Policy Research Working Paper 9088

Can Medium-Resolution Satellite Imagery Measure Economic Activity at Small Geographies?

Evidence from Landsat in Vietnam

Ran Goldblatt
Kilian Heilmann
Yonatan Vaizman

WORLD BANK GROUP
Development Economics
Knowledge and Strategy Team
December 2019

Policy Research Working Paper 9088

Abstract

This study explores the potential and the limits of medium-resolution satellite data as a proxy for economic activity at small geographic units. Using a commune-level dataset from Vietnam, it compares the performance of commonly used nightlight data and higher resolution Landsat imagery which measures daytime light reflection. The analysis suggests that Landsat outperforms nighttime lights at predicting enterprise counts, employment, and expenditure in simple regression models. A parsimonious combination of the first two moments of the Landsat spectral bands can explain a reasonable share of the variation in economic activity in the cross-section. There is however poor prediction power of either satellite measure for changes over time.

This paper is a product of the Knowledge and Strategy Team, Development Economics. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at heilmann@usc.edu.

Can Medium-Resolution Satellite Imagery Measure Economic Activity at Small Geographies? Evidence from Landsat in Vietnam

Ran Goldblatt, Kilian Heilmann, Yonatan Vaizman[†]

1 Introduction

The analysis of satellite imagery is now a key methodology in economics and other applied scientific research. Coming straight from impartial satellites, remotely sensed data has the advantage of not being filtered through national data agencies that are potentially inefficient or biased. As Donaldson and Storeygard (2016) lay out its main benefits, remote sensing allows researchers to access information that would otherwise be difficult to obtain due to low state capacity, provides high spatial resolution, and a wide (if not global) geographic coverage. Since the marginal cost of collecting more data is low, repeated consistent samples are often available to researchers to learn about the world. Lastly, satellite imagery ignores administrative boundaries and can therefore be flexibly combined with other data at any

[†]Ran Goldblatt is Chief Scientist at New Light Technologies, Washington, DC; his email address is ran.goldblatt@nltgis.com. Kilian Heilmann (corresponding author) is a Postdoctoral Researcher at the Dornsife Institute for New Economic Thinking (INET) at the University of Southern California, Los Angeles, CA; his email address is heilmann@usc.edu. Yonatan Vaizman is a Senior Researcher at Comcast Applied AI Research, Washington, DC; his email address is yonatan_vaizman@comcast.com. The authors are grateful to Gordon Hanson, Amit Khandelwal, Klaus Deininger, Fang Xia, and Anh Pham for invaluable advice. We also thank participants at the USC Dornsife INET workshops, the Annual Meeting of the Urban Economics Association in Vancouver, and the Pacific Conference for Development Economics for helpful comments.

geographical unit. Thus, satellites have enabled development economists to study geographic entities that were previously inaccessible because of insufficient data coverage. In future, new commercial satellite projects with continually improving spatial and temporal coverage will only reinforce the importance of remote sensing for academic studies.[1]

Especially the use of nighttime lights as a proxy for economic activity is important to scholars in economics and other social sciences. Nightlight intensity has been used to approximate economic activity as light is believed to be a normal good with its consumption increasing at higher incomes. Henderson et al. (2012) have found strong correlation with GDP at the national level and researchers have gone on to use nightlights at smaller geographic units. For example, Bleakley and Lin (2012) use nighttime light intensity to measure economic activity in their study on the economic persistence of defunct portage sites and Alesina et al. (2016) combines them with sub-national ethnolinguistic maps. Similarly, Storeygard (2016) employs nightlights to study intercity transport costs and their impact on income of sub-Saharan African cities, while Harari (2016) uses them to measure shapes of urban areas in India.

However, the predictive power of nighttime light at geographies smaller than the national level has recently been disputed in the literature. For example, while Mellander et al. (2015) conclude that nightlight data have a reasonably high correlation with economic activity in the developed world context of Sweden, Bickenbach et al. (2016) find severe parameter instability between regions in India, Brazil, and the US, and cast doubt that the correlation between nightlights and GDP at the country level carries over to the subnational level. Chen and Nordhaus (2011) find similar issues when scaling down from national to grid cell levels. The reasons for this are threefold: The common nightlight data produced by the National Oceanic and Atmosphere Administration (NOAA) is only available at a geospatial resolution of 30 arc seconds (about 1km). This resolution might be too coarse to capture

[1]Private satellite ventures likes Planet and Digital Globe have recently pushed the limits to spatial resolution of up to 0.5 meters/pixel.

the characteristics of small geographical units. In addition, nighttime lights have a tendency to extend into neighboring regions, a phenomenon called blooming effect (Small et al., 2005; Abrahams et al., 2018), which complicates the identification of the actual source of lights. Lastly, nighttime light data is saturated at a certain threshold of light intensity which is often exceeded in very bright urban cores, thus making the analysis of nighttime light data at small geographies difficult. The degree to which these shortcomings of the data are relevant for studies in developing counties is however largely unknown, as economic data at small geographies is hard to find in the relevant context.

This paper presents an alternative to nightlights and use Landsat imagery to predict economic activity. It is in the tradition of other papers that have explored the potential of other remotely sensed data products that are not based on nightlights. A variety of studies use high-resolution imagery and convolutional neural network models to extract object features that are then used to predict economic outcomes such as poverty (Jean et al., 2016; Engstrom et al., 2016; Babenko et al., 2017; Perez et al., 2017). The drawback to these approaches is that they tend to use computationally intensive algorithms and rely on expensive proprietary satellite imagery. While these studies have made remarkable advances in the use of remote sensing in economics, applied researchers interested in measuring economic variables remotely might lack the skills or resources to use such methods or data.

This study therefore follows a different route and rely on medium-resolution imagery and simple regression models to predict economic outcomes from sky. It looks at imagery from the Landsat program, a public domain satellite product that has been used to survey Earth over the last thirty years, and investigate its prediction power for economic activity. In contrast to nightlights, Landsat satellites measure light reflectance during daytime at different wavelengths of the spectral band. Having a much finer spatial (30m) and temporal resolution (16 days) than nightlights, Landsat imagery have the potential to detect very different correlates of economic activity and income on the ground. Yet unlike high-resolution imagery that often operates at scales of less than one meter, the medium resolution Landsat images do not allow for feasible object recognition. The empirical exercise explores the small-scale properties of Landsat imagery to predict the distribution of economic activity and compare its performance to nightlights in Vietnam. A finely-geocoded nationwide dataset on enterprise counts, employment, and expenditure, allows to test the potential and limits of both nightlight and Landsat imagery at small geographic scales. The results indicate that Landsat outperforms nighttime lights at predicting measures of economic activity and consumption in parsimoniously specified regression models in the aggregate. Using only two simple Landsat spectral indices that are designed to capture certain land use categories can explain 25-35% of the variation in formal employment and enterprise counts as well as expenditure in the cross-section. More flexible specifications making use of more characteristics of the distribution of the raw Landsat data can improve this fit to 40-50%. Employing LASSO regularization and machine-learning type cross-validation techniques suggests that these relationships are stable and not just fitting statistical noise. In contrast to the cross-sectional results, the remotely sensed data has virtually no prediction power for changes in the time series.

The study adds to the literature on remote sensing by showing that parsimonious linear models of medium-resolution satellite data can capture the spatial variation of economic activity to a fair degree and are vastly superior to simple specifications of nightlight data in the context of a developing country. This provides a viable and inexpensive alternative for applied researchers to study areas with inferior data availability. The findings further substantiate concerns about the predictive power of nighttime lights at small geographies raised in previous studies. The negative result in the time-series cautions against the use of remotely sensed data in simple models to predict economic growth over time, as cross-section and time-series parameters are highly unstable in the setting of Vietnam.

The paper proceeds as follows: Section 2 describes the economic and satellite data used in the study, Section 3 compares the predictive power of nightlights and Landsat in the cross-section while Section 4 does the same in the time series. Section 5 concludes and discusses the results.

2 Data and Descriptive Statistics

This paper uses different satellite imagery products to predict commune-level economic data from Vietnam. This section introduces the sources of the satellite measures and the economic data.

2.1 Economic Data

This study uses economic data at the production commune/ward level in Vietnam for the years between 2004 and 2012. In Vietnam, around 9,000 communes and wards make up the third-level administrative subdivision after districts and provinces, and cover the whole country. In cities, these are called wards and in more rural areas they are called communes. For simplicity these are referred to as communes in the following. These units have a large variation in size and cover areas between 0.05 and 1,567 square kilometers, with a median area of 14.7 square kilometers. Communes therefore represent very different entities. While in cities, communes tend to be small and represent neighborhoods, there are several large communes which are sparsely populated and often sit along the western border with Laos. Even though the exact decision process of outlining commune borders is unknown, the formation of these units is highly endogenous. Unfortunately, the data does not come with population data from which a mechanism (such as a target population size within a commune) could be inferred. Figure 1 provides a graphical introduction into the shape of the commune units in relation to district boundaries. Because communes are too small to sensibly plot them for the whole country, two map inlets display the heterogeneity of their boundaries. While in rural areas, boundaries can be large with diameters of more than ten kilometers, they can be very small in urban areas.

This study uses geocoded economic data for communes from the Enterprise Survey conducted annually by the Vietnam General Statistics Office.[2] This survey covers characteristics of firms such as employment, revenue, profits and industry classification. The firms are geographically identified at the commune level. This data is aggregated to the commune in order to work with the total count of enterprises and counts of employment at this small geographic scale. Economic activity of each commune is then approximated by the number of workers and the number of enterprises employing them. The data comes with several quality concerns which are discussed here. At first, it is limited by the fact that only formal employment is recorded. In a country with a potentially very high share of the informal economy (Cling et al., 2011), these numbers can severely underestimate the actual employment numbers and results should therefore be interpreted as measures of the formal economy. Second, it excludes self-operating farmers and only accounts for employment in agricultural enterprises. Further, the sampling scheme is designed to fully survey enterprises with more than 10 employees only, while a random sample is used for smaller firms, and is thus not fully representative. There are also inconsistencies over time as boundaries occasionally change.[3] Lastly, an issue with the geocoding arises due to the fact that the enterprise data is not based on plant location but rather on firm location. This potentially exaggerates measures of economic activity at the headquarter location and underestimates it at the branch location.

[2]The authors thank Klaus Deininger at the World Bank and Fang Xia at the University of International Business and Economics, Beijing for providing access and invaluable guidance relating to the data to us.

[3]This study deals with this issue by aggregating up communes that have split or merged to create boundaries consistent over time.

This study further employs real per-capita expenditure data (measured in 2010 US dollars) from the Vietnam Household Living Standards Survey (VHLSS). This survey is conducted bi-annually and has several advantages: It follows the same communes over time and is not affected by boundary changes. It is generally believed to be nationally representative and has been used to compute official poverty statistics. However, there are certain limitations due to the survey design. While the household survey maintains a representative sample at the province level, it might not be so at the commune level (Hansen and Le, 2013) and might undercount migrant workers (Pincus and Sender, 2008). The household survey has a smaller geographical coverage than the enterprise data and comprises of about 3,000 communes. The expenditure data also suffers from small sample size as only three households are randomly drawn to be surveyed within each enumeration area. Nevertheless, the sampling issue should not be correlated with the remote sensing measure and is therefore unlikely to bias the results.

