Sociological Methodology
2016, Vol. 46(1) 319–344

DOI: 10.1177/0081175016654737
http://sm.sagepub.com

10

# ADDRESSING MEASUREMENT ERROR BIAS IN GDP WITH NIGHTTIME LIGHTS AND AN APPLICATION TO INFANT MORTALITY WITH CHINESE COUNTY DATA

***Xi Chen****

**Abstract**

*As an emerging research area, application of satellite-based nighttime lights data in the social sciences has increased rapidly in recent years. This study, building on the recent surge in the use of satellite-based lights data, explores whether information provided by such data can be used to address attenuation bias in the estimated coefficient when the regressor variable, Gross Domestic Product (GDP), is measured with large error. Using an example of a study on infant mortality rates (IMRs) in the People's Republic of China (PRC), this paper compares four models with different indicators of GDP as the regressor of IMR: (1) observed GDP alone, (2) lights variable as a substitute, (3) a synthetic measure based on weighted observed GDP and lights, and (4) GDP with lights as an instrumental variable. The results show that the inclusion of nighttime lights can reduce the bias in coefficient estimates compared with the model using observed GDP. Among the three approaches discussed, the instrumental-variable approach proves to be the best approach*

*Quinnipiac University, Hamden, CT, USA

**Corresponding Author:**
Xi Chen, Quinnipiac University, 275 Mount Carmel Avenue, Hamden, CT 06518, USA
Email: xi.chen@quinnipiac.edu

*in correcting the bias caused by GDP measurement error and estimates the effect of GDP much higher than do the models using observed GDP. The study concludes that beyond the topic of this study, nighttime lights data have great potential to be used in other sociological research areas facing estimation bias problems due to measurement errors in economic indicators. The potential is especially great for those focusing on developing regions or small areas lacking high-quality measures of economic and demographic variables.*

**Keywords**

*measurement error, nighttime lights, GDP, demography*

## 1. INTRODUCTION

Conducting empirical studies with local-level economic or demographic data is often problematic for developing countries, which tend to have relatively poor statistical systems. As a result, observed indexes created by using conventional methods tend to have large systematic errors at the local level. Even if some regional economic and demographic statistics are collected regularly by national bureaus or agencies, the measurement errors in these indicators or indexes generally are not reported or are unknown. In empirical analysis, using variables with large measurement errors as a regressor can produce attenuation bias in coefficient estimates. In sociological studies, scholars often use economic indicators, such as income or Gross Domestic Product (GDP), as a regressor or control variable in hypotheses testing. Attenuation bias caused by large measurement error in such variables can lead scholars to underestimate the true effect of economic conditions on the subject of interest. As a result, researchers often focus on analyzing data at the country or province/state level, as national information can be obtained easily from international agencies (Babones 2013). A proxy variable or instrumental variable can provide a potential solution for attenuation bias problems. However, valid proxy or instrumental variables themselves often cannot be found at the local level either.

Take China as an example. Economists have been uncertain of the true value of China's GDP, and it has recently been reported that the real value of its GDP per capita was likely 30% higher in 2005 than the World Bank estimated (Feenstra et al. 2013). Among the reasons that cause the real GDP value to be underestimated, Feenstra and colleagues

emphasize the discrepancy in price data provided by China for urban and rural areas, different index number methods, and different concepts of real GDP that were adopted in GDP calculations. If at the national level the conventional method of collecting and calculating GDP is problematic, it is not hard to imagine that at the local level the data involve even more significant problems of estimation errors. Often, the true extent of such GDP error at the county or city level in China is unknown, as no literature has been identified that has investigated such topics. It is assumed in this research that data collected on the local level has greater measurement errors than data collected at the national level. This inevitably leads to problems in empirical analysis that uses Chinese GDP as a regressor to estimate the impact of economic conditions on other social phenomena in China. In China's case, a valid proxy measure or instrumental variable may not be available, as the People's Republic of China (PRC) publishes few or no local data. Thus, for studies that use local-level data, such as county or city level data, the true effect theorized may be considerably underestimated.

To address this issue related to a lack of high-quality data, the use of information collected through unconventional, nonsurvey methods, including remote sensing information, has been suggested as a possible solution (Babones 2013; Goodchild and Janelle 2004). One such source that is emerging as a promising alternative to conventional measures is nighttime lights data. This source provides great potential in estimating regional, local development, and demographic changes, as the data have been shown to highly correlate with main economic indicators (Chen and Nordhaus 2011; Henderson, Storeygard, and Weil 2011). The conceptual explanation of the close association between lights and economic development is quite straightforward and self-evident. The degree of luminosity captured from space depends on artificial light on the ground, which is generated by electricity (Mellander et al. 2015). Regions with high economic activities—especially business-dense and population-dense commercial and industrial areas—tend to use more electricity and therefore are brighter at night. The advantage of lights data collected from space is that such information covers almost the entire globe, it is updated almost instantly by satellites, and it can provide information at very small scales and is less likely to be affected by reporting or coding errors due to discrepancies in definition, collection, and calculation across regions or countries—all problems often encountered with conventional data.

This article is a first attempt to investigate whether nighttime lights can be used to address the miscalculation in conventional GDP by providing additional information or can be used as an instrumental variable in a model that predicts infant mortality rate (IMR). In theory, IMR is associated with many development indicators at the local level, such as income level, income inequality, per capita GDP, or urban population (Frisbie 2006). With China's rapid economic growth since the 1970s, IMR declined from 63 in the 1970s to 18 during the period between 2005 and 2010 (United Nations 2013). Yet, the economic gains have been geographically uneven, occurring largely in coastal areas and larger cities. Much less is known about China's regional local economic development over the last 20 years and its links to IMR at the county or city level. Data limitation has hindered research efforts in such studies. An exception to this was work conducted by Poston (1996) about 20 years ago. Using 1,441 cases from Census 1980, the author found that the development index is one of the strongest predictors of local IMR. Despite the availability of more recent China census data, no other empirical work has been identified to further investigate this topic.

The primary goal of this article is to explore how lights data can be included in a regression model that reduces the bias in coefficient estimates of GDP. Different statistical approaches are introduced and discussed, and then the empirical analyses based on these approaches are conducted with Chinese county-level data. Through evaluating the results of the analyses, I recommend the use of nighttime lights data as a general method potentially applicable to many fields of sociology in addressing the attenuation bias due to measurement error in observed economic indicators.

## 2. THE USE OF NIGHTTIME LIGHTS IN SOCIAL SCIENCE STUDIES

Nighttime lights data are imagery data collected by the U.S. Department of Defense, and they are processed from satellite-based remote sensing information. The publicly accessible lights data were developed by the Defense Meteorological Satellite Program—Operational Linescan System (DMSP-OLS 2015). The earliest satellite-based information has existed since the 1970s, but its application within social science research dates back only to the 1990s.

