CESifo WORKING PAPERS

9701
2022

April 2022

Measuring Firm Activity from Outer Space

Katarzyna Bilicka, André Seidel

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9701*

Measuring Firm Activity from Outer Space

Abstract

To understand how global firm networks operate, we need consistent information on their activities, unbiased by their reporting choices. In this paper, we collect a novel dataset on the light that factories emit at night for a large sample of car manufacturing plants. We show that nightlight data can measure activity at such a granular level, using annual firm financial data and high-frequency data related to Covid-19 pandemic production shocks. We use this data to quantify the extent of misreported global operations of these car manufacturing firms and examine differences between sources of nightlight.

JEL-Codes: H320, H260, F230.

Keywords: multinational firms, nightlight data, global firm networks.

Katarzyna Bilicka
Utah State University / Logan / UT / USA
kat.bilicka@usu.edu

André Seidel
University of Bergen / Norway
andre.seidel@uib.no

This draft: April 2022
We thank Roland Hodler, Luis Martinez, and Stelios Michalopoulos for their comments. We would also like to thank seminar participants at University of Exeter, TARC, University of Vigo, and the participants at the Urban Economic Association 2021 North American Meeting, the IIPF 2020 Annual Congress and the NTA 2020 Annual meeting for their helpful suggestions. We gratefully acknowledge funding from the Utah State University Center for Growth and Opportunity. We would also like to thank Felix Kintru, Collin Butterfield, Erin Smith, Gabriel Adams, Ellie McDonald, McKlayne Marshall, Whitney Yardley, David Mann, Jamison Jones, Abraham Alhomadi, Samuel Berry, Caleb Smith and Hayden Clegg who helped in hand-collecting the data.

1 Introduction

In recent decades, the use of remote sensing data to monitor human activity has dramatically increased. The advantage of this data is that it is consistently available on a global scale and, as a result, allows for studying human activity in areas where no reliable statistical data exists. With the development of satellites and technological progress, the question is whether this data also allows us to study the economic activity of small spatial units, such as neighborhoods or firms. This would allow us to analyze firm activities outside of developed countries, where the data is harder to obtain, and would help create datasets free of biases coming from different data reporting standards and firm reporting choices.

In this paper, we hand collect a novel dataset of car manufacturing plants that belong to the 18 largest car manufacturers in the world. We match the information on their footprints with nightlight and firm financial data to study the global distribution of their activities from space.[2]

In the first part of the paper, we focus on understanding whether we can use nightlight data to proxy for such granular firm-level activities. We do so in two steps. First, we analyze the effects of the Covid-19 pandemic closures of all car manufacturing businesses announced in spring 2020 around the world. For each car manufacturing plant, we collect daily data on nightlight emitted by those factories 4 weeks before and after the Covid-19 pandemic closure dates. These firms constitute our treated group. To control for differences in light emitted at each latitude and longitude at each time of the year, we use as a control group these same factories on those same dates, but in 2019. We find that a complete factory closure results in up to a 14% reduction in nightlight emitted by these factories. These results suggest that light emission is *causally* linked with production activities, such as the use of labor, but a far larger part of factory light emission comes from fixed infrastructure.

Second, we match our remote sensing data with annual firm financial data for the years 2013- 2018. We show strong positive correlations between firm turnover, assets, employment, reported profits, and nightlight data, in both a cross-section and across time. Nightlight data explains a large, 70%, variation in firm activities across factories, and a much smaller variation across time. The magnitude of the estimated coefficients suggests that a 1% change in nightlight emitted by a factory is correlated with a 0.29% change in firm turnover. We conclude that nightlight is a good predictor of firm activity, especially proxying for fixed

[2]The focus on manufacturing firms allows us to pin down the fixed geographical location of each production site and measure nightlight emissions and output related to that particular site. This is not possible for service firms with more mobile capital, such as Google or Amazon.

infrastructure.

In the second part of the paper, we turn to applying our dataset in two contexts. First, we use it to measure the scale of misalignment between activities predicted using nightlight data and those reported by our car manufacturers across all locations where they produce cars. To do so, we use the total aggregated nightlight data together with consolidated manufacturing turnover data to apportion firm activity to each factory using the proportion of nightlight that each factory emits. We then compare the predicted and actual turnover to calculate the scale of misallocated and misreported turnover. We find that around 50% of firm activities are not reported in places where that activity occurs and for almost 38% of turnover, we do not actually know where the real activity occurs when looking at financial data. In that, we contribute to the literature that estimates the role of tax havens in distorting financial flows (Coppola et al.; 2021) and profit reporting (Gumpert et al.; 2016; Hines and Rice; 1994; Tørsløv et al.; 2020).

We find no correlation between the size of misallocation and corporate tax rates at the country level, but we show that *missing* financial information is positively correlated with corporate tax rates of countries in which activities are not reported, even after controlling for firm size using our nightlight data. This suggests that firms would rather not report any activity than report a biased number in high-tax rate countries. As such, tax arbitrage opportunities, such as profit-shifting, may play an important role in how transparent firms decide to be about reporting their financial information (Bilicka; 2019; Bilicka et al.; 2021; Desai et al.; 2006; Suárez Serrato; 2018). This novel data can help us gain a better understanding of the global operations of large multinationals, that are not biased by reporting rules and strategic decisions.

The second application is more technical in nature. We explore *sources* of changes in nightlight emissions to reconcile differences in the predictive ability of nightlight data between urban and rural areas (Gibson et al.; 2021). Conceptually, there are two possible sources of changes in light: changes in the area occupied by a factory or the intensity of light production. Our data allows studying both effects separately for the first time. As such, this paper also offers a technical contribution to the nightlight data literature in explaining where and how the nightlight is generated and whether this matters for how we interpret the nightlight measures (Henderson et al.; 2012, p. 999). We split the total light emitted into the factory area and the intensity of light emitted by that area to show that area is a relatively stronger predictor of changes in firm activities over time. In contrast, the literature on growth within *urban* agglomerations finds nightlight intensity to be a good proxy for changes in urban productivity. We reconcile these findings by using a dataset that classifies land use in

Europe into residential and non-residential matched with nightlight emitted by those regions and regional GDP data. Similar to firm-level data findings, we show that light emitted by industrial production carries little additional information beyond overall land use. At the same time, light emitted by non-industrial areas carries a similar level of information to land use. Hence, if areas systematically differ in industry presence, we would expect nightlight intensity to have a differential predictive power between the two.

Broadly, we contribute to a large literature on the links between nightlight data and economic activity, by providing novel evidence on how these links work using small spatial units. While the nightlight is linked to economic activity at the national (Henderson et al.; 2012) and sub-national level (Bluhm and Krause; 2018; Lessmann and Seidel; 2017), there is little evidence on how this data performs at the firm level. Studies that utilize the newest generation of nightlight data obtained from the Infrared Imaging Suite (VIIRS) Day-Night Band (DNB), still focus on regional applications (Gibson; 2020; Gibson et al.; 2021). A sub-set of this literature uses nightlight data to specifically study economic shocks due to, for example, war (Li et al.; 2017), natural catastrophes (Fabian et al.; 2019; Mohan and Strobl; 2017), power grid failures (Elvidge, Hsu, Zhizhin, Ghosh, Taneja and Bazilian; 2020) or the Covid-19 pandemic (Bustamante-Calabria et al.; 2021; Elvidge, Ghosh, Hsu, Zhizhin and Bazilian; 2020; Ghosh et al.; 2020; Liu et al.; 2020; Straka et al.; 2021). However, these studies still focus on fairly large spatial aggregate units of economic activity like sub-national regions, cities, or larger neighborhoods.[3]

2 Remote sensing datasets

In this section, we describe the remote sensing datasets that we use in this paper. We focus on the two key datasets related to the firm-level activity that we use throughout the paper. We provide details of the cleaning and tagging process in Appendix A.

To identify the factory footprints, we predominately rely on daylight satellite images provided by Google maps, which we supplement with information from OpenStreetMaps and aerial shots of factories from firm websites. For each factory area, we also collect data

[3]The older generations of nightlight data have been applied in various economic contexts, such as, for instance, origins of urbanization (Henderson et al.; 2017) or its consequences (Bluhm and Krause; 2018), the origins of ethnic inequality (Alesina et al.; 2016), the growth effects of human capital formation (Gennaioli et al.; 2013), rescue conflicts (Berman et al.; 2017) favoritism (Hodler and Raschky; 2014), trade (Hirte et al.; 2020), institutions (Michalopoulos and Papaioannou; 2013, 2014) or growth of dictatorships (Martinez; 2019). For a detailed review on the use of nightlight data see Donaldson and Storeygard (2016), or more recently Levin et al. (2020).

on nightlight emitted by those areas using the Visible and Infrared Imaging Suite (VIIRS) Day-Night Band (DNB). We collect daily data from the raw Nightly DNB Mosaic and Cloud imagery and the annual data from the Annual VNL V2. For both sources, the resolution of the data is 15 arc seconds, which corresponds to an area of approximately 500m x 500m at the Equator. Given the small spatial units that we analyze, we take great care in cleaning this data to avoid any potential problems coming from neighboring blooming effects and the volatility of this data with respect to daily changes in cloud coverage.

