INSTITUTE OF DEVELOPING ECONOMIES

IDE Discussion Papers are preliminary materials circulated
to stimulate discussions and critical comments

IDE DISCUSSION PAPER No. 531

Shedding Light on the Shadow Economy: A Nighttime Light Approach

Kiyoyasu TANAKA* and Souknilanh KEOLA

August 2015

Abstract
Measuring the shadow economy is inherently difficult, but critical for understanding economic development. Using census data on formal and informal sectors in Cambodia, we document that 96.6% of non-farm establishments do not formally register with the government, and their sales accounted for 76.6% of total sales in 2011. We estimate a relationship between nighttime light and sales across regions separately for formal and informal firms for 2011, and estimate their past sales from changes in nighttime light for 1993-2010. Both formal and informal firms increased their estimated sales, and the share of informal sales increased from 68.8% in 1993 to 76.6% in 2011, suggesting that the informal sector increased quantitatively in both absolute and relative terms throughout the economic development of the Cambodian economy.

Keywords: Informal Sector, Informal Employment, Cambodia, Satellite Data

JEL classification: O17, E26, H26

* *Corresponding author*: Inter-Disciplinary Studies Center, Institute of Developing Economies; address: 3-2-2 Wakaba, Mihama-ku, Chiba-shi, Chiba, 261-8545, Japan; e-mail: kiyoyasu_tanaka@ide.go.jp

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, the Middle East, Africa, Latin America, Oceania, and Eastern Europe.

INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO
3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI
CHIBA 261-8545, JAPAN

Shedding Light on the Shadow Economy: A Nighttime Light Approach[†]

Kiyoyasu Tanaka[ζ]
Institute of Developing Economies

Souknilanh Keola[§]
Institute of Developing Economies

August 2015

Abstract:
Measuring the shadow economy is inherently difficult, but critical for understanding economic development. Using census data on formal and informal sectors in Cambodia, we document that 96.6% of non-farm establishments do not formally register with the government, and their sales accounted for 76.6% of total sales in 2011. We estimate a relationship between nighttime light and sales across regions separately for formal and informal firms for 2011, and estimate their past sales from changes in nighttime light for 1993-2010. Both formal and informal firms increased their estimated sales, and the share of informal sales increased from 68.8% in 1993 to 76.6% in 2011, suggesting that the informal sector increased quantitatively in both absolute and relative terms throughout the economic development of the Cambodian economy.

Keywords: Informal Sector, Informal Employment, Cambodia, Satellite Data
JEL classification: O17, E26, H26

[†] We acknowledge the financial support of JSPS Grant-in-Aid for Young Scientists (B) Grant Number 24730232 and by the Institute of Developing Economies. We are grateful to Fumihiko Nishi for data assistance, and Yukata Arimoto, Momoe Makino, Yoichi Sugita and participants at the IDE seminar for useful comments and discussions. All remaining errors are our own.
[ζ] *Corresponding author*: Inter-Disciplinary Studies Center, Institute of Developing Economies, JETRO; Address: 3-2-2 Wakaba, Mihama-ku, Chiba-shi, Chiba 261-8545, Japan; E-mail: kiyoyasu_tanaka@ide.go.jp
[§] Inter-Disciplinary Studies Center, Institute of Developing Economies, JETRO; Address: 3-2-2 Wakaba, Mihama-ku, Chiba-shi, Chiba 261-8545, Japan. E-mail: souknilanh_keola@ide.go.jp

1. Introduction

Small businesses and family-run shops are found on the street from morning until night in developing economies. These business establishments produce goods and services, just as global companies in developed economies do. Self-employment and small enterprises create major employment opportunities and contribute substantially to produce value added in the overall economy. However, a large number of these establishments are not likely to register formally with the government and report to the tax authority. Unregistered economic activities, termed the *shadow* economy, are not only missed by official economic surveys, but also excluded from the official calculation of national income and output. Consequently, it remains unsettled whether the shadow economy diminishes during the process of economic development. On the one hand, the share of self-employment in total employment is significantly smaller in developed economies than developing economies, suggesting that the shadow economy should decline as economies develop (La Porta and Shleifer, 2014). On the other hand, informal employment increased in Latin America and the Caribbean countries during the 1990s (Perry et al., 2007). Even for the high-growth Asia-Pacific regions, Bajada and Schneider (2005) estimate that the shadow economy increased on average by 5.1% of official GDP from 1989 to 2001. As the mixed evidence leads to mixed views of informality, Farrell (2004) argues that informal businesses even depress economic growth by preventing more productive formal firms from gaining a market share. Thus, accurate measurement of the shadow economy is critical for understanding economic development.

In this paper, we shed light on the shadow economy in two ways. First, we document the size of informal economic activity in Cambodia using the Economic Census of Cambodia in 2011 (EC2011).[1] The dataset is novel in that it covers all non-farm establishments and enterprises across all industrial sectors in all areas of Cambodia and asks whether individual establishments are registered with the Ministry of Commerce. Unregistered economic activity is a commonly used definition of informality, and we can use business registration as an objective criterion to distinguish between formal and informal firms. The comprehensive coverage enables us to overcome sample-selection bias in measuring a large number of small informal businesses, which causes survey-based studies to suffer seriously from underestimated bias in measurement (Dabla-Norris et al., 2008). Additionally, employment in the EC2011 is classified into self-employed proprietors, unpaid family workers, and regular paid employees. In addition to registration-based classification between formal and informal firms, we can shed further light on the size of informal employment. Taken together, our dataset allows for accurate

[1] The Cambodian economy was devastated under the Pol Pot regime in 1975-79 and the subsequent civil war. The Paris Conference on Cambodia in 1991 led to agreement on a comprehensive political settlement of the conflict and set the stage for economic reconstruction. The economic growth has averaged 6% for the last decade, with per capita GDP of 931.2 U.S. dollars in 2012 (IMF, 2012).

measurement of aggregate informal activities, thereby serving as a credible reference for cross-country comparisons of shadow economies.

Second, we propose a satellite-imagery approach based on nighttime light viewed from outer space to estimate the development of the shadow economy across regions in Cambodia. In using nighttime light data for prediction, there is a systematic change in the scaling of satellite sensor readings across years, which results from changes in sensor settings, such as sensor degradation and satellite renewal (Henderson et al., 2012). We set up a statistical framework in which observed light intensity at the country level is related to the observed GDP and a time-varying scaling factor, the latter of which is estimated from a panel dataset covering 52 countries for 1992-2011. Using the light-scaling factors across years, we carefully correct the systematic change in the measured intensity of nighttime light. Next, we estimate the relationship between nighttime light and sales across regions for 2011 separately for formal and informal firms. Assuming that the estimated relationships at the regional level have been persistent over time, we use past change in corrected nighttime light to obtain the predicated values for formal and informal sales during the period 1993-2010. The predicted values are exploited to construct a chain of time-series indexes of formal and informal sales in each region. By multiplying these indexes with the observed sales of formal and informal firms in 2011, we estimate formal and informal sales across regions for past years.