Most early studies using nighttime lights were conducted and published by geoscience and remote sensing field scholars to estimate a wide range of demographic and economic indicators, including population density, GDP, income per capita, wealth, urbanization, and energy use (Doll, Muller, and Elvidge 2000; Ebener et al. 2005; Elvidge et al. 1997; Elvidge et al. 2001; Elvidge et al. 2007; Elvidge et al. 2009; Noor et al. 2008; Sanderson et al. 2000; Sutton, Elvidge, and Ghosh 2007). Over the last decade, the number of studies exploring the usefulness of lights has increased, and among these studies the topics have expanded to include electricity usage rates (Elvidge et al. 2011), global distribution of economic activity (Ghosh et al. 2010a), and fossil fuel carbon dioxide emissions (Ghosh et al. 2010b; Oda and Maksyutov 2011).

Social scientists have only recently begun paying attention to the potential of nighttime lights data, especially in the field of economics. Over the last five years, increasing numbers of economists have been looking to the lights data for economic output estimates at both national and subnational levels with formal statistical approaches (Chen and Nordhaus 2011; Henderson et al. 2011; Henderson, Storeygard, and Weil 2012; Nordhaus and Chen 2015). Nordhaus and Chen's analyses show that lights can provide more additional information in estimating GDP for poor countries than for middle-income and rich countries, as the poor countries often report GDP with large errors. Lights data are also found to closely correlate with small area poverty for many poor and developing countries (Chen 2015, 2016). Likely due to China's increasing importance on the world stage and concurrently the lack of high-quality measures on some variables, China has also been the focus of at least a handful of studies using lights to estimate GDP, regional economic development, and electric power consumption (Li et al. 2013; Shi et al. 2014; Zhao, Currit, and Samson 2011). It is estimated that lights data have been used in almost 3,000 studies to investigate economic phenomena alone (Nordhaus and Chen 2015). These studies, establishing the strong association between lights and economic indicators, undoubtedly have now formed a foundation for further applications of lights in disciplines beyond economics.

In terms of different approaches to using lights, there are largely three main categories. The first primarily explores the association between lights and other variables. Most early work on nighttime lights belongs to this category. Often in these studies, correlation statistics are used to demonstrate the close association between lights and other variables.

The second category focuses on how to use lights data information to create new indicators, such as the human footprint index (Sanderson et al. 2000) and Night Light Development Index (NLDI) (Elvidge et al. 2012), or to improve a current measurement of indicators, such as GDP and grid cell output (Chen and Nordhaus 2011; Nordhaus and Chen 2015) and national income (Henderson et al. 2012; Henderson et al. 2011). The final category uses lights data as a regressor in hypothesis testing. The rationale for using lights data in regression is the large body of existing literature that has demonstrated a strong correlation between lights and demographic and economic indicators. For example, Bharti et al. (2011) used lights directly as a population density measure in a model that predicted seasonal fluctuations of measles in urban areas in Niger, where direct population measures are not available. The lights variable is also used as a proxy for economic shocks to predict civil conflict in African countries (Hodler and Raschky 2014), and it has been used as a development indicator to predict steel stock for buildings and civil engineering infrastructure in China (Liang et al. 2014).

Although using lights as a regressor in a model is innovative and may indeed be able to address the poor data quality and measurement error issues in many research areas, the strength and weakness of the different methods of using lights in hypothesis-testing have not been carefully explored. To address issues regarding measurement error bias, there are three potential approaches, among which two have been previously suggested or implemented in the literature reviewed above. The first approach uses lights directly as a regressor to substitute the variable that researchers are interested in testing but that is often unavailable or measured with a large error. The variable being substituted can include a population and economic development index, as lights are indeed highly correlated with these variables. The second approach is to replace the theoretical variable with a modified, improved measure based on additional information from lights. Although the formal statistical procedure of how to use lights to improve current measures of GDP has been thoroughly discussed (Chen and Nordhaus 2011; Nordhaus and Chen 2015), its further applications in hypothesis testing have not yet been explored. Finally, the conventional solution for measurement error bias is the instrumental variable approach, and lights can potentially be a valid instrumental variable, as these data are highly correlated with GDP; in addition, because lights data are collected by NASA satellites instead of by local or regional government agencies, they do not suffer

the same problems of conventional GDP measures. To simplify the terms in the analysis that follows, I refer to the first approach as the *direct substitution approach*, the second as the *synthetic measure approach*, and the third as the *instrumental variable approach*.

Section 3 discusses each approach in detail in terms of whether it can reduce the measurement error bias in regression coefficient estimates. In the context of predicting county IMR with GDP values, four models are specified: (1) those using observed GDP values published by the PRC, (2) those using the lights variable as a direct substitute for GDP, (3) those using synthetic GDP measures based on weighted lights and observed GDP, and (4) those using lights as an instrumental variable for GDP. The primary goal is to see whether lights-based measures, in one way or another, can reduce the bias in GDP coefficient estimators. The results of four regression models using a Chinese county sample in the year 2000 are compared to show the improvement of estimated GDP coefficients with the different methods.

## 3. METHODS

The main question in the analysis is how GDP affects IMR. To simplify the expressions, we consider only the case of regression with a single regressor:

$$w = c + Y\mu + u, \tag{1}$$

where $w$ is IMR. We can assume that it is measured without errors, $c$ denotes a constant, $u$ is an error term, and $Y$ is the theoretical variable, GDP. The true GDP, $Y$, is measured with error $\varepsilon$:

$$y = Y + \varepsilon. \tag{2}$$

The ordinary least squares (OLS) estimator via the regression of $w$ on $y$ yields a biased estimator of $\mu$:

$$w = c_{gdp} + y\mu_{gdp} + u_{gdp}$$

and

$$\hat{\mu}_{gdp} = \left(\frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_\varepsilon^2}\right)\mu. \tag{3}$$

In the hope that lights can provide a solution to reduce bias in the OLS estimator, the three different approaches are presented and discussed below.

### 3.1. *Direct Substitution Approach*

The direct substitution approach uses the lights variable as a simple substitute for the GDP measure in the equation. In such a case, the relationship between lights and GDP is overlooked. Although lights data are highly correlated with many variables, including GDP, they can also be influenced by other physical and social factors not related to GDP. Thus, the correlation coefficient between lights and GDP is not equal to 1, and using lights directly as a regressor can still cause a bias in estimating the true effect of GDP on the dependent variable, IMR.

The equation for such an approach can be expressed as

$$w = c_{light} + m\mu_{light} + u_{light}, \tag{4}$$

where $m$ is an observed light measure. The relationship between true GDP and the observed lights variable can be expressed with the following equation:

$$Y = m\gamma + v. \tag{5}$$

Here, $\hat{\mu}_{lights}$is still a biased estimator of true GDP, as the relationship between $\hat{\mu}_{lights}$ and $\mu$ can be expressed with the equation

$$\hat{\mu}_{lights} = \mu * \gamma. \tag{6}$$

In equation (5), $\gamma$ is the standardized coefficient. Unless $\gamma$ is equal to 1, the direct substitution approach will lead to a downward bias, which is determined by $\gamma$.