Figure in the appendix plots the distribution of the three variables that will serve as the main measures of economic activity together with commune size on a log scale. While commune size, the number of employees, and per-capita expenditure approximately follow a log-normal distribution, the number of enterprises is highly skewed and has the majority of its mass in the left tail.

2.2 Satellite Imagery

This study uses remote sensing imagery from the DMSP-OLS nighttime light dataset and the Landsat program. This section introduces the Landsat program and describes the processing of the data.

Landsat Data The Landsat program is managed by the United States Geological Survey (USGS) and consists of several satellites that capture global imagery at frequent intervals. Unlike satellites that measure nightlights, the Landsat satellites records daytime reflection of sunlight at different spectral bands. These bands include wavelengths of visible and invisible light, and thus rich information on objects on the ground. Depending on their structure, objects on earth reflect different portions of the electromagnetic spectrum, and therefore form a "spectral signature". Landsat data is organized into individual georeferenced and time-coded images and covers all of Earth except for the polar regions.

The Landsat program dates back to the early 1970s and has launched several satellites since then. The currently active satellites (Landsat 7 and Landsat 8) provide a spatial resolution of 30x30 meters, and their output can therefore be classified as "medium-resolution" imagery. While Landsat imagery lacks the detail of commercial mapping services such as Google Maps, the spatial resolution is fine enough to visually recognize certain features like cities and other land use. Besides visible light that can be recognized by the human eye, the satellites also capture non-visible light such as infrared and thermal heat. While widely used in environmental sciences, the vast Landsat dataset has only recently been introduced into the social sciences, mostly in urban studies. Several papers have examined the predictive power of Landsat imagery for population counts and urban boundaries, e.g. Goldblatt et al. (2016). There is very little usage of Landsat in Economics despite the fact that Landsat imagery is available easily at no cost.[4]

This paper uses Landsat 7 which launched in 1999 and is still operating to produce data.[5] Landsat 7 records eight different spectral bands and has a temporal resolution (the time until the satellite revisits a certain position on earth) of 16 days. Table 1 provides an overview of the recorded bands and their resolution. The analysis uses annual simple composite images

[4]The major exception is Burchfield et al. (2006)'s study employing Landsat 5 imagery to measure urban sprawl.

[5]The newest satellite, Landsat 8, collects data for additional bands, but was only launched in 2013.

of Vietnam. The simple composite algorithm corrects for the disturbance by cloud-coverage that is perennial in tropical regions and stitches together cloud-free images collected at different times. This comes at the loss of temporal accuracy as the data is not sourced from a single snapshot in time. In specific, a standard Top-of-Atmosphere (TOA) calibration is employed on all USGS Landsat 7 Raw Scenes in one year with less than 10% cloud coverage and then use the median of each pixel that satisfies this restriction. Several characteristics of the band distribution within each commune are extracted. For the empirical analysis, the mean, median, and standard deviation, as well as the 25th and 75th percentile of all bands listed in Table 1 are calculated.

Nightlights Data For nighttime lights, the stable light band of the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS) is used. In contrast to daytime Landsat imagery, these satellites measure light emitted from the globe at night. The dataset has a spatial resolution of 30 arc-seconds which, depending on the distance to the equator, is about 1 kilometer. Each pixel is coded with an integer value between zero (no light) and 63 (maximum light). This top coding due to saturation is an issue in bright city areas that easily hit the maximum sensitivity of the satellite sensor. While nightlights are in principle measured every 24 hours, the raw data requires elaborate ex-post processing and datasets are typically released as annual composites.[6] The paper uses the stable lights product that removes unstable light sources such as moonlight, clouds, fires, and gas flares that create large outliers in the data (Baugh et al., 2010). Annual images are extracted to calculate the sum of light (SOL) of each commune. This commonly used measure simply adds up all pixel values within a commune.

[6] The Visible Infrared Imaging Radiometer Suite (VIIRS) data product released by NASA now provides monthly releases from 2012 onward.

3 Predictive Power of Satellite Imagery in the Cross-Section

This section presents the methodology to correlate satellite data with economic activity data and report the empirical results in the cross-section.

3.1 Econometric Setup

This study starts by exploring the predictive power of satellite imagery in the cross-section. It estimates simple linear ordinary least squares (OLS) regressions to predict economic variables using the remotely sensed satellite data for each sample year separately:[7]

$$\log y_i = \alpha + \beta \; satellitedata_i + \epsilon_i \tag{1}$$

where y_i denotes the economic outcome of commune i and vector *satellitedata* contains commune-specific statistics from the remotely sensed satellite products. The parameters of interested are β and the R-squared measure as an indicator of the predictive power of remotely sensed data for economic outcomes.

Later specifications control for the area of the communes which turns out to be an important predictor for economic outcomes. Commune area is not random and accounting for the endogenous choice of commune boundaries is important. depicts the strong negative relationship between the number of enterprises and employees, expenditure, and commune size (See binned scatterplot in the appendix). On average, small communes have higher economic activity with estimated elasticities of -.52 (enterprises) and -.65 (employment)

[7]The appendix explores the sensitivity of the results to alternative estimation specifications. Poisson models for count data yield measures of predictive power very similar to OLS.

respectively. The same negative relationship holds for per capita expenditure with an albeit smaller elasticity of -.22.

3.2 Empirical Results

3.2.1 Nighttime Lights

The nighttime lights serve as the baseline of the study. The analysis is based on regressions of the measures of economic activity (employment, number of enterprises, expenditure) on nightlight data from DMSP-OLS. This exercise focuses on one single year in 2012. Since commune boundaries change, the shapefile does not perfectly match the economic data provided for all years. The highest matching rate is for the year 2012 and therefore the analysis proceeds with this year.

This paper follow the sum of lights (SOL) approach that is commonly applied for indicating light intensity. The sum of lights approach counts up all nightlight sensor values (coded from 0-63) of pixels that fall within the area of a commune. To account for non-linearities in the relationship, log-log models are estimated.[8] In the case of the nightlight approach with a singular regressor, the coefficient β can therefore be interpreted as the elasticity of the economic variable with respect to the total light emitted.

Table 2, Panel A reports the results from the simple linear regression model in (1). In column (1), the estimated elasticity of the number of enterprises with respect to nightlights is 0.393 in the year 2012, indicating a strong relationship between economic activity and nightlights even at small geographies. The coefficient is significant at the 99.9% confidence level and measured with high precision. The R-squared measure of 0.09 however indicates a rather low predictive power of the nightlights in the cross-section. The employment regression

[8]The log-log specification drops communes with zero nightlights of which there are 850. Using a log-linear prediction model does not alter the main results (see robustness checks in the appendix).

yields similar results. The elasticity of the number of formal employees per commune with respect to the sum of lights is 0.563 and again measured with a small standard error. Again the nightlights explain only a rather small share of the between variation in employment ($R^2 = 0.103$). The elasticity of per capita expenditure with nightlights is much smaller and estimated with a β of 0.126. This indicates that (at least in the development country setting of Vietnam) nightlights are more responsive to differences in production but less so to consumption. The R-squared measure of expenditure is even lower at $R^2 = 0.023$, suggesting that nightlights are not very predictive of overall cross-section consumption patterns in the setting of Vietnam. This finding is in contrast to other studies that find higher prediction power for small geographic units of up to 40-50% in Africa (Jean et al., 2016) and of up to 60% in Sri Lanka (Engstrom et al., 2016).

Controlling for potentially endogenous commune sizes confirms the strong positive relationship between nightlights and economic activity. Column (2) includes the log area (in square kilometers) as a regressor. The coefficient is strongly negative and indicates an elasticity of the number enterprise with respect to commune size of -0.631. The coefficient on log SOL increases to 0.589 and is still strongly significant. This is a natural consequence of the negative correlation between commune size and economic activity as outlined before. Similarly, in column (4) the employment elasticity increases to 0.806. For expenditure, the elasticity increases only slightly from 0.065 to 0.126. The considerable increase in the R-squared measure in all regressions indicates the predictive power of commune size independently of nightlights and confirms that commune boundaries are endogenous.

3.2.2 Landsat Spectral Indices

The focus is now on the question whether Landsat band values have similar predictive power for ground truth data of economic activity as nightlights. Since Landsat bands measure the reflectance of light of a certain wavelength and are difficult to interpret numerically, the analysis first proceed by using spectral indices derived from these bands. These measure certain land use patterns and can be easily interpreted. The regressions use the normalized difference built-up index (NDBI) and the normalized difference vegetation index (NDVI) which are non-linear combinations of two Landsat bands each.[9] These indices are defined as

$$NDBI = \frac{NIR - SWIR}{NIR + SWIR} \quad \text{and} \quad NDVI = \frac{NIR - Red}{NIR + Red}$$

where *Red* corresponds to the Landsat band 3 (red light), NIR is the Near Infrared measurement of band 4, and SWIR is the value of the Shortwave Infrared band 5. These indices are designed to measure urban areas and vegetation by capturing typical spectral signatures of these features. Their values range between -1 and 1, and a higher index value corresponds to more vegetation and built-up area presence respectively. For example, the NDVI measure is designed to capture live green vegetation on the ground. For photosynthesis, live plants absorb visible light (low wavelengths) but reflect infrared light (higher wavelengths), thus a higher value of $NIR - Red$ indicates presence of vegetation. The NDVI has been used in the environmental sciences to distinguish vegetation from other land uses. Similarly, NDBI is designed to detect land cover typical of urban areas (Zha et al., 2003).

These indices are chosen for several reasons: 1) They are simple transformations of the raw Landsat data that can be easily calculated by applied researchers. 2) They directly correspond to land uses that are man-made (cities, agriculture) and are plausibly correlated with economic activity. 3) They have been commonly used in other fields and can in principle be replicated with other satellite products that measure spectral light such as Sentinel-2. All measures follow a distribution which can be approximated by a normal distribution, although there are thick tails in the right (NDBI) and the left (nightlights, NDVI) end of the distributions (see density figures in the appendix).

[9] A third commonly used indicator is the normalized difference water index (NDWI) measuring water bodies. In this setting, NDVI and NDWI are highly negatively correlated. The latter is excluded to avoid issues of multicollinearity.

The analysis proceeds by estimating equation (1) with the two Landsat indices above. Since these indices are already normalized and well bounded, there is no need to apply the natural logarithm. Table 2, Panel B summarizes the results. In column (1), both indices are statistically significant in explaining differences in the (log) number of enterprises between communes. While a higher measure of built-up areas is positively correlated, the coefficient on NDVI is negative and indicates that the presence of vegetation predicts fewer enterprise in the cross-section. This is not surprising as high NDBI and low NDVI values indicate the presence of cities. NDVI is often used as a measure of agricultural productivity, but as self-employed farmers are not covered by the enterprise survey, the vegetation measure does not reflect this. The R-squared measure of the model is 0.341, suggesting a rather high prediction power with the use of only two indices. This is about four times larger than in the nightlights model.

The coefficients in the regression on employment are largely similar, indicating that employment is higher in more urban places. The Landsat indices explain 25.7% of the variation in employment and thus about 2.5 times as much as the nightlight approach (10.3%). The coefficients in the expenditure model are much smaller in absolute value, but still highly significant at the 0.1% level. Again, the difference in predictive power of both remote sensing approaches is striking. The simple Landsat model explains about 10 times as much of the variation as the sum of light measure ($R^2 = 0.274$ versus $R^2 = 0.023$). Figure 2 summarizes the predictive power of the models by plotting the actual values versus the predicted values for all three economic outcome variables.