Measuring the shadow economy is inherently difficult, and prior estimation methods draw on a wide range of economic indicators such as electricity consumption, currency circulation, and national income statistics (Schneider and Enste, 2013, Chap. 3). The prior approaches rely on the official statistics reported by the government in developing economies although developing economies tend to lack both resources and capacity to produce reliable statistics. By contrast, nighttime light on the earth can be observed from outer space, and comparable measures of observed nighttime light are widely available in administrative areas of virtually any size for a long span of periods in both developed and developing economies. These advantages have led to a burgeoning literature that exploits satellite-imagery data for measuring human economic activity, such as economic growth (Ghosh et al., 2010; Chen and Nordhaus, 2011; Henderson et al., 2012; Keola et al., 2015). However, there are limited studies, such as Ghosh et al. (2009) and Harati and Hardy (2013), to exploit the satellite data for measuring the shadow economy. As compared with the prior works, our approach is advantageous in that we correct a systematic component of measurement errors in nighttime light data and estimate an accurate relationship between nighttime light and economic activity for both formal and informal sectors.

The main findings can be summarized as follows. First, we find that informal sector plays a substantially large role in industrial activity of Cambodia. While there were 505,093

establishments in February 2011, 96.6% of them were classified as informal firms. Given total sales of 600 million USD, the informal firms accounted for 76.6%. There are 1.67 million workers, and 66.4% of them belong to informal firms. On the other hand, informal employment as measured by self-employed proprietors and unpaid family workers accounted for 49.7% of employment. These results suggest that a precise share of the informal sector depends crucially on the measurement of informal activity. Second, we find that light intensity and lighted area can account for a substantial variation in both formal and informal sales across regions. The results are robust to an alternative dependent variable and alternative specifications. We find that estimated sales increase for both formal and informal sectors for 1993-2011. The percentage share of informal sales in total sales increases from 68.8% in 1993 up to 76.6% in 2011, suggesting that informal sectors increased over time quantitatively in both absolute and relative terms. Consistent with results based on the multiple-indicators and multiple-causes model in Schneider et al. (2010), we find a growing role for the informal sector in Cambodia.

The rest of this paper is organized as follows. Section 2 discusses related studies on the size of the informal sector and estimation using satellite data. Section 3 explains an empirical strategy for predicting informal activity, followed by a description and correction methodology for using nighttime light data. Section 4 describes data sources on informal firms in Cambodia and documents the size of informal activity. Section 5 shows estimation results and the estimated size of informal activity in Cambodia in past years. Section 6 concludes.

2. Review of Related Studies

In this section, we discuss two branches of related literature: the estimation of informal economic activity and the application of satellite-imagery data for measuring aggregate economic activity. We extend these distinct branches of literature by employing a unique dataset of informal sector in Cambodia and proposing a satellite-imagery approach for measuring informal activity.[2] There exists a large body of literature on the informal sector, and our review of related studies is selective. Comprehensive surveys on a variety of issues in informal sector are provided by Schneider and Enste (2000), Eilat and Zinnes (2002), Maloney (2004), Bruhn and McKenzie (2014), and La Porta and Shleifer (2014).

As is surveyed in Schneider and Enste (2013, Chap. 3), there are direct and indirect approaches to estimate the shadow economy. Using the direct approach, previous studies exploit survey information on informal firms, as determined by the lack of formal business registration and/or tax authority registration.[3] Because the direct approach is based on formal business

[2] Regarding informal activity of formal firms, Slemrod and Weber (2012) discuss prior work on tax evasion, and Zhou and Oostendorp (2014) analyze the underreporting of sales by formal firms in Mongolia.

[3] Dabla-Norris et al. (2008) use firm-level surveys across countries and report that 4,477 sample

registers for sampling, sample-selection bias may exist in the estimated proportion of informal firms.[4] de Paula and Scheinkman (2011) use a survey of small firms in Brazil for 2003 and define formal firms as those firms that register with the Brazilian tax authority. They show that 86.9% of 47,920 sample firms with five or fewer employees do not have tax registration. In the case of Vietnam, Rand and Torm (2012) show that 60% of non-state-owned firms were informal in 2007. In addition, Bosch et al. (2012) use employment surveys in Brazil for 1983-2002 and define informal employment as having informal salaried employees with no registered labor contract and informal self-employed workers.[5] They show that informal employment accounted for 36% of the work force in 1983, which increased up to 45% in 2002. Finally, an indirect approach exploits a wide range of official economic indicators, such as currency circulation and electricity consumption, to estimate the magnitudes and development of shadow economies.[6] A recent approach takes into account multiple causes that lead to multiple indicators of the shadow economy. This method is termed as the multiple-indicators multiple-causes (MIMIC) model (Frey and Weck-Hanneman, 1984). Based on the MIMIC method, Chaudhuri et al. (2006) estimate state-level shadow economies in India, whereas Schneider et al. (2010) estimate country-level shadow economies in 162 countries for the period 1999-2007.

Another branch of related studies is the remote-sensing approach to use satellite-imagery data. Croft (1978) was among the first to point out that nighttime light reflects human economic activities on the ground. Elvidge et al. (1997) estimate a relationship between lighted area and economic variables such as population, GDP, and electricity usage across countries. In contrast, Sutton and Costanza (2002) use the strength of nighttime light to estimate GDP, and land cover information to estimate ecosystem services product. While these studies focus on country-level data, Doll et al. (2006) estimate a relationship between nighttime light and the gross regional product (GRP) for 11 countries in the EU and states in the U.S. Their findings suggest that the elasticity of nighttime light with respect to GDP varies widely within most countries. Ghosh et al. (2009) divide economic activities into commerce/industry and agriculture, and estimate specific elasticities across industries and regions. By reviewing the related studies, Ghosh et al. (2013) illustrate that nighttime satellite imagery has also been applied to measure such indicators as poverty, human development index, and electrification rates, with global coverage. In the economics literature, Chen and Nordhaus (2011) and Henderson et al. (2012) propose a statistical framework to use nighttime light for estimating economic growth. Both arrive at the

---

firms do not report 2.9% of their sales, on average, as a proxy for informal activity.

[4] de Mel et al. (2009) and Joshi et al. (2011) discuss a variety of issues in the survey on economic activities of informal small enterprises, including sales, expenses, and profits.

[5] A contract entitles the employees to labor rights and benefits.

[6] Giles (1999) and Thomas (1999) discuss measurement issues in the indirect methods, such as currency ratio and monetary transactions approaches.

same conclusion that intensity of nighttime light is useful in estimating the 'true' level of economic activities for those economies with less reliable official statistics. Finally, Keola et al. (2015) extend their approach by accounting for agriculture and forestry, which emit little nighttime light.

There are few studies that exploit satellite data for measuring the shadow economy. Ghosh et al. (2009) estimate a relationship between nighttime light and gross state income across the U.S. states and predict gross state income of the Mexican states by using the U.S. regression parameters and nighttime light in Mexico. While the difference between estimated and official gross national incomes in Mexico is attributed to the magnitude of the informal economy and inbound remittances, it is difficult to justify the assumption that U.S.-based regression models can be applied to Mexico. On the other hand, Harati and Hardy (2013) assume that economic activity predicted by nighttime light should capture both formal and informal activity, suggesting that the difference between light-predicted and official GDPs should be attributed to a share of informal activity. Their approach is limited in that it produces a negative fraction of informal activity for some countries. While the satellite-imagery data provide a novel source of useful information for estimating the shadow economy, there are limitations in the estimation methods used in prior studies.