### 3.2. *Synthetic Measure Approach*

This approach first combines information from lights and observed GDP to create an improved, synthetic measure of GDP that is then used as a regressor in the model. The study adopts Nordhaus and Chen's (2015) method, as the authors have specified a statistical procedure to use lights to improve current measures of GDP. According to this method, the precise amount of information provided by lights can be

calculated. Theoretically, the synthetic measure created through optimal weights of lights and observed GDP is the best estimate of true GDP. However, its further implementation in regression models has not been explored in detail. One caveat related to this method is that the value of synthetic GDP is determined by three parameters that can potentially cause bias in the final GDP coefficient estimate in IMR regression models, although it may be better than the OLS estimator based on observed GDP values.

The equation to be estimated with a synthetic value can be expressed as the following:

$$w = c_{syn} + s\mu_{syn} + u_{syn}. \tag{7}$$

A synthetic GDP measure, $s$, can be calculated based on weighted information from both observed GDP, $y$, and the observed lights variable $m$. The section that follows illustrates how to calculate synthetic GDP, according to Chen and Nordhaus (2011) and Nordhaus and Chen (2015). We can begin with an assumption that there is a structural relationship between observed lights and true GDP:

$$m = \alpha + Y * \beta + \tau. \tag{8}$$

A proxy measure of GDP can be calculated by inverting the coefficient $\beta$. However, because the true values of $Y$ are unknown, the true value of $\beta$ is unknown. We can use observed GDP and lights to estimate $\beta$ in equation (8), with the resultant estimation referred to as $\hat{\beta}$. This coefficient is biased due to measurement error in $y$, $\varepsilon$. To get the error-corrected estimates of the structural coefficient, $\tilde{\beta}$, we can use the classic errors-in-variable correction, as shown below:

$$\tilde{\beta} = \left( \frac{\sigma_y^2 + \sigma_\varepsilon^2}{\sigma_y^2} \right) \hat{\beta}.$$

Here, $\hat{\beta}$ is the regression coefficient based on observed data. $\sigma_y^2$ and $\sigma_\varepsilon^2$ can be determined if $\varepsilon$ is known.

The proxy GDP variable $z$ is calculated by inverting $\tilde{\beta}$:

$$\hat{z} = (1/\tilde{\beta})m.$$

The proxy measure of GDP, $z$, provides an alternative measure to the observed GDP value. To improve the current observed GDP measure,

we need information from both observed value $y$ and the lights-based proxy measure $z$. A synthetic measure of $Y$ can then be calculated by taking weighted averages of $y$ and $\hat{z}$:

$$\hat{s} = (1 - \theta)y + \theta\hat{z}. \tag{9}$$

The optimal weight, $\theta$, is important as it indicates how much information is from lights-based GDP (see the detailed discussion in Nordhaus and Chen [2015] and Appendix A in the online journal for $\theta$ estimation):

$$\theta = \frac{\beta^2\sigma_\varepsilon^2}{\beta^2\sigma_\varepsilon^2 + \sigma_\tau^2}.$$

Here $\sigma_\tau^2$ and $\beta$ can be consistently estimated with equation (8). The value of $\sigma_\varepsilon^2$is based on a prior estimate of measurement error in Chinese county-level GDP. Let us assume it is 40%, as it is the value used for a 1 degree latitude by 1 degree longitude grid cell of economic output in Chen and Nordhaus (2011), and the measurement error at the national level can be up to 30% (Feenstra et al. 2013). The synthetic measure of GDP, $s$, can be calculated once $\theta$ is known, and it can then be used as a regressor in equation (7). The $\mu_{syn}$ in equation (7) is an OLS estimator, and it can be expressed as a function of $\mu$ (see details in Appendix B in the online journal):

$$\mu_{syn} = \frac{(1 - \theta + \theta * \frac{\beta}{\tilde{\beta}}) * \sigma_Y^2}{(1 - \theta + \theta * \frac{\beta}{\tilde{\beta}}) * \sigma_Y^2 + (1 - \theta)^2\sigma_\varepsilon^2 + (\frac{\theta}{\tilde{\beta}})^2\sigma_\tau^2} * \mu. \tag{10}$$

According to the above equation, $\mu_{syn}$ has the following three properties: First, when $\theta$ equals 0—that is, when no weight is provided by lights—$\mu_{syn}$ equals $\mu_{gdp}$ in equation (3), and again $\mu_{syn}$ is a downward-biased estimator of $\mu$. Second, under the assumption that $\beta = \tilde{\beta}$, the equation can be written as

$$\mu_{syn} = \frac{\sigma_Y^2}{\sigma_Y^2 + (1 - \theta)^2\sigma_\varepsilon^2 + (\frac{\theta}{\tilde{\beta}})^2\sigma_\tau^2} * \mu.$$

Here, the estimation bias is not determined by the error in $y$, $\sigma_\varepsilon$, as in equation (3), but rather is determined by the weighted error terms of $\sigma_\varepsilon$ and $\sigma_\tau$. In addition, because the final synthetic value is a product of weighted observed and lights-based measure of GDP, the extent of bias

in $\mu_{syn}$ is also determined by the weight, $\theta$. Third, unless $\sigma_\varepsilon$ is zero, $\mu_{syn}$ will always be a downward-biased estimator of $\mu$, because a part of the denominator $(1-\theta)^2\sigma_\varepsilon^2 + (\frac{\theta}{\beta})^2\sigma_\tau^2$ is always larger than zero. However, this bias can be smaller than the model using observed GDP, particularly when $\sigma_\varepsilon^2 < \frac{\theta\sigma_\tau^2}{\beta(2-\theta)}$. Furthermore, when assumptions of the two-stage least squares (2SLS) method are not met, the synthetic approach probably is a better solution than the substitution approach, as it considers the structural relationship between lights and GDP ($\sigma_\tau^2$ and $\beta$).

### 3.3. *Instrumental Variable Approach*

The final possible way of using lights information is to treat lights as an instrumental variable and estimate the GDP coefficient with a 2SLS regression. The benefit of using lights as an instrumental variable is that it not only meets the two criteria (explained below) of being a valid instrumental variable for GDP but it is also available almost anywhere in most time periods and such data can be aggregated to various scales. The lights variable is highly correlated with observed GDP (criteria 1), as shown in Table 1 and Figure 1, and it is less likely to correlate with the error term (criteria 2) in the original regression equation using observed GDP, or $u_{gdp}$ in equation (3).