3 Is nightlight a good proxy for firm-level activity?

To understand whether nightlight data is a good measure of firm-level activity, we need to test whether nightlight can proxy for such a granular level of operations. We do so in two steps. First, we use the exogenous shock to short-run firm activities that the Covid-19 pandemic generated. Second, we use correlations with detailed unconsolidated financial data to uncover relationships between nightlight and reported operations of firms.

3.1 Is nightlight causally linked with firm activities?

In Spring 2020 almost all of the car manufacturers around the globe closed their factories to prevent the spread of Covid-19. This shock offers a perfect laboratory to examine the effects of short-term changes in activity in, and around, these factories on the emission of nightlight. These factory closures should not have any immediate effect on the production capacities of these firms. Hence, any change in the light emitted by factories around these events is likely related to changes in short-run production activities, such as running machines, employees coming to work, or transporting final and intermediate products. We use this shock to measure how a complete shutdown of a factory affects the amount of light emitted by that factory.

Sample and estimation approach We hand collect the dates when car manufacturing factories first closed due to Covid-19 from marklines.com and just-auto.com. We summarize these in Figure A2. Note that majority of car manufacturing factories in Europe and the US closed around March 17th - 20th. Later closure dates come mainly from Asian countries that closed due to part shortages from European and US factories, while January closure dates refer to factories in China.[4] We provide descriptive statistics for this sample in Table A1.

We use the difference-in-differences (DID) approach to investigate the causal effect of Covid-19 factory closures on nightlight emitted by those factories. We compare nightlight 4 weeks prior and post the factory closures. We do not have traditional control and treated groups as almost all factories around the world closed due to Covid-19 at some point during 2020. Instead, we construct our control group using the same factory on the same day in 2019. This approach controls for the fact that light is different at each latitude and longitude at each time of the year, i.e. for the systematic influence of so-called "stray lights". As such, in our estimations, we compare the effects of factory closure on light emitted in 2020 relative to light emitted by the same factory on the same day in 2019.

We further account for the potential blooming lights effect coming from the nearby neighborhood areas and other local omitted factors. In our context, the most important of those other factors is the effect of the general lockdown measures that countries have introduced in response to Covid-19. Because we are dealing with very small units of observations relative to previous literature, the concern we have is that lights emitted by our factories may be contaminated by the effect of the blooming lights coming from the nearby neighborhood areas. To make sure that we pick up the effect of factory closure, rather than nearby lockdown and stay-at-home orders, we control for the average nightlight emission within a 5km radius around each production site excluding the factory itself.

Our identification strategy relies on the assumption that in the absence of Covid-19 closure, the amount of light emitted by factories would evolve in the same way in 2020 as in 2019 during non-cloudy nights. Consequently, to test this assumption *and* to estimate the dynamic effect of closure on nightlight emitted by these factories, we use an event study design. We use a week before Covid closure as a benchmark and normalize all coefficients to zero in that week. Hence, we estimate the following equation:

$$\begin{aligned} ln(light\_sum_{i,w}) =& \alpha + \sum_{\kappa=-4}^{4} \gamma_w 1[w=\kappa] + \sum_{\kappa=-4}^{4} \delta_w treated_i \times 1[w=\kappa] \\ &+ \delta \times X'_{iw} + \psi_i + \mu_j + \epsilon_{i,w} \end{aligned} \tag{1}$$

where, i is a factory, w is weeks. $ln(light\_sum_{i,w})$ is light emitted by a factory in each week;

[4] We also collected the official dates when factories announced that they restarted their car production. However, some factories, especially in the US and Europe, reopened earlier, producing face masks or ventilators. We do not have information on when these partial activities have restarted.

$\sum_{\kappa=-4}^{4} 1[w = \kappa]$ is a series of week dummies that equal one in each of the κ weeks away from the closure date, with the dummy variable corresponding to $\kappa = -1$ as the omitted category. $treated_i$ is a dummy variable that equals 1 in 2020 and zero in 2019. X'_{iw} is a set of factory-level control variables, such as general weather conditions around the factory; ψ_i is the factory-specific fixed effect, μ_j are weekend fixed effects and $\epsilon_{i,w}$ is the error term. We control for weekend fixed effects, because the light emitted by factories may be related to weekday operation schedules. We restrict our sample to observations with no cloud coverage within a 5km radius. We estimate the model using 4 weeks before and after each factory closure, and consequently, we bin coefficients at those endpoints and do not plot them. The coefficients of interest are the δ_w, which measure the average difference in lights in each week relative to the week before Covid closure in 2020.

Summary of results We plot the coefficients from the dynamic estimation in Figure 1, with the corresponding coefficients reported in Table A3 in the Appendix. Blue hollow diamonds correspond to coefficient estimates, while vertical lines are 95% confidence intervals. We find that in the week that factories close due to Covid-19, there is an almost 14% reduction in nightlight emitted by those factories. This effect persists in week 1 and gradually declines. By week 3, a lot of factories in our sample started filling out government orders for masks or ventilators production and ground activity started picking back up. Further, we find no difference in the light emitted by the affected factories relative to 2019 before the Covid closure, which suggests the effect is due to the closure itself.[5] The magnitude of the estimated effect suggests that production activity has a significant, but small effect on the light emitted by production sites. Note that factory closures were accompanied by an almost 100% reduction in the labor force. Hence, a 14% reduction in nightlight emitted indicates that lights are mostly related to already existing production facilities, rather than short-run activities.

3.2 Is nightlight correlated with firm activities?

To quantify the correlations between nightlight and firm-level reported operations, we match our factory-level data with firm-level annual accounting data. Firm-level data was collected using Orbis Bureau van Dijk unconsolidated information on total assets, fixed assets, em-

[5]In Table A3 we report results from the general estimation using daily, rather than weekly, data and all the observations until the official opening date. We find that after the factory closure, there was, on average, an 8% reduction in the nightlights emitted by those factories in 2020 relative to 2019 before they officially reopened.

ployment, profits and losses, and turnover. We use the hand-collected factory footprints information from Google maps to manually identify firms in Orbis using firm names. In our dataset, some subsidiaries own several factories and we aggregate factory areas and the light they emit at the subsidiary level to be able to compare firm financial information with our remote sensing data. We test the validity of this aggregation by keeping only the observations where each factory belongs to one subsidiary and our results remain unchanged. The unit of observation in this analysis is a reporting subsidiary. We provide descriptive statistics for this sample in Table A2.

In this part of the analysis, we rely on the cross-sectional *and* panel variation in the firm financial data. We create a cross sectional dataset by collapsing the data at the firm level across all time periods and use averages of all financial and nightlight variables. Both approaches use the following general specifications:

$$ln(output)_i = \alpha + \beta \times ln(light\_sum_i) + \delta \times X_i^{'} + \epsilon_i \tag{2}$$

where, $ln(output)_i$ is the logarithm of turnover, total assets, fixed assets, profits and losses, or employment; $ln(light\_sum_i)$ is the logarithm of the total light emitted by a factory; $X_i^{'}$ includes a set of control variables. In the cross-sectional analysis that includes parent and country fixed effects; in the panel analysis, year and firm fixed effects. $\epsilon_{i,d}$ is the error term. We present results using both cross-sectional and panel estimates for three reasons. First, both the variation across firms and across time is informative in our setting, as they offer different information about the predictive power of the nightlight data. Second, given that we only have 6 years of financial information available, large changes are likely to be rare during our analysis period, giving us relatively small time variation for identification purposes. Third, it is entirely plausible that annual changes in financial data may not reflect fundamental changes in the economic activity of firms.

Summary of results We summarize the correlation between the remote sensing indicators for firm activity and output in Table 1. In Panel A, we present results using a cross-sectional variation, while in Panel B, we use panel-level data. We show that a 1% difference in nightlight emissions between firms is associated with a 0.8% difference in firm turnover (Panel A). The results from Panel B column (1) suggest that a 1% increase in light is associated with a 0.29% increase in firm turnover. Since these estimates rely on the variation in light across time, we can directly compare them to previous findings; for example, (Henderson et al.; 2012) finds this correlation to be a similar, 0.28, at the country level.