When controlling for commune size, the coefficients on NDBI and NDVI are reduced in absolute value, but remain highly statistically significant. This indicates some correlation between commune area and vegetation and built-up area respectively, highlighting the endogenous nature of commune boundaries. However, the correlation is less pronounced than in the nightlight regression. The increase in R-squared measures after controlling for commune size is smaller than for nightlights, suggesting that the Landsat measure is rather independent of the factors that drive commune size. In case a researcher is not comfortable using commune size as a predictor, for example in out-of-sample prediction in a different country, the results suggest that Landsat is the more powerful correlate of economic activity.

3.2.3 Combining Nightlights and Landsat

The focus is now on exploring combinations of nighttime light data and Landsat imagery as predictors for the economic outcome variables. Due to their different observation period during the day, the two satellites might pick up very different correlates of economic activity. Combining them as independent sources of variation could increase the predictive power. Initial analysis shows that at the commune level, nighttime lights are positively correlated with NDBI (ρ = 0.17) but negatively with NDVI (ρ = -0.05). This would be a natural conclusion if nighttime lights are indicative of urban settlements that are picked up by the NDBI measure whereas rural areas with high share of vegetation emit less light.

Table 2, Panel C summarizes the regression results for the combinations of Landsat indices and nighttime light. Independent of the Landsat indices, the log sum of light measure remains a strong predictor of economic activity with a p-value of less than 0.001. Compared to the base regressions of the two measures in Panel A and B, the coefficients on all three predictors are very similar and nested in each other's confidence band. This suggests that nightlights and Landsat pick up very different correlates of economic activity. Naturally, model fit measures increase when incorporating both satellite datasets. Percentage increases in R-squared are 23% for enterprises, 37% for employment and 6.7% for expenditure. Combining the two data sources thus improves prediction accuracy for economic differences between communes.

3.2.4 Exploiting all Landsat Spectral Bands

Next, the predictive power of the full set of spectral Landsat bands in the cross-section is explored. We regress a "kitchen sink" specification that includes all Landsat bands as explanatory variables and then compare the fit with the results from the regressions with the simple indices only. Band averages as well as other characteristics (such as the standard deviation, median, interquartile range) of the band distributions within a geographical unit are used as predictors. While the Landsat indices such as NDBI provide an easy to interpret measure of ground characteristics correlated with economic outcomes, a researcher who is interested in remote sensing economic activity needs to form a prior on which spectral signatures to use as predictors. In contrast, using all spectral Landsat bands as predictive variables allows for an agnostic and flexible (yet difficult to interpret) way of recovering the statistical relationship between measures of economic activity and remotely sensed data.

However, the large number of potential variables poses the risk of overfitting the data and of mistaking noise as a valuable signal. To alleviate this issue, two kinds of exercises are performed to guide the econometric approach: LASSO techniques (Tibshirani, 1996; Belloni et al., 2014) to restrict the variable space and cross-validation methods from the machine learning literature. The sample is divided into a training and a testing dataset to judge out-of-sample validity of the estimated parameters. The algorithm randomly attributes 70% of the sample to a training dataset, estimates the coefficients, and then examine the out-of-sample fit for the remaining 30% testing dataset. This exercise is repeated 500 times to calculate the cross-validated R-squared as the average out-of-sample R-squared of these 500 draws.

Table 3 compares the regression results for the Landsat indices (column 1) with the ones of more flexible models. Column 2 first only uses the mean band values of each commune and then augments them with the standard deviation (column 3). Skewness of the distribution of each band is accounted for by including medians (column 4) and the interquartile range (25th and 75th percentile, column 5). Comparing specification (1) and (2) for enterprise, using only the band means as predictors ($R^2 = 0.416$) yields already a better fit than NDBI and NDVI only ($R^2 = 0.340$), indicating that other band values than those to calculate the indices provide valuable information. The LASSO regression picks all of the band means and the cross-validated R-squared is reduced only marginally compared to the full in-sample R-squared. Adding the standard deviations as predictors further increases both in-sample and out-of-sample R-squared measures ($R^2 = 0.492$), while adding further information about the distribution yields only very marginal improvements in the fit.

For employment and per capita expenditure, the results are similar. Using band means and standard deviations yields higher prediction power than the Landsat indices, while adding further distribution characteristics beyond the first two moments yields only small improvements in the R-squared measure. Throughout the analysis, the LASSO technique tends to select almost all potential regressors as informative. This is the case despite the very high correlation between certain Landsat bands. The cross-validation exercise yields very little evidence of overfitting and only in the expenditure regression are there small differences between in-sample and out-of-sample fit. This suggests that in the cross-section, the parameters of the Landsat bands are highly stable throughout the whole country for predicting economic activity.[10]

3.2.5 Exploring Heterogeneity of the Prediction Power

The analysis next explores sources of heterogeneity in the prediction power of both satellite products and address measurement errors that are likely correlated with the satellite

[10]We confirm an earlier result that SOL is an independent predictor of economic activity in unreported regressions. For each model specification, the LASSO approach picks the sum of lights measure as a predictor. Inclusion of this regressor can push the cross-validated R-squared measures of column (5) to 0.5499, 0.4538, and 0.4077 for enterprises, employment, and expenditure respectively.

measures. As stated above, the original motivation to introduce Landsat imagery was due to the expected difficulty of nightlights to capture very small but highly productive units. This section lends evidence to this by performing additional analysis of the prediction power of Landsat and nightlights by dividing the data into subsamples.

At first, the sample is divided by the share of formal employment. Since the data only measures economic activity of formal enterprises, there is concern that the share of informal employment will be correlated with the satellite data and thus will introduce a non-classical measurement error. To check this, the paper follows McCaig and Pavcnik (2015) and calculates the share of informal workers at the district level using the 2009 Population and Housing Census.[11] The sample is then split into deciles of formal employment and run separate regressions for log employment at the commune level for each sub-sample.[12] Figure 3 shows the resulting R-squared measures for three specifications: 1) Sum of Lights, 2) a specification with only the two Landsat spectral indices (NDBI and NDVI), 3) the first two moments of all Landsat spectral bands. In the upper panel of Figure 3 one can see that, as expected, the Landsat measures perform extraordinarily well in communes with a high share of formal employment. These communes tend to be cities with smaller area size. The flexible Landsat specification dominates the nightlight approach at all deciles, while the simple indices are superior to the SOL measure only for communes with a high share of formal employment.

Next, the sample is split up into deciles of commune size and again regress individual models for each decile. The lower panel of Figure 3 shows the resulting R-squared measures according to commune size. As expected, the sum of lights approach does extremely poorly at predicting economic activity of very small communes. The prediction power then increases sharply for the next decile and then decays almost monotonously with size. In contrast, both

[11] McCaig and Pavcnik (2015) define the share of informal employment by the share of all workers either being self-employed or working in family and farm businesses.

[12] Due to the uneven number of communes within a district, the decile bins do not contain exactly the same number of communes.

Landsat specifications have high prediction power for the smallest communes. One potential reason is that Landsat does well at capturing the built-up area. There is similar decay of prediction power with increasing commune size where the full Landsat band specification dominates the sum of light approach for the lower half of the sample. The simple Landsat indices specification however performs less well than the nightlights beyond the smallest communes.

The analysis concludes that, consistent with earlier studies, nightlights perform poorly at predicting economic activity at very small geographic units. This drives the overall prediction power of typical nightlight regressions down, suggesting that the nightlight-economic activity relationship is unstable. The prediction of Landsat instead is poor for very large communes. For practical purposes, the results suggest that Landsat is a good predictor if researchers are interested in studying formal growth in urban areas, while nightlights are superior for the purpose of measuring economic activity in rural areas. Regardless of which setting is more important for the practitioner, both products can be a useful complement to each other and, if available, should both be used increase prediction power for economic activity.

4 Evaluating Changes in the Time Series

The paper next examines the predictive power of satellite imagery in the time series and look at changes in economic activity over time.

4.1 Econometric setup

We continue to use a simple linear prediction model and estimate differences in the economic outcome variables on changes in the remotely sensed satellite data. For that the estimating equation in (1) is differenced to arrive at

$$\Delta \log y_{i,t} = \alpha + \beta \ \Delta satellitedata_{i,t} + \nu_{i,t} \tag{2}$$

where the operator Δ denotes the change between t and $t-1$ and $\nu_{i,t} = \epsilon_{i,t} - \epsilon_{i,t-1}$ is the potentially serially correlated error term. The key questions of interest are whether the parameter β is stable between the cross-sectional regressions in Section 3 and the analysis in the time series and whether the reduced variation in the satellite measure has enough predictive power for economic activity at the commune level.

The data contains the number of enterprises and employment for the years 2004 to 2012, and expenditure for the years 2004, 2008, 2010, and 2012. However, inconsistencies in the commune boundaries and changes in the sampling frame of the surveys lead to problems in the comparability of the data. The analysis therefore explores the predictive power for changes of different horizons. Potential serial correlation is dealt with by estimating equation (2) both in a panel framework as well as in long differences of different length.The analysis focuses on medium-run changes of enterprise counts and employment from 2004 to 2012. For the change in expenditure, results are reported for two time frames that correspond to the sampling frameworks of the VHLSS: a four-year period between 2004 and 2008 (based on the 1999 population census) and a shorter horizon from 2012 to 2014 (which both were based on the 2009 population census).

4.2 Empirical Results

4.2.1 Nightlights and Landsat Indices

Table 4 summarizes the results for the different specifications of equation (2) for the combined nighttime lights and the simple Landsat indices. In contrast to the cross-section regressions, the results show very low predictive power and parameter instability for predicting changes in the economic outcome variables of employment and enterprise counts. For the percentage change in enterprise counts for the whole period from 2004-2012 (column 1), the estimated elasticity with respect to nighttime lights is -0.0356 which is statistically significant. This finding differs vastly from the cross-sectional regression where the elasticity was positive. The employment regression in column (2) yields a similarly small SOL coefficient. This suggests that the enterprise growth is negatively correlated with increases in nighttime light output, which is counterintuitive. The analysis focuses now on the simple Landsat indices and assesses the changes in the mean NDBI and NDVI as predictors for changes in the economic outcome variables. The estimated coefficients for both measures are positive and highly significant. Thus an increase in urban areas and vegetation is associated with expansion of enterprises and employment at the commune level. Overall, the model performs poorly at capturing the overall variation over time and yields very little predictive power with R-squared measures of close to zero.

To control for serial correlation, a long-difference regression is estimated for the same period in columns (3) and (4). This yields positive coefficients on nighttime light intensity, although this parameter is imprecisely measured in the enterprise regression. Again there are strong positive effects for NDBI and a negative effect of NDVI in the enterprise regression, but an insignificantly positive effect of NDVI in the employment model. The sign changes in these regressions suggest the presence of large measurement error. Again, there is very low predictive power with R-squared statistics of at most 0.017. In summary, the time series models perform poorly and the parameter instability suggests that the model can not nail down the relationship between changes in satellite measures and economic activity well.

The expenditure data is limited by the lower sample size and the shorter time period of consistent data. The results indicate negative coefficients for the SOL measure which are imprecisely estimated. While the coefficients on NDBI are largely similar in both the 2010-2012 and 2004-2008 regressions and estimated to be around 0.3, they are only statistically significant in the former. The NDVI coefficients are statistically not significant and fluctuate. Similar to the production measure regressions, the models for expenditure can explain virtually no part of the overall variation in the data.