The measurement error in GDP provided by the Chinese government is more likely to correlate with the error term in the regression model, because measurement error in GDP, $\varepsilon$, is strongly influenced by the incompetency of local statistical systems (Feenstra et al. 2013), which can also relate to $u_{gdp}$, the error term in the OLS regression. Satellite lights data collected and processed by U.S. agencies are not influenced by the data collection procedures of the Chinese government, and they are therefore less likely to have the error that is correlated with $u_{gdp}$. In the first-stage regression, the observed GDP, $y$, is regressed on lights variable $m$. The OLS estimate of the lights coefficient can then be used to project $y$. The projected value, $\hat{Y}_{2sls}$, is used as the regressor in the second-stage regression:

$$w = c_{2sls} + \hat{Y}_{2sls}\mu_{2sls} + u_{2sls}$$

and

**Table 1.** Correlation Coefficients of Variables

| | IMR | GDP | Synthetic GDP | Lights | Total Population | Illiterate Females (%) | Female Workers (%) | Agricultural Population (%) | General Hospital | Maternal and Child Health Service |
|---|---|---|---|---|---|---|---|---|---|---|
| GDP | −.5195 | | | | | | | | | |
| Synthetic GDP | −.5481 | .9965 | | | | | | | | |
| Lights | −.6301 | .8779 | .9149 | | | | | | | |
| Total population | −.3378 | .7045 | .7007 | .6106 | | | | | | |
| Illiterate females (%) | .4032 | −.3911 | −.4107 | −.4635 | −.3172 | | | | | |
| Female workers (%) | .1231 | .0304 | .0157 | −.0568 | .1409 | .2059 | | | | |
| Agricultural population (%) | .1089 | −.0265 | −.0475 | −.1438 | .1697 | .2363 | .4393 | | | |
| General hospital | .1628 | −.3617 | −.3571 | −.2986 | −.478 | .0942 | −.1223 | −.2501 | | |
| Maternal and child health service | .1175 | −.3428 | −.3373 | −.2765 | −.4355 | .1157 | −.0519 | −.1037 | .2919 | |
| Family planning service | .2258 | −.319 | −.328 | −.3377 | −.3017 | .0957 | .0262 | .0301 | .0769 | .1865 |

*Note:* GDP = Gross Domestic Product; IMR = infant mortality rate.

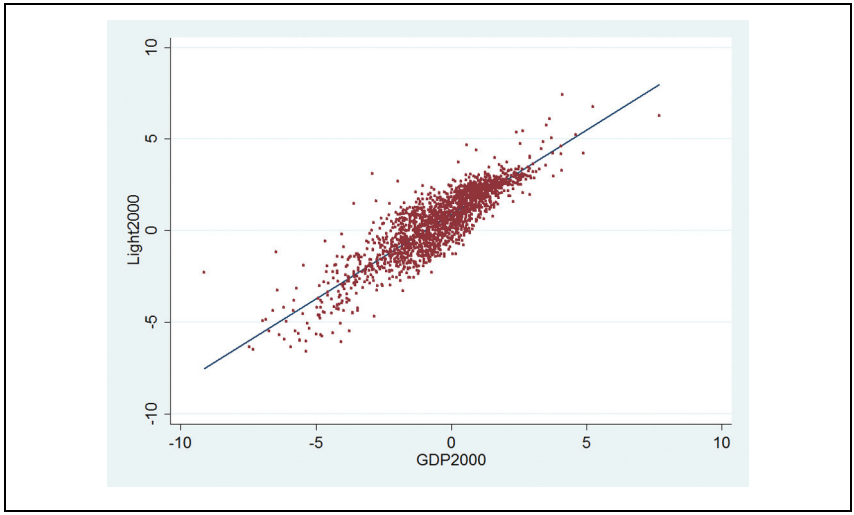


**Figure 1.** Scatterplot of log Gross Domestic Product (GDP) and log lights for Chinese counties, 2000.

$$\mu_{2sls} = \mu + \frac{Corr(m, u_{2sls})}{Corr(m, y)} * \frac{\sigma_{u_{2sls}}}{\sigma_y}. \tag{11}$$

The instrumental variable (IV) estimator is consistent if $Corr(m, u_{2sls}) = 0$. The bias occurs when $Corr(m, u_{2sls})$ is large and $Corr(m, y)$ is very small. In this study, $Corr(m, y)$ is .88. Thus, if $Corr(m, u_{2sls})$ is zero or very small, the IV estimator will have a very small bias. In the specific case of China, air pollution in Chinese cities has increased with GDP growth over the last decade and could correlate with model errors in the IMR model. Air pollution could also have a dimming effect on lights. However, no studies have been identified investigating to what extent air pollution influences nighttime lights quality in China. Considering that the data are from the year 2000, a period when air pollution in China was not as severe as it is today, we can assume it does not pose a serious problem for the IV approach here. In the above three approaches, estimators are derived in the case of a simple linear regression with one regressor. In the case of multiple regression, the conditional variances and covariance of variables should be considered. In Section 4, the results of three approaches using Chinese county-level data are compared and the difference in outcomes is further explored.

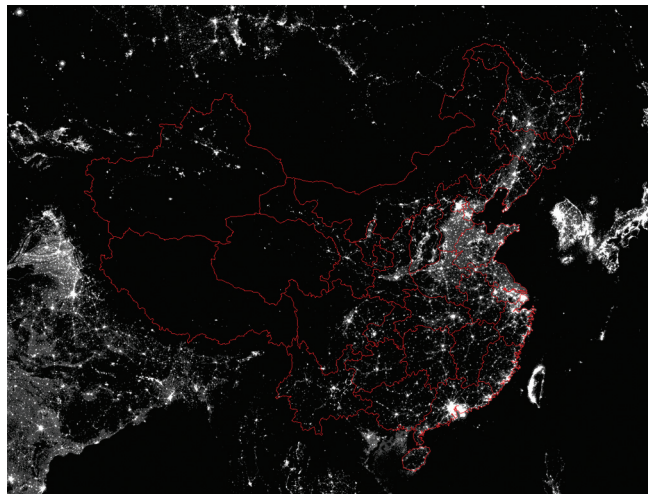

**Figure 2.** Radiance (calibrated) nighttime light map of China.
*Data source:* DMSP-OLS 2014.

## 4. DATA AND ANALYSIS

### 4.1. *Lights Data*

The nighttime lights data used in this study are obtained from the National Geophysical Data Center (NGDC) of the National Oceanic and Atmospheric Administration (http://ngdc.noaa.gov/eog/). NGDC provides multiple lights data sets, among which stable and raw lights are annual averaged lights stored in 30 arc-second grids and available from 1992 to 2013 (DMSP-OLS 2015). Because stable lights are processed from raw lights with multiple clean-up steps, they are considered a better quality and are used most often in the lights studies mentioned in Section 2. Most lights data cover the entire globe, except areas with high latitude—that is, beyond 75°N and 65°S latitude. In addition to time series data, DMSP also generates "radiance" or "calibrated" lights (Figure 2). In comparison to stable lights, calibrated lights introduce less measurement error caused by saturation and overglowing, but they are only available for certain years. Because the calibrated lights are available for the year 2000, which is the same year that China county and city IMR was measured, the analysis that follows uses calibrated

lights. Log lights density is the main variable used below. It is measured with the logarithmic value of average digital numbers (DNs) of lights per square kilometer, capturing the average intensity of lighting at night. (Additional information on digital numbers can be found at the NOAA DMSP website, http://ngdc.noaa.gov/eog/.)

