IMPROVED ESTIMATES OF USING LUMINOSITY AS A PROXY FOR ECONOMIC STATISTICS: NEW RESULTS AND ESTIMATES OF PRECISION

By

William Nordhaus and Xi Chen

March 2012

COWLES FOUNDATION DISCUSSION PAPER NO. 1857

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Improved Estimates of Using Luminosity as a Proxy for Economic Statistics: New Results and Estimates of Precision [1, 2]

March 29, 2012

William Nordhaus [3], Xi Chen [4]

---

[1] The research reported here was supported by the Glaser Foundation, the U.S. National Science Foundation, and the U.S. Department of Energy. The authors declare that they have no relevant and material financial conflicts with the research described in this paper.

[2] We thank participants in the National Bureau of Economic Research–Conference on Research in Income and Wealth Workshop and the Yale Environmental Economics Workshop for comments. We are grateful for comments on the work by Ben Jones, Dale Jorgenson, and Steve Landefeld as well as for comments on the statistical approach from Zhipeng Liao. We have benefited from extensive advice from Chris Elvidge and his team and thank them particularly for their work in developing data on nighttime lights and making them available to the scientific community.

[3] Department of Economics and Cowles Foundation, Yale University and National Bureau of Economic Research (william.nordhaus@yale.edu).

[4] Department of Sociology, Quinnipiac University (Xi.Chen@quinnipiac.edu).

Abstract

Previous work has analyzed whether luminosity data contain useful information for estimating economic output and concluded that there was significant promise for regions with poor quality economic statistics. The present paper examines alternative measures of the precision of the estimates using bootstrap and prior estimates of the errors for both the luminosity quality and the national accounts quality. Based on the new results, we conclude: First, for countries with high quality systems, there is no reason to use luminosity data as a supplement to standard data in any context where standard data are available. Second, we find that there is no advantage at present of using lights data for time-series corrections for any purposes where standard data are available. Third, for countries with low quality statistical systems, the estimates suggest that there may be substantial information in the luminosity data for cross-sectional estimates of output. Fourth, the major concerns about the use of lights as a proxy involve uncertainties about the precision of standard national accounts data. Finally, we recommend that future work be concentrated on integrating luminosity data into the cross sectional estimates of national and regional output primarily for countries with poor quality statistical systems.

JEL classification: E01, O47, O5, Q4

Keywords: luminosity, output measurement, national accounts, proxy measures

I. Introduction

Measures of national output and income are the major social indicators used to evaluate the relative performance of countries over space and time. Richard Froyen describes economic policy in the era before economic accounts were developed as follows (Richard Froyen, 1996):

> One reads with dismay of Presidents Hoover and then Roosevelt designing policies to combat the Great Depression of the 1930's on the basis of such sketchy data as stock price indices, freight car loadings, and incomplete indices of industrial production. The fact was that comprehensive measures of national income and output did not exist at the time. The Depression, and with it the growing role of government in the economy, emphasized the need for such measures and led to the development of a comprehensive set of national income accounts.

Development of a full set of national economic accounts has been a major accomplishment of national statistical systems since the 1930s, but there is much further work needed to integrate output, income, and wealth accounts (Dale Jorgenson, J. Steven Landefeld, and William Nordhaus, 2006).

While economic statistics in wealthy countries have improved greatly in recent decades, statistical data are often of low quality in other countries. This is especially true for the poorest countries in sub-Saharan Africa, many of which have no reliable censuses of population and only rudimentary economic statistics. A few countries (Iraq, Afghanistan, and Somalia being examples) have virtually no statistical systems.

A promising approach to estimating output in countries with poor data systems is to use alternative data sources. A small number of studies have addressed this question using a new and independent data set – nighttime luminosity (Chen and Nordhaus, 2011, Henderson, Storeygard, and Weil 2011, 2011a). In our previous work, we analyzed whether luminosity data contains useful information for estimating national and grid cell (1°x1°) economic output.

Figure 1 shows a scatter plot of log luminosity density and log output density for all grid cells with positive output for 2006 (N = 12,393, log refers to natural logarithms). It is clear that luminosity and output have a strong positive correlation at high output densities, but the relationship is less apparent at low output densities. To extract the information from the luminosity data, we construct a synthetic measure of output (blending luminosity data and standard national-accounts measures) and calculate optimal weights that minimize the expected error of that synthetic measure. The optimal weight on the luminosity-based proxy (θ^*) measures how much useful information luminosity data contains as a proxy for measuring national or regional output.

Our study divides countries on the basis of the "grades" of their statistical systems, from A through E, with A being the highest. The conclusion from the first round of studies in this area was that luminosity is likely to have informational value primarily for countries with poor quality statistical systems (countries receive a D and an E grade), but has very limited value added for high or middle income countries (countries receive an A, a B or a C grade). The weight on luminosity-based proxy is close to zero for A, B or C countries, while it ranges from 10% to 70% for D or E countries. Similar studies using a different statistical approach were undertaken by Henderson et al. (2011a, 2011b). For countries with poor national income statistics, they found that the luminosity-based proxy and conventional growth measure should have roughly equal weights.

Our earlier study used cross-sectional as well as time-series output estimates (Chen and Nordhaus, 2011). Additionally, we looked at a comparison not only for countries but also for a more disaggregated level (grid cells), which helped to remove any country effects and increased the sample by a factor of approximately 100. We examined two output concepts, the growth rate measure of output from 1992 to 2008 and annual output density measured as constant-price output per unit area. We also examined three different luminosity measures – raw, stable, and intercalibrated lights.

None of the first round of studies presented formal tests of the precision of the estimates of the contribution of luminosity, and that is the purpose of the present study. It is well established that weighting estimates for proxies need to be treated in a statistical manner (North, et al. 2006). As we show below, the statistical model for deriving the optimal weights on conventional GDP measures and luminosity is underidentified and requires estimates of three parameters: the measurement error of conventional GDP measures, the measurement error of nighttime lights, and the coefficient in the regression equation of output and luminosity. The challenge in the present study is that the estimates are a mixture of prior estimates of errors as well as statistically based estimates, so we need to combine both approaches. In addition, there is very little evidence on the reliability of the national accounts data outside the high-income countries.

In the remainder of the paper, we first discuss the parameters we use to calculate the luminosity weight, next present the analytic model underlying the estimation, and then present the results from bootstrap and sensitivity analysis.

II. The Analytic Model

A. Data

The primary nighttime image data were gathered by US Department of Defense satellites starting in the mid-1960s to determine the extent of worldwide cloud cover. The data were later declassified and made publicly available as the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS). The raw data can be acquired in two spatial resolution modes. The full resolution data, also referred to as "fine" data, have nominal spatial resolution of 0.5 km. The "smoothed" data are an average of 5 × 5 blocks of fine data and have a nominal spatial resolution of 2.7 km. The data that we obtained from the National Oceanic and Atmospheric Administration–National Geophysical Data Center are constructed using the smoothed spatial resolution mode, at a resolution of 30 arc-seconds, covering 180° W to 180° E longitude and 75° N to 65° S latitude. There are different versions of the data; three of particular importance are the “raw,” the “stable lights,” and the “calibrated” versions. After considerable testing, we have relied on the stable lights version.

For standard output data, at the country level, we used GDP purchasing power parity (PPP) values at constant 2005 international US dollars from the World Bank from 1992 to 2008. For disaggregated output data, we used the GEcon data set, available at gecon.yale.edu. This is available at 1° x 1° latitude and longitude resolution for all terrestrial grid cells for 1990. 1995, 2000, and 2005 using PPP values at constant 2005 international US dollars. For intermediate years, we interpolated cell GDP using national output and population numbers. For more details, see Chen and Nordhaus (2011).

B. Derivation of the Optimal Weights

We next explain the basic approach and derive the optimal weighting of conventional output and luminosity. This section draws upon the Supplementary Information from Chen and Nordhaus (2011). For this purpose, we define the different variables as follows:

Y = output from national accounts (GDP in constant 2005 international U.S. $)

Y^* = true output (GDP in constant 2005 international U.S. $)

X = synthetic measure of output (GDP in constant 2005 international U.S. $)

M = measured luminosity (index value)

Z = luminosity-based measure of output (GDP in constant 2005 international U.S. $)

i = grid cell (here 1° latitude by 1° longitude)

j = country

k= country grade (A, B, C, D, E)

t = year

y = log (Y) and similarly for other upper case variables

ε_i = measurement error in GDP
ξ_i = measurement error in luminosity
u_i = error in output-luminosity relationship
$\alpha, \beta, \ \mu$ = structural parameters

For notational purposes, we define $x_i(t)$ as the value of variable x in grid cell i averaged over a year. We omit the time variables when they are inessential to the exposition. Begin by assuming that there is an unknown true level of output for each country and grid cell, which is measured with error.

$$y_i = (1-\mu) E\left(y_i^*\right) + \mu y_i^* + \varepsilon_i$$

For the present study, we assume that there is no bias in measured output, so $\mu = 1$. This assumption is not completely innocuous as there may be systematic growth mismeasurement due, say, to incomplete source data or infrequent observations. The important issues raised by $\mu \neq 1$ have not been solved. Assuming that $\mu = 1$ yields:

(1) $$y_i = y_i^* + \varepsilon_i$$

Luminosity is subject to measurement error (due to satellite, calibration, and other sources):

(2) $$m_i = m_i^* + \xi_i$$

There is assumed to be a structural relationship between luminosity and true output as follows:

(3) $$m_i = \alpha + \beta y_i^* + u_i$$

The error in equation (3) arises from several sources. One important error is that luminosity is sampled at night, whereas economic activity is generally concentrated in the daytime. More important, the light intensity differs greatly across sectors. Often, lights are associated with electricity use. The use of electricity per dollar of output in different sectors provides a rough idea of how light-intensities might vary. In the 2002 U.S. input-output tables, the electricity used per unit output of real estate was 200 times greater than that of software. (See the input-output tables at www.bea.gov for the underlying data.) Similar differences are seen across other sectors. This example suggests that industrial composition across countries and regions is likely to make the output-luminosity relationship in equation (3) relatively noisy.

