Nightlights as measure of local development: the case of Italy Very preliminary version

Davide Fiaschi
University of Pisa

Angela Parenti
University of Pisa

Cristiano Ricci
University of Pisa

March 15, 2023

To respect the size limit in the submission, figures are available at
`https://poisson.phc.dm.unipi.it/~cricci/fiaschi_parenti_ricci_NLIGHT.pdf`

Abstract

In this paper we show that, while there exists a strong cross-sectional correlation between nightlights and personal income and population, the dynamics of personal income and population can only partially proxied by nightlights at municipal level in Italy in the period 2012-2019 due to the presence of a downward trend in the intensity of nightlights, whose slope is positively correlated with the intensity of nightlights in 2012. We also discuss how the aggregation at NUTS 3 and 2 level can (falsely) increase the capacity of nightlights to proxy for the local dynamics of personal income and population. Our findings point to a possible miss-use of nightlights for the study of local development, at least for developed countries.

JEL Classification Numbers: C23, R12, R15
Keywords: spatial economic agglomeration, spatial distribution of income and population, aggregation failure

Contents

1 Introduction

This paper explores the possibility to use for Italy the nightlights as proxy for economic activity and population at different levels of geographical aggregation (municipalities, NUTS 3 and NUTS 2 regions).

Starting from Nordhaus (2006), a recent literature proposes nightlights to "augment official income growth measures", also stressing the possibility to "measure growth for sub- and supranational region" (Henderson et al., 2012). In particular, Donaldson and Storeygard (2016) and Michalopoulos and Papaioannou (2018) contain a review of the increasing number of contributions in economics which use nightlights in order to have a proxy for local economic development. In another perspective, Martinez (2022) argue that nightlights can be use to check the trustworthiness of the official statistics of countries with weak institutions (in his case autocracies).

From the analysis of the relation between nightlights, personal income and population at municipal, NUTS 3 and NUTS 2 level in Italy in the period 2012-2019 we get good and bed news for the use of nightlights to study the economic development at local level. The good news is that nightlights appears a very good proxy for local population at NUTS 2 level, by explaining about 94% of total spatial distribution. At finer geographical level such capacity drop to 90% for NUTS 3 and to 86% for municipalities. As proxy for personal income nightlights maintain a substantial power, but lower than population, about 88% for NUTS2, 80% for NUTS 3 and 83% municipal level. The bad news, instead, regards the capacity of nightlights to proxy for the local dynamics of population and income. In particular, we find a convincing evidence that nightlights have a declining time trend for municipalities with initial high level of nightlights, which produces a decoupling between the dynamics of nightlights and personal income for these municipalities. These findings cast several doubts on the possibility to use nightlights to study local economic growth. We find a similar evidence also for the relationship between the dynamics of nightlights and population.

The paper is organized as follows: Section 2 describes the sources of data used in the analysis; Section 3 reports the maps of nightlights, personal income and population at municipal levels for driving the econometric analysis; Section 4 contains the analysis on the capacity of nightlights to proxy for local personal income and population; finally, Section 5 contains some concluding remarks.

2 Datasets used in the analysis

In the analysis we will use nightlights, personal income and population at municipal level, NUTS 2 and NUTS 3 regions. Nightlights are taken from VIIRS 2.1 database (`https://eogdata.mines.edu/products/vnl/`), which provides the average intensity of nightlight for cells of 500 x 500 meters, for a total of about 1.9 millions of cells for the Italian territory. Version 2.1 represents the state of art in terms of the quality of nightlights, fixing very important bugs and inconsistencies present in the previous versions (see, `https://eogdata.mines.edu/products/vnl/VNL_v21_readme_20220713.pptx`). Part of the discrepancies of our findings with respect to other contributes can be traced to the use of the improved version of VIRS (Gibson et al., 2020). In 2019, the cells with the highest intensive nightlights corresponds to Ilva of Taranto (a big industrial plant in the South), San Siro Stadium (in Milan), Malpensa airport (close to Milan), and, finally, the harbour of Salerno (close to Naples). The cells with zero nightlights are mainly concentrated in the mountain part of the country (see Figure 1).