Columns (2) - (5) show that nightlight is also positively correlated with other measures of firm activity, such as total assets, employment, fixed assets, and profits and loss before taxes. Further, we consider the overall power of our proxies. When looking at the R^2 reported in Panel A Table 1, nightlight explains a large amount of the overall variation in firm activity, over 70%. However, this predictive power is driven by the ability of light to explain the cross-sectional variation. In Panel B, comparing the overall and within R^2, we find that financial proxies do poorly in predicting the variation in firm activity over time, explaining 5-14% of that variation. This is in sharp contrast to the previous literature that promotes nightlight data primarily as a proxy for growth. In the case of firm-level data, the strength of nightlight proxies is related to differences between, rather than within, firms, at least in the short panel that we analyze here. Further, the relatively large standard errors in column (3) in Panel B suggest that the relationship between employment and nightlight is less precise than that for total assets and turnover. This is in line with the causal estimates that point towards stronger explanatory power of infrastructure over labor for firm-level nightlight data.[6]

4 Economic application: A bird's-eye view of the global activities of MNCs

Given the strong correlation between factory activities and nightlight data, we can use our new dataset to analyze the global activities of MNCs, especially in places where no financial reporting of those activities exists. We first focus on understanding the scale of unreported turnover and profits, relative to factory operations as measured by nightlight data. Second, we consider the scale of misallocated turnover, profits, and assets by comparing the reported firm activities with the ones predicted using nightlight data. For this analysis, we use the cross-sectional data given that nightlight is a much better predictor of firm activity in that context.

4.1 Missing accounting data

We start by considering what determines the extent of missing financial information at the firm level. To do that, we estimate the following model:

$$miss_i = \alpha + \beta \times ln(light\_sum_i) + \delta \times X_k^{'} + \epsilon_i \tag{3}$$

[6]When we include a control for potential blooming effect in Table A4 the results remain the same.

where $miss_i$ is a dummy equal to 1 if a firm never reports any financial information, $ln(light\_sum_i)$ is the logarithm of the total light emitted by a factory, $X_k^{'}$ includes all the country level (k) explanatory variables that may be correlated with firm propensity to report financial information and ϵ_i is an error term. The novelty of this approach is that we can use our nightlight measure to proxy for firm size directly without having financial information for each firm.

We summarize our findings in columns (1)-(5) of Table 2, where we show what determines the likelihood of not reporting financial information. We consider each of the following, in turn: turnover, assets, profits, fixed assets, and employment. We find no relationship between firm size (proxied by nightlight) and the propensity to report financial information. Instead, we find that there is a large and positive correlation between corporate tax rates and the likelihood of not reporting financial information even after controlling for GDP per capita in each country. A 1% increase in the corporate tax rate decreases the likelihood of reporting financial information by approximately 20-30%. If firms want to minimize their tax bill and engage in tax avoidance practices, higher corporate tax rates may lead to lower publicly reported profits.

Further, we show that there is no consistent relationship between the level of country development, proxied by GDP per capita, and the likelihood of not reporting financial information. If a higher level of economic development provided more capacity to enforce transparency, we would expect the correlation between GDP per capita and missing financial data to be negative. In turn, in our data, we find that firms are less likely to report financial information for factories located in their home countries.[7]

Note that these results do not measure the amount of profits *hidden* from tax authorities, but the amount of profits not publicly reported. Hence, we document the difficulty that tax authorities face when they want to verify the information provided in tax statements using external data. As transparent documentation of firm activity is mandatory in almost all countries for large and publicly listed firms[8], we can interpret this lack of transparency for private subsidiaries, as an indicator of the existence of tax avoidance opportunities.

4.2 The scale of misallocated turnover

In this section, we use the nightlight data to predict firm activities and compare the predictions with the actual financial data. We use three sources for consolidated financial in-

[7]We visualize these results at the country level in Figure A3, where we map the share of reported profits relative to factory area for firms in our sample, aggregated at the country level.

[8]Most of the firms in our sample are located in countries that have implemented the BEPS Action 13 and are required to prepare country-by-country reports on the global allocation of their profits. Source: OECD

formation, (a) hand-collected consolidated turnover data for all car manufacturers for 2019, (b) manufacturing-related turnover for a subset of those firms, and (c) Orbis consolidated data for the manufacturing business segment for turnover, profits, and assets. For each subsidiary, we proxy the extent of the true economic activity by the total nightlight that the factory produces. We assume that the productivity of factory floor space is uniform across subsidiaries belonging to the same manufacturer.[9] We sum up the total nightlight that all factories belonging to the same car manufacturer produce and calculate the share of the nightlight of each factory. We multiply that by the consolidated firm turnover to obtain "*predicted turnover*". We also sum up unconsolidated turnover from financial statements across all subsidiaries that report this information. We summarize the results in Table 3.

First, the sum of reported turnover in the first two columns is higher than that of predicted turnover, which suggests that subsidiaries tend to overreport turnover. Given that the missing turnover result suggests firms are more likely to report turnover in lower-tax countries, this is consistent with the notion of shifting operations to low-tax countries to reduce the tax burden in high-tax countries. Second, we subtract the sum of all reported turnover from the consolidated number, to get 577 billion USD of missing turnover. This may be an underestimate since we know that MNCs may also *shift* turnover and profits between subsidiaries. Hence, we use the predicted turnover and subtract it from consolidated turnover, to find that 680 billion USD turnover has not been apportioned to subsidiaries of car manufacturing firms globally. This means that 38% of all turnover of car manufacturers is generated in subsidiaries that do not report their economic activity in their financial statements.

Third, for subsidiaries for which we have both the actual and predicted turnover, we calculate the aggregate size of the difference between the two to be 991 billion USD. This suggests that over 56% of the consolidated turnover of those MNCs is misallocated relative to where the economic activity takes place. These results are qualitatively comparable when we use the subset of turnover data or turnover from Orbis. For profits, we find a much larger fraction of naive missings relative to the fraction of those that are apportioned to any location according to our calculations. We show the opposite for assets. We explain these results by calculating the reporting bias, which we define as the difference between predicted and reported turnover, scaled by predicted turnover (or assets or profits). We

[9]For example, Bloom et al. (2019) show that the largest variation in management practices is attributed to the differences between firms, rather than across establishments within firms. Similarly, Bilicka and Scur (2021) demonstrate small, mostly within 0.5 point, variation of management practices within manufacturing multinationals.

plot the distribution of this bias across the three measures of firm operations in Figure A4. Positive bias suggests that subsidiaries in our sample report less than our nightlight data predict. Negative bias suggests that they report more. We find a large and negative bias for total assets. Firms have very little physical activity by our measures, but report to have a large amount of assets. With firms using intangible assets to shift profits, this may be evidence of profit shifting Dischinger and Riedel (2011).[10]

We complement these findings by showing correlations between calculated bias and country-level characteristics. We summarize these results in columns (6)-(8) of Table 2, where we show little systematic correlation between the calculated bias and country-level characteristics. GDP per capita and tax rates do not determine the size of the bias. Hence, our results suggest that firms may choose to strategically not report the financial information rather than report the amounts that are not consistent with their actual operations. We also find no correlation between home and reporting biases. Again, this means that firms may choose not to report in their home countries, rather than misreport.

Summary These results shed light on a systematic misreporting of turnover and profits of subsidiaries that belong to car manufacturing MNCs. We find that firm turnover is not reported where the economic activity occurs for about 50% of that activity. Combined with evidence that when faced with higher tax rates, firms tend not to report any financial information at all, this suggests that we do not know much about where MNC operations occur when using financial subsidiary-level data.

5 Economic application: Industry nightlights

5.1 Land use or nightlight intensity?

Previous literature has shown that nightlight emitted by regions or countries is a good proxy for overall regional productivity. Conceptually, there are two possible sources of change in nightlight emissions. The first is the increase in the intensity of light per pixel; the second is the increase in the share of pixels with positive light emission in the observed area. Light is a good proxy for economic activity because both of these sources are good indicators of economic development. Higher land use (extensive margin) and more intensive land use (intensive margin) are commonly associated with higher productivity. The distinction between

[10]We do not have consolidated financial information for fixed assets, intangible assets and employment for manufacturing business segment to be able to distinguish further.

these two sources is key in understanding what drives changes in GDP and the literature so far has not been able to separate the two.