We want to construct a luminosity-based output proxy from these relationships. We have measurements of all variables over time and space at the national and grid cell levels. However, we need to develop measures of the error of measurement of national and grid cell output, the coefficient on luminosity, as well as the error in the structural relationship in equation (3).

Our procedure is first to estimate equation (3) using measured output and luminosity. This provides a biased estimate of the coefficient, $\hat{\beta}$, because output is measured with error. We then do an errors-in-variable correction using our prior estimates of the measurement error of GDP to get a corrected estimate of the structural coefficient, which we denote $\tilde{\beta}$. The corrected coefficient is calculated as:

$$(4) \qquad \tilde{\beta} = \left(\frac{\sigma_{y^*}^2 + \sigma_{\varepsilon}^2}{\sigma_{y^*}^2} \right) \hat{\beta}$$

Here, $\hat{\beta}$ is the estimated coefficient in equation (3); σ_{ε}^2 is the a priori measurement-error variance of true output; and $\sigma_{y^*}^2$ is the estimated variance of true output. The consistent estimate of $\tilde{\beta}$ follows immediately.

We then estimate the luminosity-based output proxy as follows by inverting equation (3):

$$(5) \qquad \hat{z}_i = (1/\tilde{\beta})(m_i - \tilde{\alpha}_i)$$

where $\hat{z}_i$ is the log of our luminosity-output proxy and $\tilde{\beta}$ and $\tilde{\alpha}$ are the corrected coefficients from equation (4).

Next, we construct a combined measure of output by taking weighted averages of conventional measures of output and our luminosity-based output proxy:

(6) $\hat{x}_i = (1 - \theta) y_i + \theta \hat{z}_i$

where

$\hat{X}_i$ = new synthetic measure of output
$\hat{x}_i$ = $\ln(\hat{X}_i)$
θ = weighting fraction on luminosity.

The key variable of this study is θ, which is the share (or weight) of the luminosity-based output proxy. The central question we address is whether we can significantly improve conventional measures of output using luminosity. If the measurement error of the synthetic luminosity-based output proxy is low relative to the measurement error of conventional output estimates, then luminosity can be a useful proxy.

The procedure for calculating the optimal weight on luminosity is as follows. Define $V(\theta)$ as the mean squared error (MSE) of $\hat{x}_i$ as a function of the weight, θ. We proceed intuitively by first assuming that all parameters are known. In this case, we can derive $V(\theta)$ as a function of θ as follows (we omit the grade superscript in the derivation until the next section):

$$V(\theta) = E\,[(1-\theta)y + \theta\hat{z} - y^*]^2$$

$$= E\,[(1-\theta)(y^* + \varepsilon) + \frac{\theta}{\beta}(\beta y^* + u) - y^*]^2$$

$$= E\,[(1-\theta)\varepsilon + \frac{\theta}{\beta}u\,]^2$$

$$= (1-\theta)^2\sigma_\varepsilon^2 + \frac{\theta^2}{\beta^2}\sigma_u^2$$

If all the parameters are known, then minimizing $V(\theta)$ with respect to θ yields a unique value for the optimal weight, θ^*:

$$V'(\theta^*) = 0 = -2(1-\theta^*)\sigma_\varepsilon^2 + \frac{2\theta^*}{\beta^2}\sigma_u^2$$

or

(7) $$\theta^* = \frac{\beta^2\sigma_\varepsilon^2}{\beta^2\sigma_\varepsilon^2 + \sigma_u^2}$$

However, because the parameters in (7) are unknown, we need to find an appropriate estimator of θ. We assume σ_ε^2 is known from external evidence (see below). Further, assume that that σ_u^2 and β can be consistently estimated as $\tilde{\sigma}_u^2$ and $\tilde{\beta}$, respectively. It can be shown that θ^* is the uniformly consistent estimator of θ.

A sketch of the proof is the following. Define $\tilde{\theta}_n^*$ as the estimator of the optimal weight θ^* for sample size n:

(8) $$\tilde{\theta}_n^* = \frac{\tilde{\beta}^2 \sigma_\varepsilon^2}{\tilde{\beta}^2 \sigma_\varepsilon^2 + \tilde{\sigma}_u^2} \ .$$

Because $\tilde{\beta}$ and $\tilde{\sigma}_u^2$ are consistent estimators, $\tilde{\theta}_n^* \to \theta^*$ in the probability limit. So the estimator in (8) is a consistent estimator of $\tilde{\theta}_n^*$ that minimizes the asymptotic MSE of the synthetic output measure. Note that $\tilde{\theta}_n^*$ is not necessarily an unbiased estimator in small samples. The next section addresses the reliability of the estimates in finite samples.

C. Estimating the reliability of the optimal weight

It is tempting to construct new measures of output based on luminosity *(θ)* when the estimated optimal weight on luminosity is large. This is not advisable unless we have a clear idea of the reliability of the estimates of the optimal weights. However, because the procedure used to estimate the optimal weight is complex and the estimator is only consistent, we cannot use standard confidence techniques.

We therefore use a bootstrap procedure to estimate the precision of the weighting parameter, $\tilde{\theta}_n^*$ (see Efron and Tibshirani, 1986 as well as the survey in Davison, Hinkley, and Young, 2003). Bootstrapping is a procedure that uses resampling of the data to determine the accuracy of sample estimates. We use Monte Carlo resampling because of the size of the sample. This requires resampling the data with replacement, where the size of the resample is equal to the size of the original data set. We then calculate the distribution of the statistic of interest by taking multiple replications. We interpret the estimation as random observations in our calculations although a fixed-variables interpretation might be more natural in this context. We do not do statistical tests because we are primarily concerned with the overall reliability, which can be best understood by dispersions and boxplots, shown below.

The precision of the estimate of θ in equation (8) depends upon three parameters. Two of them ($\widehat{\beta}$ and $\tilde{\sigma}_u^2$) come from the regression analyses. The other (σ_ε^2) comes from a priori estimates for measurement errors in standard output. We discuss each of these in turn.

1. *Bootstrap estimates for statistical parameters*

The parameters for the error variance of luminosity and the coefficient in the luminosity equation can be estimated by standard bootstrap techniques. Applying bootstrap procedures to the regression equation can generate a set of $\tilde{\beta}_m$ and $(\tilde{\sigma}_u^2)_m$, where the subscript m after estimates indicate that they are bootstrapped replications. These will provide consistent estimates of the errors of these parameters.

2. *Measurement errors of the national accounts data*

In general, estimates of GDP and other national accounting measures do not have associated statistical errors. Unlike other data series (such as the unemployment rate), GDP is not a statistically based estimate but is built up from multiple sources of data and several ad hoc procedures.

III. Errors of measurement of output

A. Overview

The thorniest issue in estimating the reliability of the optimal weights is determining the errors of standard national accounts GDP measures. We begin with a discussion of "measurement error" in this area (see Fixler, 2008 for a useful discussion). It is general practice for statistical offices deriving national economic accounts not to provide estimates of the measurement error. Instead, accountants generally discuss "reliability," which is the inverse of measurement error (Fixler and Grimm, 2008). Fixler and Grimm note that "total measurement error… in the [national income and product accounts] is never observed."[5] The major focus in estimating reliability is to determine the size of revisions, which is a component of total measurement error but will generally be smaller.

We define measurement error for standard national output estimates as the error of estimate of output or output growth relative to an ideal measure of national output. For our purpose, we define the "ideal" measure of output as that one corresponding to the definition of national output in the System of National Accounts (1993).

Fixler (2009) described measurement error as arising from six sources: sampling error, response error, non-response error, coverage error, processing error, improperly designed source data, and non-statistically related errors. It is likely that in our framework the last of these (non-statistically related errors) may be most important. Non-statistical errors include imputations, conceptual differences, index construction, sectoral definitions, and the scope of the exclusions (such as home production, subsistence farming, illegal activity, and smuggling).

In some cases, surprisingly, luminosity will more closely track the actual location of economic activity than conventional accounts. For example, the output of ocean fisheries is generally taken to be onshore in national accounts; similarly, the output of off-shore oil production usually shows up in the regional accounts of national capitals (in the case of Egypt).

We can distinguish two different kinds of errors. The first are *time-series errors*. Measures of output growth generally keep the conceptual basis of the measures as well as the data sources constant over time (at least for short periods). Time-series or growth-rate errors will arise primarily from errors in the source data

[5] Dennis Fixler and Bruce Grimm, 2008.

and errors in aggregation. Moreover, since there are two or three alternative methods of constructing national output (e.g., income and expenditures), we can examine the statistical discrepancy to make an initial estimate of the size of the measurement error.

A second kind of measurement error is *cross-sectional level* or *density errors.* These would apply to comparisons among countries or regions, such as a comparison of the per capita output of the US and Mali. Cross-sectional errors will encompass a broader set of concerns than time series errors. They will include most of the ones mentioned above. In addition, they will reflect differences in source data, concepts, and price measurement by country, as well as errors in measuring the effective exchange rates among different currencies. Moreover, there are no identities that can be relied upon to provide alternative estimates of the kind that produce the statistical discrepancy in time-series measures. We would therefore expect the cross-sectional errors to be larger than the time-series errors.

In addition, the present study examines both country output data and grid-cell output data. We will therefore need to consider errors in both countries and grid-cells as well as time-series and cross-sectional (density) estimates.