### 4.2. *China County/City Data*

The small area IMR data are obtained from China Data Online (chinadataonline.org). The IMR is a crude measure, and it is defined as 1 minus the percentage of the number of living children to live births in the year 2000. There are 2,869 valid observations for this variable, including almost all counties defined by the China 2000 census. The corresponding administrative boundary digital map can also be downloaded from China Data Online. Other county variables downloaded from this website include total population, percentage of illiterate female to total population of 15 and over, percentage of employed female population to total employed population, percentage of agricultural population over total population, general hospitals per 100,000 persons, maternal and child health services per 100,000 persons, and family planning services per 100,000 persons. Population density is defined as the total population residing in a county area divided by the area size in square kilometers. The general economy indicator, GDP, is downloaded from "County Statistics" from China Data Online. The data provide county-level GDP (in 100 million yuan) in 2000 for 2,065 observations. Based on GB (Guobiao) code and county names, only 1,964 observations are merged into IMR data. Then all China demographic and GDP data are merged with lights data. The sample, without missing values, is 1,915. All variables are expressed in natural logarithms, except three local health condition control variables—general hospital, maternal and child health services, and family planning services, as many counties have none of these facilities.

Four models are analyzed: model 1 uses observed GDP as the regressor in OLS; model 2 uses lights data as a substitute for a GDP measure; model 3 uses a lights-based synthetic measure of GDP; and model 4 uses a lights variable as an instrumental variable and 2SLS regression. A set of covariates are also included in all four models. The comparison of the results can help determine whether inclusion of lights data can improve the GDP coefficient estimates compared with the model using observed

values of GDP, and it can determine which model produces the estimates with the least attenuation bias in GDP coefficient estimates. The standard errors of estimated coefficients are adjusted with provincial clusters in all models to address potential heteroscedasticity problems. In model 4 (IV method), the standard error reported is adjusted by taking into account original GDP measures.

## 5. RESULTS

As shown earlier, Table 1 reports correlation coefficients for all variables used in regression models. Observed GDP and lights are strongly correlated ($r$ = .878). This is also shown in the scatterplot of log GDP and log lights in Figure 1. The strong association between GDP and lights does not seem to vary across more or less developed counties. The scatterplot (Figure 1) shows that their association is quite consistent across all counties, and correlation coefficients for counties with GDP above and below the GDP mean are .75 and .76, respectively. Differences in correlation coefficients are also small between the poorest county group ($p$ = .75 for counties in lowest quantile of GDP) and the most developed county group ($p$ = .71 for counties in the highest quantile of GDP). Thus, although economic conditions of Chinese counties are highly uneven, the lights and GDP relationship is quite consistent across all counties.

The synthetic GDP using an optimal weight method is also calculated based on equation (9). The value of $\theta$ is predicted based on sample estimates and the prior estimate of measurement error in $y$. Specifically, for the sample data, the regression coefficient of lights on observed $y$, $\hat{\beta}$ is .923 ($p < .0001$). The measurement error in reported Chinese county-level GDP in 2000 is estimated to be 40%, and the estimated value of $\tilde{\beta}$ is therefore equal to .973. According to equation (9), the synthetic measure of GDP can be calculated as

$$\mathrm{GDP}_{syn} = .839 * \mathrm{GDP} + .161 * light * 1.028.$$

Here, .839 is the weight for observed GDP, and .161 is the weight for lights-based proxy GDP. The error-corrected, inverted coefficient in the GDP-light structural model is 1.028. In other words, in the synthetic measure of GDP, about 84% of information is from the observed GDP and 16% of information is from lights-based estimates of GDP. Table 1

shows that the synthetic measure is strongly correlated with both observed GDP ($r$ = .997) and lights ($r$ = .915).

Table 2 reports regression results for four models that predict IMR at the county level. Again, all variables are converted to logarithmic values, except for general hospital, maternal and child health, and family planning services. Because both the theoretical regressor and dependent variable are log-transformed, we can compare the coefficients across models, as they reflect the ratio of the percentage change in IMR to the percentage change in GDP measures. The coefficient value of GDP varies substantially across the four models. It is the lowest in model 1 (−.195), which uses observed GDP. It is the highest in model 4, which uses lights as an instrumental variable of GDP (−.354). The values of the coefficients in model 2 (using direct substitution) and model 3 (using synthetic GDP) are very close and in between the values in model 1 and model 4. Comparing the goodness-of-model-fit index, $R$-squared, model 1 also has the lowest value, while model 4 has the highest.

Table 3 reports the Wald tests for equal coefficients across any two models. The tests reject the null hypothesis in all pairs, except for the pair in models 2 and 3. These results indicate significant improvement in coefficient estimates in all lights-based models (models 2, 3, and 4) over the model using observed GDP (model 1). Furthermore, the coefficient of the 2SLS approach is significantly higher than the coefficients in the other three models.

These results suggest that the measurement error in current Chinese county GDP leads to substantial downward bias in OLS estimates. Most likely such error is caused by mistakes or inconsistency in coding, reporting, and calculating in Chinese local or national statistical data. Subsequently, when lights information is included, the downward bias is reduced significantly in models 2, 3, and 4, most likely due to the fact that lights data are not collected by conventional methods. However, comparing models that use lights information, model 4 shows the largest increase in both GDP coefficient and IMR variance explained. The estimation bias in model 2 is probably caused by the imperfect association between lights and true GDP, as lights can be influenced by other factors that are unrelated to GDP, as illustrated in equation (5). We can expect that when the correlation between true GDP and lights is very high, the downward bias in the coefficient estimator (model 2) could be relatively small. The estimator in model 3 is also a biased one, as the