Taken together, we contribute to the related literature in two ways. First, our dataset is unique in that it covers all the formal and informal establishments in Cambodia with information on unpaid family workers and regular paid employees. These features allow us to address sample-selection bias in the prior studies, based on survey data, and document the detailed structure of informal sectors in Cambodia. Second, we propose a nighttime-light approach to estimate the development of informal activity. We correct a systematic component of measurement errors in nighttime light data, and our census data allow us to estimate an accurate relationship between nighttime light and economic activity for both formal and informal sectors across regions. Compared with the official indicators used in the prior indirect approach, satellite data allow us to exploit regional variations in nighttime light for estimating regional patterns of informal activity within a country.

3. Empirical Framework

In this section, we discuss an empirical framework for estimating informal activity in the Cambodian economy from nighttime light data. We explain the empirical specification and describe the satellite-imagery dataset. Because nighttime light data are subject to measurement issues, we carefully discuss a correction method for using the nighttime light in prediction.

3.1. Empirical Specification

Because reliable data on informal activity in past periods are not available, we attempt to estimate past informal activity in Cambodia. First, we estimate the relationship between nighttime light and formal/informal activities across regions for the year 2011. Second, we assume that the estimated relationship at the regional level has been *persistent* over time, suggesting that past change in nighttime light can be used to estimate the development of formal/informal activities for past periods. From the estimated values, we construct a chain of time-series indexes on formal/informal activities in each region for 1993-2011, with the indexes set at unity for year 2011. Finally, we multiply these indexes with formal/informal activities in 2011 to estimate their past trends.

Following prior literature such as Ghosh et al. (2009), Chen and Nordhaus (2011) and Henderson et al. (2012), we assume that nighttime light reflects the volume of economic activity in corresponding regions, as both formal and informal firms can produce light at nighttime. This implies that greater economic activity is associated with higher intensity of nighttime light. Moreover, nighttime light indicates bright and dark areas in each region, and we hypothesize that greater economic activity is associated with a larger share of lighted area in each administrative region. Additionally, we allow for the formal and informal sectors to have different relationships between nighttime light and their economic activities.[7] It should be noted that we follow the prior literature in abstracting from a complex relationship between observable nighttime light and industrial activity. Given the above assumptions, we consider, as a benchmark, a simple constant elasticity model for the relationship between total observable light and total economic activity across regions. Accounting for the geographic area of nighttime light, we estimate the following specification for region j and sector $s \in \{Formal, Informal\}$.

$$\ln(Y_{js}) = \beta_{s0} \ln(L_j) + \beta_{s1} LA_j + \beta_{s2} \ln A_j + \boldsymbol{Z'}_j \boldsymbol{\beta}_{s3} + u_{js} \quad (1)$$

where Y_{js} indicates the aggregate economic activity, which we measure by aggregate sales or expenses. L_j is the intensity of nighttime light. LA_j is the share of lighted area in region j. A_j is the land area. $\boldsymbol{Z}_j$ is a vector of additional control variables, including squared terms of explanatory variables. If these additional variables affect economic activity, the inclusion of these variables should improve the fit of our empirical model to the data. Finally, u_{js} is an error term. Note that we do not include a constant term in equation (1) for the assumption that economic activity should not occur when all observed measures, such as nighttime light and land area, are zero.

[7] Formal and informal firms are different in terms of economic characteristics such as productivity, profitability, and size. After controlling for self-selection effects, formal registration could affect firm performance (McKenzie and Seynabou Sakho, 2010; Fajnzylber et al., 2011).

The next task is to obtain predicted values of economic activity in the past periods from the regression model, which should well capture the relationship between nighttime light and economic activity. Given the best-fit estimates, we use the data on nighttime light to estimate the predicted volume of economic activity for year *t* as follows.

$$\ln\widehat{(Y_{jst})} = \hat{\beta}_{s0}\ln(L_{jt}) + \hat{\beta}_{s1}LA_{jt} + \hat{\beta}_{s2}\ln A_j + \boldsymbol{Z}'_{jt}\widehat{\boldsymbol{\beta}}_{s3} \quad (2)$$

Using these predicted values, we can construct a chain of indexes on changes in economic activity of informal and formal sectors over time for each region. To take advantage of the credible information on economic activity for the year 2011, we define the predicted value of the dependent variable in 2011 as the base value, implying that the index is defined to take on unity across regions for 2011. We then calculate the following index from the predicted values.

$$Index_{jst} \equiv \ln\widehat{(Y_{jst})}/\ln\widehat{(Y_{js,2011})}.$$

Finally, we multiply the actual data on formal and informal activities in 2011 with the chain of these indexes to estimate economic activities in each region for past years.

3.2. Nighttime Light Data

We now describe data on nighttime light. The U.S. Air Force has operated the Defense Meteorological Satellite Program (DMSP). The program is based on a series of orbiting satellites with the primary objective of monitoring weather. When it is night on the surface, the electromagnetic energy sensed by the system is mostly a product of man-made light-emitting activities. The DMSP operational line scan sensors operate at an altitude of 830 km, which can generate the global daytime and nighttime coverage of the Earth. Images are processed at the National Geophysical Data Center (NGDC) of the National Oceanic and Atmospheric Administration (NOAA). Processing removes pixels for observations on the locations affected by the moon-lit half of the lunar cycle, late sunset in the summer (e.g., the Scandinavian countries), auroral activity, and clouds. The remaining observations from each operating satellite sensor are averaged to produce a satellite dataset for each year. Images from the system have been recorded since 1973, but were periodically overwritten to free storage capacity. Data are digitally archived and recently made available for the period from 1992. The processed data are available to the public through the NOAA and NGDC Internet portals. The intensity of light is coded in a grid format as six-bit digital numbers (DIs), which gives integers between 0 (no light) and 63. Using this dataset, we construct the intensity and geographic share of nighttime light in each region. Specifically, we aggregate the DIs of light intensity over regions to measure the strength of nighttime light. We approximate the presence of nighttime light with light intensity of more than one, $\mathrm{DI} \geq 1$, to calculate the proportion of lighted area in each region.

In using data on nighttime light, we must be concerned about possible changes in the scaling of satellite sensor readings across years, as is emphasized in Henderson et al. (2012). As

the sensor settings change with the age of a satellite and across satellites, raw DIs may not be simply compared across years. Even if there were little growth of economic activity, an improvement in sensor readings of a new satellite may overestimate the intensity of nighttime light. On the other hand, sensor degradation of a satellite would underestimate the strength of nighttime light. The former should lead to an upward bias in predicted economic activity whereas the latter suggests a downward bias. Because we aim to predict the past level of economic activities from changes in nighttime light, we have to carefully correct for systematic changes in the intensity of nighttime light.