B. Errors in national-level data

Time series errors: general

Estimation of errors in standard national accounts is a vast and largely uncharted enterprise, and we can only suggest estimated errors in the current study. We begin by surveying estimation errors for the U.S. The Bureau of Economic Analysis (BEA), which produces the accounts, has devoted considerable attention to reliability issues. The most systematic measure of error is the statistical discrepancy (SD) between income and expenditure accounts. The SD is generally "unmanaged" and is therefore a relatively reliable measure of the measurement error (conditional on the definitions). The absolute value of the change in the ratio of the annual SD to GDP averages around 0.43 percentage points (per year) of GDP for the 1929-2010 period and 0.35 percentage points for the 1948-2010 period. If the true value is the average of income and product measures (which is suggested by some studies), this would indicate that the error in the growth rate is 0.20 to 0.24 percentage points per year.

As an alternative estimate of the errors, BEA has examined the change from the third annual estimates to current methodology and found an average absolute revision of the annual growth rate of real GDP of 0.41 percentage points for 1983-2006 with 0.29 for nominal GDP (Fixler and Grimm, 2008).[6] This second calculation includes some changes in methodology as well. From these two calculations, this indicates a lower bound for the measurement error of the growth rate of real output of around 0.3 percentage points per year for the U.S. We apply this number to other countries with high-quality statistical systems.

Time-series errors from index-number differences

One of the methodological differences among countries involves the index-number techniques used in determining growth rates. Most high-quality systems currently use superlative techniques (such as Fisher's Ideal index), while other countries (such as China) continue to use Laspeyres indexes. BEA's calculations indicate that for the U.S., the average error due to using Laspeyres rather than Fisher indexes is around 0.3 percentage points per year (this being always positive). Larger biases would be expected in countries with particularly rapid structural change.

Cross-sectional errors from revisions and methodological differences

A second set of estimates concerns the level of GDP (or GDP density per square km when used in conjunction with luminosity). For the U.S., we can use the average ratio of SD to GDP as a lower bound estimate of the measurement error conditional on the methodology. Over the period 1929-2008, the SD was 0.50 percent of GDP. Again, if the average of income and product sides is the correct

[6] All data comparisons are based on data from the BEA web site at www.bea.gov as of February 2010.

estimate, this would indicate an average measurement error of 0.35 percent of GDP as a lower bound for high-quality statistical systems.

However, it is likely that cross-sectional errors will arise from other issues such as measurement error, sectoral inclusion, exchange rates, and even idiosyncratic country effects. One way to estimate the methodological differences is to examine the change in the level of nominal GDP in different vintages of estimates. Comparing current estimates with those of 1964 and 1973, BEA found an average error of between 3.1 and 3.3 percent. Most of these are probably definitional (such as the inclusion of software in investment) rather than measurement error, however. If we go back to the earliest estimates of national income by Simon Kuznets published in the 1930s, we find much more substantial differences, however (Kuznets 1937). The difference between the original estimates of national income by Kuznets and the current estimates by the BEA is 17 percent for 1929-32. The average absolute difference in the logarithmic growth rates of nominal national income was 4.5 percentage points for this period. This is a very demanding test, of course, because these first estimates were produced at the dawn of national income accounting, and the period was the descent into the Great Depression. These are suggestive of the very substantial cross-sectional differences that can arise in immature accounting systems as well as the measurement problems that can arise in economic crises.

Cross-section differences from exchange-rate calculations

One of the thorniest issues in country comparisons is the conversion from national currencies into a common unit. Common practice today is to use PPP, or purchasing-power parity, exchange rates rather than market exchange rate. While there is (in our view) no question about the appropriateness of PPP measures conceptually, the practice of calculating them has proven extremely difficult, and in some cases, such as the appropriate multilateral weights, unresolved.[7]

---

[7] There is a vast literature on the subject. For a recent review, see the article by Deaton and Heston (2008).

A system of country grades from A to D was introduced by Robert Summers and Alan Heston (1991). These were judgmental margins of error (actually defined as the root mean squared error). Table 1 shows the margins of error as defined in the original Summers Heston study (1991).

The system of grading has been adopted in the current Penn World Table estimates of national output. Countries are assigned subjective quality grades from A to D by the authors of PWT based on several criteria of the data. We will add grade E for those countries with essentially no statistical system and ones that are missing from the PWT and other standard sources. In our estimates of measurement errors, we will rely on the country grading system defined by the Penn World Table and earlier authors. These are most usefully interpreted as the cross-sectional errors that arise from price and aggregation estimates involved in moving from market exchange rates to PPP exchange rates.

Very few countries receive the grade of A and a substantial number are C or D. The A countries would be representative of countries such as the United States. Note that the margin of error is much greater than the average statistical discrepancy, but not as large as the difference in the Kuznets-BEA estimates. We will adopt the margin of errors in Table 1 for our estimates of the cross-sectional errors for countries.

The errors in the international data have been recently examined in a comprehensive study by Simon Johnson, William Larson, Chris Papageorgiou, and Arvind Subramanian (2009). They examined the revisions of estimates of both the level of the price index and the growth in real GDP across countries between Penn World Table 6.1 and 6.2, shown in Table 2. One feature of the study was to examine the changes or revisions in cross-sectional differences due to price changes. These revisions arise from several sources: revisions in source data, methodological changes, and most importantly from changes in the international price data. These can be interpreted as an approximation of the error in PWT 6.1. For example, if the PWT 6.2 was exactly correct, then the errors would be the errors in the PWT 6.1.

There is no theoretical way to determine whether these should be higher or lower than the Summers-Heston margins of error, but they do tend to be considerably smaller. Note that the Summer-Heston grades applied to an earlier version of PWT and should presumably be smaller on average in the latest version.

Johnson et al. (2009) focus primarily on the revisions of the growth rates. Based on their results, we have compiled estimates of the revisions to the growth rates of real GDP by country grade, and the results are shown in Table 3. This table shows the revisions in growth rates across countries between Penn World Table 6.1 and 6.2. We transformed the 29-year growth rate differences to 1-year growth rate differences by assuming that the differences by year were independent. As with the cross-sectional differences in Table 2, there is no necessary relationship between the errors and the revisions.

The number for the U.S. using this methodology is 0.3 percentage points per year, which is virtually identical to the number derived above. The errors for the A countries are consistent with the estimates for the U.S., indicating about double the measurement error for the U.S. Part of the difference between the US and other countries, however, is probably due to the fact that the U.S. dollar is the numeraire. The revision numbers for the D countries are astoundingly high, indicating a revision that is actually more than double the mean growth rate. However, note that 135 of the poor quality countries are C while only 13 are D. At the same time, some of the worst statistical systems – Iraq, West Bank and Gaza, and Afghanistan – are not even in the PWT data set.

Differences from compilation errors

A final issue comes from transmission errors that come when those who compile databases make errors of omission or commission. We have generally relied on the World Bank compilation of national accounts statistics because it is the most complete. One set of errors would arise because the data are collected from earlier vintages of publications from the national statistical offices. For data available on the same date, the ratio of the World Bank measure of U.S. nominal GDP to the official version erred by an average of 1.1 percent over the 1960-2006 period. The average growth rate of real GDP differed by 0.18 percentage points. By contrast, the IMF database associated with the *World Economic Outlook* had essentially no errors.

Grid-cell estimates

In our estimates below, we use grid cell as well as country output estimates. The grid cell output data have higher errors than the national data, but they are an important source because of the much higher resolution than country data. We have about 20,000 non-zero grid cells as compared to somewhat less than 200 countries.

The tradeoff is that estimating the grid-cell errors is more challenging because their estimation is in its infancy. We have used grid-cell output data based on the GEcon data base (*gecon.yale.edu*).We consider the national level and growth estimates to be a lower bound for our grid-cell estimates. The major approach available to estimate potential error is similar to that used above – to examine changes in estimates of levels of PPP output for individual grid cells across revisions of the GEcon data sets. The revisions have added considerable accuracy by using improved maps, better population estimates, and improved imputations. In addition, the GEcon estimates have added output estimates for E quality countries for which data are not generally available, such as Somalia and Afghanistan. Revisions have been completed for 43 of the countries in the data set, with 8670 grid cells.

The revisions considered here are from the first published version (GEcon 1.3 from 2005) to GEcon 3.4 in 2011. This includes one comprehensive set of revisions in the gridded population data. The economic data have been thoroughly revised for most major countries.

Table 4 shows the revisions measured as the standard deviation of the logarithm of first to last estimates for GDP per grid cell and population per grid cell for countries with revisions. These should be compared with the cross-sectional results above. The results suggest very high potential error for the grid cell output estimates, even for grade A countries.

We will also be using growth rate estimates for the grid cell data. These have been developed from a combination of gridded population for different years and country data on regional or national per capita GDP. At present there are no revisions of the data for comparison purposes. Work on regional GDP estimates suggest, however, that relative per capita incomes in most countries are relatively stable over time, so the errors in the growth rates of gridded data are likely to be only slightly above the estimates of the errors for the population and for the national GDP data. We have made tentative estimates here, but we recognize that these error estimates have only a sparse empirical basis.

The last column in Table 4 provides the estimates of the estimated errors for grid cell output data for each of the five groups of countries. These are clearly very high, and are still tentative given the tentative nature of the GEcon data. The question is whether, given the very high potential errors in these data, the luminosity information can be used to improve the estimates.

Table 5 summarizes our estimates of the output-measurement errors for different countries and concepts used in Chen and Nordhaus (2011). The estimates for country time-series and cross-sectional (density) errors are based on a variety of studies. For countries, the cross-sectional error estimates are largely consistent with the PWT grades, and the growth error estimates are largely drawn from results reported by Johnson et al. (2009). The assumption for grid cells is based on the change for the G-Econ data between revisions. We also note that the errors for the E countries are particularly uncertain, because they are not found in other studies.