In conclusion, the analysis finds large parameter instability not only between the cross-section and the time series regressions, but also within different periods of the time series. The conclusion is that regardless of the time period used, neither changes in nightlights nor Landsat indices individually can predict changes of economic outcome variables in a useful way. The results confirm the cross-sectional results that the NDBI measure is a strong correlate of economic activity, even though it explains only a marginal share of its variation over time.

4.2.2 Spectral Landsat Bands

This section examines the predictive power of the spectral Landsat bands over time. Table 5 summarizes the findings. This exercise restricts attention to the first two band moments that showed the highest predictive power in the cross-section. T his table only reports estimates on coefficients that were significant at the 5% level in any of the specifications.

Column (1) reports the results of a linear regression of the number of enterprises on changes in the raw band moments. The R-squared from this augmented regression is still very low at $R^2 = 0.015$ although several band moments are highly significant predictors. The employment regression yields largely similar results: All the coefficient estimates have the same sign as in the enterprise regression although the statistical significance varies. The prediction power is considerably higher for this variable and the full Landsat model can explain 5.6% of the squared variation in log employment over time. This is of course still a much smaller share than in the cross-section. For expenditure, the model performs poorly and has barely any prediction power. Only two of the many coefficients are measured with some statistical precision while other estimates are often very different from the enterprise and employment regressions.

The analysis suggests that even when augmenting the Landsat models flexibly with moments of all spectral bands, the prediction power of the simple OLS model in the time series is very low. While Landsat performed well at capturing the spatial variation of economic activity in the cross-section, changes over time cannot be easily modeled using simple variables from either satellite measure. There are several possibilities that could lead to this result. Measurement error over time in either the economic or satellite data could increase the noise-to-signal ratio in the regression. Another speculative reason is that while land use patterns and economic activity are strongly correlated in the long term, land use change might lag behind short-term economic growth and decline. For example, an increase in production might only manifest itself in recognizable changes on the ground through construction of factories and roads after several years. Likewise, a potential asymmetric response of land use between growth and decline could not be properly captured in a linear model.

5 Discussion and Conclusion

This paper has introduced remotely sensed imagery data from the Landsat program and evaluated its usefulness for prediction of economic activity. Unlike the commonly used nightlight data, Landsat imagery comes at a much higher spatial resolution, is measured more frequently, and provides data on a multitude of spectral bands during daytime. While nighttime lights from the DMSP-OLS measures human activity in form of light consumption, Landsat imagery captures a bigger picture of Earth shaped by both nature and man. Thus, there is potential for detecting a variety of relevant features on the ground that correlate with socioeconomic data.

Using small-scale economic data, the analysis shows that Landsat imagery can act as a strong predictor for counts of enterprises and employment as well as expenditure in the cross-section in the context of Vietnamese communes. Simple combinations of Landsat bands that were developed to detect urban areas and vegetation already have reasonable predictive power, while flexible combinations of band means and standard deviations can explain a large share of the differences in production and consumption measures between communes. Cross-validation exercises and LASSO regressions indicate strong parameter stability and do not suggest overfitting due to the large number of potential parameters.

Comparing the Landsat measures to the often-used nighttime light approach, the results indicate that Landsat outperforms nightlights at predicting economic outcomes in the cross-section aggregate. This is especially true in very small areas and in communes with a high share of formal employment. These units tend to be urban areas and this confirms concerns that the DMSP-OLS nightlight data is too coarse and saturated to be applied on a very small geographic scale. However, nightlights perform reasonably well at larger commune sizes. This suggests that Landsat and nightlights are complements rather than substitutes for applied researchers interested in capturing the spatial variation of economic activity. Given their very different nature of data collection, including both satellite measures into prediction models can improve their precision. This is especially attractive as both nightlights and Landsat data are freely accessible.

When looking at changes over time, neither nightlights nor Landsat have much prediction power in the simple models. Although some coefficients of Landsat bands are significant in the regression, they can only explain a small share of the variation over time and the R-squared of the linear models is close to zero. Likewise, the results show large differences between the nightlight coefficients in the cross-section and time-series regressions. This casts doubt on the usefulness of simple Landsat or nightlight models for predicting economic growth at very small entities and confirms Bickenbach et al. (2016)'s findings of parameter instability in the SOL approach.

The analysis showed that the strength of such simple models lies in capturing stable features on the ground that correlate with economic activity but that they are unable to capture changes over time. This does not rule out that more sophisticated models and more advanced signal extraction can further improve the value of Landsat or nightlight data in the time series. This is an active research area that is constantly expanding the understanding of remote sensing tools for measuring economic growth.

References

Abrahams, Alexei, Christopher Oram, and Nancy Lozano-Gracia, "Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination," *Remote Sensing of Environment*, 2018, *210*, 242–258.

Alesina, Alberto, Stelios Michalopoulos, and Elias Papaioannou, "Ethnic inequality," *Journal of Political Economy*, 2016, *124* (2), 428–488.

Babenko, Boris, Jonathan Hersh, David Newhouse, Anusha Ramakrishnan, and Tom Swartz, "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico," *arXiv preprint arXiv:1711.06323*, 2017.

Baugh, Kimberly, Christopher D. Elvidge, Tilottama Ghosh, and Daniel Ziskin, "Development of a 2009 Stable Lights Product using DMSPOLS data," *Proceedings of the Asia-Pacific Advanced Network*, 2010, *30*, 114–130.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "High-dimensional methods and inference on structural and treatment effects," *The Journal of Economic Perspectives*, 2014, *28* (2), 29–50.

Bickenbach, Frank, Eckhardt Bode, Peter Nunnenkamp, and Mareike Söder, "Night lights and regional GDP," *Review of World Economics*, 2016, *152* (2), 425–447.

Bleakley, Hoyt and Jeffrey Lin, "Portage and Path Dependence," *The Quarterly Journal of Economics*, 2012, *127* (2), 587–644.

Burchfield, Marcy, Henry G Overman, Diego Puga, and Matthew A Turner, "Causes of sprawl: A portrait from space," *The Quarterly Journal of Economics*, 2006, *121* (2), 587–633.

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

Cling, Jean-Pierre, Mireille Razafindrakoto, and Francois Roubaus, "The Informal Sector in Vietnam," *Report of the International Labour Organization. Accessed December*, 2011, *15*, 2013.

Donaldson, Dave and Adam Storeygard, "The View from Above: Applications of Satellite Data in Economics," *Journal of Economic Perspectives*, November 2016, *30* (4), 171–98.

Engstrom, Ryan, Jonathan Hersh, and David Newhouse, "Poverty from Space: Using High Resolution Satellite Imagery for Estimating Economic Well-being and Geographic Targeting," *Job market paper*, 2016.

Goldblatt, Ran, Wei You, Gordon Hanson, and Amit K. Khandelwal, "Detecting the Boundaries of Urban Areas in India: A Dataset for Pixel-Based Image Classification in Google Earth Engine," *Remote Sensing*, 2016, *8.*

Hansen, Henrik and Trung Dang Le, "The importance of being surveyed: The representativeness and impact of the Vietnam Household Living Standards Surveys," *Unpublished Manuscript*, 2013.

Harari, Mariaflavia, "Cities in Bad Shape. Urban Geometry in India," *Working Paper*, 2016.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil, "Measuring economic growth from outer space," *The American Economic Review*, 2012, *102* (2), 994–1028.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, 2016, *353* (6301), 790–794.

McCaig, Brian and Nina Pavcnik, "Informal employment in a growing and globalizing low-income country," *American Economic Review*, 2015, *105* (5), 545–50.

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

Perez, Anthony, Christopher Yeh, George Azzari, Marshall Burke, David Lobell, and Stefano Ermon, "Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning," *arXiv preprint arXiv:1711.03654*, 2017.

Pincus, Jonathan and John Sender, "Quantifying poverty in Viet Nam: who counts?," *Journal of Vietnamese Studies*, 2008, *3* (1), 108–150.

Small, Christopher, Francesca Pozzi, and Christopher D Elvidge, "Spatial analysis of global urban extent from DMSP-OLS night lights," *Remote Sensing of Environment*, 2005, *96* (3), 277–291.

Storeygard, Adam, "Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa," *The Review of Economic Studies*, 2016, *83* (3), 1263–1295.

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

Zha, Yong, Jay Gao, and Shaoxiang Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *International Journal of Remote Sensing*, 2003, *24* (3), 583–594.