In this paper, we assume the following statistical framework to estimate an adjustment factor for the measured intensity of nighttime light over time. First, measured light intensity is related to true light with the scaling factor of the satellite sensor and a measurement error for country *i* and year *t*:

$$Light_{it}^{Data} = Light_{it}^{True} exp(\lambda_t + e_{it}), \tag{3}$$

where λ_t is the scaling factor in year *t*, and e_{it} is the measurement error of observed light in country *i* and year *t*. We need to estimate the magnitude of the time-varying scaling factor to compute true light from observed light. Because true light is unknown, we need an alternative measure. We assume that true light is related to true GDP through a conversion factor μ: $Light_{it}^{True} = \left(GDP_{it}^{True}\right)^{\mu}$. We also assume that observed GDP is a measure of true GDP with measurement error ε: $GDP_{it}^{True} = GDP_{it}^{Data} exp(\varepsilon_{it})$. Based on these relationships, we can write true light as a function of observed GDP:

$$Light_{it}^{True} = \left[GDP_{it}^{Data} exp(\varepsilon_{it})\right]^{\mu}. \tag{4}$$

Using Equations (3) and (4) and taking logs, we can express observed light as follows:

$$\ln Light_{it}^{Data} = \mu \ln GDP_{it}^{Data} + \lambda_t + \omega_{it}, \tag{5}$$

where $\omega_{it} = \mu\varepsilon_{it} + e_{it}$. Equation (5) shows that log observed light depends on log observed GDP, the time-varying scaling factor, and the composite of measurement errors.

Our next step is to estimate the scaling factor λ_t for the regression model. Specifically, we approximate the scaling factor by a set of year-dummy variables with the corresponding coefficient of the year-dummy π_t. Ordinary-least squares (OLS) estimation of Equation (5) gives us a set of coefficients of aggregate year effects, except for the benchmark year. Denoting these year effects as $\hat{\pi}_t^{OLS}$, we approximate $exp(\lambda_t)$ by $exp(\hat{\pi}_t^{OLS})$. We then normalize the variable $exp(\hat{\pi}_t^{OLS})$ according to the base value of $exp\left(\hat{\pi}_{2000}^{OLS}\right)$, implying that true and observed light data are set as equal in the normalization year, 2000. For estimation, we construct the country-level panel data on nighttime light intensity and real GDP during the period 1992-2011. Data on real GDP are taken from the World Development Indicators (WDI) published by the World Bank. While our objective is to obtain a set of coefficients for year-dummy variables, there is concern about the reliability of real GDP data, which tend to

contain larger measurement errors in developing economies. To exclude the sample countries with less reliable statistics, we exploit the statistical capacity indicator from the WDI, which is a composite score to assess the capacity of a country's statistical system, including methodology, data sources, and periodicity and timeliness. The indicator lies between 0 and 100, with larger values indicating better capacity. Based on the indicator, we exclude the remaining sample countries with a score less than 80. Note that we keep the OECD members in the sample.

Table 1 presents the summary statistics of the country-level sample during the periods 1992-2011 with 52 countries. Table 2 presents the estimation results of Equation (5) by an OLS estimator. Column (1) presents the estimates for the year-dummy variables and log real GDP. As expected, the coefficient of GDP is significantly positive at the 1% level, consistent with the prior literature. As R-squared is 0.997, these explanatory variables explain almost all variation in log light intensity. We compute $1/exp(\hat{\pi}_t^{OLS})$ for year-dummy variables and normalize the computed values by the year 2011 values. Light factor in column (1) shows the estimated discount factors of the light intensity across years. In column (2), we exclude sample countries in which territorial land is situated in the northern Arctic zone, doing so because the observable nighttime light may be strongly influenced by the data cleaning process to address auroral activity and midnight sun in summer months.[8] We calculate light factors from the estimation result and compare these with the light factors in column (1). The correlation coefficient is 0.92 and is significant at the 1% level, indicating that the estimated light factors are robust to alternative sampling of countries.[9]

---Tables 1 and 2---

In sum, we find that the scaling factors of satellite sensors change across years. We use the light factor in column (1) of Table 2 to correct for the systematic change in the intensity of nighttime light in Cambodia.[10] Another issue is that there are unsystematic changes in nighttime light across regions for various reasons such as weather conditions, natural disasters, and temporary construction sites. Regardless of actual changes in industrial activity, local conditions may lead to a sharp fluctuation in local nighttime light across years. To mitigate such unsystematic changes in nighttime light, we take the 3-year averages of the light intensity and lighted area in each region, using 1-year lag and forward variables.

[8] Specifically, we exclude Canada, Estonia, Finland, Iceland, Norway, and Sweden.

[9] We also estimate with country fixed effects because cross-country differences in the calculation of measured GDP may affect the assumption that $GDP_{it}^{True} = GDP_{it}^{Data} exp(\varepsilon_{it})$. Computing light factors with this specification, we find a positive correlation coefficient of 0.97, significant at the 1%level, with the light factors in column (1).

[10] Light deflators in 1992 are approximated by the data in 1993.

4. Data on the Informal Sector

4.1. Data Source

Our main dataset is *the Economic Census* of Cambodia in 2011 (EC2011). The census was conducted in March 2011 to survey economic activities of all the non-farm establishments and enterprises over the entire territory of Cambodia. EC2011 was mainly funded by the Japanese Official Development Assistance and implemented by the National Institute of Statistics, the Cambodian Ministry of Planning, in cooperation with the Japan International Cooperation Agency. The survey aims to collect information on firm activities, including financial statements, persons engaged in the business, main line of business, and so on. The administrative geographic units consist of 193 districts in 24 provinces, including the municipality of Phnom Penh.

The questionnaire of EC2011 includes a question about registration with administrative agencies and the names of ministries regarding licenses and approval for operating. Specifically, establishments/enterprises must answer whether they have registered with the Ministry of Commerce or Provincial Department of Commerce.[11] We exploit this question to define the formal sector as the business activities of registered firms and the informal sector as those of unregistered firms. This definition implies that formal firms have completed several procedures for registration: (1) depositing the legally required initial capital in a bank and obtaining deposit evidence, (2) conducting an initial check for uniqueness of the company name at the Intellectual Property Department and the Business Registration Office, and (3) publishing an abstract of the company organization documents and incorporating the company with the Business Registration Department in the Ministry of Commerce (World Bank, 2014). These procedures are estimated to cost at least 400 USD and take 1 month.

Although the terms and definitions of informal activity are not harmonized in the literature, Schneider and Enste (2013, Chp. 2) defines the national economy as the dual economy of an official sector and an underground sector. Both sectors contribute to gross national production, but the latter is not included in calculations. Furthermore, the underground sector consists of the shadow economy, based on market transactions, and the household sector, based on non-market transactions to satisfy individual needs. In this paper, the shadow economy in their definition closely corresponds to our data on informal activity in Cambodia as defined by unregistered economic activity in manufacturing and services industries. However, it should be mentioned that the EC2011 data do not cover establishments in agriculture, forestry and fishing; the agricultural sector accounted for 36.7% of the GDP at current prices in 2011 (Asian

[11] In Cambodia, the "Law Bearing upon Commercial Regulations and the Commercial Register" was established in 1995 and stipulates the obligation of companies to register and the formal procedures for commercial registration.

Development Bank, 2014). Although the survey covers street vendors that operate at a fixed location, it does not include mobile establishments, such as bike taxis and mobile street vendors. Finally, our data do not cover illegal economic activities, including tax evasion, abuse of public finance, smuggling, fraud, and so on. Thus, the coverage of informal activity in this paper is a part of the entire informal activity in the Cambodian economy.