IV. Methods and Results

Chen and Nordhaus (2011) provided estimates of the optimal weighting fractions of conventional GDP measures and lights-based measures. We now provide estimates of the weighting fraction based on revised data along with the estimated error for the optimal weights for each of the different approaches (time series and cross section, country and grid-cell, for each country grade). Table 6 lists the numbers of country and cell observations and representative countries for each country grade used in this study. A complete list of countries and grades is available in the SI to the original article at www.pnas.org.

Our formal analysis is in two steps. The first step is a standard bootstrap analysis of the standard error of the estimated parameters. This first step can be performed for all parameters except the errors of measurement of standard national-account output measures. For the standard output measures, we do not have a statistical method to generate errors; therefore for the second step we use sensitivity analysis.

Bootstrap analysis: background

We are concerned with the precision of the point estimates of the optimal weight on luminosity *(θ)* provided in Chen and Nordhaus (2010). The value of θ is determined by three parameters (β, σ_u^2 , and σ_ε^2) as shown in equation (8). Therefore, the reliability of the optimal weight is influenced by the reliability of these parameters.

Because the procedure contains multiple steps and assumptions, we cannot estimate the precision using standard techniques. So, as a first step, we use bootstrap techniques to determine the precision from those parameters that are statistically derived (in the present section) and sensitivity analysis to estimate the precision (in the next section).

In the exposition that follows, we simplify the notation by substituting β, σ_u^2, and θ for $\tilde{\beta}_n$, $(\tilde{\sigma}_u^2)_n$, and $\tilde{\theta}_n^*$, respectively. These short-hand expressions are used to keep the text intelligible, but readers are reminded of the formal definitions provided in the first part of the paper.

Bootstrap analysis: parameters

In the first step, we apply standard bootstrap techniques (Freeman, 1982; Efron and Tibshirani, 1986) to estimate the uncertainty of β and σ_u^2, and ultimately θ, caused by sampling error in the regression model. We first generate a set of estimates of β and σ_u^2 by resampling the data with replacement for equation (3), and then combine the bootstrap estimates with the baseline estimates of σ_ε^2 from Table 5 to calculate the corrected regression coefficient β. Putting error-corrected β, σ_u^2 and baseline σ_ε^2 together, we finally calculate the optimal θ with equation (8). For all calculations, we set the number of bootstrapped replications at N = 1000.

We present the results for the bootstrapped estimates in both tables and box plot figures by country grade. The statistics of 1000 replications of regression coefficient (β) and root mean squared error (σ_u) for country data are shown in Tables 7a and 7b, while those for cell data are shown in Tables 7c and 7d.

The results show that the β parameter is reliable estimated for the cross sections for both countries and cells. By contrast, the time-series coefficients of β are much less reliable, particularly for countries. In Tables 7a and 7b (country data analysis), the standard deviation of β for cross sections (column 3 in Table7a) ranges from 0.006 to 0.043, but it is much larger for time series (column 3 in Table7b), ranging from 0.157 to 0.420. These results indicate that for country data the β estimate is more reliable for cross sections than for time series.

Tables 7c and 7d show the similar results for cells. The standard deviation for β is much smaller for cross sections than for times. This is most easily seen by the fact that the interquartile range (IQR) for the bootstrapped β is much smaller for cross sections than for time series for both country and cell data. Comparing all specifications (cell versus country and cross sections versus time series), we find the standard deviation and the IQR of β estimate are largest for time series country data (Table 7b).

The results are more easily visualized in box plot figures, which are a convenient method for displaying the dispersion of an estimate. Figure 2 and 3 present the distribution of the β parameters in box plots for country and cell data respectively. The upper and lower edges of the box indicate the value at the 75th and 25th percentile, and the difference is the IQR. The upper hash mark indicates the 3rd quartile plus 1.5 IQR, and the lower hash mark indicates the 1st quartile minus 1.5 IQR. These figures clearly show that the box sizes (the IQR) are much larger for time series (the bottom panels in the Figures 2 and 3) than for cross sections (the top panels in Figures 2 and 3). The same point is seen for the hash marks. We observe the similar patterns in both country and cell figures. The difference between cross section and time series is not surprising given the vast difference in the cross-sectional levels of output across regions as compared to the relatively limited difference in the growth rates among regions.

Next, we examine the bootstrapped results for the root mean squared error (RMSE) or σ_u shown in the column 5 to 7 in Tables 7a to 7d and in box plot Figures 4 and 5. Again, the results are based on one thousand replications of regressions for countries and cells for each grade, and for all countries and all cells. The results here are similar to the results for bootstrapped β. The RMSE estimator of cross sections is more reliable than for time series, particularly for the cell data. The only exception is the grade A countries.

Taking all the results together, we see that the estimates of the parameters are most reliable for cross section cells, and are least reliable for time series countries. However, we also find the difference between cross sections and time series shown in RMSE estimates is not as large as shown in β estimates. Using country analysis as an example, we see the difference between cross sections and time series for standard deviation of β is more than a factor of 10 (comparing column 3 in Table 7a and 7b), but the difference in RMSE estimates is less than a factor of 4 (comparing column 5 in Table 7a and 7b).

Bootstrap analysis: optimal weights

Our final steps is to estimate the precision of the optimal weighting coefficient on the lights-based proxy measure, or θ. Recall that $\theta = 0$ when all the weight is on standard national-accounts measures, and $\theta = 1$ when the entire weight is on the light-based proxy measures. To calculate θ, we use equation (4) and the baseline σ_{ε}^{2}. We take the error-adjusted coefficient for β for each replication of β, and calculate θ based on the error-adjusted β, mean squared error or σ_{u}^{2}, and baseline σ_{ε}^{2}. We do this for each country grade and for each specification (time series and cross sections, and country and grid cell data). Again N = 1000 for each of the different versions.

Tables 8a and 8b present the θ estimates without bootstrap procedure (the "baseline" value), and the statistics for bootstrapped θ (mean, standard deviation, and interquartile range) for country and cell analysis. The baseline values of θ (column 2 in Tables 8a and 8b) are very close to the results of θ estimator published in our early work (Chen and Nordhaus, 2011). We updated the GEcon data in summer 2011 and the present analysis is based on the latest version of the GEcon data. The results are consistent with our previous findings that the luminosity signal adds considerable information for D and E country grades, but adds very little information for A, B, and C country grades. This conclusion holds for both cross sections and time series, and for both country and cell analysis as well. We do not include a discussion of the country time-series estimates for E countries because the sample is too small.

Next, we focus on the bootstrapped θ results. Comparing the baseline θ (column 2 in Tables 8a and 8b) and the mean of bootstrapped θ (column 3 in Tables 8a and 8b), we find the baseline θ is generally slightly lower than the mean of bootstrapped θ, especially for country time series data. In instance, for C country grade (Table 8a), the base θ for time series is 0.022, while the mean of θ estimates is 0.047. For D country grade, the values are 0.272 and 0.350. The underestimation of θ is probably caused the non-linearity of the estimate in equation (8). For parameters that are relatively well determined, the non-linearity is unimportant, and the bootstrapped mean will be close to the baseline estimate. Table 8b shows for cell cross sections the baseline θ and mean of θ estimates are identical for all different country grades.

Tables 8a and 8b show that the distribution of the θ estimator varies by country grade and model specification. By country grade, we find the standard deviation and the (IQR) for θ estimators are much larger for E country grade than for other countries. Column 4 in Table 8a shows that the standard deviation and IQR of θ for cross section for E country grade is 0.033 and 0.043, while the highest numbers for other country grades are 0.006 and 0.008. This result indicates that the θ estimate is least reliable for countries with poorest statistics system. This is a discouraging result as these are the countries that could benefit most for an independent data source.

In addition to difference across country grade, we also found standard deviation and the IQR of θ are larger for time series than for cross sectional data, especially for D and E grades. The results of cell analysis (Table 8b) show that for D grade the standard deviation and IQR for time series data are 0.014 and 0.017, while for cross sections are 0.002 and 0.002. Similarly, for E grade, standard deviation and IQR for time series is 0.187 and 0.217, while for cross sections are only 0.008 and 0.011. This result is consistent with the results of β and RMSE estimates shown in Tables 7a-d and Figures 2 to 5, that the θ estimate is more reliable for cross section than time series data. We will discuss a graphical presentation of the results in the next section.

Sensitivity analysis for output measurement errors

In the second step, we need to test the sensitivity of the estimated optimal weight on luminosity proxy (θ) on the prior estimate of the measurement error of standard national-accounts output (σ_{ε}^{2}). We do not have reliable ways to estimate the precision of this measurement error. To test this question, we take values of the measurement errors of output that are one-half and two times the base values estimated above. This would seem a plausible bound on the measurement errors given the procedures used to derive them described above. However, we cannot place a statistical interpretation on the upper and lower numbers, and we therefore interpret these as sensitivity analyses.

For these calculations, we perform two more sets of bootstrap analysis. In each, we use the same bootstrap sample for β and σ_u^2 as used for the earlier calculations so that the only difference in the estimated θ is the error of measurement of standard GDP. Assuming that σ_ε is equal to, half, or double of the baseline σ_ε, we use equation (4) to calculate adjusted β, and then use equation (8) to calculate the new value of θ. Using doubled baseline σ_ε in equation (4) caused a problem in many replications because the new σ_ε^2 is larger than the value of σ_y^2. This is theoretically impossible in our specification because $\sigma_{y^*}^2$ cannot be negative (i.e., σ_ε^2 should be always less than σ_y^2). To deal with this problem, we set the upper bound for σ_ε^2 at 95 percent of σ_y^2 in the cases where σ_ε^2 is larger than σ_y^2. In the first step of analysis, we generate one set of β and σ_u^2 (N=1000) through bootstrapping regression and derive one set of θ (N=1000) based on baseline σ_ε. In the second step, we generate two additional sets of θ based on different value of σ_ε, and N is equal to 1000 for each set.