Using our firm-level data, we measure the land occupied by factories *and* the intensity of light produced by that given area. Previous literature simply used nightlight intensity within an area irrespective of whether the land was occupied or not.[11] As such, we estimate the following model, which closely follows equation (2):

$$ln(output)_i = \alpha + \beta_1 \times ln(area)_i + \beta_2 \times ln(light)_i + \delta \times X_i^{'} + \epsilon \tag{4}$$

where $ln(area)_i$ is the logarithm of the area that a factory occupies and $ln(light)_i$ is the logarithm of the average light emitted by a factory. $ln(output)_i$ is the logarithm of firm turnover. $X_i^{'}$ is a set of factory-level control variables. We summarize the results in Table 4, where column (1) reports results using cross-sectional data and column (2) uses panel data. We find that both land use and land intensity are similarly good predictors of differences in turnover between firms in a cross section.[12] Using panel data, we find that changes in the area occupied by factories predict changes in turnover to a much larger extent than light intensity does. The magnitude of the coefficient suggests that a 1% increase in the area occupied by a factory increases turnover by 0.8%. This is similar in size to both coefficient estimates from the cross-sectional regression. In turn, we show that a 1% increase in light intensity increases turnover only by 0.2%.

These findings are in line with our baseline results in which we show that changes in light are a good predictor of production activities, but explain a small variation in those activities over time. The major part of light emissions is linked with land that our factories occupy rather than production activities. Given that the car manufacturing industry is very homogeneous, the infrastructure is highly correlated with land use. As such, these results stand in sharp contrast to the previous literature studying growth within urban agglomerations (e.g Bluhm and Krause (2018). These studies typically find that nightlight intensity explains differences in *urban* productivity extremely well, even when keeping land consumption constant. Our factories belong to a particular type of industry, hence, in the next section, we use industry-level data, to reconcile our findings.

[11] Note that the land occupied by a region is typically fixed over time. Consequently, with the traditionally used log-log estimates and panel data, findings based on nightlight intensity are conceptually identical to estimates based on the sum of light.

[12] Note that the standard error for the light coefficient estimate is larger.

5.2 Are industry lights different?

To generalize our firm-level findings to industrial production, we use Global Human Settlement European Settlement Map. This data classifies the land use in Europe using a 10-by-10 meter resolution for the year 2015. The land is classified into residential or non-residential buildup. We assume that non-residential buildup is a proxy for industrial production areas and measure the area and the nightlight emitted by this area for the municipalities (NUTS3 regions) in Europe. We use these two to proxy for land use and the intensity of land use, similar to the firm-level data. To reduce the noise in the light date we restrict the sample to regions with their center below a latitude of 60° North. To measure regional industry production, we use regional GDP data from Eurostat.

We follow the same framework as in the firm-level regressions and estimate an equivalent of equation (4). Now, $ln(output)_i$ is the logarithm of the total industry output; $ln(light\_sum_i)$ is the logarithm of the total light emitted by a regional industry; $ln(area)_i$ is the logarithm of the area that the regional industry occupies and $ln(light)_i$ is the logarithm of the average light emitted by a regional industry divided by the area it occupies.

We summarize the results in columns (3) and (4) of Table 4. First, in column (3) we show that the sum of light emitted over an area is significantly correlated with the GDP generated by this area. Specifically, a 1% difference in the sum of light emitted over areas with non-residential buildup between regions is associated with a 0.5% difference in regional industry-related GDP. In column (4), we split the total nightlight into light intensity and land use. Again, we find that area is more strongly correlated with industry output than the intensity of nightlight is. These findings are consistent with our firm-level findings. When we consider the residential land use and the intensity of nightlight over residential areas in columns (5) and (6), we find results consistent with the previous literature: both land use and intensity of nightlight contribute similarly to explaining the differences in output.

Summary Our findings suggest that light emitted by industrial production mainly carries information on land use. This might explain differences between the predictive abilities of nightlight data in urban and rural areas, documented by previous literature. If areas systematically differ in industry presence, we would expect that light explains regional variation differently between the two types of regions. As such, industry lights do not contain the same information as urban lights.

6 Conclusion

Our results offer a new dataset and a methodology that can be used to track firm-level activities more consistently, especially in places where information on such activities does not exist or is not systematically reported. We provide a nuanced perspective on what nightlight data is able to measure and how we should go about using it to quantify the growth of smaller geographical units. We leave collecting such data for other industries with a large ground presence for future research. However, we acknowledge that such data collected on a larger scale would allow researchers to understand the activities of large firms that operate across different countries with different reporting standards. Further, this data offers a unique ability to measure the allocation of profits and assets across firms with international presence and to quantify the scale of global misreporting. This may be an extremely useful source of information in the context of understanding the implications of global proposals for minimum taxes that rely on our ability to measure the scale and location of firm activities accurately.

References

Alesina, A., Michalopoulos, S. and Papaioannou, E. (2016). Ethnic inequality, *Journal of Political Economy* **124**(2): 428–488.

Berman, N., Couttenier, M., Rohner, D. and Thoenig, M. (2017). This mine is mine! how minerals fuel conflicts in africa, *American Economic Review* **107**(6): 1564–1610.

Bilicka, K. A. (2019). Comparing UK Tax Returns of Foreign Multinationals to Matched Domestic Firms, *American Economic Review* **109**(8): 2921–2953.

Bilicka, K. A., Qi, Y. and Xing, J. (2021). Real Responses to Anti-Tax Avoidance: Evidence from the UK Worldwide Debt Cap, *CESifo Working Paper Series 9044*, CESifo.

Bilicka, K. A. and Scur, D. (2021). Organizational capacity and profit shifting, *Working Paper 29225*, National Bureau of Economic Research.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I. and Van Reenen, J. (2019). What drives differences in management practices?, *American Economic Review* **109**(5): 1648–83.

Bluhm, R. and Krause, M. (2018). Top lights-bright cities and their contribution to economic development.

Bustamante-Calabria, M., Sánchez de Miguel, A., Martín-Ruiz, S., Ortiz, J.-L., Vílchez, J. M., Pelegrina, A., García, A., Zamorano, J., Bennie, J. and Gaston, K. J. (2021). Effects of the covid-19 lockdown on urban light emissions: ground and satellite comparison, *Remote Sensing* **13**(2): 258.

Coesfeld, J., Anderson, S. J., Baugh, K., Elvidge, C. D., Schernthanner, H. and Kyba, C. C. M. (2018). Variation of individual location radiance in viirs dnb monthly composite images, *Remote Sensing* **10**(12).

Coppola, A., Maggiori, M., Neiman, B. and Schreger, J. (2021). Redrawing the map of global capital flows: The role of cross-border financing and tax havens, *The Quarterly Journal of Economics* **136**(3): 1499–1556.

Desai, M. A., Foley, C. F. and Hines, J. J. (2006). The demand for tax haven operations, *Journal of Public Economics* **90**(3): 513–531.

Dischinger, M. and Riedel, N. (2011). Corporate taxes and the location of intangible assets within multinational firms, *Journal of Public Economics* **95**(7-8): 691–707.

Donaldson, D. and Storeygard, A. (2016). The view from above: Applications of satellite data in economics, *Journal of Economic Perspectives* **30**(4): 171–98.

Elvidge, C. D., Ghosh, T., Hsu, F.-C., Zhizhin, M. and Bazilian, M. (2020). The dimming of lights in china during the covid-19 pandemic, *Remote Sensing* **12**(17): 2851.

Elvidge, C. D., Hsu, F.-C., Zhizhin, M., Ghosh, T., Taneja, J. and Bazilian, M. (2020). Indicators of electric power instability from satellite observed nighttime lights, *Remote Sensing* **12**(19): 3194.

Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C. and Taneja, J. (2021). Annual time series of global viirs nighttime lights derived from monthly averages: 2012 to 2019, *Remote Sensing* **13**(5): 922.

Fabian, M., Lessmann, C. and Sofke, T. (2019). Natural disasters and regional development–the case of earthquakes, *Environment and Development Economics* **24**(5): 479–505.

Gennaioli, N., La Porta, R., Lopez-de Silanes, F. and Shleifer, A. (2013). Human capital and regional development, *The Quarterly journal of economics* **128**(1): 105–164.

Ghosh, T., Elvidge, C. D., Hsu, F.-C., Zhizhin, M. and Bazilian, M. (2020). The dimming of lights in india during the covid-19 pandemic, *Remote Sensing* **12**(20).

Gibson, J. (2020). Better night lights data, for longer, *Oxford Bulletin of Economics and Statistics* .