4.2. The Size of Informal Activity

Table 3 presents aggregate figures of formal and informal activity in Cambodia, with financial figures measured for 1 month (February 2011). The total number of establishments was 505,093, which consists of 17,374 formal firms and 487,719 informal firms. In terms of the number, 96.6% of the establishments can be classified as informal. This figure about informality is much larger than those from the survey samples reported in de Paula and Scheinkman (2011) and Rand and Torm (2012) using the tax-registration definition of informality. Despite slightly different definitions of informality, our data suggest that the sample surveys are likely to miss a large number of informal firms.

---Table 3---

Turning to financial status, we find that 76.6% of total sales of 600 million USD were accounted for by informal firms. The percentage share is similar when we measure with expenses including purchases of products, costs for providing services, rents and employees' wages. In terms of financial aspects, informal activity remains prominent in the manufacturing and services industries. However, the prominence of informal activity appears to be weaker than suggested by the number of informal establishments. These results should reflect the fact that formal firms tend to exhibit greater economic activity than informal firms do. Additionally, we see that 59.2% of total wages of 35.4 million USD were paid by informal firms. Although total wage payments point to the dominant role of informal activity, the informal sector appears to be less prominent. This pattern is consistent with the causal observation that formal sectors provide employment opportunities through formally paid contracts.

Finally, we shed light on informal employment. When formal registration is used to define informality, 66.4% of 1.67 million workers are associated with informal firms, whereas 33.6% of them are associated with formal firms. These figures indicate that a majority of industrial workers belong to the informal sector. On the other hand, informal employment can be defined to include self-employed proprietors and unpaid family workers, regardless of formal registration of their establishments. In this case, 49.7% of employment is classified as informal. Comparing with the findings, based on a similar definition, in Bosch et al. (2012), we find that

our percentage share of informal employment is similar to their survey results in Brazil.[12] Nevertheless, the figures for informal employment are sensitive to its definition. If we include regular paid employees of informal firms as part of informal employment, the share of informal employment increases to 67.6%.[13] By way of comparison, Cling et al. (2011) use the Vietnamese labor force survey for 2007, which was conducted at a nationwide scale, to survey individual workers and their establishments. By defining unregistered private enterprises as informal, they find that the informal sector explains 47.1% of 21.9 million jobs in non-farm sectors. These findings suggest that the proportion of informal employment is larger in Cambodia than in Vietnam.

Taken together, our data suggest generally that the informal sector plays a substantially large role in the industrial activity of Cambodia, consistent with prior evidence. However, the precise share of informal activity depends crucially on the measurement and definition of informal activity. In the case of Brazil, Henley et al. (2009) show that different definitions of informal employment (e.g., employment status) change the interpretation of whether the proportion of the informal sector declines or increases for 1992-2004. Together with our findings, we conclude that employment status is a sensitive definition to measure informality. Therefore, the following analysis will use only government registration to distinguish between formal and informal firms.

5. Results

5.1. Estimation Results

Table 4 shows the summary statistics of the sample used for estimation. For the dependent variable, we use the log of sales and expenses for formal and informal firms. Because the sales and expenses are reported for 1 month in the EC2011, we multiply these financial figures by 12 to match annual satellite data. *Light* is defined as the log of total observable light intensity, whereas *Light share* is the share of lighted area relative to total land area. *Land area* is the log of land area in square kilometers.

---Table 4---

Table 5 presents the estimation results of Equation (2). To account for a possible nonlinear relationship between economic activities and explanatory variables across regions, we include

[12] Perry et al. (2007) report a comprehensive analysis of informal employment in Latin America and the Caribbean countries.

[13] See Williams and Lansky (2013) for more discussion on the definition and measurement of informal employment. Schneider and Enste (2013, Chapter 5) provide results on informal employment in developed and developing economies.

squared terms of *Light*, *Light share*, and *Land area* in the specification. Column (1) shows the result for informal sales. We find that light intensity has a significantly positive association with informal sales, and the significantly negative correlation with the square of *Light* shows a diminishing effect. Lighted area is significantly and positively associated with informal sales. Additionally, informal sales significantly increase with the land area of districts, with diminishing effect. This specification shows an R-squared value of 0.9945, implying that these explanatory variables explain almost all variations in informal sales across districts. On the other hand, we find that formal sales are not significantly correlated with light intensity, but increase significantly with lighted area. Formal sales also increase significantly with land area, which also exhibits diminishing effects. The R-squared value is 0.9773, suggesting a fairly good fit of the model with data on formal sales. However, the root-mean-squared errors across specifications suggest that prediction errors are larger for formal sales than for informal sales.

---Table 5---

We conduct several robustness checks. First, a possible concern is that nighttime lights may be more strongly reflected in input measure than in output measure because light can serve as an input of economic activity. We estimate the specification with total expenses as a dependent variable. These results are presented in columns (2) and (4) of Table 5 for informal and formal expenses, respectively. We find that the results are qualitatively and quantitatively similar to the results based on sales. While informal expenses increase with light intensity and lighted area, formal expenses increase with lighted area only. Both formal and informal expenses significantly increase with land area. Second, we include the cubes of *Light*, *Light share*, and *Land area* in the specification. In unreported results, we find that these higher order terms are not significant across alternative specifications, and the inclusion of these terms makes little contribution to improving the fit of the model. Finally, we examine the issue of saturated pixels in light intensity (i.e., $\mathrm{DI} = 63$). On examination, we find no saturated pixels in the year 2011 data. In past data, there are some saturated pixels, but the lighted area with these saturated pixels is a very tiny share of total land area in Cambodia. These results lead us to conclude that saturated pixels should have little influence on the main results.

5.2. Estimated Trends in Formal and Informal Sales

Using the results shown in columns (1) and (3) of Table 5, we estimate past sales in the formal and informal sectors at the district level in Cambodia. Before proceeding to discuss the analysis, it must be emphasized that the following results should be carefully interpreted for potential measurement errors. In particular, our estimates are derived from *observable* nighttime

lights in each district, implying that we do not estimate sales in districts without nighttime light data. Such districts are not included in the estimated sales.

Table 6 presents the estimated sales at the country level for the period 1993-2010 and the EC2011 data for 2011. Consistent with the observed trend in economic growth of the Cambodian economy for the past decades, we find that estimated total sales increased from 4.41 billion USD in 1993 up to 7.20 billion USD in 2011. This increase is accounted for by the formal and informal sectors together. Estimated formal sales increase from 1.37 billion USD in 1993 up to 1.68 billion USD in 2011, whereas estimated informal sales increase from 3.04 billion USD in 1993 up to 5.51 billion USD in 2011. Thus, our estimates suggest that both formal and informal sectors contribute to economic growth in Cambodia over time. Additionally, we examine a trend in the relative importance of informal activity. Computing the percentage share of informal sales in total sales, we find that it is 68.8% in 1993, 72.3% in 2000, and 75.4% in 2010. Taken together, our estimates suggest that informal sectors have increased over time quantitatively in both absolute and relative terms. Schneider et al. (2010) use the MIMIC model and estimate that the shadow economy in Cambodia accounted for 49.8% of the GDP in 1999 and 54.2% in 2007. Consistent with their results, the Cambodian economy appears to have experienced a growing informal sector over the course of its economic development.