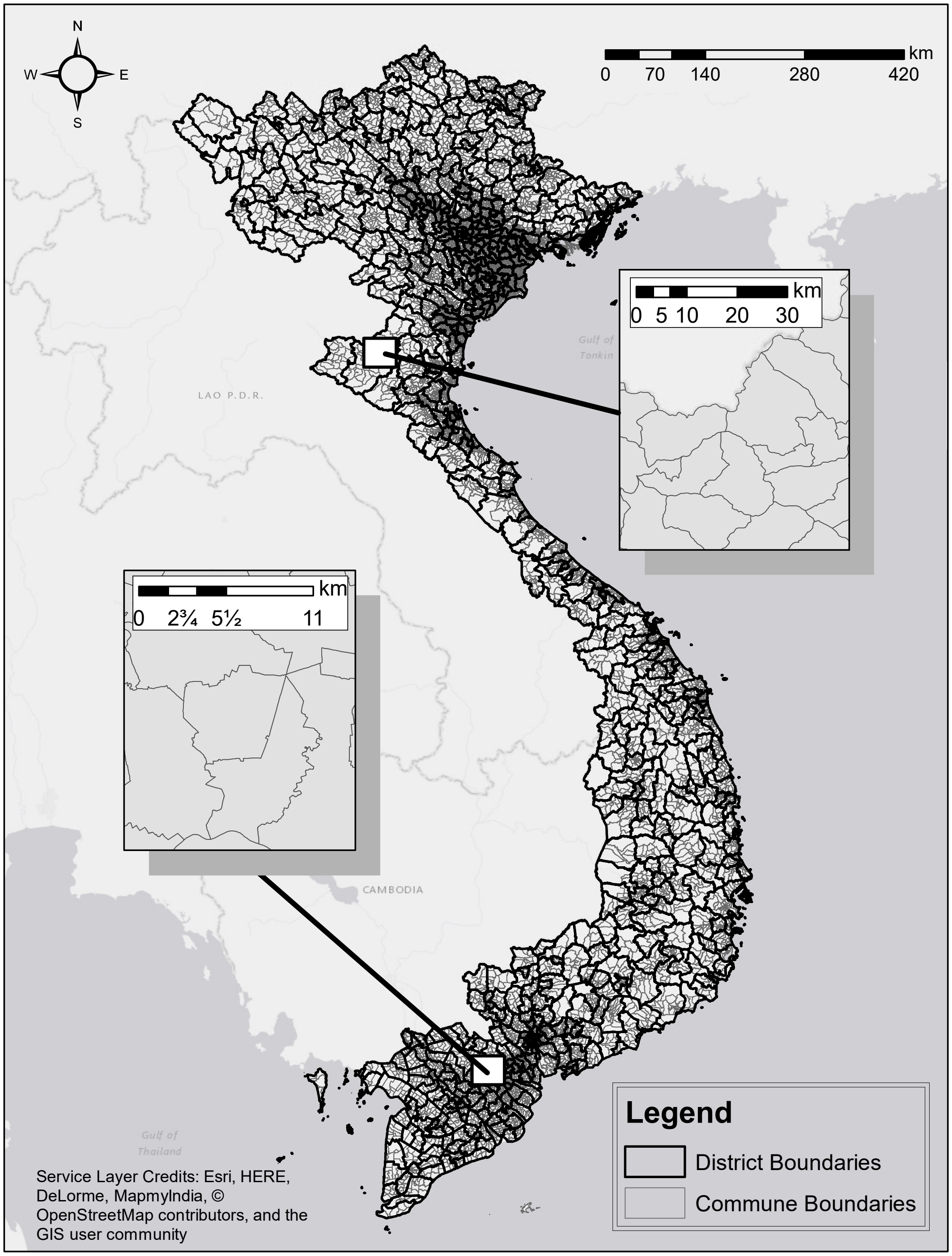


Figure 1: Administrative Boundaries of Vietnam

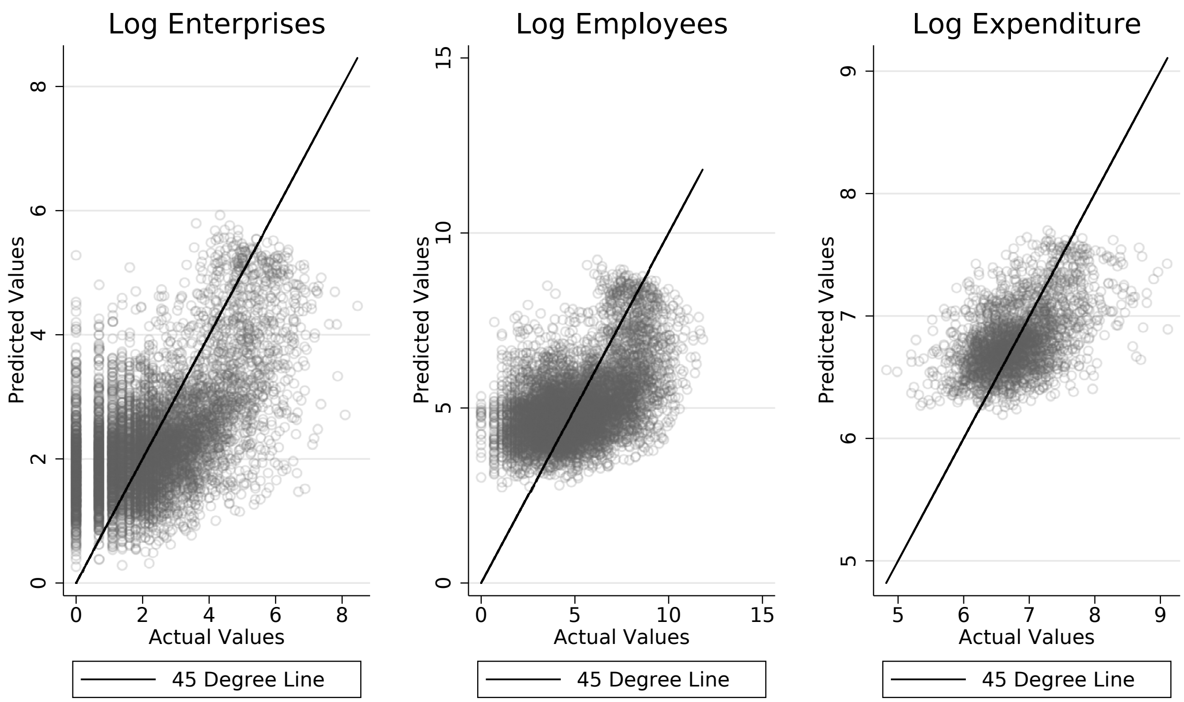


Figure 2: Scatterplots: Predicted vs Actual Values. Note: These scatterplots show the predicted versus the actual values from the simple model with two Landsat spectral indices (NDBI and NDVI) for log enterprise, log employment, and log per capita expenditure in the cross-section of the year 2012. These prediction models create some overestimation of low-activity communes and tend to underestimate high-activity neighborhoods

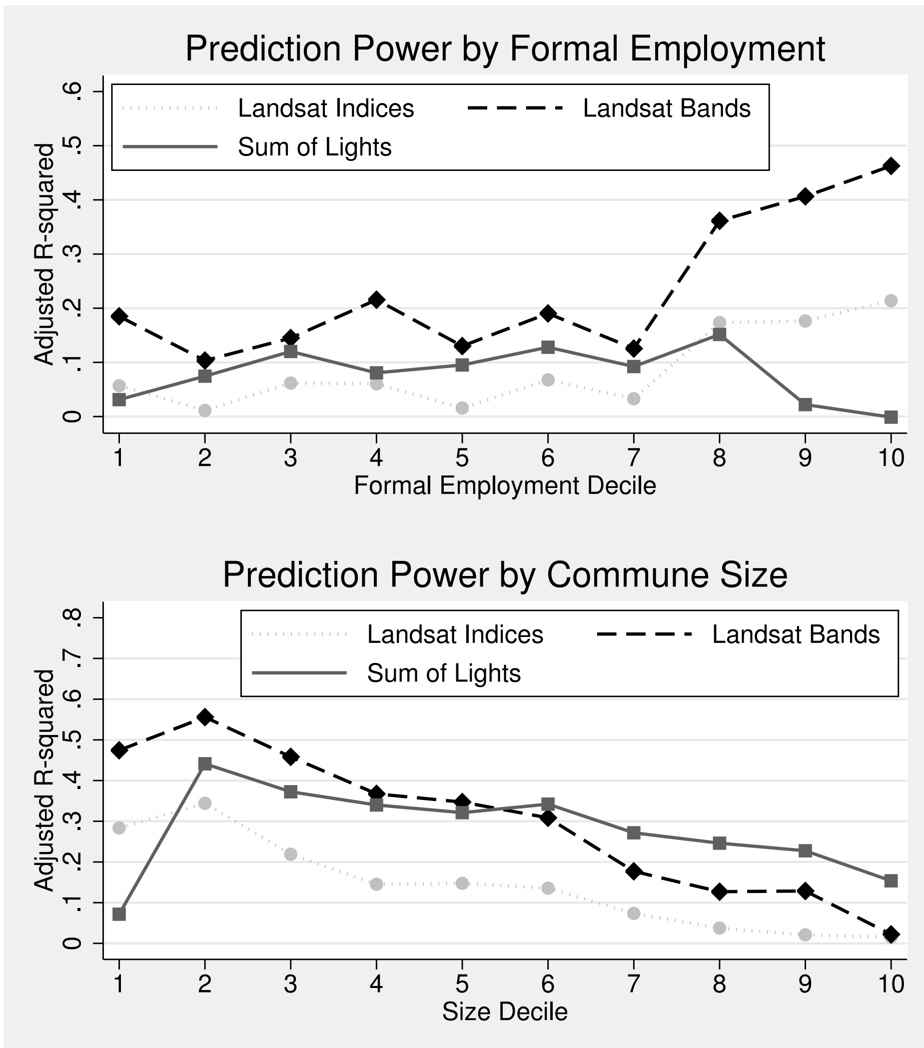


Figure 3: Heterogeneity of the Prediction Power.

Table 1: Bands of Landsat 7

Bands	**Wavelength**	**Resolution**
Band 1 - Blue	0.45-0.52μm	30m
Band 2 - Green	0.52-0.60μm	30m
Band 3 - Red	0.63-0.69μm	30m
Band 4 - Near Infrared (NIR)	0.77-0.90μm	30m
Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75μm	30m
Band 6 - Thermal	10.40-12.50μm	60m
Band 7 - Shortwave Infrared (SWIR) 2	2.09-2.35μm	30m
Band 8 - Panchromatic	0.52-0.90μm	15m

Note: Landsat records data for Band 6 at two different sensitivities. Since they are highly correlated we proceed with only the less sensitive setting.
Source: USGS Landsat

Table 2: Regression Results in the Cross-Section

Panel A: Nightlights

Dep. Var.	Enterprises	Enterprises	Employment	Employment	Expenditure	Expenditure
Log SOL	0.393***	0.589***	0.563***	0.806***	0.0655***	0.126***
	(0.0148)	(0.0194)	(0.0197)	(0.0254)	(0.00779)	(0.00844)
Log Area		-0.631***		-0.784***		-0.225***
		(0.0146)		(0.0180)		(0.00762)
Observations	7643	7643	7643	7643	2629	2629
Adjusted R^2	0.090	0.348	0.103	0.325	0.023	0.328

Panel B: Landsat Spectral Indices

Dep. Var.	Enterprises	Enterprises	Employment	Employment	Expenditure	Expenditure
Average NDBI	6.191***	5.817***	7.227***	6.678***	1.043***	0.752***
	(0.212)	(0.208)	(0.266)	(0.262)	(0.118)	(0.110)
Average NDVI	-4.129***	-2.926***	-4.894***	-3.131***	-1.732***	-0.916***
	(0.124)	(0.133)	(0.164)	(0.184)	(0.0784)	(0.0799)
Log Area		-0.219***		-0.321***		-0.148***
		(0.0137)		(0.0193)		(0.00822)
Observations	8493	8493	8493	8493	2990	2990
Adjusted R^2	0.341	0.362	0.257	0.282	0.274	0.349

Panel C: Combination

Dep. Var.	Enterprises	Enterprises	Employment	Employment	Expenditure	Expenditure
Average NDBI	6.543***	6.099***	7.534***	6.887***	1.289***	1.193***
	(0.222)	(0.222)	(0.276)	(0.272)	(0.122)	(0.123)
Average NDVI	-3.678***	-3.815***	-4.191***	-4.392***	-1.405***	-1.429***
	(0.138)	(0.139)	(0.179)	(0.178)	(0.0856)	(0.0855)
Log SOL		0.364***		0.530***		0.0566***
		(0.0127)		(0.0176)		(0.00682)
Observations	7643	7643	7643	7643	2629	2629
Adjusted R^2	0.330	0.406	0.242	0.332	0.253	0.270

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Authors' calculations

Table 3: Landsat Kitchen Sink Regressions

Dependent Variable:	**Log(Enterprises per Commune)**				
	(1)	(2)	(3)	(4)	(5)
Variables included	Indices	Means	+Std. Dev.	+Median	+Interquartile Range
Observations	8493	8493	8493	8493	8493
Adjusted R-squared	0.341	0.416	0.493	0.505	0.512
Crossvalidated R-squared	0.340	0.416	0.492	0.503	0.508
Variables selected by LASSO	2/2	8/8	16/16	23/24	37/40

Dependent Variable:	**Log(Employment per Commune)**				
	(1)	(2)	(3)	(4)	(5)
Variables included	Indices	Means	+Std. Dev.	+Median	+Interquartile Range
Observations	8493	8493	8493	8493	8493
Adjusted R-squared	0.257	0.319	0.403	0.413	0.419
Crossvalidated R-squared	0.257	0.318	0.401	0.410	0.414
Variables selected by LASSO	2/2	8/8	16/16	22/24	39/40