Gibson, J., Olivia, S., Boe-Gibson, G. and Li, C. (2021). Which night lights data should we use in economics, and where?, *Journal of Development Economics* **149**: 102602.

Gumpert, A., Hines, J. R. and Schnitzer, M. (2016). Multinational firms and tax havens, *The Review of Economics and Statistics* **98**(4): 713–727.

Henderson, J. V., Storeygard, A. and Deichmann, U. (2017). Has climate change driven urbanization in africa?, *Journal of development economics* **124**: 60–82.

Henderson, J. V., Storeygard, A. and Weil, D. N. (2012). Measuring economic growth from outer space, *American economic review* **102**(2): 994–1028.

Hines, J. R. and Rice, E. M. (1994). Fiscal Paradise: Foreign Tax Havens and American Business, *The Quarterly Journal of Economics* **109**(1): 149–182.

Hirte, G., Lessmann, C. and Seidel, A. (2020). International trade, geographic heterogeneity and interregional inequality, *European Economic Review* **127**: 103427.

Hodler, R. and Raschky, P. A. (2014). Regional favoritism, *The Quarterly Journal of Economics* **129**(2): 995–1033.

Lessmann, C. and Seidel, A. (2017). Regional inequality, convergence, and its determinants–a view from outer space, *European Economic Review* **92**: 110–132.

Levin, N., Kyba, C. C., Zhang, Q., de Miguel, A. S., Román, M. O., Li, X., Portnov, B. A., Molthan, A. L., Jechow, A., Miller, S. D. et al. (2020). Remote sensing of night lights: A review and an outlook for the future, *Remote Sensing of Environment* **237**: 111443.

Li, X., Li, D., Xu, H. and Wu, C. (2017). Intercalibration between dmsp/ols and viirs night-time light images to evaluate city light dynamics of syria's major human settlement during syrian civil war, *International Journal of Remote Sensing* **38**(21): 5934–5951.

Liu, Q., Sha, D., Liu, W., Houser, P., Zhang, L., Hou, R., Lan, H., Flynn, C., Lu, M., Hu, T. et al. (2020). Spatiotemporal patterns of covid-19 impact on human activities and environment in mainland china using nighttime light and air quality data, *Remote Sensing* **12**(10): 1576.

Martinez, L. R. (2019). How much should we trust the dictator's gdp growth estimates?, *Available at SSRN 3093296* .

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

Michalopoulos, S. and Papaioannou, E. (2014). National institutions and subnational development in africa, *The Quarterly journal of economics* **129**(1): 151–213.

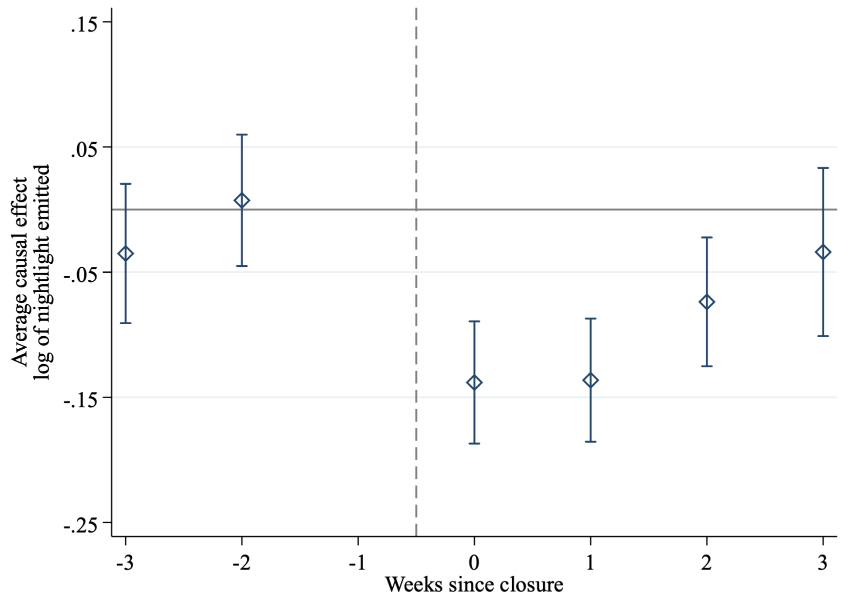
Mohan, P. and Strobl, E. (2017). The short-term economic impact of tropical cyclone pam: An analysis using viirs nightlight satellite imagery, *International journal of remote sensing* **38**(21): 5992–6006.

Straka, W., Kondragunta, S., Wei, Z., Zhang, H., Miller, S. D. and Watts, A. (2021). Examining the economic and environmental impacts of covid-19 using earth observation data, *Remote Sensing* **13**(1): 5.

Suárez Serrato, J. C. (2018). Unintended Consequences of Eliminating Tax Havens, *NBER Working Papers 24850*, National Bureau of Economic Research, Inc.

Tørsløv, T., Wier, L. and Zucman, G. (2020). The Missing Profits of Nations, *National Bureau of Economic Research Working Paper* (24071).

Figure 1: Dynamic plot: the effect of factory closures on nightlight emitted.



Note: In this figure, we plot the coefficient estimates for the effect of Covid-19 closure on light emitted by factories using a dynamic specification. Dots represent coefficient estimates, δ_w, from a regression that takes the form $ln(light\_sum_{i,w}) = \alpha + \sum_{\kappa=-4}^{4} \gamma_w 1[w = \kappa] + \sum_{\kappa=-5}^{5} \delta_w treated_i \times 1[w = \kappa] + \delta \times X_{iw}^{'} + \psi_i + \mu_j + \epsilon_{i,w}$, where, i is a factory, w is weeks. $ln(light\_sum_{i,w})$ is light emitted by a factory in each week; $\sum_{\kappa=-4}^{4} 1[w = \kappa]$ is a series of week dummies that equal one in each of the κ weeks away from the closure date, with the dummy variable corresponding to $\kappa = -1$ as the omitted category. $treated_i$ is a dummy variable that equals 1 in 2020 and zero in 2019. $X_{iw}^{'}$ is a set of factory-level control variables, such as general weather conditions around the factory; ψ_i is the factory-specific fixed effect, μ_j are weekend fixed effects and $\epsilon_{i,w}$ is the error term. δ_w coefficients, plotted as hollow diamonds here, represent the difference in nightlight emitted in each week relative to the closure week, relative to year 2019. The vertical lines represent the 95% confidence intervals. Each specification includes factory fixed effects.

Table 1: Correlations between firm activities and nightlight.

Panel A: Cross sectional data					
	(1) ln(turnover)	(2) ln(assets)	(3) ln(employment)	(4) ln(fixed assets)	(5) log(plbt)
$ln(light\_sum_i)$	0.796***	0.708***	0.643***	0.742***	0.745***
	(0.085)	(0.052)	(0.069)	(0.127)	(0.087)
Observations	316	267	276	271	204
R^2	0.720	0.783	0.743	0.720	0.722
Panel B: Panel data					
$ln(light\_sum_{i,t})$	0.284**	0.169***	0.290*	0.162**	0.310*
	(0.143)	(0.064)	(0.162)	(0.071)	(0.166)
Observations	1389	1314	941	1309	987
R^2	0.955	0.984	0.950	0.973	0.900
R^2 within	0.0866	0.139	0.0627	0.0473	0.0527

Note: The table presents the correlation between firm activities and nightlight emissions. In Panel A, we show results for collapsed cross-sectional averages for 2013-2018 and include country and parent fixed effects. In Panel B, we show results using annual data for 2013-2018 and include year and firm fixed effects. The dependent variable is the logarithm of turnover in column (1), the logarithm of total assets in column (2), the logarithm of the number of employees in column (3), the logarithm of fixed assets in column (4) and the logarithm of profit and loss before taxes in column (5). $ln(light\_sum_i)$ is the logarithm of the total light emitted by a factory. These estimates are robust to keeping the sample size constant between specifications, Table A5. Robust standard errors are clustered at the ultimate owner and country level in Panel A and the firm level in Panel B. ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. R^2 within refers to the within-firm variation in Panel B.