---Table 6---

To shed light on the distribution of informal sector shares across districts, Figure 1 presents percentiles of the informal sector shares across years. The median value increases from 91.9% in 1993 up to 96.7% in 2011, whereas the mean value also increases from 85.7% in 1993 up to 91.6% in 2011. The 25th and 75th percentiles appear to increase steadily between 1993 and 2011. Thus, the relative size of informal sales is significantly large and increases over time. Compared with the country-level measures, the district-level shares are relatively high on average. A possible reason is that a majority of informal firms are relatively small in terms of sales and their presence is dominant in a large number of districts. A minority of formal firms, such as foreign-owned companies, tends to concentrate in certain regions, and they have a substantial volume of sales. As a result, country-level aggregation tends to mask wide regional variations in informality. Additionally, it should be emphasized that the relative size varies widely by districts. For instance, the minimum and maximum of informal shares in 2011 are 24.2% and 100%, respectively.

We conclude this section by discussing possible explanations for why informal activity increased significantly in Cambodia. First, the demographic structure is a fundamental reason. La Porta and Shleifer (2014) highlight the dual view of informality, in which formal and

informal sectors produce different products for different customers using different inputs in labor, capital and entrepreneurial skill. While efficient formal firms replace inefficient informal firms to reduce informal economies in the process of economic development, this replacement process may be discouraged by high population growth and a shortage of educated entrepreneurs. These explanations are consistent with the demographic trend in Cambodia. The WDI data show that the total population increased from 9.05 million in 1990 up to 14.6 million in 2011, and the average annual growth rate of population is over 2.3% during the period 1990-2011. Additionally, 58.7% of literate persons had not finished primary education in 1998, which declined only slightly to 50.4% in 2008 according to the Population Census in 1998 and 2008 by the Cambodian National Institute of Statistics. Although educational attainment of the labor force improved, only 4.0% of literate persons aged 25 years and older had completed secondary school in 2008. Thus, the high population growth may increase demand for products and supply of workers in the informal sector, whereas the shortage of entrepreneurial skill may not lead to enough growth of formal-sector businesses to absorb the growing number of new workers.

Second, low institutional quality is also a fundamental reason. While formal firms incur larger costs to pay taxes and comply with regulations than informal firms, weak institutions in Cambodia may magnify the formalization costs for informal firms by making them more vulnerable to corruption and unequal treatment by government bureaucrats (Hill and Menon, 2013). Together with business registration costs, these formalization costs could significantly discourage a large number of small businesses from becoming formal. According to the Worldwide Governance Indicators by the World Bank, institutional quality in Cambodia has not improved over past decades. The control of corruption indicator captures the extent to which public power is exercised for private gain, which ranges from -2.5 (weak) to 2.5 (strong). This indicator deteriorated from -0.96 in 1996 to -1.22 in 2011. The regulatory quality indicator represents the ability of the government to provide sound policies and regulations for private sector development. This indicator also declined from -0.05 in 1996 to -0.57 in 2011. As the political stability indicator improved from -1.15 in 1996 to -0.33 in 2011, the recent political stability in Cambodia should contribute to rapid economic growth. However, there is little improvement in institutions that support the formal private sector, and high formalization costs may not induce growing informal firms to become formal.

6. Conclusion

Informal activity is prominent in developing economies, and accurate measurement of informal activity is critical for understanding economic development. In this paper, we exploit a novel dataset on the formal and informal sectors in Cambodia for 2011 to document the detailed

structure of the informal sector. Unregistered informal firms play a large role in industrial activity in terms of the establishment number and sales. Informal employment as measured by self-employed proprietors and unpaid family workers accounts for a substantial share of workers. To examine development in the informal sector, we exploit satellite-imagery data to estimate formal and informal activity across regions and over time. Our results show that the informal sector plays a substantial role in Cambodian industry in terms of various measures, such as sales, expenses, and the number of establishments. Our estimates suggest that both formal and informal activities have increased significantly over time. However, the growth of informal activity is larger than that of formal activity, thereby leading to a growing share of informal activity over time in Cambodia. As is predicted by the dual view of informality in La Porta and Shleifer (2014), high population growth and low educational attainment in Cambodia may discourage the process of inefficient informal firms being replaced by efficient informal firms. Weak institutions in Cambodia may also magnify the formalization costs and discourage growing informal firms from becoming formal. Thus, shadow economies may not necessarily shrink during early stages of economic growth.

We conclude to discuss some remaining issues in this paper. Our approach relies crucially on the quality of nighttime lights to estimate the shadow economy. As is emphasized by Nordhaus and Chen (2014), the intensity of nighttime lights may be set at zero in the lights dataset for data filtering when the level of light intensity is too low to be distinguished from the background lights. Some regions with positive population and output may have zero lights. Because our data show zero lights for some regions with positive sales for 2011, there is a need to improve the lights data with very low resolution in the least developing economies such as Cambodia. Additionally, our approach assumes that the estimated relationship between lights and sales at the regional level should have been persistent over time for prediction. However, the estimated relationship could change in the past periods, and thus a deviation from the current estimated relationship is likely to generate measurement errors in predicted past sales. The lack of reliable data on past sales makes it difficult to estimate such a deviation, it would be useful to examine this assumption with future datasets.

References

Asian Development Bank, 2014. Key Indicators for Asia and the Pacific 2014. Manila.

Bajada, C., Schneider, F., 2005. The shadow economies of the Asia-Pacific. Pacific Economic Review 10 (3), 379-401.

Bosch, M., Goni, E., Maloney, W., 2012. Trade liberalization, labor reforms, and formal-informal employment dynamics. Labour Economics 19 (5), 653-667.

Bruhn, M., McKenzie, D., 2014. Entry regulation and the formalization of microenterprises in developing countries. The World Bank Research Observer 30 (1), 186-201

Chaudhuri, K., Schneider, F., Chattopadhyay, S., 2006. The size and development of the shadow economy: an empirical investigation from states of India. Journal of Development Economics 80 (2), 428-443.

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

Cling, J. P., Razafindrakoto, M., Roubaud, F., 2011. The Informal Economy in Vietnam. International Labour Organization, Geneva.

Croft, T. A. 1978. Night-time images of the earth from space. Scientific American 239 (1), 86–98.

Dabla-Norris, E., Gradstein, M., Inchauste, G., 2008. What causes firms to hide output?: the determinants of informality. Journal of Development Economics 85 (1-2), 1-27.

de Mel, S., McKenzie, D. J., Woodruff, C., 2009. Measuring microenterprises profits: must we ask how the sausage is made? Journal of Development Economics 88 (1), 19-31.

de Paula, Á., Scheinkman, J. A., 2011. The informal sector: an equilibrium model and some empirical evidence from Brazil. Review of Income and Wealth 57, S8-S26.

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

Eilat, Y., Zinnes, C., 2002. The shadow economy in transition countries: friend or foe? A policy perspective. World Development 30 (7), 1233-1254.

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

Fajnzylber, P., Maloney, W. F., Montes-Rojas, G. V., 2011. Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program. Journal of Development Economics 94 (2), 262-276.