Dependent Variable:	**Log(Per Capita Expenditure)**				
	(1)	(2)	(3)	(4)	(5)
Variables included	Indices	Means	+Std. Dev.	+Median	+Interquartile Range
Observations	2990	2990	2990	2990	2990
Adjusted R-squared	0.274	0.356	0.397	0.404	0.413
Crossvalidated R-squared	0.274	0.320	0.378	0.388	0.396
Variables selected by LASSO	2/2	8/8	16/16	24/24	35/40

Notes: Crossvalidated R-squared is the average R-squared from 500 replications of splitting the data into a 70% training sample to estimate the model parameters and a 30% testing dataset to calculate the out-of-sample R^2.
Source: Authors' calculations

Table 4: Regression Results for Changes over Time

	Panel 2004-2012		Long Differences 2004-2012		2010-2012	2004-2008
	Enterprises	Employment	Enterprises	Employment	Expenditure	Expenditure
Change in SOL	-0.0356***	-0.0410***	0.000496	0.114***	-0.00263	-0.00283
	(0.00311)	(0.00509)	(0.0128)	(0.0231)	(0.0108)	(0.0130)
Change in NDBI	0.492***	0.685***	1.049***	1.523***	0.304*	0.333
	(0.0432)	(0.0708)	(0.161)	(0.290)	(0.123)	(0.215)
Change in NDVI	0.218***	0.469***	-0.891***	0.390	0.106	-0.163
	(0.0243)	(0.0398)	(0.132)	(0.237)	(0.112)	(0.133)
Observations	28768	28768	5308	5307	2614	2324
Adjusted R^2	0.010	0.009	0.017	0.009	0.001	0.001

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Source: Authors' calculations

Table 5: Long Differences Regression

	(1) Change in Enterprises	(2) Change in Employment	(3) Change in Expenditure
Change in Average B1	0.0388	0.0741***	-0.00323
	(0.0272)	(0.0146)	(0.0126)
Change in Average B3	-0.0897**	-0.0890***	0.0180
	(0.0333)	(0.0182)	(0.0173)
Change in Average B4	-0.0512***	-0.0498***	0.00699
	(0.0154)	(0.00765)	(0.00672)
Change in Average B5	0.109***	0.0717***	-0.0154
	(0.0192)	(0.0103)	(0.0106)
Change in Average B7	-0.0996***	-0.0556***	0.0201
	(0.0232)	(0.0126)	(0.0144)
Change in Average B8	0.0326	0.0349*	-0.0184
	(0.0359)	(0.0171)	(0.0159)
Change in Standard Deviation B2	-0.0937*	-0.0851***	-0.0180
	(0.0447)	(0.0247)	(0.0248)
Change in Standard Deviation B3	0.0781*	0.0901***	0.00771
	(0.0333)	(0.0183)	(0.0177)
Change in Standard Deviation B4	0.0178	0.0192***	0.0104
	(0.00978)	(0.00561)	(0.00535)
Change in Standard Deviation B5	-0.0412*	-0.0552***	-0.0224*
	(0.0190)	(0.0107)	(0.00999)
Change in Standard Deviation B7	0.0151	0.0320**	0.0271*
	(0.0215)	(0.0122)	(0.0115)
Change in Standard Deviation B8	0.0422**	0.000126	-0.00390
	(0.0143)	(0.00798)	(0.00827)
Observations	5794	5795	2677
Adjusted R^2	0.015	0.056	0.005

Standard errors in parentheses

Table only reports coefficients that were significant in at least one specification

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Source: Authors' calculations

Appendix

S1 Robustness Checks and Extended Analysis

S1.1 Robustness to Specification with Count Models

In the main analysis, we have used log transformations of our economic variables in linear models to capture non-linear effects. For our economic production measures, we however work with count data of enterprises and employees that are never below zero and are thus positively skewed. OLS models are unable to account for that and might lead to negative predictions of these non-negative variables. In this section, we control for robustness of our findings to different specifications and perform the same analysis using non-linear count models. To do so, we estimate Poisson regressions of the untransformed enterprise and employment measures. Table S.1 and reports the results for nighttime lights and Landsat spectral indices as predictors of economic activity.

The results are largely in line with the OLS regressions. We document a highly significant relationship between nightlights and the number of enterprises as well as total employment, and a strong negative relationship with the size of the commune. In Panel B, we again find that NDBI is a positive predictor of economic activity at the commune level while average NDVI values are negatively correlated with enterprise counts and total employment. When combining both satellite sources, the estimated coefficients do not change much and sign and significance remain the same. We also report McFadden Pseudo R-squared measures to judge the model fit. While these likelihood ratios are not directly comparable to the R-squared measures of the OLS regressions, we can nevertheless compare model fit between the Poisson regressions. Again, we find similar patterns: 1) The Landsat spectral indices dominate the sum of lights measures for both the enterprise and the employment model. 2) Including area of the commune massively improves the model fit in the nightlight regression, but less so in the specification with spectral indices. 3) Including both satellite measures improves model fit suggesting that both measures pick up independent correlates of economic activity. We therefore are confident that our conclusions are not dependent on the specific regression model used.

Table S.1: Poisson Regression Results in the Cross-Section

Panel A: Nightlights

Dep. Var.	Enterprises	Enterprises	Employment	Employment
Log SOL	0.339***	1.713***	0.650***	1.908***
	(0.0408)	(0.0571)	(0.0512)	(0.0685)
Log Area		-1.596***		-1.547***
		(0.0376)		(0.0457)
Observations	7643	7643	7643	7643
Pseudo R^2	0.046	0.627	0.140	0.566

Panel B: Landsat Spectral Indices

Dep. Var.	Enterprises	Enterprises	Employment	Employment
Average NDBI	3.949***	3.811***	4.238***	4.822***
	(0.382)	(0.498)	(0.408)	(0.570)
Average NDVI	-5.524***	-5.403***	-4.305***	-4.849***
	(0.275)	(0.277)	(0.294)	(0.258)
Log Area		-0.0243		0.108**
		(0.0341)		(0.0402)
Observations	8493	8493	8493	8493
Pseudo R^2	0.433	0.433	0.284	0.289

Panel C: Combination

Dep. Var.	Enterprises	Enterprises	Employment	Employment
Average NDBI	3.935***	4.935***	4.250***	5.208***
	(0.389)	(0.416)	(0.417)	(0.425)
Average NDVI	-5.324***	-6.334***	-3.989***	-5.453***
	(0.289)	(0.345)	(0.311)	(0.371)
Log SOL		0.528***		0.744***
		(0.0411)		(0.0476)
Observations	7643	7643	7643	7643
Pseudo R^2	0.417	0.531	0.266	0.470

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

S1.2 Functional Form Issues

In the main analysis, we have followed the literature and estimated log-log regressions to quantify the prediction power of nighttime lights for economic indicators. Previous studies that have analyzed larger regions have shown that the effect is not linear and the log transformation is preferred. In our study with very small geographic units however, taking the log of the sum of light measures creates a problem as there are communes that have zero light and are subsequently not included in the regression. In this section we experiment with different specifications and explore whether there is information in zero values of nightlights. Table S.2 reports the results for log-linear specifications of

$$\log y_i = \alpha + \beta \ SOL_i + \nu_i \tag{1}$$

and compares the model fit to the log-log specification from equation (1). Panels A and B show the results for the two specifications for nightlights. Using the linear regressor of SOL yields a higher R-squared in the enterprise (0.094 v 0.090) and expenditure regressions (0.039 v 0.023), but a lower R-squared in the employment regression (0.099 v 0.103). Once controlling for the log size of the area, the linear specification always has a higher R-squared than the log-log specification. Panels C and D compare the two specifications for the combination of Landsat indices and nightlights. In the main analysis, we forced the sample size of the Landsat regression to be the same as the non-zero nightlight set. Including the zero-light communes into this regression yields a higher R-squared in all three regressions. When using the linear SOL measure as an additional regressor, R-squared increases in the expenditure regression, but decreases in the enterprise and employment regression.

Putting these results together, we conclude that there is evidence that zero light has some additional explanatory power in certain regressions. Thus taking the log of nightlight measures at very small geographical entities might discard this valuable information. However, the model fit in the log-log regressions does not change the relative performance of the separate models. Regardless of the nightlight specification, the Landsat regressions have R-squared measures order of magnitude larger than the ones using only SOL. Similarly, the inclusion of both satellite data products always increases model fit for either functional form.

Table S.2: Regression Results in the Cross-Section. Function Forms

Panel A: Nightlights (log-log)

Dep. Var.	Enterprises	Enterprises	Employment	Employment	Expenditure	Expenditure
Log SOL	0.393***	0.589***	0.563***	0.806***	0.0655***	0.126***
	(0.0148)	(0.0194)	(0.0197)	(0.0254)	(0.00779)	(0.00844)
Log Area		-0.631***		-0.784***		-0.225***
		(0.0146)		(0.0180)		(0.00762)
Observations	7643	7643	7643	7643	2629	2629
Adjusted R^2	0.090	0.348	0.103	0.325	0.023	0.328

Panel B: Nightlights (log-linear)

Dep. Var.	Enterprises	Enterprises	Employment	Employment	Expenditure	Expenditure
SOL	0.0213***	0.0289***	29.47***	39.16***	4141.6***	6392.7***
	(0.00194)	(0.00238)	(2.733)	(3.300)	(538.2)	(693.7)
Log Area		-0.624***		-0.791***		-0.248***
		(0.0119)		(0.0152)		(0.00675)
Observations	8493	8493	8493	8493	2990	2990
Adjusted R^2	0.094	0.366	0.099	0.337	0.039	0.386

Panel C: Combination (log-log)

Dep. Var.	Enterprises	Enterprises	Employment	Employment	Expenditure	Expenditure
Average NDBI	6.543***	6.099***	7.534***	6.887***	1.289***	1.193***
	(0.222)	(0.222)	(0.276)	(0.272)	(0.122)	(0.123)
Average NDVI	-3.678***	-3.815***	-4.191***	-4.392***	-1.405***	-1.429***
	(0.138)	(0.139)	(0.179)	(0.178)	(0.0856)	(0.0855)
Log SOL		0.364***		0.530***		0.0566***
		(0.0127)		(0.0176)		(0.00682)
Observations	7643	7643	7643	7643	2629	2629
Adjusted R^2	0.330	0.406	0.242	0.332	0.253	0.270

Panel D: Combination (log-linear)

Dep. Var.	Enterprises	Enterprises	Employment	Employment	Expenditure	Expenditure
Average NDBI	6.191***	5.474***	7.227***	6.198***	1.043***	0.881***
	(0.212)	(0.212)	(0.266)	(0.263)	(0.118)	(0.120)
Average NDVI	-4.129***	-4.210***	-4.894***	-5.012***	-1.732***	-1.739***
	(0.124)	(0.123)	(0.164)	(0.160)	(0.0784)	(0.0779)
SOL		0.0177***		0.0254***		0.00331***
		(0.00172)		(0.00248)		(0.000459)
Observations	8493	8493	8493	8493	2990	2990
Adjusted R^2	0.341	0.405	0.257	0.329	0.274	0.298

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

S1.3 Additional Economic Variables

In this section, we present regression results for additional outcome measures. We construct an income measure from the 2010 VHLSS and a household asset measure from the 2009 Population and Housing Survey. For our income measure, we include all reported wage/salary income in cash and kind and augment this with income from bonuses and other payments. The income question asked in the VHLSS has a higher sample size per enumeration area than the expenditure item. For the household asset measure, we calculate the share of households that are in possession of TV sets, refrigerators, washers, and computers. Unfortunately, this data is only available at the district level and therefore at a coarser geography than the income data at the commune level.[1]

Table S.3 reports the results for these regressions. In Panel A, we regress log labor income on the sum of lights and Landsat spectral indices at the commune level. Although the SOL coefficient is positive and statistically significant at the 10% significance level, we find virtually no predictive power of nightlights for income. In contrast, the two Landsat indices NDBI and NDVI can explain about 12% of the squared variation. Similar to the previous analysis, NDBI is positively and NDVI is negatively correlated with household labor income.[2]

In Panel B and C, we compare the performance of nightlights and Landsat spectral indices for predicting the share of household assets at the district level. The sum of lights approach is a consistently positive correlate of the presence of TV sets, refrigerators, washers, and computers. However, the R-squared measures of these regressions are rather low and vary from 0.04 for computers to 0.123 for refrigerators. The two Landsat spectral indices in contrast provide higher predictive power for all household assets. With exception of the TV regression, NDBI and NDVI are statistically significant and have the expected sign. R-squared measures range from a low of 0.139 in the model for TVs to 0.492 in the computer regression.[3]

[1] Vietnam is divided in to 687 districts of which we were able to match 551. The average size of a district is 484 square meters and is therefore almost 17 larger than that of the average commune.