**Table 2.** Regression Results on Infant Mortality Rate

| | Model 1: OLS Using Observed GDP | Model 2: OLS Using Lights as GDP Measure | Model 3: OLS Using Synthetic GDP | Model 4: 2SLS |
|---|---|---|---|---|
| GDP | −.195** | | | −.354*** |
| | (−3.55) | | | (−5.58) |
| Lights | | −.247*** | | |
| | | (−5.99) | | |
| Synthetic GDP | | | −.220*** | |
| | | | (−4.03) | |
| Total population | .048 | .066 | .072 | .237* |
| | (.50) | (.87) | (.76) | (2.28) |
| Illiterate females (%) | .286* | .186+ | .263* | .208* |
| | (2.50) | (1.79) | (2.38) | (1.99) |
| Female workers (%) | .304 | .265 | .302 | .337 |
| | (1.11) | (1.10) | (1.14) | (1.33) |
| Agricultural population (%) | −.014 | −.180 | −.052 | −.140 |
| | (−.07) | (−.88) | (−.25) | (−.60) |
| General hospital | .007 | .002 | .005 | −.002 |
| | (.54) | (.20) | (.42) | (−.14) |
| Maternal and child health service | −.051* | −.040+ | −.051* | −.063* |
| | (−2.41) | (−1.87) | (−2.35) | (−2.26) |
| Family planning service | .017 | .007 | .014 | .008 |
| | (1.32) | (.71) | (1.20) | (.80) |
| Constant | 4.792** | 5.669*** | 4.748** | 3.230* |
| | (3.20) | (3.97) | (3.21) | (2.00) |
| *N* | 1915 | 1915 | 1915 | 1915 |
| Adjusted *R*-squared | .332 | .422 | .354 | .425 |

*Note:* GDP = Gross Domestic Product; OLS = ordinary least squares; 2SLS = two-stage least squares.

+$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$ in two-tailed tests.

coefficient, $\mu_{syn}$, is determined by the error term in observed GDP ($\varepsilon$) as well as the structural parameters $\beta$ and $\tau$, as shown in equation (10). In comparison with other estimated coefficients, model 4, using lights as an instrumental variable, produces a better result. In this particular study, the Durbin and Wu-Hausman test rejects the null hypothesis (both statistics are significant at the .0001 level) that the observed GDP is uncorrelated with error terms in the structural model. Therefore, an instrumental-variables estimator is really needed here. The method section also proves that if assumptions of a valid instrumental variable are satisfied, the 2SLS estimator is unbiased. For Chinese county data, the

**Table 3.** The Values of Chi-square in the Wald Tests of Equal Coefficients across Models

| | Model 1: OLS Using Observed GDP | Model 2: OLS Using Lights as GDP Measure | Model 3: OLS Using Synthetic GDP | Model 4: 2SLS |
|---|---|---|---|---|
| Unstandardized coefficient | −.195 | −.247 | −.22 | −.354 |
| Robust standard error | .055 | .041 | .055 | .063 |
| Wald test for equal coefficient. The null hypothesis is $b1 - b2 = 0$. | | | | |
| Model 2 | 3.45$^{+}$ | | | |
| Model 3 | 17.71*** | 1.31 | | |
| Model 4 | 28.96*** | 36.02*** | 29.31*** | |

*Note:* The above test considers the covariance of the estimators in addition to standard errors of estimators, as the models are fit on the same sample, and therefore the estimators are not stochastically dependent. GDP = Gross Domestic Product; OLS = ordinary least squares; 2SLS = two-stage least squares.

$+p < .10$. $*p < .05$. $**p < .01$. $***p < .001$ in two-tailed tests.

test on a weak instrumental variable reports the minimum eigenvalue statistic of 2895.01, which rejects the null hypothesis that the nighttime lights variable is a weak instrument.

In short, to address GDP measurement error issues in regression analysis, the instrumental variable approach seems to provide a better solution, if assumptions for the IV regressions are met. The changes in the values of coefficient and goodness-of-fit statistics of IMR regressions illustrate this. Using observed GDP measures, the model 1 results indicate that for a 1% of increase in GDP, the IMR is reduced by .196%, while with lights as an instrumental variable for GDP, the results in model 4 actually show that county-level GDP has a much larger impact on reducing IMR: A 1% increase in GDP leads to a .354% decrease in IMR. The effect estimated with the input of lights information almost doubles the effect estimated with observed GDP data. The IV method can potentially address multiple sources of endogeneity bias including measurement error, omitted variables, and reverse causation. In this particular analysis, additional explanatory variables are added to reduce bias caused by omitted variables, and reverse causation seems counterintuitive, so the IV approach here most likely corrects the bias due to measurement error in GDP.

The coefficients of other independent variables are also shown in Table 2, among which female education level and maternal and child health services are significant. Specifically, the percentage of illiterate women is positively associated with IMR and is significant in all four models. The number of maternal and child health services per 1,000 persons is negatively associated with local IMR, and the coefficient signs are consistent in all models and significant at $\alpha = .05$ in three models. Because the primary goal of this study is to determine the most appropriate approach of using lights in addressing estimation bias, literature and results related to other predictors will not be discussed in detail here. However, more studies on these topics are needed because there is only a small body of literature on *local* IMR and its determinants.

## 6. DISCUSSION

The primary goal of this study is to demonstrate how nighttime lights can be used to address measurement error issues in economic indicators, particularly in GDP, when economic indicators are used to predict other social patterns or changes. The Chinese IMR study presented here is used as an empirical example for introducing lights data and statistical approaches, which can have wide implications for future sociological studies. First, economic index is crucial to understanding many subjects in sociology, including studies dealing with political systems, demographic processes, health outcomes, cultural adaption, inequality, and other social problems, just to name few. In almost every subfield of sociology, we can identify some midrange theory hypothesizing the effect of economic conditions on the topic that is central to this field, such as economic push-pull factors in immigration studies. In addition, economic conditions are often treated as control variables when other theoretical variables are the subject of research interest. If the index of economic condition is measured with large error, the results of empirical analysis problematize our findings, due to the attenuation bias. That is, the effect of economic conditions may not appear significant or will appear substantially smaller than what is actually the case. The existence of such a problem not only undermines the methodological rigor of the discipline but also provides misleading information to scholars for theory building and empirical forecasting and to governmental agencies for policymaking. Although satellite-based nighttime lights data have made breakthrough methodological contributions in economics

over the last 10 years, with more than 3,000 articles discussing their use since 2000 (Nordhaus and Chen 2015), their implication and further contribution to sociology studies have not yet been fully investigated. With formal statistical approaches, this study shows both mathematically and empirically that lights data are able to address attenuation bias when economic measures are used as a regressor.

Furthermore, sociological research on smaller areas, developing regions, or nations with restricted information flow can benefit from findings in this study. The lights data and methodology approaches presented here provide a potential solution for data-quality problems. Data of small areas or from developing regions, such as sub-Saharan Africa, in general have large measurement errors. With appropriate methods, including primary analysis on the reliability of lights,[1] researchers can improve the empirical results substantially. In other cases, lights, as a source for additional information for both economic and demographic indicators, can help researchers working in areas where governments are not willing to publish such information. China is such an example. Fortunately, Chinese local-level information cannot hide from satellites, and the finest resolution lights data published by NOAA (VIIRS lights) can provide information by 15 arc-second (450 meters) geographic grids, and they are not affected by flaws in the collection process or limited by political administrative boundaries. As shown in IMR studies, lights as an instrumental variable for Chinese county GDP, where no other high-quality data are available, improve the estimated coefficient on IMR significantly. Similar approaches can be applied to study other topics in China—for example, labor migration, protests, wealth inequality—or to study those same topics in other developing regions of the world.