We present three sets of bootstrapped θ estimates from both step one and two in box plots in Figures 6 and 7. Figure 6 presents the results of three sets θ for each country grade, with the top panel for cross sections and the bottom one for time series. Figure 7 presents the comparable results for the grid cells. These box plots show that the value of estimated θ (the optimal weight on luminosity-based proxy) is in some cases quite sensitive to the prior estimate of the measurement errors in conventional GDP data.

We can explain the results using the result for the grade C countries in the top panel in Figure 6 as an example – this being the country cross sectional analysis. We see that the values of three sets of bootstrapped θ are extremely sensitive to the value of σ_ε. The middle box shows the distribution of θ when we apply the baseline value σ_ε (Table 5) to equation (4) and (8). For cross sectional country analysis for C grade countries in particular, our baseline value for measurement error for output level in standard output data is 20%. Using this number and the bootstrapped regression results, β and σ_u^2 (N = 1000), we calculate 1000 replications of error-corrected β and final θ (N = 1000). The middle line in the box indicates the median value of θ (N=1000). It is around 0.1 in the middle box for C countries here, which means the median optimal weight for luminosity-based proxy is 10%, and the median optimal weight for conventional measure is around 90%.

Similarly, the left-hand box plot shows the distribution of θ calculated with one-half baseline σ_ε. Here σ_ε = 10% for grade C countries. The median value of bootstrapped θ for this specification is close to zero with very little dispersion. Finally, the right-hand box plot shows the distribution of θ calculated with two times baseline σ_ε, that is, σ_ε = 40% for grade C countries. The corresponding median values for θ estimator increase to 0.30.

The box plots in Figures 6 and 7 also confirm our conclusion on sampling errors of θ estimator based on Table 8: Comparing time series and cross sectional results, we find the box size (the interquartile range) for cross sectional output is much smaller than for time series data, which indicates the θ estimator is more reliable for cross sections than for time series. The box sizes for cell cross sectional analysis are smallest among all specifications, suggesting the estimate of θ has highest reliability for this specification. On the other hand, the box plot for θ based on time series country data shows largest box size, that is, lowest reliability.

Finally, we find the only case that the value of θ is not sensitive to σ_ε is for time series cell analysis for the highest grade countries, particularly grade A. The box plot for this analysis in Figure 7 (the bottom panel) shows that changing the value of baseline σ_ε leads to very little change in θ. The precision of the output estimates is sufficiently high that even doubling the measurement error in conventional output still leads no change in the value of the luminosity proxy for these countries.

V. Conclusions

The purpose of present paper is to determine the value of night-time lights for measuring output. More specifically, we present revised estimates and examine the precision of the estimates of the optimal weights on conventional GDP and night-time lights data for estimating "true output" in countries and grid cells over the period 1992-2008.

Our major findings and recommendations are as follows. First, *for grade A and B countries*, there is no reason to use luminosity data as a supplement to standard data in any context where standard data are available. We found virtually no value added in these countries for either country or cell aggregates for either time series or cross sections. The lights data are not useful for A and B countries because the standard data are sufficiently reliable. These results are reasonably robust to statistical and sensitivity tests.

Second, we find that there is no advantage at present of using lights data for *time-series corrections* for countries or grid cells for any countries. Again, for A and B countries, there is no value added in the time-series lights data. For D and E and most C countries, the uncertainties in the estimates of the weights are too large at present to allow their use in construction of time-series estimates based on lights. We conclude that, without further refinement of the lights data (for example, developing a careful intercalibration of the data over years and satellites), lights data are not a reliable proxy for time-series measures of output growth. The one possible exception is that lights data may have use for grade C countries, but this would require further refinement.

Third, *for D and E countries for cross-sectional estimates of output*, the estimates suggest that there may be substantial information in the luminosity data. Our results indicate that the cross-sectional errors in estimating the optimal weights for D and E countries come primarily from uncertainty about the error in the standard output data and not in the measurement errors for lights or in the lights-output coefficient. Therefore, if the measurement errors in the cross section could be more precisely determined, there would be substantial information in the lights data that could be used to supplement current estimates of the level of output for both countries and grid cells.

Fourth, *the major concerns about use of lights as a proxy involve uncertainties about the precision of standard national accounts data*. We can derive satisfactory estimates of the uncertainties involved in using lights deriving from errors in lights and in the lights-output relationship. However, because we do not have techniques to estimate the reliability of standard national and regional output data, we cannot judge the degree of imprecision coming from this source. Our sensitivity analysis suggests that this source of uncertainty is likely to dominate the overall imprecision in the optimal weighting fraction between lights and conventional output. This conclusion reminds us of the admonition of Josiah Stamp (1929), "The Government are very keen on amassing statistics - they collect them, add them, raise them to the nth power, take the cube root and prepare wonderfu1 diagrams. But what you must never forget is that every one of the figures comes in the first instance from the village watchman, who just puts down what he damn pleases."

Finally, given these results, *we recommend that future work be concentrated on integrating luminosity data into the cross sectional estimates of national and regional output for D and E countries*. The main open issues in integrating lights with economic output in these cases involve estimating the reliability of national accounts data and not the reliability of the night-time lights data.

Estimated margin of error in PPP cross section	
Country grade	Error in level
A	9%
B	15%
C	21%
D	30%

Table 1. Cross-sectional errors estimated by Summers and Heston (1991)

Estimated margin of change in PPP price between PWT 6.1 and PWT 6.2	
Country grade	Change in level
A	2.9%
B	2.9%
C	10.9%
D	22.6%

Table 2. Cross-sectional errors in conventional national-accounts measures

Source: Errors are estimated by the change in country PPP level between PWT 6.1 and PWT 6.2 by Simon Johnson, William Larson, Chris Papageorgiou, and Arvind Subramanian (2009)

Difference between PWT 6.1 and 6.2			
Country grade	29-year growth rate	1-year growth rate*	
		Average	Median
A	0.108%	0.583%	0.350%
B	0.134%	0.720%	0.700%
C	0.576%	3.104%	1.642%
D	1.644%	8.852%	6.731%

* One-year growth rate is 29-year multiplied by square root of 29.

Table 3. Revision to growth rate of real GDP in Penn World Table

Source: Simon Johnson, William Larson, Chris Papageorgiou, and Arvind Subramanian (2009)

Country Grade	Standard deviation of ln(first estimate/last estimate)	
	Output	Population
A	36.1	33.4
B	24.4	20.9
C	79.5	33.4
D	57.4	41.8
E	na	na

Table 4. Revisions in estimates of gross cell product data from GEcon data

Source: Authors.

Country grade	Estimates for country output		Estimates for grid-cell output	
	1-year growth rate	Output level	1-year growth rate	Output level
A	0.6%	10%	1.2%	20%
B	0.8%	15%	1.6%	30%
C	3.0%	20%	4.0%	40%
D	5.0%	30%	5.0%	60%
E	6.0%	50%	8.0%	100%

Table 5. Estimates of errors of national and gridded GDP data used in estimates of combined measures of output

Source: Authors.

Grade level	Number of countries	Number of cells	Representative country
A	16	2,838	Australia, Canada, U.S.
B	13	880	Argentina, Germany, Spain
C	103	6,606	Bangladesh, Egypt, Mexico, Russia
D	29	859	Algeria, Cambodia, D.R. Congo, Libya
E	9	285	Iraq, North Korea, West Bank and Gaza
Total	170	11,468	

Table 6. Distribution of countries and cells without missing values by grade

Cells are at the 1° x 1° resolution. The sample of cells used in the growth rate analysis includes all available observations after merging the GEcon dataset (4.0) and DMSP-OLS Nighttime Stable Lights Time Series (Version 4) and taking logarithm of both variables.

Source: Authors.

Grade	Regression coefficient (β)			Root mean squared error (σu)		
	Mean	SD	IQR	Mean	SD	IQR
Grade A	0.777	0.022	0.030	0.509	0.050	0.074
Grade B	0.742	0.013	0.018	0.513	0.023	0.031
Grade C	0.990	0.009	0.013	0.707	0.009	0.012
Grade D	0.952	0.013	0.017	0.800	0.017	0.024
Grade E	1.318	0.043	0.059	0.638	0.030	0.040
All grades	0.957	0.006	0.008	0.727	0.008	0.011

Table 7a. Results for country cross sectional analysis.

Grade	Regression coefficient (β)			Root mean squared error (σu)		
	Mean	SD	IQR	Mean	SD	IQR
Grade A	0.161	0.420	0.567	0.170	0.021	0.028
Grade B	0.251	0.306	0.388	0.187	0.034	0.048
Grade C	0.424	0.213	0.292	0.328	0.030	0.041
Grade D	0.882	0.206	0.234	0.369	0.045	0.063
Grade E	0.011	0.389	0.443	0.287	0.067	0.081
All grades	0.525	0.157	0.221	0.339	0.023	0.032

Table 7b. Results for country time series analysis

Grade	Regression coefficient (β)			Root mean squared error (σu)		
	Mean	SD	IQR	Mean	SD	IQR
Grade A	0.764	0.002	0.003	1.285	0.006	0.008
Grade B	0.913	0.003	0.004	1.182	0.008	0.011
Grade C	0.896	0.002	0.002	1.584	0.003	0.005
Grade D	0.943	0.005	0.007	1.738	0.008	0.012
Grade E	0.727	0.013	0.018	1.930	0.013	0.017
All grades	0.875	0.001	0.002	1.591	0.003	0.004

Table 7c. Results for cell cross sectional analysis.