[2] In regressions not reported, we find that the more flexible Landsat specification using the first two moments of all Landsat spectral bands leads to R-squared measures of up to 0.24.

[3] The poor performance in the regression for TV sets could be a result of the very high penetration of television sets in Vietnam. In an average district, 86.6% of all households possess at least one, while the penetration rates for the other household assets is at most 27%.

Table S.3: Regression Results: Additional Outcome Variables

Panel A: Labor Income (Nightlights and Landsat)

Dependent Variable	Labor Income	Labor Income
Log SOL	0.0235	
	(0.0134)	
Average NDBI		1.915***
		(0.130)
Average NDVI		-0.559***
		(0.0968)
Observations	2322	2322
Adjusted R^2	0.001	0.119

Panel B: Household Assets (Nightlights)

Dependent Variable	TV	Refrigerator	Washer	Computer
Log SOL	0.0000292***	0.0000498***	0.0000272***	0.0000195***
	(0.00000560)	(0.00000713)	(0.00000569)	(0.00000416)
Observations	551	551	551	551
Adjusted R^2	0.100	0.123	0.057	0.040

Panel C: Household Assets (Landsat Spectral Indices)

Dependent Variable	TV	Refrigerator	Washer	Computer
Average NDBI	-6.367	101.5***	100.3***	86.90***
	(7.107)	(9.414)	(8.555)	(8.088)
Average NDVI	-34.87***	-62.28***	-41.83***	-37.20***
	(3.900)	(5.041)	(4.212)	(3.598)
Observations	551	551	551	551
Adjusted R^2	0.139	0.437	0.464	0.492

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

S1.4 The Role of Commune Size

In this section explore the role of commune size. Commune boundaries are almost certainly endogeneous in our context of Vietnam, however the process of delineating is not known to us. Commune size is a strong predictor of economic outcome variables (Figure S.1). Thus researchers and practitioners that are interested in predicting economic variables over space would want to make use of this predictor. Yet, the endogeneous nature might be very context specific. This creates a problem if one wanted to predict the economic geography of a neighboring country using satellite data calibrated on Vietnamese data. The same issue arises if researchers want to study areas that do not follow commune boundaries.

In this robustness check, we explore how much of our economic outcome measures commune size can still explain once we control for both satellite measures. In Table S.4, we augment the results from Panel C of Table 2 with commune size. For each outcome variable, we find that including log area increases the R-squared measure with the regressor being significant at the 0.1% significance level. Similarly, we find that commune size is always picked by the LASSO regression as a predictor even after controlling flexibly for all Landsat bands. This suggests that commune size is still a predictor for economic activity independent of the satellite measures. In regressions not reported, we find that this is also true if we specify nightlights as a linear regressor.

Lastly, we explore the prediction power of commune size for changes over time. Table S.5 replicates the results from Table 4 with log commune area as an additional regressor. Commune size is only significant in the long differences regressions where the coefficient is negative for enterprises but positive for employment. In the other regressions, we cannot distinguish the effect from being zero. Most notably, the coefficients on the the changes of satellite measures do not change in a statistically distinguishable way. This implies that commune size is not highly correlated with changes in SOL or Landsat. R-squared measures increase in the long difference regressions, but are still at most 0.019, indicating that overall prediction power of the time series models remain low.

Figure S.1: Bin Scatter Plot for Economic Activity and Commune Size (Year 2012)

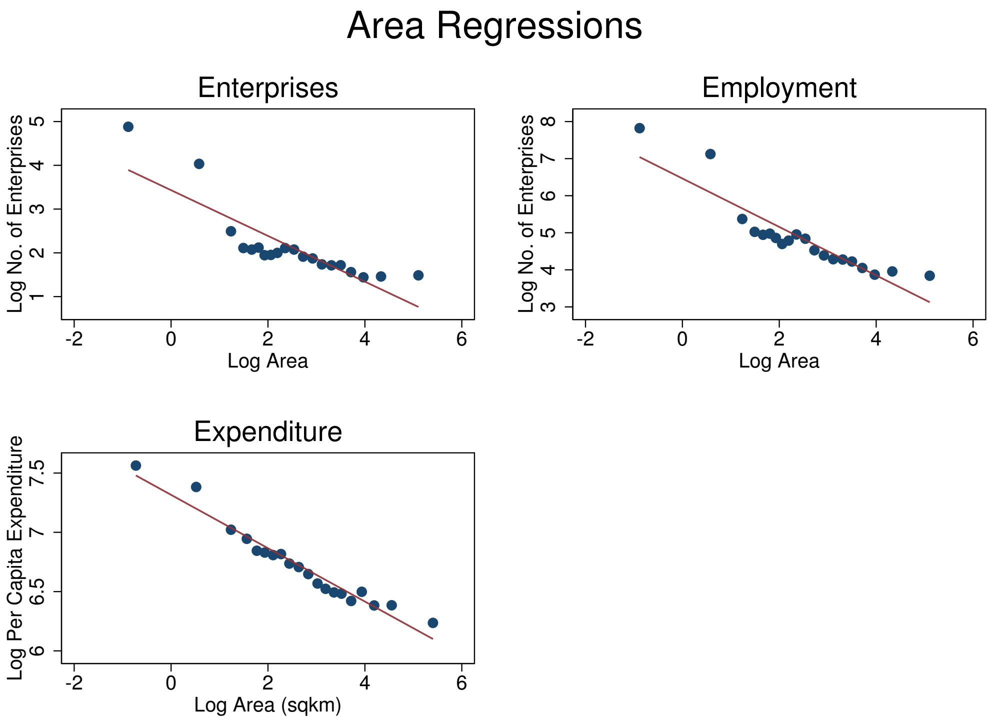


Table S.4: Landsat and Nightlight Combination with Commune Size

	(1) Enterprises	(2) Enterprises	(3) Employment	(4) Employment	(5) Expenditure	(6) Expenditure
Average NDBI	6.099***	5.274***	6.887***	5.706***	1.193***	0.750***
	(0.222)	(0.208)	(0.272)	(0.254)	(0.123)	(0.116)
Average NDVI	-3.815***	-1.869***	-4.392***	-1.601***	-1.429***	-0.600***
	(0.139)	(0.139)	(0.178)	(0.184)	(0.0855)	(0.0813)
Log SOL	0.364***	0.481***	0.530***	0.697***	0.0566***	0.103***
	(0.0127)	(0.0168)	(0.0176)	(0.0231)	(0.00682)	(0.00800)
Log Area		-0.376***		-0.539***		-0.164***
		(0.0168)		(0.0224)		(0.00923)
Observations	7643	7643	7643	7643	2629	2629
R^2	0.406	0.459	0.332	0.393	0.271	0.363

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S.5: Regression Results for Changes over Time with Commune Size

	Panel 2004-2012		Long Differences 2004-2012		2010-2012	2004-2008
	Enterprises	Employment	Enterprises	Employment	Expenditure	Expenditure
Change in SOL	-0.0356***	-0.0410***	0.0117	0.0827***	0.00293	-0.00272
	(0.00311)	(0.00509)	(0.0132)	(0.0237)	(0.0112)	(0.0130)
Change in NDBI	0.494***	0.681***	1.009***	1.637***	0.275*	0.473*
	(0.0433)	(0.0708)	(0.162)	(0.290)	(0.123)	(0.229)
Change in NDVI	0.219***	0.466***	-0.812***	0.167	0.0962	-0.155
	(0.0243)	(0.0398)	(0.133)	(0.240)	(0.112)	(0.133)
Log Area	-0.00231	0.00456	-0.0256***	0.0728***	0.0111	0.0129
	(0.00182)	(0.00298)	(0.00746)	(0.0134)	(0.00617)	(0.00734)
Observations	28768	28768	5308	5307	2614	2324
Adjusted R^2	0.010	0.009	0.019	0.015	0.002	0.001

Standard errors in parentheses

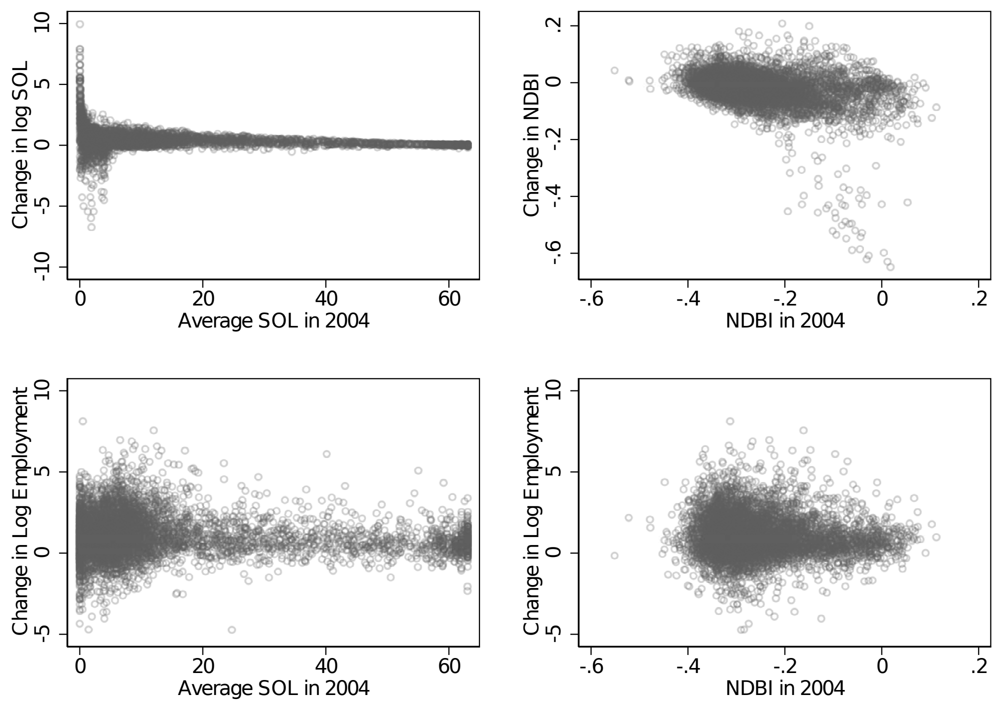
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

S1.5 Saturation

In this section, we describe the dynamics of the different satellite measures with respect to their base levels. We focus on SOL and NDBI and compare their proneness to saturation. The nightlights measure suffers from the top-coding of the satellite sensor and the threshold is easily reached in very bright cities. In our sample, 2.6% of the communes had a maximum nightlight value of 63 per pixel in 2004 and thus cannot increase their nightlight intensity anymore. However, these communes still observe positive employment growth between 2004 and 2012. To visualize the evolution of the satellite measures, we plot the changes in SOL and NDBI with respect to their base level in 2004, and then also display changes in log employment. The upper left panel in Figure S.2 shows how the the growth in log SOL flattens out over time as communes are increasingly light saturated.