Another area that can potentially benefit from the lights data and methods presented here is cross-national comparison research. Measurement issues are crucial in cross-national research (Singh 1995). The attenuation bias caused by measurement error in cross-national indicators can even lead to unexpected results predicted by otherwise valid theoretical propositions (You and Khagram 2005). Measurement errors in such studies commonly exist due to data incompatibility in time frame, different variable definitions and operationalizations, and inconsistent units of analysis across countries or regions. Lights data can potentially provide solutions to these problems, as they cover the

entire globe and collect data across regions at the same time, and they can be aggregated to the scale needed by particular research. With rigorous statistical methods, such as the synthetic variables and 2SLS suggested in this study, lights can be used to effectively address attenuation bias problems in cross-national comparison studies, especially in those using economic indicators such as GDP, income, and business establishment. Studies using population density and urbanization can also consider the applicability of lights, as lights data are highly correlated with these variables as well (see Section 2). Additional studies on how to apply lights in cross-national comparison studies are needed. In short, nighttime lights data, coupled with rigorous methods, can address a wide range of methodological problems that are currently unanswered in sociology studies.

## 7. CONCLUSION

This study is the first to show that nighttime lights data can be used to address estimator bias issues due to measurement error in a regressor. It further points out the wide range of possible applications of nighttime lights in sociological studies. In addition to methodological expositions, the particular methods are explained with an illustration of a GDP and IMR study at the Chinese county level. The methodology section shows that the lights variable can be used to address attenuation bias through three approaches: using lights as a direct substitute measure for GDP, creating a synthetic variable with optimal weights from lights and observed GDP, and using lights as an instrumental variable. The three estimators and their potential biases are discussed thoroughly. Compared with an OLS estimator with observed GDP, the substitution and synthetic measure approaches also cause attenuation bias, but the magnitude of their bias is mostly or partially determined by the parameters of the structural equation between lights and true GDP. Both approaches could potentially reduce the coefficient estimate bias in the regression models in the case that measurement error in observed GDP is very large. However, the degree of reduction is determined by other parameters (see again the methods discussion in Section 3). The analysis using Chinese county-level data in the year 2000 shows that the estimated effects of GDP are higher in substitution and synthetization models than in the model that uses observed GDP. Although using

lights as an instrumental variable for economic indicators has not been widely practiced, the methodological and empirical exposition in this study suggests that such an approach can be very effective in addressing bias in coefficient estimates caused by large measurement error in observed economic data. In short, when lights information is added in one form or another, improvement in the estimated economic effect and overall model fit can be substantial, compared with models that use only observed economic data. In the case of Chinese IMR studies, the improvement in the instrumental variable approach is most significant.

The approaches discussed here can be implemented in many other sociology research areas. As widely used predictors or controls in sociology studies, the economic, demographic, and urbanization variables measured with conventional survey methods can cause biased estimation when they are measured with large errors and included in regression models. Theoretically these variables have a great impact on a wide spectrum of subjects of interest to the field. Prior to this study, there were no effective solutions to the empirical problem of attenuation bias across many fields of sociology. Lights data hold a great potential to address such a significant problem. In addition, such data can also meet the needs of researchers with interests in developing regions of the world, small areas, or comparison studies. Coupled with statistical approaches, lights data can improve methodological rigor in these studies, given that data quality is a quite common problem, often becoming an obstacle for advanced empirical analysis.

Even though research into the use of satellite-based nighttime lights data, or other remote-sensing information in the social sciences, is still at an early stage, there now appears little doubt that with appropriate methods, such information can potentially address some limitations or deficiencies in research based on conventional data collection or calculation. With more satellite imagery data made available to researchers over the last several decades, there is a possibility that sociology and other social science disciplines will expand research beyond the limitations imposed by current data problems for some parts of the world. This not only benefits scholars, by opening avenues of new information and hence new studies, but also aids policymakers, by providing more precise regionalized information with which to create and assess policy.

## Acknowledgments

I thank the following individuals for feedback on earlier drafts of this paper: Salvatore Babones (University of Sydney), Zhipeng Liao (UCLA), William Nordhaus (Yale University), Dudley Poston (Texas A&M University), and Joel M. Vaughan and Keith Kerr (Quinnipiac University).

## Notes

1. Researchers should check the association between lights and the variable of their interest prior to using lights data in a formal statistical model, as such association can vary by country or vary across subsamples, even though this is not the case for GDP and lights for Chinese counties. See Chen and Nordhaus (2011) for more explanation on the variation of the relationship between lights and GDP.

## References

Babones, Salvatore J. 2013. *Methods for Quantitative Macro-comparative Research*. Thousand Oaks, CA: Sage Publications.

Bharti, Nita, Andrew J. Tatem, Matthew Ferrari, Rebeeca Grais, Ali Djibo, and Bryan Grenfell. 2011. "Explaining Seasonal Fluctuations of Measles in Niger Using Nighttime Lights Imagery." *Science* 334(6061):1424–27.

Chen, Xi, and William Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *The Proceedings of National Academy of Sciences* 108(21): 8589–94.

Chen, Xi. 2015. "Explaining Subnational Infant Morality and Poverty Rates: What Can We Learn from Night-time Lights?" *Spatial Demography* 3(1):27–53.

Chen, Xi. 2016. "Using Nighttime Lights Data as a Proxy in Social Scientific Research." Pp. 301–23 in *Recapturing Space: New Middle-range Theory in Spatial Demography*, edited by F. Howell, J. Porter, and S. Matthews. New York: Springer International.

DMSP-OLS. 2015. "National Defense Meteorological Satellite Program, Image and Data processing by NOAA's National Geophysical Data Center. DMSP data collected by the US Air Force Weather Agency." Retrieved April 10, 2015 (http://www.ngdc.noaa.gov/dmsp/dmsp.html).

Doll, Christopher, Jan-Peter Muller, and Christopher Elvidge. 2000. "Nighttime Imagery as a Tool for Global Mapping of Socio-economic Parameters and Greenhouse Gas Emissions." *Ambio* 29(3):157–62.

Ebener, Steve, Christopher Murray, Ajay Tandon, and Christopher Elvidge. 2005. "From Wealth to Health: Modeling the Distribution of Income per Capita at the Sub-national Level Using Night-time Light Imagery." *International Journal Health Geographics* 4(1):5–14.