Table 2: Country-level correlations - missing and biased accounting data.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Missing accounting data					Biased accounting data		
	miss turnover	miss assets	miss profits	miss fx assets	miss empl	bias turnover	bias assets	bias profits
ln(light_sum$_i$)	-0.001	-0.010	-0.003	-0.017	-0.013			
	(0.010)	(0.012)	(0.009)	(0.011)	(0.013)			
ln(GDP pc)	-0.003	0.022	0.025*	0.022	-0.105***	(0.043)	(0.377)	(0.043)
	(0.015)	(0.014)	(0.013)	(0.014)	(0.022)	(0.082)	(0.301)	(0.073)
ln(corptax)	0.077	0.216**	0.213**	0.213**	0.319***	-0.335	-1.124	-0.243
	(0.082)	(0.090)	(0.085)	(0.088)	(0.098)	(0.238)	(1.135)	(0.256)
home country=1	0.528***	0.603***	0.589***	0.589***	0.240***	0.034	-0.608	-0.202*
	(0.035)	(0.045)	(0.040)	(0.044)	(0.045)	(0.103)	(0.386)	(0.106)
ln(light surrounding 5k)						0.029	0.571	-0.024
						(0.043)	(0.377)	(0.043)
Observations	475	475	475	475	475	244	140	186
R^2	0.326	0.427	0.411	0.407	0.151	0.038	0.131	0.047

Note: In columns 1-5 of this table, we report the regression results using as a dependent variable a dummy equal to 1 when turnover, assets, profits, fixed assets, and employment are missing. In columns 6-8 of this table, we report the regression results using the size of bias in accounting data relative to nightlight data as an outcome variable. Bias turnover is bias using turnover that is calculated as predicted turnover minus actual turnover divided by predicted turnover, where predicted turnover is calculated using consolidated car manufacturing firm turnover apportioned by a proportion of nightlight emitted by each factory in all nightlight emitted by all factories of that car manufacturer. Similar calculations are done for total assets and profits and loss before taxes. $ln(light\_sum_i)$ is the total nightlight emitted by each factory, ln(light surrounding 5k) is the total nightlight emitted by a 5km radius area around the factory, excluding the factory itself, ln(GDP pc) is the natural logarithm of GDP per capita, ln(corptax) is the natural logarithm of the statutory corporate tax rate, home country=1 is a dummy equal to 1 when a factory is located in a country where firm headquarters are. Robust standard errors clustered at the parent level are in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 3: Summary of misallocated turnover, profits and assets.

	turnover hand collected	turnover subset	turnover	profits Orbis	assets
total	1.77	1.46	2.96	0.172	0.877
sum of reported	1.19	0.896	0.922	0.0432	0.775
sum of predicted	1.09	0.837	1.63	0.0727	0.348
naive missing	0.58	0.564	2.038	0.1288	0.102
(total - reported)					
as %	33%	39%	69%	75%	12%
not apportioned	0.68	0.623	1.33	0.0993	0.529
(total - predicted)					
as %	38%	43%	45%	58%	60%
total missallocated	0.991	0.735	1.29	0.0764	0.777
(predicted - reported)					
as %	56%	50%	44%	44%	89%

Note: This table summarizes the calculations for misallocated and not apportioned turnover. Hand-collected information comes from 2019. The turnover subset is a subset of the hand-collected data for which we have information for only manufacturing operations. Turnover, profits, and assets in columns 3-5 come from Orbis business line items that made specific reference to the automobile or manufacturing industry, but not to finance or services. These are an average of 2005 - 2020 data. Total is consolidated turnover, profits, or assets, the sum of reported adds up all unconsolidated reported turnover, while the sum of predicted adds up all turnover apportioned from consolidated turnover and weighted according to the nightlight emitted by each factory.

Table 4: Economic activity: comparison between intensity of nightlight and land use.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep.Var.:	ln(turnover)		ln(industry GDP)		ln(non industry GDP)	
Unit of Obs.	firms		EU NUTS3 regions			
Time:	means 12-19	panel 12-19	mean 2015			
ln(light_sum)			0.536***		0.850***	
			(0.036)		(0.029)	
ln(light)	0.796***	0.195*		0.184*		0.825***
	(0.154)	(0.114)		(0.107)		(0.040)
ln(area)	0.811***	0.777**		0.675***		0.875***
	(0.099)	(0.341)		(0.036)		(0.036)
Country FE	✓		✓	✓	✓	✓
Parent FE	✓		✓	✓	✓	✓
Firm FE		✓				
Year FE		✓				
Observations	316	1402	1409	1409	1414	1414
R^2	0.723	0.956	0.702	0.731	0.866	0.866
R^2 within	0.694	0.105	0.574	0.616	0.766	0.767

Note: This table shows the correlation between production, land use, and land-use intensity. The dependent variable in columns (1) and (2) is the log of total turnover at the firm level, in columns (3) and (4) the industry production, and in columns (5) and (6) the non-industry production of NUTS3 regions in Europe. The area is defined as a factory area in columns (1) and (2), a regional area with non-residential buildup in columns (3) and (4), and areas with residential buildup in columns (5) and (6). ln(light_sum) is the total nightlight emitted by these areas and ln(light) is the average light. Column (1) presents cross-sectional results on the collapsed dataset for 2013-2018. Column (2) presents panel estimates for 2013-2018. Columns (3) - (6) present regional cross-sectional results for 2015. Robust standard errors clustered at the ultimate owner and country level in column (1), at the firm level in column (2), and at the regional level in columns (3) - (6) are reported in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. R^2 within refers to the within ultimate owner variation in Column (1), within-firm variation in Column (2), and within NUTS3 region variation in Column (3) - (6).

Appendices

A Data cleaning and selection procedures

A.1 Factory land consumption

Factory footprints were tagged by research assistants based on a set of predefined rules. These include, that only buildings were tagged, parking lots and racetracks were excluded and the fence around the production site was used to mark the factory line. We did not tag subsidiaries that were on the same site but there were two distinct subsidiaries in Orbis. This means that we aggregate the financial data for those subsidiaries to match with the light data. If there was a joint venture, we copied the same polygon for both firms. We apportion depending on the joint ownership share or drop joint ventures.

The data were collected in two waves. First, factory footprints were tagged based on the 2019 images by multiple assistants, and the data was cross-validated using a subset of data collected by all the researchers. In the second wave, the change in footprints since 2012 was extracted by one research assistant. During this second wave, a second quality check occurred, where 10% of firms with low-quality tagging were excluded. There were two main reasons for this sample selection: first, the resolution of daylight images was not high enough to credibly identify the location of factories; second, we did not find another data source to cross-validate the information collected. For the latter reason, factories in China are more likely to be classified as unreliable, as, for example, Google street view and maps or OpenStreetMaps data were not available or company websites were inaccessible.

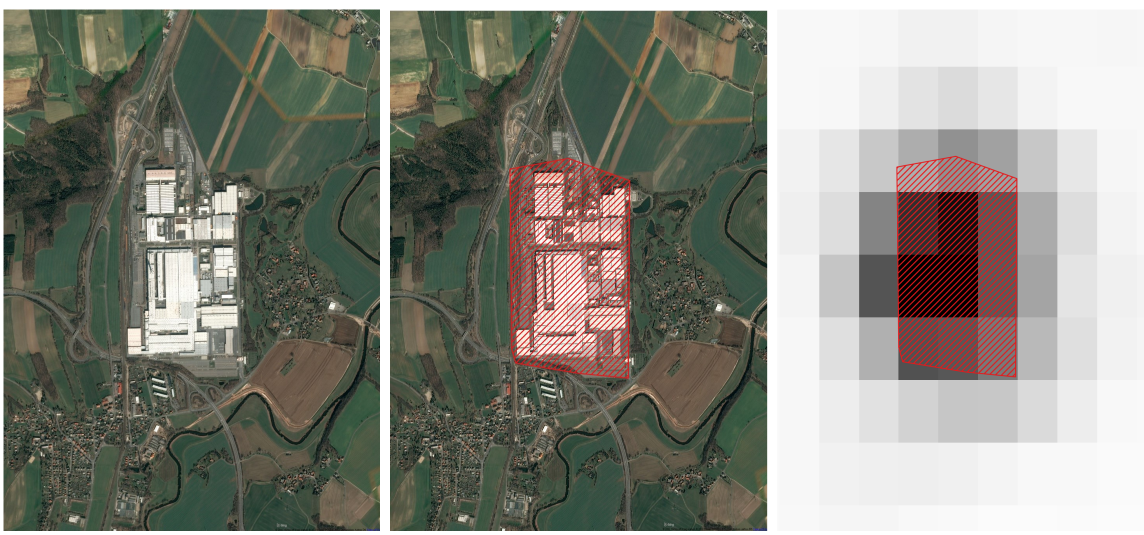
A.2 Nightlight data

We collect nightlight data from the Visible and Infrared Imaging Suite (VIIRS) Day-Night Band (DNB) that sits on board of satellites of the Joint Polar-orbiting Satellite System provided by the Earth Observation Group (EOG). We collect daily data from the raw Nightly DNB Mosaic and Cloud imagery and the annual data comes from the Annual VNL V2. For both sources, the resolution of the data is 15 arc seconds, which corresponds to an area of approximately 500m x 500m at the Equator. This data measures light emitted by these areas in watt per steradian per square meter, $nW/cm^2/sr$.

