82
TGO	1.28	2.90	1.62	3.02	1.74	3.67	4.84	1.16	4.81	1.14
TZA	6.52	3.58	-2.95	3.76	-2.76	6.58	4.97	-1.61	4.41	-2.17
UGA	6.78	3.29	-3.49	3.53	-3.25	6.70	6.74	0.04	5.51	-1.19
ZAF	4.14	3.97	-0.17	4.12	-0.02	2.53	2.28	-0.25	2.77	0.24
ZMB	6.12	4.30	-1.82	4.37	-1.75	7.92	4.80	-3.12	4.16	-3.75
ZWE	-6.07	1.47	7.54	2.70	8.77	3.28	4.45	1.17	2.89	-0.40

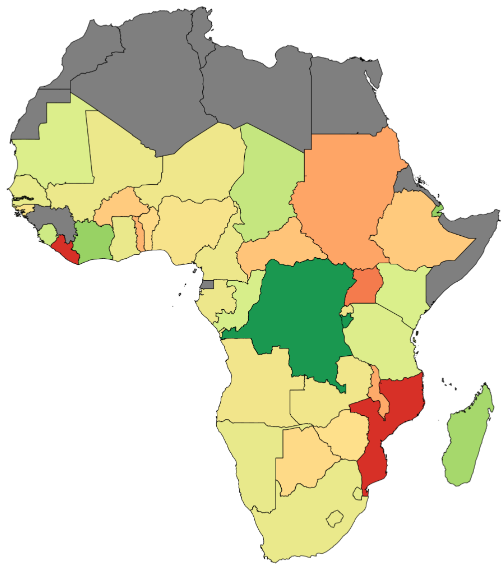
Notes: See table 10.

H.2 Suggested corrections with single $\hat{\gamma}$

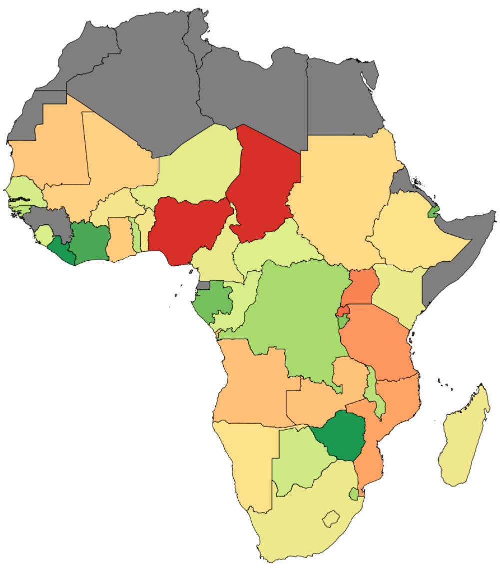
Figure 17: Growth according to official vs. lights data (common slope)



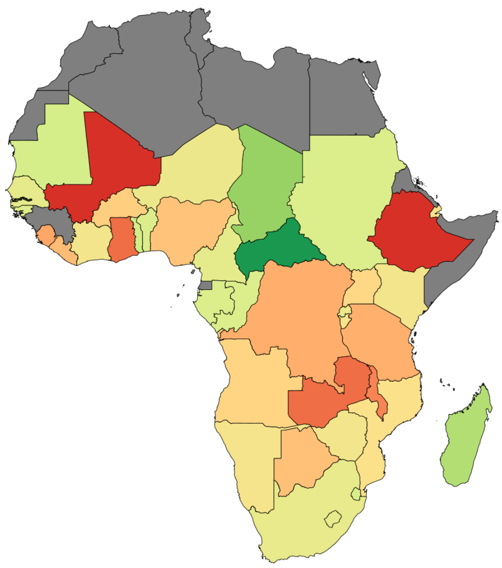
(a) Whole period 1993-2013



(b) 1993-1999



(c) 2000-2006



(d) 2007-2013

Notes: See figure 11