Farrell, D., 2004. The hidden dangers of the informal economy. The McKinsey Quarterly, no. 3, 26-37.

Frey B. S., Weck-Hanneman, H., 1984. The hidden economy as an 'unobserved variable'. European Economic Review, 26 (1-2), 33-53.

Ghosh, T., Anderson, S., Powell, R. L., Sutton, P. C., Elvidge C. D., 2009. Estimation of Mexico's informal economy and remittances using nighttime imagery. Remote Sensing 1 (3), 418-444.

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

Ghosh, T., Anderson, S. J., Elvidge C. D., Sutton, P. C., 2013. Using nighttime satellite imagery as a proxy measure of human well-being. Sustainability 5 (12), 4988-5019.

Giles, D. E., 1999. Measuring the hidden economy: implications for econometric modelling. The Economic Journal 109 (456), 370-380.

Harati, R., Hardy, M., 2013. Illuminating the shadow economy: using lights from space to estimate a new measure of informal economic activity. Centre d'Economie de la Sorbonne: mimeo.

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

Henley, A., Arabsheibani, G. R., Carneiro, F. G., 2009. On defining and measuring the informal sector: evidence from Brazil. World Development 37 (5), 992-1003.

Hill, H., Menon, J., 2013. Cambodia: rapid growth with weak institutions. Asian Economic Policy Review 8 (1), 46-65.

IMF, 2012. World Economic Outlook 2012. Washington, D.C.

Joshi, K., Amoranto, G., Hasan, R., 2011. Informal sector enterprises: some measurement issues. Review of Income and Wealth 57 (S), S143-S165.

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

La Porta, R, Shleifer, A., 2014. Informality and development. Journal of Economic Perspectives 28 (3), 109-26.

Maloney, W. F., 2004. Informality revisited. World Development 32 (7), 1159-1178.

McKenzie, D., Seynabou Sakho, Y., 2010. Does it pay firms to register for taxes? The impact of formality on firm profitability. Journal of Development Economics 91 (1), 15-24.

Nordhaus, W., Chen, X., 2014. A sharper image? Estimates of the precision of nighttime lights as a proxy for economic statistics. Journal of Economic Geography, forthcoming.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A., Saavedra-Chanduvi, J., 2007. Informality: exit and exclusion. World Bank Latin American and Caribbean Studies, The World Bank, Washington, D.C.

Rand, J., Torm, N., 2012. The benefits of formalization: evidence from Vietnamese manufacturing SMEs. World Development 40 (5), 983-998.

Schneider, F., Buehn A., Montenegro, C. E., 2010. Shadow economies all over the world: new estimates for 162 countries from 1999 to 2007. Policy Research Working Paper 5356, World Bank.

Schneider, F., Enste, D. H., 2000. Shadow economies: size, causes, and consequences. Journal of Economic literature 38 (1), 77-114.

Schneider, F., Enste, D. H., 2013. *The Shadow Economy: An International Survey*, Cambridge University Press, Cambridge.

Slemrod, J., Weber, C., 2012. Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. International Tax and Public Finance 19 (1), 25-53.

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

Thomas, J., 1999. Quantifying the black economy: ‘measurement without theory’ yet again? The Economic Journal 109 (456), F381-F389.

Williams, C. C., Lansky, M. A., 2013. Informal employment in developed and developing economies: perspectives and policy responses. International Labour Review 152 (3-4), 355-80.

World Bank, 2014. Doing business 2014: Understanding Regulations for Small and Medium-size Enterprises, Washington D.C.

Zhou, F., Oostendorp, R., 2014. Measuring true sales and underreporting with matched firm-level survey and tax office data. The Review of Economics and Statistics, 96 (3), 563-76.

Table 1. Summary Statistics of the Country-level Sample

Variable	No. of Obs.	Mean	Std. Dev.	Min	Max
Log of light intensity	1,036	13.98	1.45	8.69	18.24
Log of real GDP in 2005 USD	1,036	25.73	1.83	21.16	30.26

Note: The sample economies include Argentina, Armenia, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Colombia, the Czech Republic, Denmark, Ecuador, Estonia, Finland, France, Georgia, Germany, Greece, Guatemala, Hungary, Iceland, Indonesia, Ireland, Israel, Italy, Japan, Kazakhstan, the Kyrgyz Republic, Lithuania, Luxembourg, Malaysia, Mexico, Moldova, Netherlands, New Zealand, Norway, Peru, the Philippines, Poland, Portugal, the Slovak Republic, Slovenia, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, the United Kingdom, the U.S., and Venezuela.

Table 2. Results for Nighttime Light Scaling Factors

Dependent variable: log light intensity

	(1)			(2)		
Variable	Coef.	Std. Err.	Light factor	Coef.	Std. Err.	Light factor
Year 1993	-0.28*	(0.14)	1.202	-0.26+	(0.15)	1.235
Year 1994	-0.33*	(0.15)	1.264	-0.32+	(0.16)	1.311
Year 1995	-0.14	(0.15)	1.045	-0.11	(0.16)	1.063
Year 1996	-0.16	(0.14)	1.066	-0.13	(0.16)	1.084
Year 1997	-0.15	(0.15)	1.055	-0.11	(0.16)	1.063
Year 1998	-0.041	(0.15)	0.946	-0.0037	(0.16)	0.956
Year 1999	-0.16	(0.16)	1.066	-0.085	(0.16)	1.037
Year 2000	-0.31*	(0.15)	1.239	-0.24	(0.16)	1.210
Year 2001	-0.28+	(0.15)	1.202	-0.23	(0.16)	1.198
Year 2002	-0.47*	(0.20)	1.454	-0.20	(0.16)	1.163
Year 2003	-0.30*	(0.15)	1.226	-0.24	(0.16)	1.210
Year 2004	-0.43**	(0.15)	1.397	-0.38*	(0.16)	1.392
Year 2005	-0.37*	(0.15)	1.315	-0.32+	(0.16)	1.311
Year 2006	-0.41**	(0.15)	1.369	-0.36*	(0.16)	1.365
Year 2007	-0.46**	(0.14)	1.439	-0.42**	(0.16)	1.449
Year 2008	-0.26+	(0.14)	1.178	-0.21	(0.16)	1.175
Year 2009	-0.31*	(0.15)	1.239	-0.25	(0.16)	1.223
Year 2010	0.16	(0.15)	0.774	0.20	(0.16)	0.780
Year 2011	-0.096	(0.15)	1.000	-0.049	(0.16)	1.000
Log real GDP	0.55**	(0.0067)		0.55**	(0.0071)	
No. of obs.	1,036			920		
R-squared	0.9975			0.9979		

Notes: We report robust standard errors clustered by country; light factor is calculated by computing $1/\exp(\pi)$ for the estimated coefficients of year-dummy variables and normalizing the values by the year 2011 value; column (2) excludes sample economies whose territorial land is situated in the northern Arctic zone.

** Significant at the 1% level

* Significant at the 5% level

+ Significant at the 10% level

Table 3. Aggregate Figures of Formal and Informal Sectors in Feb. 2011.