Grade	Regression coefficient (β)			Root mean squared error (σu)		
	Mean	SD	IQR	Mean	SD	IQR
Grade A	0.358	0.070	0.093	0.662	0.027	0.037
Grade B	0.526	0.110	0.155	0.641	0.036	0.049
Grade C	0.696	0.028	0.039	0.968	0.018	0.025
Grade D	0.451	0.117	0.153	1.036	0.039	0.051
Grade E	0.623	0.235	0.322	1.148	0.057	0.076
All grades	0.729	0.024	0.035	0.908	0.013	0.017

Table 7d. Results for cell time series analysis.

Note: The above statistics in Tables 7a to 7d are mean, standard deviation (SD), and interquartile range (IQR) for bootstrapped regression results for β and root mean squared error (σ_u) by grade. These use 1000 replications for each analysis.

	Grade	Baseline θ	Sample distriubtion of θ		
			Mean	SD	IQR
Cross sectional	A	0.022	0.024	0.004	0.007
	B	0.043	0.046	0.005	0.006
	C	0.074	0.074	0.002	0.003
	D	0.116	0.119	0.006	0.008
	E	0.620	0.637	0.033	0.043
	all	0.078	0.079	0.002	0.002
Time series	A	0.002	0.005	0.011	0.005
	B	0.002	0.007	0.012	0.008
	C	0.022	0.047	0.055	0.054
	D	0.272	0.350	0.159	0.186
	E	0.000	0.351	0.356	0.666
	all	0.043	0.059	0.041	0.053

Table 8a. Results for country data analysis

	Grade	Baseline θ	Sample distriubtion of θ		
			Mean	SD	IQR
Cross sectional	A	0.014	0.014	0.000	0.000
	B	0.053	0.053	0.001	0.001
	C	0.053	0.053	0.000	0.000
	D	0.114	0.114	0.002	0.002
	E	0.246	0.246	0.008	0.011
	all	0.043	0.043	0.000	0.000
Time series	A	0.001	0.001	0.000	0.000
	B	0.003	0.004	0.002	0.002
	C	0.017	0.017	0.002	0.002
	D	0.025	0.028	0.014	0.017
	E	0.499	0.672	0.184	0.217
	all	0.015	0.015	0.001	0.001

Table 8b. Results for cell data analysis

Note: Tables 8a and 8b list the initial estimation of θ based on the baseline σ_{ε}^{2} from Table 5. The last three columns in the table list the statistical properties (mean, standard deviation, and interquartile range) of sample distribution of θ through bootstrapping procedure (N = 1000).

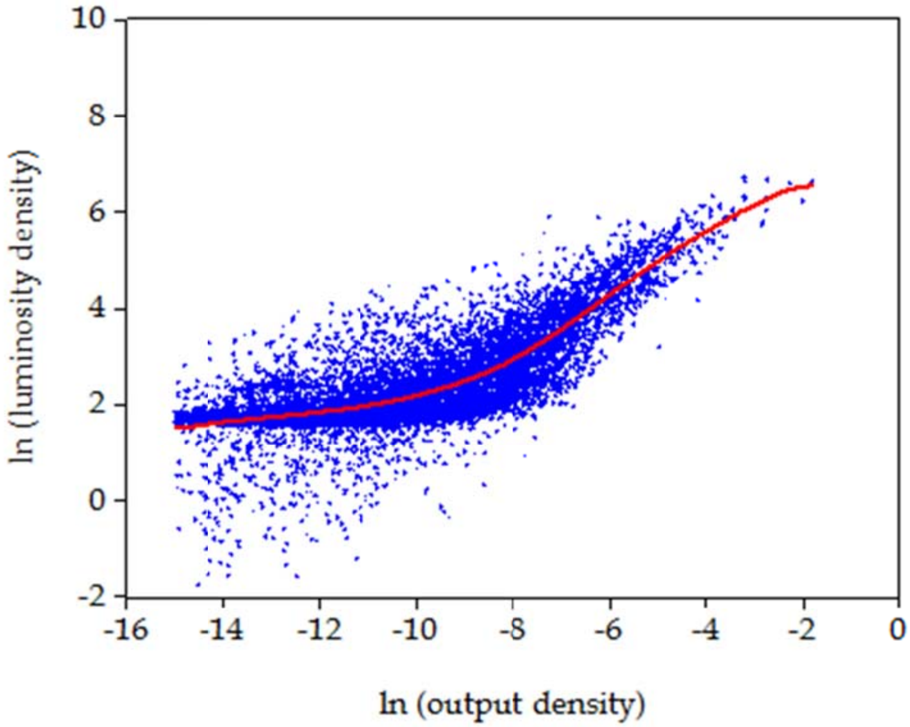


Figure 1. Gross cell product (GCP) and luminosity data, all cells

Figure shows the scatter plot of log calibrated luminosity for 2006 and log of gross cell product for all cells at the 1° x 1° resolution. Output density is gross cell product (PPP in billions in 2005 international $) per km^2. Luminosity density per km^2 is the radiance calibrated luminosity for 2006. They include all grid cells (N = 12,393). The solid line is the kernel estimator using an Epanechnikov kernel and 100 grid points per kernel. (Source: Chen and Nordhaus, 2011)

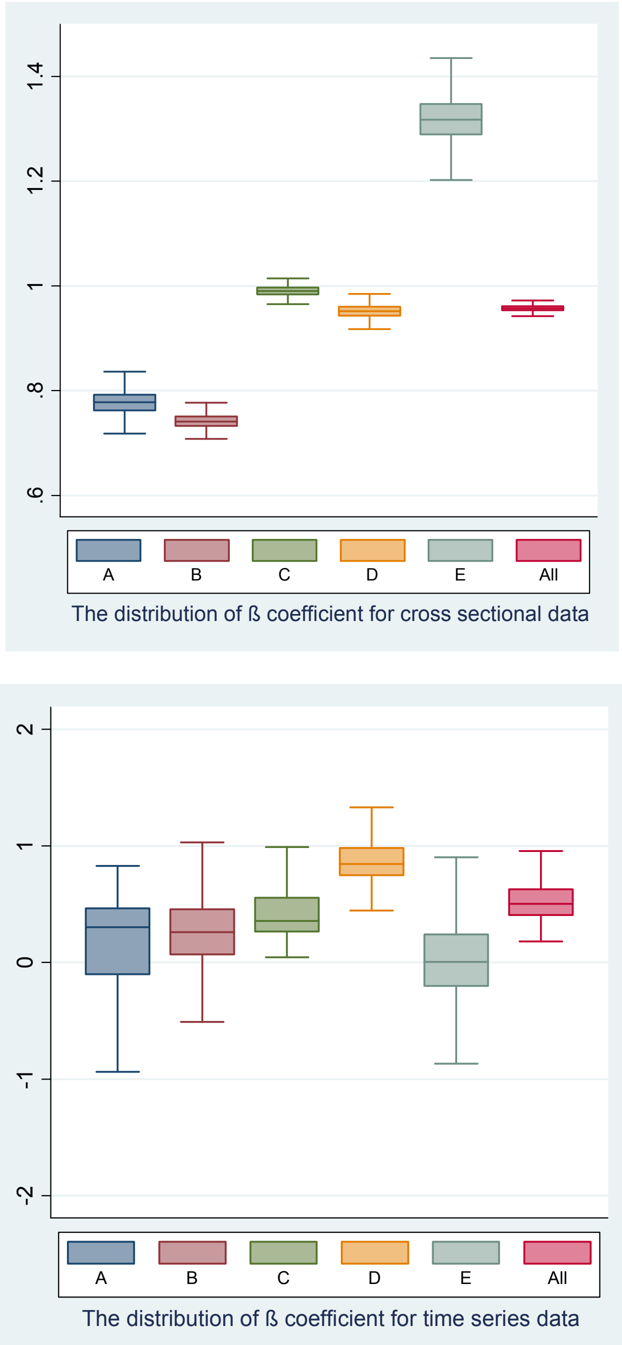


Figure 2. Box plot for bootstrapped regression coefficient (β) for each grade for country data

Figure shows the estimated elasticity of luminosity with respect to true output. The middle line in the box is median value, and the upper and lower edges of the box indicate the value at the 75th and 25th percentile. The upper hash mark indicates the 3rd quartile plus 1.5 IQR (interquartile range). The lower hash mark indicates the 1st quartile minus 1.5 IQR. The different boxes are different country grades as indicated.