For the annual data, we use the cleaning procedure proposed by Elvidge et al. (2021). As such, we filter the data to remove sunlit, moonlit, and cloudy pixels. Further, we remove outliers to discard biomass burning pixels and isolate the background. For daily data, we have to rely on the raw data, but using the VIIRS Cloud Mask we identify cloud-free observations. We calculate light emitted by factories on cloud-free days only and we also exclude days when the sky was cloudy over the 5km radius around factories. Further, studies have shown that even the cleaned monthly VIIRS data is relatively volatile. This can occur due to, for example, cyclically local phenomena such as whether or moonlight (Coesfeld et al.; 2018). To account for this, we either control for (in cross-sectional estimates), or estimate (in daily estimates) the difference to the light emitted around a factory.

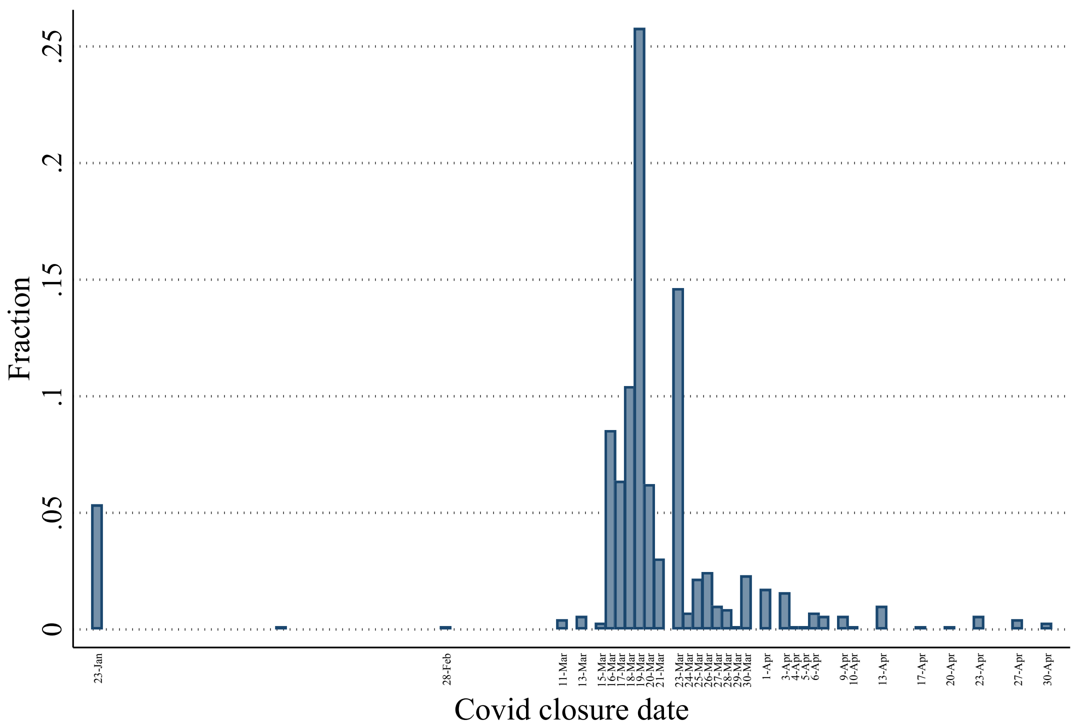
B Supplementary Tables and Figures

Figure A1: Factory day and nightlight images: example.



Note: This figure summarizes the tagging process. The first image is a daylight image from bing.com (© 2020 Microsoft). The second image shows how each factory is tagged, the last image is the nightlight produced by that factory that comes from VIIRS monthly average for 2014 in this example.

Figure A2: Distribution of Covid closing dates.



Note: This figure plots the distribution of Covid-19 closing dates for 2020. The dataset was hand-collected using marklines.com portal followed by news updates provided by just-auto.com. We only collected information for the first Covid-19 related closure, not the subsequent supply-related closures.

Table A1: Descriptive statistics: plant level data.

variable	mean	median	sd	min	max
ln(light sum)	3.19	3.31	1.89	-3.38	13.59
ln(light mean)	3.49	3.55	0.87	-1.61	13.09
ln(light 5km ring)	2.45	2.46	1.05	-3.06	13.14
area	0.28	0.14	0.39	0.00	5.95
Cloud cover	0.03	0.00	0.11	0.00	1.72

Note: This table provides summary statistics related to plant-level nightlight data. ln(light sum) is the logarithm of the total light emitted by a factory. ln(light mean) is the logarithm of the light emitted by a factory divided by the factory area. ln(light 5km ring) measures the mean light within a 5 km radius around the production sites excluding factory lights. Cloud cover is the percentage of cloud coverage within a 10 km radius. Those nightlight-related variables come from the VIIRS dataset. The area is a factory area in square kilometers, measured by the footprint of the factory buildings.

Table A2: Descriptive statistics: financial and lights data.

variable	mean	median	sd	min	max
Panel A: Panel Data					
ln(light sum)	3.43	3.58	1.85	-2.03	8.03
ln(light mean)	3.33	3.39	0.73	-1.93	5.13
ln(light 5km ring)	2.16	2.29	1.03	-5.67	4.46
area	0.47	0.21	0.85	0.00	7.33
ln(turnover)	13.33	13.31	2.12	7.91	18.42
ln(total assets)	12.99	12.90	1.96	9.05	18.87
ln(employment)	7.28	7.24	2.00	0	12.11
ln(fixed assets)	12.02	11.89	2.35	6.54	18.73
ln(profit and loss before tax)	10.27	10.13	2.29	4.96	15.85
Panel B: Cross sectional Data					
ln(light sum)	3.39	3.60	1.78	-1.10	7.96
ln(light mean)	3.34	3.38	0.69	0.90	4.96
ln(light 5km ring)	2.16	2.29	0.98	-1.22	4.41
area	0.47	0.21	0.84	0.00	7.32
ln(turnover)	12.99	13.04	2.29	6.33	18.04
ln(total assets)	12.92	12.84	1.87	9.10	17.93
ln(employment)	6.99	6.91	1.87	1.10	11.28
ln(fixed assets)	11.88	11.84	2.37	5.70	18.45
ln(profit and loss before tax)	10.09	10.07	2.27	4.93	15.63

Note: This table provides summary statistics related to firm-level data. In Panel A, we show nightlight and financial data summary for the panel data sample we use and in panel B, we show data summary for the cross-sectional data sample we use. ln(light sum) is the logarithm of the total light emitted by a factory. ln(light mean) is the logarithm of the light emitted by a factory divided by the factory area. ln(light 5km ring) measures the mean light within a 5 km radius around the production sites excluding factory lights. Those nightlight-related variables come from VIIRS dataset. The area is a factory area in square kilometers, measured by the footprint of the factory buildings. Financial data related to turnover, total assets, employment, fixed assets and profit and loss before taxes comes from Orbis Bureau van Dijk. All these financial variables are reported at the unconsolidated subsidiary level. When a subsidiary has more than one factory, we aggregate nightlight information for that subsidiary across all factories that it owns.

Table A3: Nightlight and firm activity: Covid-19 production shock.

	(1) log(lights)	(2) log(lights)	(3) log(lights)
covid close=1	0.006	-0.001	
	(0.013)	(0.008)	
year 2020=1	-0.049***	-0.021**	
	(0.010)	(0.008)	
covid close=1 × year 2020=1	-0.080***	-0.046***	
	(0.014)	(0.010)	
week=-3			-0.035
			(0.028)
week=-2			0.007
			(0.027)
week=0			-0.138***
			(0.025)
week=1			-0.136***
			(0.025)
week=2			-0.074***
			(0.026)
week=3			-0.034
			(0.034)
Cloud cover	✓	✓	✓
ln(light 5km ring)		✓	✓
Observations	57582	57326	11879
R^2	0.614	0.837	0.817
R^2 within	0.00954	0.584	0.0112

Note: This table reports coefficient estimates from regressions looking at the effects of Covid-19 production shock on factory nightlight emissions. The unit of observation in all columns is a factory and the dependent variable is the mean of light emitted by each factory. Columns 1 and 2 show results using covid close dummy, which is equal to one after a factory closed down due to the Covid pandemic. Column 3 presents the difference in difference coefficient estimates for each week, as plotted in Figure 1. All estimates include firm, week, and weekend fixed effects. The sample is restricted to observations with no cloud coverage in a 5km radius and observations between the first of January and the official time of reopening of factories. ln(light 5km ring) measures the mean light within a 5 km radius around the production sites excluding the factory lights. All estimates control for the percentage of cloud coverage within a 10 km radius (Cloud cover). Robust standard errors clustered at the firm level in parentheses. ***, **, * denote significance at the 1%, 5%, and 10% level, respectively. R^2 within refers to within-firm variation.