	Formal	Informal	Total
Number of establishment	17,374	487,719	505,093
	(3.4)	(96.6)	
Sales (mil. USD)	140.31	459.75	600.07
	(23.4)	(76.6)	
Expenses (mil. USD)	112.67	364.13	476.80
	(23.6)	(76.4)	
Wages (mil. USD)	14.46	20.97	35.42
	(40.8)	(59.2)	
Employment (mil. people)	0.561	1.112	1.673
	(33.6)	(66.4)	
Self-employed proprietors	0.010	0.476	0.486
	(2.2)	(97.8)	
Unpaid family workers	0.009	0.336	0.345
	(2.6)	(97.4)	
Regular employees	0.542	0.300	0.842
	(63.9)	(36.1)	

Notes: Formal indicates registered firms, whereas Informal indicates unregistered firms; figures in parentheses indicate percentage share of each sector in the corresponding variable; expenses include purchases of products, costs for providing services, rents and employees' wages.

Table 4. Summary Statistics

Variable	Description	No. of Obs.	Mean	Std. Dev.	Min	Max
Informal sales	Log of sales by informal firms	162	16.58	1.21	13.05	19.87
Informal expenses	Log of expenses by informal firms	162	16.33	1.24	12.58	19.60
Formal sales	Log of sales by formal firms	148	13.57	2.55	7.27	19.94
Formal Expenses	Log of expenses by formal firms	148	13.39	2.55	7.27	19.61
Light	Log of lights	1,728	4.16	1.81	0.18	8.73
Light share	A share of lightened area	1,728	0.13	0.24	0.0002	1.00
Land area	Log of land area in km^2	1,728	5.93	1.40	0.80	8.60

Note: Sales and expenses are measured in USD.

Table 5. Estimation Results

	(1)	(2)	(3)	(4)
	Informal Firms		Formal Firms	
Variable	Sales	Expenses	Sales	Expenses
Light	1.05+	1.03+	0.75	0.75
	(0.41)	(0.40)	(0.48)	(0.48)
Light squared	-0.13*	-0.13*	-0.085	-0.082
	(0.048)	(0.048)	(0.055)	(0.055)
Light share	7.63*	7.52*	9.66+	9.20+
	(2.34)	(2.34)	(3.89)	(3.89)
Light share squared	0.56	0.55	-1.30	-1.10
	(2.53)	(2.52)	(3.52)	(3.52)
Land area	4.31**	4.27**	3.82**	3.76**
	(0.34)	(0.33)	(0.43)	(0.43)
Land area squared	-0.32**	-0.32**	-0.32**	-0.32**
	(0.035)	(0.035)	(0.053)	(0.052)
No. of obs.	162	162	148	148
R-squared	0.9945	0.9942	0.9773	0.9763
Root Mean Squared Error	1.26	1.27	2.12	2.14

Note: Figures in parenthesis indicate robust standard errors.

** Significant at the 1% level

* Significant at the 5% level

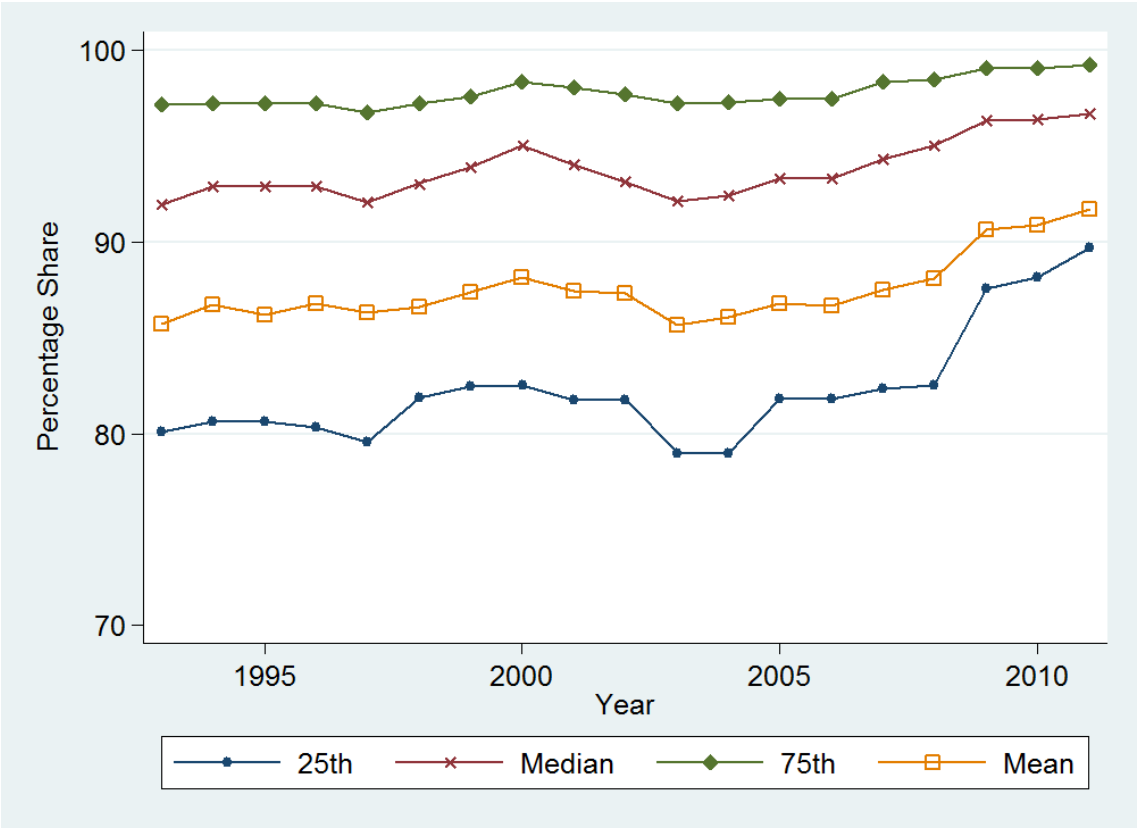
+ Significant at the 10% level

Table 6. Estimated Sales of Formal and Informal Firms

Year	Total	Formal Sector	Informal Sector	Informal Share (%)
1993	4,416	1,377	3,040	68.82
1994	4,540	1,387	3,153	69.46
1995	4,335	1,350	2,985	68.87
1996	4,767	1,425	3,342	70.11
1997	4,722	1,424	3,298	69.84
1998	5,094	1,487	3,607	70.82
1999	5,394	1,516	3,878	71.90
2000	5,549	1,533	4,017	72.38
2001	5,512	1,538	3,974	72.10
2002	5,445	1,531	3,914	71.89
2003	5,331	1,546	3,784	70.99
2004	5,390	1,553	3,837	71.18
2005	5,334	1,542	3,793	71.10
2006	5,378	1,547	3,831	71.23
2007	5,399	1,544	3,854	71.40
2008	5,315	1,513	3,802	71.53
2009	6,422	1,606	4,816	75.00
2010	6,646	1,634	5,012	75.42
2011	7,202	1,684	5,518	76.62

Note: Sales are measured in millions of USD; the number of observations indicates the number of districts with predicted sales for 1993-2010, while the 2011 data come from the Economic Census in 2011.

Figure 1. Trends in the Share of the Informal Sector's Sales at the District Level.



Note: 25th and 75th indicate the 25th and 75th percentiles of the share of the informal sector's sales at the district-level in each year.