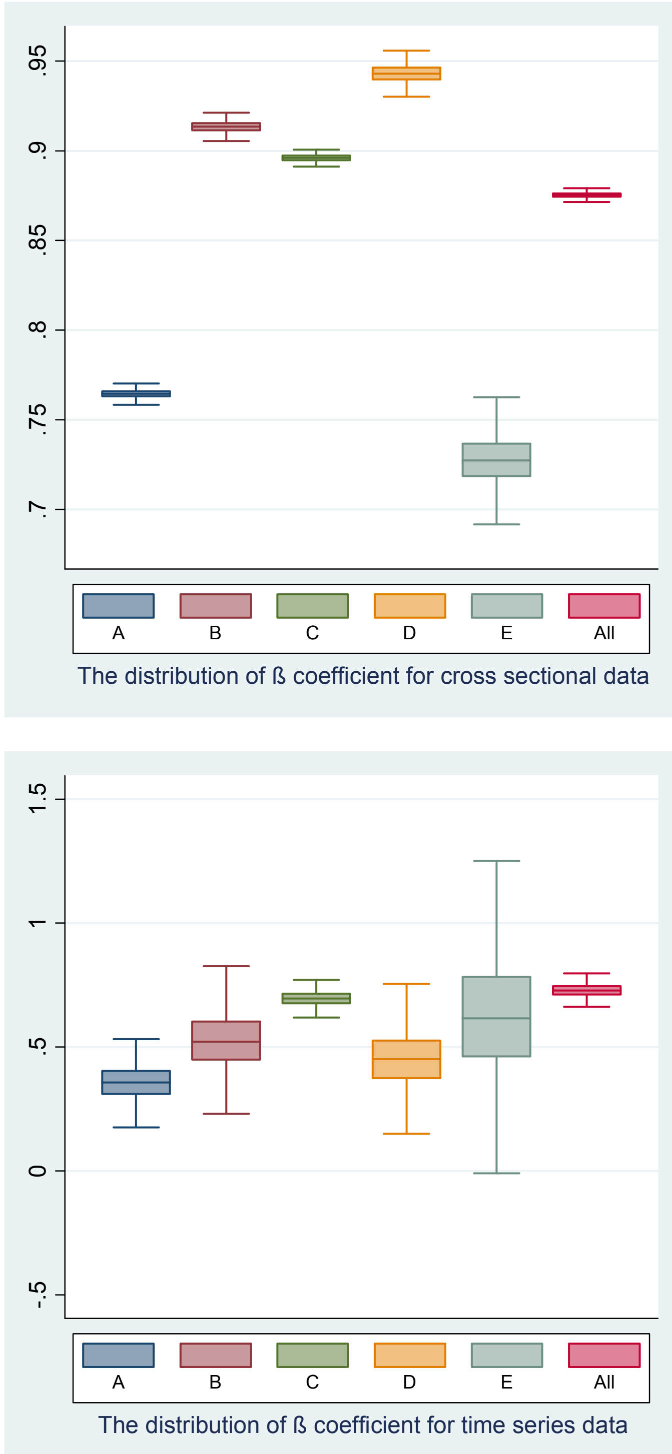
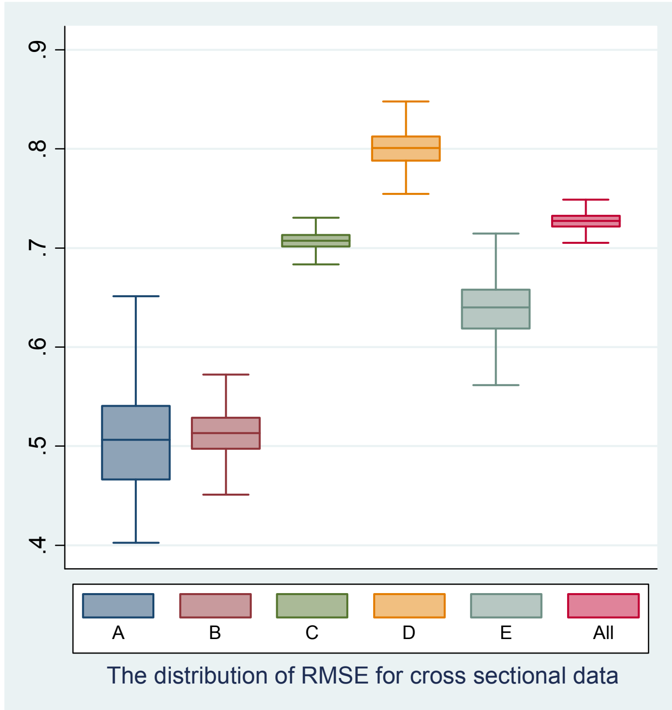


Figure 3. Box plot for bootstrapped regression coefficient (β) for each grade for grid-cell data

See Figure 2 for the explanation for the box plot and legend.



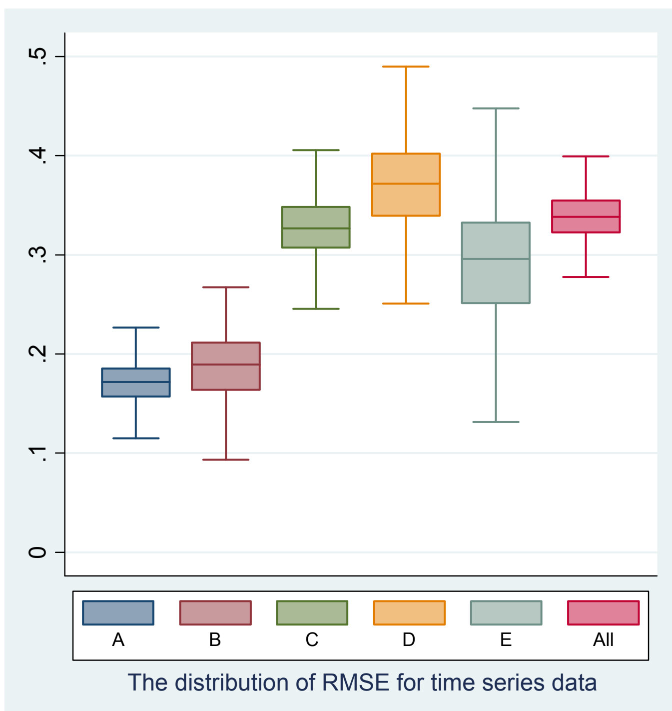
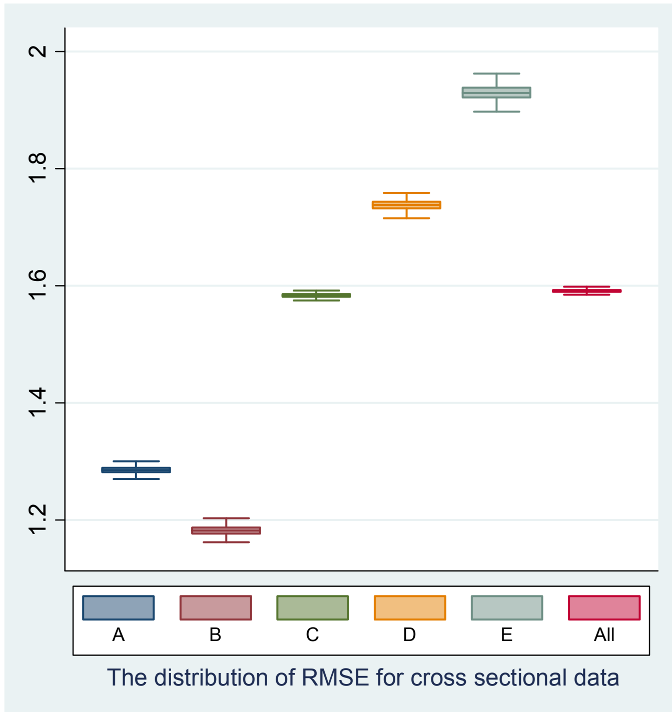


Figure 4. Box plot for bootstrapped root mean squared error (RMSE), σ_u , for each grade for country data

See Figure 2 for the explanation for the box plot and legend.



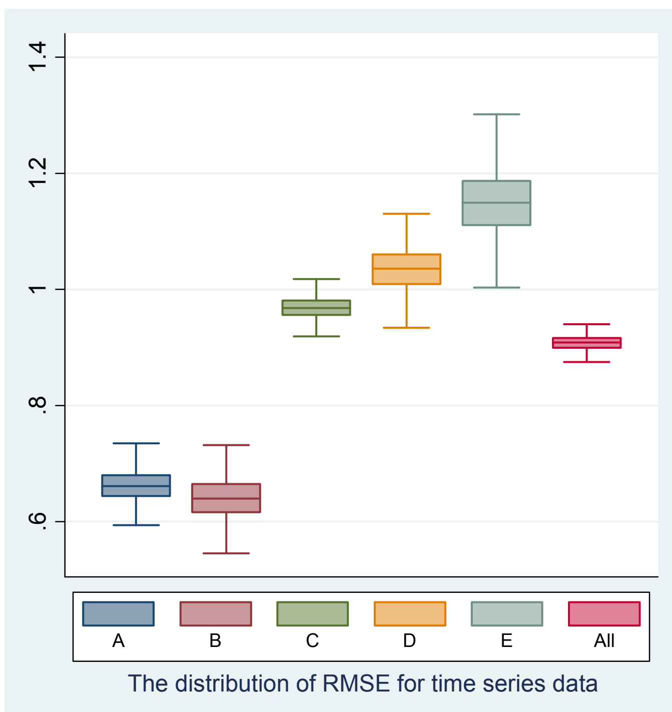
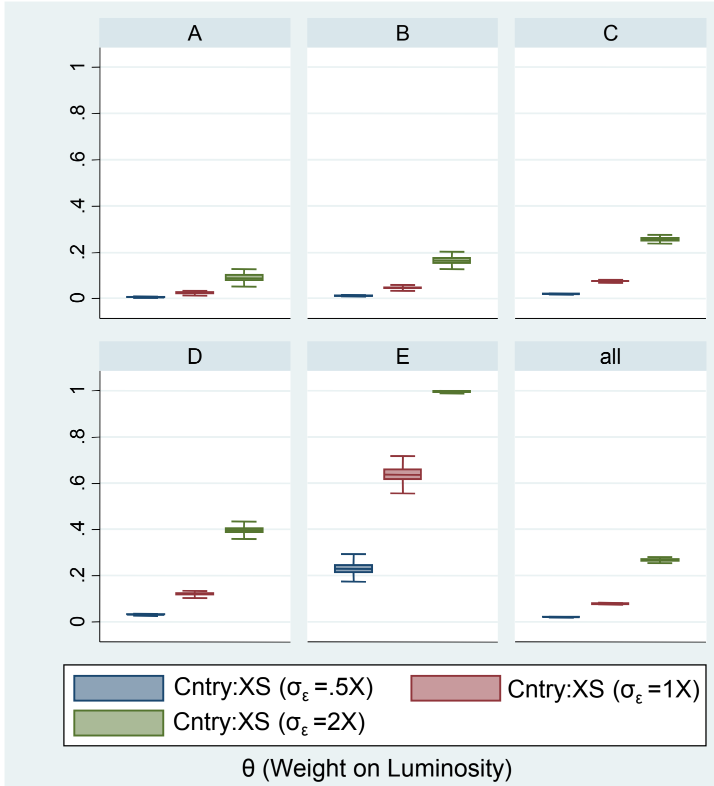


Figure 5. Box plot for bootstrapped root mean squared error (RMSE), σ_u **, for each grade for cell data**

See Figure 2 for the explanation for the box plot and legend.