Table A4: Correlations between firm activities and nightlight: controlling for blooming.

	(1) ln(turnover)	(2) ln(assets)	(3) ln(employment)	(4) ln(fixed assets)	(5) log(plbt)
Panel A: Cross sectional data					
ln(light_sum$_i$)	0.792***	0.712***	0.644***	0.741***	0.752***
	(0.098)	(0.053)	(0.071)	(0.123)	(0.088)
ln(light 5km ring)	0.044	-0.037	-0.011	0.018	-0.066
	(0.131)	(0.095)	(0.142)	(0.162)	(0.190)
Observations	316	267	276	271	204
R^2	0.720	0.783	0.743	0.720	0.722
Panel B: Panel data					
ln(light_sum$_i$)	0.341*	0.244***	0.320	0.256**	0.401
	(0.195)	(0.091)	(0.205)	(0.102)	(0.246)
ln(light 5km ring)	-0.175	-0.233**	-0.100	-0.282**	-0.244
	(0.227)	(0.111)	(0.242)	(0.133)	(0.345)
Observations	1389	1314	941	1309	987
R^2	0.955	0.984	0.950	0.973	0.901
R^2 within	0.0880	0.147	0.0633	0.0526	0.0538

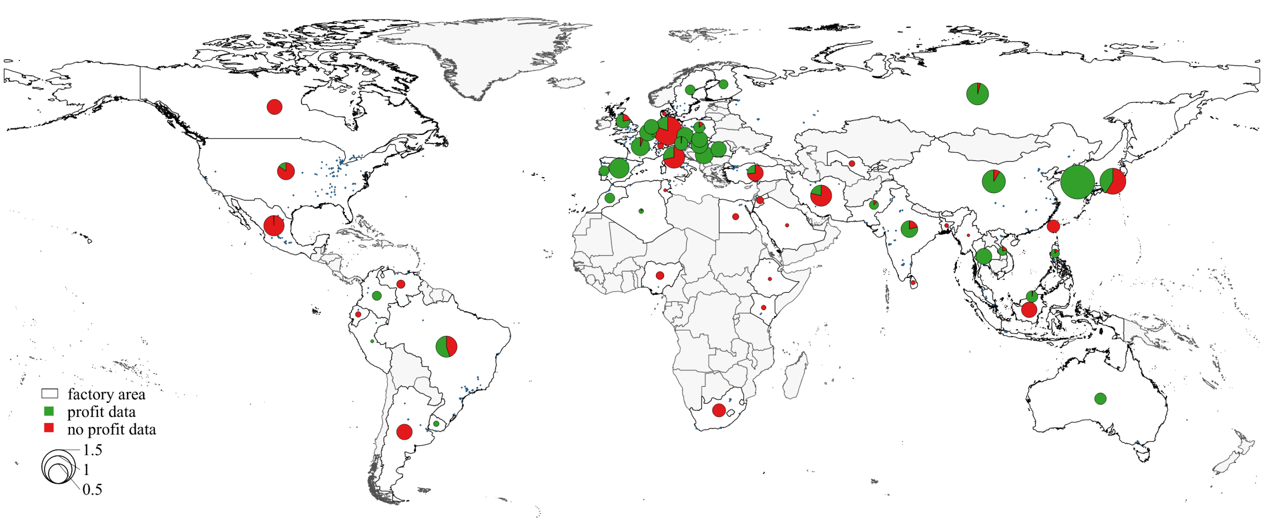
Note: This table presents the correlation between firm activities and nightlight emissions, including the average nightlight emissions within a 5km radius around the factory production site (ln(light 5km ring)) to control for the blooming effect. In Panel A, we show results for collapsed cross-sectional averages for 2013-2018 and include country and parent fixed effects. In Panel B, we show results using annual data for 2013-2018 and include year and firm fixed effects. The dependent variable is the logarithm of turnover in column (1), the logarithm of total assets in column (2), the logarithm of the number of employees in column (3), the logarithm of fixed assets in column (4) and the logarithm of profit and loss before taxes in column (5). $ln(light\_sum_i)$ is the logarithm of the total light emitted by a factory. Robust standard errors are clustered at the ultimate owner and country level in Panel A and the firm level in Panel B. ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. R^2 within refers to the within-firm variation in Panel B.

Table A5: Correlations between firm activities and nightlight: sample size robustness.

	(1) ln(assets)	(2) ln(employment)	(3) ln(assets)	(4) ln(employment)
ln(light_sum$_i$)	0.227**	0.346*	0.312**	0.404
	(0.106)	(0.200)	(0.139)	(0.255
ln(light 5km ring)			-0.274*	-0.187
			(0.160)	(0.266)
Observations	847	847	847	847
R^2	0.986	0.955	0.987	0.955
R^2 within	0.162	0.0725	0.172	0.0748

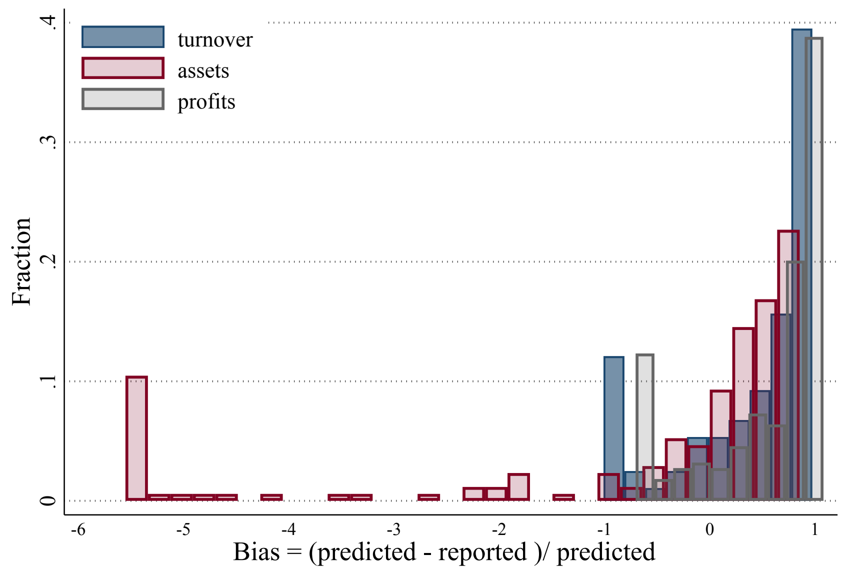
Note: The table presents the correlation between firm activities and nightlight emissions. The dependent variable is the logarithm of total assets in columns (1) and (3) and the logarithm of the number of employees in columns (2) and (4). We only show results using panel data for 2013-2018 and include year and firm fixed effects. $ln(light\_sum_i)$ is the logarithm of the total light emitted by a factory. In columns (3) and (4), we also include a control for the average nightlight emissions within a 5km radius around the factory production site (ln(light 5km ring)). Robust standard errors are clustered at the firm level. ***, **, * denote significance at the 1%, 5%, and 10% levels, respectively. R^2 within refers to the within-firm variation.

Figure A3: The share of publicly reported profits per km^2 of factory area



Note: This figure plots the country-level summary of the coverage of financial data relative to the nightlight data. Each circle size refers to the total factory area of car manufacturing plants in our sample. The larger the circle size, the larger the total factory area. In green, we have a fraction of the area for which we observe profits data in Orbis, in red, we have a fraction for which that information is missing.

Figure A4: Predicted vs reported turnover.



Note: This figure plots the distribution of the differences between apportioned and actual financial activities of firms. Bias is calculated as predicted turnover minus actual turnover divided by predicted turnover, where predicted turnover has been calculated using consolidated car manufacturing firm turnover apportioned by a proportion of nightlight emitted by each factory in all nightlight emitted by all factories of that car manufacturer. Similar calculations are done for total assets and profits and loss before taxes. For consolidated variables, we use the Orbis manufacturing segment turnover data averaged from 2009 to 2019.