The lower left panel however indicates that these communes still experience non-trivial employment growth. The NDBI measure does not suffer from this issue in the same degree. At first, none of the communes are "NDBI-saturated" as indicated by values far lower than the theoretical maximum of 1. While there is a trend towards lower NDBI growth and a tendency to have large reductions in NDBI (potentially due to measurement error) at high base values, the attenuation is much less pronounced. This is confirmed by the rather equal distribution of employment growth over the range of NDBI values. In regressions not reported, we exclude communes that had very high light intensity in 2004 from our time series regressions. This does not improve prediction power, indicating that the light saturation is not the main driver of the low prediction power of SOL over time.

Figure S.2: Changes of satellite measures and employment from 2004 to 2012 in relation to base levels of SOL and NDBI.



S2 Data Appendix

This section provides further details on the data used in the paper. We describe the data construction and report summary statistics of the basic variables in the regression analysis.

S2.1 Sample Selection

The matching of communes in the Enterprise Survey to their boundaries provided in the shapefile is not perfect. Of the 8,493 communes with positive number of enterprises and employment, we are unable to match 157 (=1.78% of communes with enterprise data and 1.61% of the expenditure sample) to their boundaries. This selection into the sample, albeit small, might bias the coefficients and inflate measures of model fit. For example, very poor communes might have simply not been geocoded. To check for this, we compare summary statistics for communes that were matched against those that were not matched in Table S.6. While communes that we were able to match have on average a higher number of enterprises and employees, these differences are not statistically significant at conventional levels due to the very high variation in both measures. For the VHLSS data where the sample is much smaller, the matched communes are characterized by lower real per capita expenditure. Again, the differences are not statistically significant. Together with the low percentage of unmatched communes, we therefore do not believe that any potentially endogeneous selection into the sample would drive the results presented in the main section. Note that not all communes present in the VHLSS have a positive number of enterprises, so the VHLSS sample is not a perfect subset of the Enterprise Survey sample.

Table S.6: Balance Table

	Matched	Not Matched	Difference
Number of Enterprises	39.8	28.0	-11.7
	(134.7)	(49.3)	(10.7)
Number of Employees	1242.05	923.99	-318.06
	(4956.3)	(1877.6)	(396.1)
Observations	8493	157	
Real Per Capita Expenditure	989.6	1147.96	158.35
	(735.9)	(643.9)	(105.8)
Observations	2990	49	

Standard errors in parentheses

S2.2 Inconsistencies in Commune Boundaries

The commune boundaries in Vietnam are not fixed and change over time. This creates problems in the time-series regression as the geographic units are not comparable over time. Since commune boundary changes are likely non-random, it could also indicate a selection into the sample if we ignore them in the cross-section. We deal with this issue by aggregating up communes that split or merged to larger units that are comparable over time. Boundary changes affect 3.2% of all the communes in our sample. To check for robustness to the aggregation of communes, we exclude observations that experienced boundary changes and report regression results for this specification in Table S.7.

The results indicate that excluding these observations does not qualitatively change the results of the previous analysis. In the cross-section regression, the prediction power of both models is very similar. In general, aggregating communes leads to larger R-squared measures in the nightlights and smaller measures in the Landsat regressions. This is expected as aggregating geographical units reduces the measurement error for the coarse nightlight data. The time-series models using only the communes that did not change boundaries still have very little prediction power, suggesting that boundary changes do not drive the non-results over time.

Table S.7: Robustness to Commune Aggregation

	Enterprises		Employment		Expenditure	
Commune aggregation	yes	no	yes	no	yes	no
	Cross-Section Regressions					
Nightlights						
Adjusted R^2	0.09	0.073	0.103	0.088	0.023	0.02
Observations	7643	7373	7643	7373	2629	2436
Landsat indices						
Adjusted R^2	0.341	0.356	0.257	0.266	0.274	0.285
Observations	8493	8209	8493	8209	2990	2788
	Long Difference Regressions					
Adjusted R^2	0.017	0.016	0.009	0.01	0.001	0.001
Observations	5308	5082	5307	5081	2324	2110

S2.3 Descriptive Statistics

Table S.8 summarizes the distribution of both the economic outcome variables as well as the satellite imagery. Panel A reports the statistics from the raw data while Panel B summarizes the regression sample for the cross section in 2012. Panel C shows the equivalent for the long-changes from 2004-2012 while Panel D reports statistics of the household assets from the 2009 Population Census. Figure S.3 shows more detail on the shape of the distribution for the economic activity variables as well as the main satellite measures. The correlation of this measure with the individual Landsat band means is presented in Figure S.4. The nightlights show very low correlation with either the visible and non-visible Landsat spectrum. The maximum correlation is with the shortwave infrared band 7 at 0.137.

Table S.8: Descriptive Statistics

Panel A: Raw Data					
	Observations	Mean	Std Dev	Min	Max
Total number of enterprises	8,493	39.8	134.7	1	4,722
Total number of employees	8,493	1242.0	4956.4	1	134,574
Per capita expenditure (2010 USD)	2,990	989.6	735.9	106.8	9,024.8
Sum of Lights (SOL)	8,757	13,473	22,551	0	573,895
Area in (sqkm)	8,757	28.6	56.2	.05	1,567.1
Panel B: Regression Sample Covariates (Commune Level, Year 2012)					
	Observations	Mean	Std Dev	Min	Max
Log Enterprises	7,643	2.28	1.57	0	8.45
Log Employment	7,643	5.05	2.10	0	11.80
Log Expenditure	2,629	6.79	.542	4.81	9.10
Log Labor Income (VND)	2,549	9.92	.703	6.21	12.4
Log SOL	7,643	9.02	1.19	1.60	13.26
Average NDVI	8,757	.405	.147	-.172	.710
Average NDBI	8,757	-.285	.089	-.653	.086
Log Area	7,643	2.34	1.31	-2.85	7.35
Panel C: Regression Sample Covariates (Changes over Time)					
	Observations	Mean	Std Dev	Min	Max
Change in Log Employment	5,307	1.10	1.29	-4.71	8.12
Change in Log Enterprises	5,307	2.83	1.50	0	8.45
Change in Log SOL	5,307	.417	.765	-6.72	9.94
Change in Average NDBI	5,307	-.022	.060	-.648	.208
Change in Average NDVI	5,307	.029	.074	-.273	.389
Panel D: Household Assets (District Level, Year 2009)					
	Observations	Mean	Std Dev	Min	Max
Share of HH with TV	551	86.58	12.24	22.03	99.21
Share of HH with Refrigerator	551	27.04	18.84	1.99	94.96
Share of HH with Washer	551	11.17	14.95	.123	76.30
Share of HH with Computer	551	10.26	12.73	.698	75.41

Figure S.3: Histograms of Economic Variables and Satellite Measures (Year 2012)

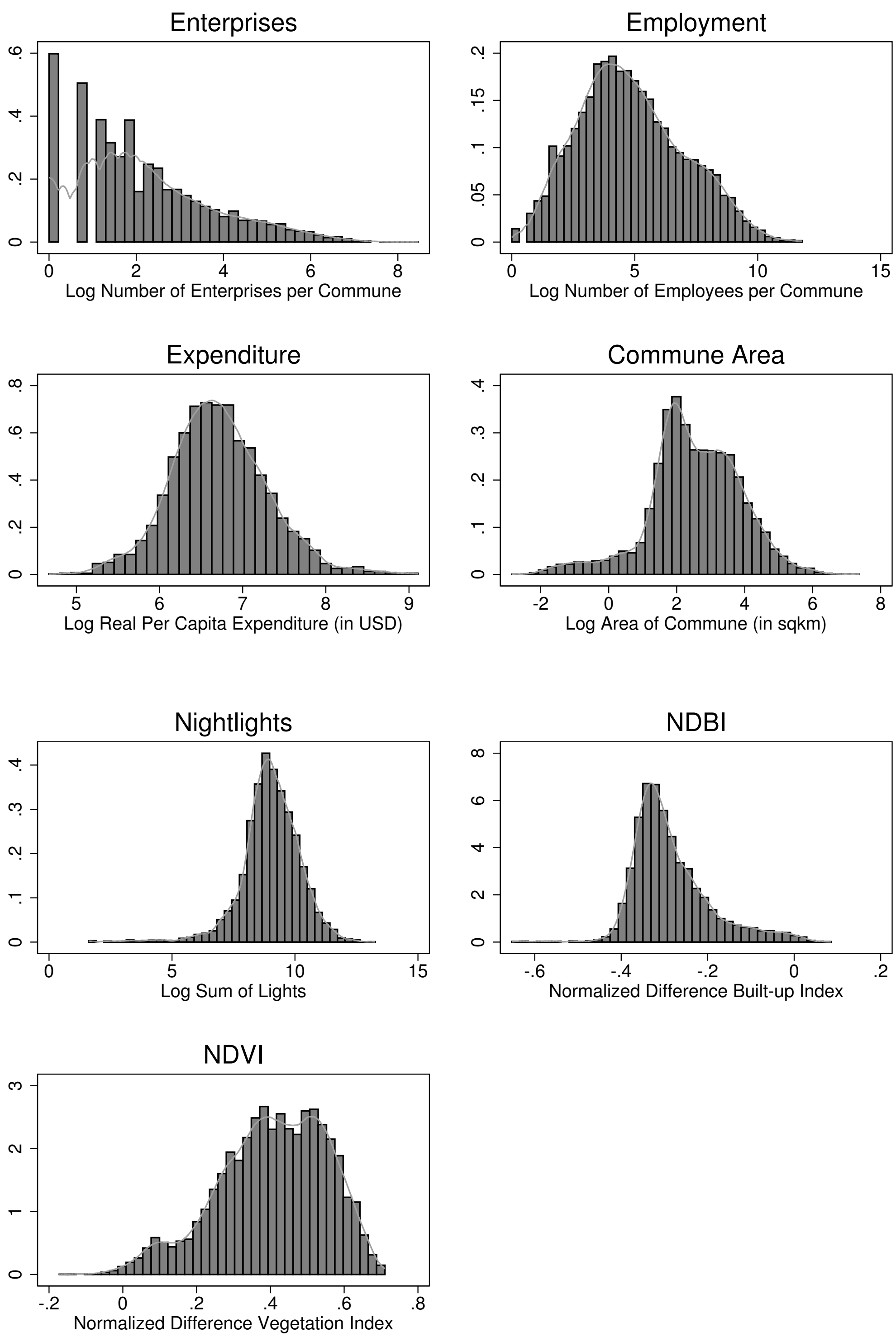


Figure S.4: Correlation Heatmap between Nighttime Light and Landsat Bands.