Elvidge, Christopher D., Kimberly E. Baugh, Eric Kihn, Herbert Kroehl, E. R. Davis, and C. W. Davis. 1997. "Relation Between Satellites Observed Visible-near Infrared Emissions, Population, Economic Activity and Electric Power Consumption." *International Journal of Remote Sensing* 18(6):1373–79.

Elvidge, Christopher D., Marc L. Imhoff, Kimberly E. Baugh, Vinita Ruth Hobson, Ingrid Nelson, Jeff Safran, John B. Dietz, and Benjamin T. Tuttle. 2001. "Night-time Lights of the World: 1994–1995." *ISPRS Journal of Photogrammetry and Remote Sensing* 56(2):81–99.

Elvidge, Christopher D., Jeffrey Safran, Benjamin Tuttle, Paul Sutton, Pierantonio Cinzano, Donald Pettit, John Arvesen, and Christopher Small. 2007. "Potential for Global Mapping of Development via a Nightsat Mission." *GeoJournal* 69 (1-2): 45–53.

Elvidge, Christopher D., Paul C. Sutton, Tilottama Ghosh, Benjamin T. Tuttle, Kimberly E. Baugh, Budhendra Bhaduri, and Edward Bright. 2009. "A Global Poverty Map Derived from Satellite Data." *Computers and Geosciences* 35(8):1652–60.

Elvidge, Christopher, Kimberly Baugh, Paul Sutton, Budhendra Bhaduri, Benjamin Tuttle, Tilottama Ghosh, Daniel Ziskin, and Edward H. Erwin. 2011. "Who's in the Dark: Satellite Based Estimates of Electrification Rates." Pp. 211–24 in *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*, edited by X. Yang. Chichester, England: Wiley-Blackwell.

Elvidge, Christopher, Kimberly Baugh, Sharolyn J. Anderson, Paul Sutton, and Tilottama Ghosh. 2012. "The Night Light Development Index (NLDI): A Spatially Explicit Measure of Human Development from Satellite Data." *Social Geography* 7(1):23–35.

Feenstra, R. C., H. Ma, J. Peter Neary, and D. S. Prasada Rao. 2013. "Who Shrunk China? Puzzles in the Measurement of Real GDP." *The Economic Journal* 123(573): 1100–29.

Frisbie, W. P. 2006. "Infant Mortality." Pp. 251–82 in *Handbook of Population*, edited by D. L. Poston Jr. and M. Micklin. New York: Springer.

Ghosh, Tilottama, Rebecca Powell, Christopher Elvidge, Kimberly Baugh, Paul Sutton, and S. Anderson. 2010a. "Shedding Light on the Global Distribution of Economic Activity." *The Open Geography Journal* 3(1):148–61.

Ghosh, Tilottama, Christopher Elvidge, Paul Sutton, Kimberly Baugh, Daniel Ziskin, and Benjamin Tuttle. 2010b. "Creating a Global Grid of Distributed Fossil Fuel $CO_2$ Emissions from Nighttime Satellite Imagery." *Energies* 3(12):1895–1913.

Goodchild, Michael F., and Donald G. Janelle. 2004. *Spatially Integrated Social Science*. New York: Oxford University Press.

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. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102(2):994–1028.

Hodler, Roland, and Paul A. Raschky. 2014 "Economic Shocks and Civil Conflict at the Regional Level." *Economics Letters* 124(3):530–33.

Li, Xi, Huimin Xu, Xiaoling Chen, and Chang Li. 2013. "Potential of NPP-VIIRS Nighttime Light Imagery for Modeling the Regional Economy of China." *Remote Sensing* 5(6):3057–81.

Liang, Hanwei, Hiroki Tanikawa, Yasunari Matsuno, and Liang Dong. 2014. "Modeling In-use Steel Stock in China's Buildings and Civil Engineering Infrastructure Using Time-series of DMSP/OLS Nighttime Lights." *Remote Sensing* 6(6):4780–4800.

Mellander, Charlotta, José Lobo, Kevin Stolarick, and Zara Matheson. 2015 "Night-time Light Data: A Good Proxy Measure for Economic Activity?" *PLoS One* 10(10): e0139779.

Noor, Abdisalan M., Victor A. Alegana, Peter W. Gething, Andrew J. Tatem, and Robert W. Snow. 2008. "Using Remotely Sensed Night-time Light as a Proxy for Poverty in Africa." *Population Health Metric* 6(1):1–13.

Nordhaus, William, and Xi Chen. "A Sharper Image? Estimates of the Precision of Nighttime Lights as a Proxy for Economic Statistics." *Journal of Economic Geography* 15(1):217–46.

Oda, Tomohiro, and Shamil Maksyutov. 2011. "A Very High-resolution (1 km × km) Global Fossil Fuel $CO_2$ Emission Inventory Derived Using a Point Source Database and Satellite Observations of Nighttime Lights." *Atmospheric Chemistry and Physics* 11(2):543–56.

Poston, Dudley. 1996. "Patterns of Infant Mortality." Pp.47–65 in *China: The Many Facets of Demographic Change*, edited by A. Goldstein and W. Feng. Boulder, CO: Westview Press.

Sanderson, Eric W., Malanding Jaiteh, Marc A. Levy, Kent H. Redford, Antoinette V. Wannebo, and Gillian Woolmer. 2000. "The Human Footprint and the Last of the Wild." *Bioscience* 52(10):891–904.

Shi, Kaifang, Bailang Yu, Yixiu Huang, Yingjie Hu, Bing Yin, Zuoqi Chen, Liujia Chen, and Jianping Wu. 2014. "Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales: A Comparison with DMSP-OLS Data." *Remote Sensing* 6(2):1705–24.

Singh, Jagdip. 1995. "Measurement Issues in Cross-national Research." *Journal of International Business Studies* 26(3):597–619.

Sutton, Paul, Christopher Elvidge, and Tilottama Ghosh. 2007. "Estimation of Gross Domestic Product at Sub-national Scales Using Nighttime Satellite Imagery." *International Journal of Ecological Economics and Statistics* 8(S07):5–21.

United Nations, Department of Economic and Social Affairs, Population Division. 2013. *World Population Prospects: The 2012 Revision*. New York: Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.

You, Jong-Sung, and Sanjeev Khagram. 2005 "A Comparative Study of Inequality and Corruption." *American Sociological Review* 70(1):136–57.

Zhao, Naizhuo, Nate Currit, and Eric Samson. 2011. "Net Primary Production and Gross Domestic Product in China Derived from Satellite Imagery." *Ecological Economics* 70(5): 921–28.

## Author Biography

**Xi Chen** is an assistant professor of sociology at Quinnipiac University. She was previously a research scientist for the Department of Economics at Yale University before joining Quinnipiac University, and she is currently still managing the Yale G-Econ project (Gecon.yale.edu). Her research specialties include quantitative methods, demography, ethnic population in China, and the application of satellite-based nighttime lights data.