A
B
C
D
E
all
0
.2
.4
.6
.8
1
Cntry:XS (σε =.5X)
Cntry:XS (σε =1X)
Cntry:XS (σε =2X)
θ (Weight on Luminosity)

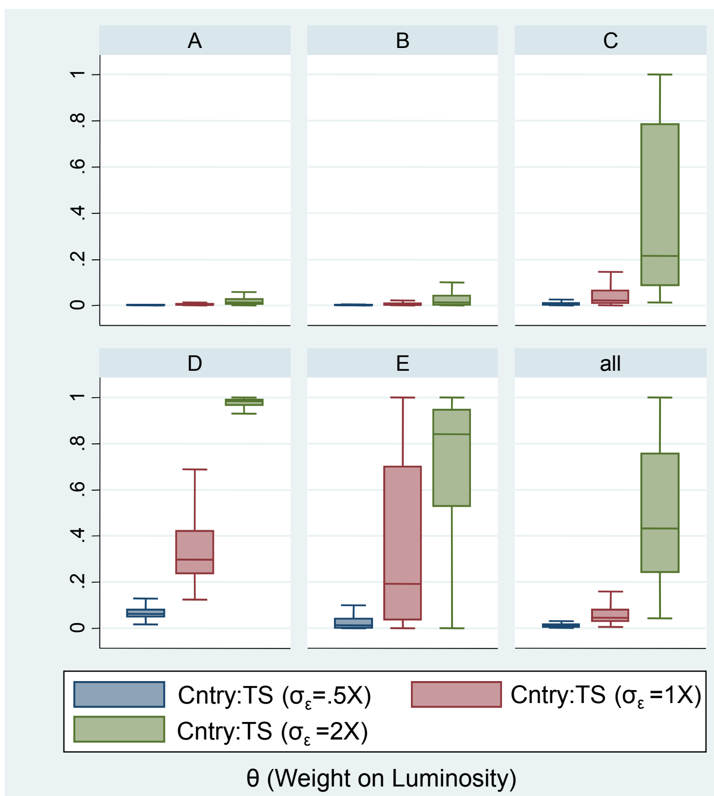
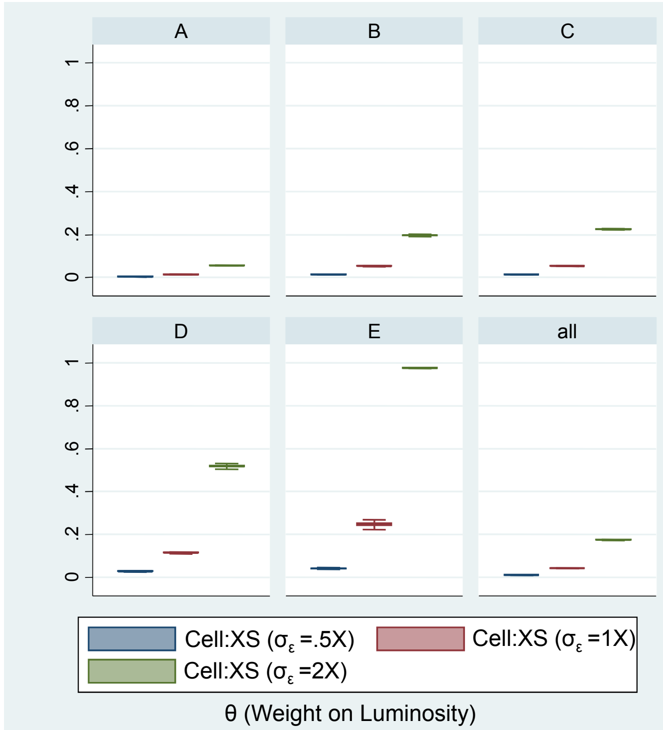

A
B
C
D
E
all
0
.2
.4
.6
.8
1
Cntry:TS (σε=.5X)
Cntry:TS (σε =1X)
Cntry:TS (σε =2X)
θ (Weight on Luminosity)

Figure 6. Box plot for bootstrapped θ estimator for cross sectional (XS) and time series (TS) data for countries.

Note to Figure 6: There are two graphs (the upper one for cross sectional and the lower one for time series analysis), and six panels in each graph above (one for each grade, A through E, and one for all observations). The figure shows three sets of θ for each panel. The left-hand box plot shows the distribution of θ estimator using an estimated error of the national accounts estimate of output equal to one-half the base value, σ_ε. The middle box plot shows the estimator using the base value of σ_ε (the results from the first step of the analysis). The right-hand box plot shows the estimator using two times the base value of σ_ε. Each box plot is based on 1000 bootstrapped observations of θ. See Figure 2 for the explanation for the box plot.



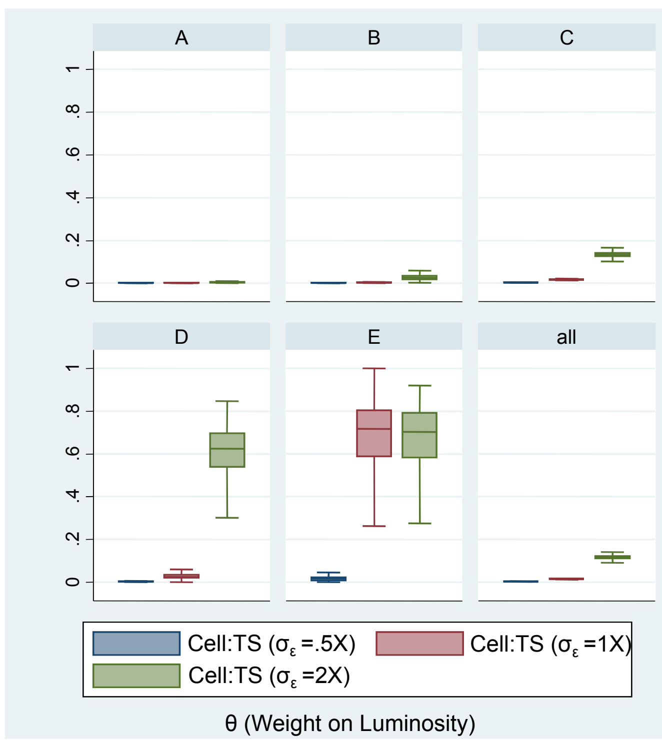


Figure 7. Box plot for bootstrapped θ estimator for cross sectional (XS) and time series (TS) data for grid cells.

For explanation, see Figure 6.

References

Chen, Xi, and William Nordhaus. 2011. "Using luminosity data as a proxy for economic statistics." *The Proceedings of National Academy of Sciences*, 108 (21): 8589-8594.

Davison, A. C., D. V. Hinkley, G. A. Young. 2003. "Recent Developments in Bootstrap Methodology." *Statistical Science,* 18(2): 141-157.

Efron, Bradley, and Robert Tibshirani. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." *Statistical Science*, 1(1): 54-77.

Fixler, Dennis. 2009. "Measurement error in the national accounts." in Jane M.

Binner, David L. Edgerton, and Thomas Elger, eds., *Measurement error: consequences, applications and solutions*, Bingley, UK, Emerald Group Publishing Ltd, 91-106.

Fixler, Dennis and Bruce Grimm. 2008. "The reliability of the GDP and GDI estimates." *Survey of Current Business*, 88(2): 16-32.

Freedman, David. 1981. "Bootstrapping Regression Models." *The Annals of Statistics*, 9(6): 1218-1228.

Froyen, Richard. 1996. *Macroeconomics: Theories and Policies,* New York, Prentice Hall.

Henderson, Vernon, Adam Storeygard, and David N. Weil. 2011. "A Bright Idea for Measuring Economic Growth." *American Economic Review*, 101(3): 194–99.

Henderson, Vernon, Adam Storeygard, and David N. Weil. 2011a. "Measuring Economic Growth from Outer Space." *American Economic Review*, forthcoming.

Johnson, Simon, William Larson, Chris Papageorgiou, and Arvind Subramanian. 2009. "Is Newer Better? Penn World Table Revisions and Their Impact on Growth Estimates." National Bureau of Economic Research Working Paper 15455.

Jorgenson, Dale, J. Steven Landefeld, and William Nordhaus, Eds. 2006. *Architecture for the National Accounts*, Chicago, Chicago University Press.

Kuznets, Simon. 1934. *National Income and Capital Formation, 1919-1935,* New York, National Bureau of Economic Research.

North, Gerald R, et al. 2006. *Surface Temperature Reconstructions for the Last 2,000 Years.* Washington, DC, National Academy Press.

Stamp, Josiah. 1929. *Some Economic Factors in Modern Life*. London, King and Son.

Summers, Robert, and Alan Heston. 1991. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988". *The Quarterly Journal of Economics*, 106: 327–68.