Regional population for the 101 NUTS 3 and 20 NUTS 2 regions are taken from EUROSTAT, while total personal income from fiscal declarations for the 7923 Italian municipalities is taken from Agenzia delle Entrate (`https://www1.finanze.gov.it/finanze/pagina_dichiarazioni/public/dichiarazioni.php`. Personal income is particularly adapted to our scopes because it appears less biased by the phenomenon which reported residence is very different from the actual place where income is produced, as it happens for big companies whose profits are fully assigned to the locations where the company has its legal residence (for Italy, the most of big companies have legal residence in the North). The period of analysis is limited to 2012-2019 for the lack of available data at municipal level before 2012 and we exclude 2020 for the burst of COVID pandemic. The building of the dataset is completed by summing the nightlights of the cells corresponding to the different Italian municipalities, NUTS 3 and NUTS 2 regions, while personal income of NUTS 3 and 2 regions is calculated starting from municipal data.

3 Some stylized facts on nightlights, income and population at municipal level

	CV			cor			corΔ			corGR			$\#\Delta < 0$		
time	N	I	P	NI	NP	PI	NI	NP	PI	NI	NP	PI	N	I	P
2012	3.99	7.27	5.40	0.95	0.96	0.98									
2013	3.97	7.23	5.45	0.95	0.96	0.98	-0.05	-0.10	0.12	0.41	0.33	0.82	2900	1817	4151
2014	3.93	7.31	5.50	0.94	0.96	0.98	0.02	0.03	0.10	-0.35	-0.36	0.92	3555	3588	4646
2015	3.90	7.29	5.52	0.93	0.95	0.98	0.06	0.07	0.09	-0.68	-0.61	0.69	4317	1436	5105
2016	3.85	7.27	5.56	0.92	0.95	0.98	0.01	0.03	0.12	-0.34	-0.18	0.75	3622	2291	5589
2017	3.37	7.33	5.58	0.91	0.94	0.98	-0.09	-0.15	0.13	0.20	-0.64	0.36	324	5102	5363
2018	3.28	7.24	5.60	0.90	0.94	0.98	-0.03	0.10	0.08	-0.60	-0.14	0.51	5524	474	5445
2019	3.24	7.20	5.62	0.91	0.94	0.98	-0.02	-0.02	0.06	0.10	0.07	0.26	1933	2534	5439
2012-2019							0.01	-0.04	0.60	-0.63	-0.79	0.83	1497	997	5815

Note: N: Nightlights; I: Income; P: Population. Δ: absolute variation; GR: growth rate.

Figure 1 reports a comparison of the spatial patterns of nightlights, personal income and population in 2012 at municipal level. Each of the three variable is reports in their absolute value is because we are interested in their relationship. It is out of the scope of the paper the cross-sectional regional inequality, which would have instead required a normalization by different municipalities' areas. Hence, for example, the very deep blue/green for Apulia does not mean a high level of income per capita and/or population density but of a large and very populated municipality.

Figure 1 suggests a strong correlation between the absolute levels of nightlights, personal income and population and strong similarities in their spatial patterns, confirming the general findings of literature (Chen and Nordhaus, 2011).

figures/map_lights_comm_2012-eps-converted-to.pdf

figures/map_income_comm_2012-eps-converted-to.pdf

figures/map_pop_comm_2012-eps-conv

(a) Nightlights 2012

(b) Personal income 2012

(c) Population 2012

Figure 1: A comparison of the spatial patterns of nightlights, personal income and population in 2012 at municipal level

figures/map_Deltalights_comm-eps-converted-to.pdf

figures/map_Deltaincome_comm-eps-converted-to.pdf

figures/map_Deltapop_comm-eps-conv

(a) Δ Nightlights 2019-2012

(b) Δ Personal income 2019-2012

(c) Δ Population 2019-2012

Figure 2: The spatial patterns of the changes in nightlights, personal income and population between 2012 and 2019 at municipal level

However, when we look at the dynamics, the big urban agglomerations (Rome, Milan, Turin, Venice, etc.) show negative changes in their intensity of nightlights, and, on the contrary, positive changes, in their levels of personal income and population. The places with the strongest positive increases of nightlights show a spatial pattern and are located in specific areas of Italy, mainly in Emilia Romagna, Tuscany, Apulia, a part of Sicily, and a part in the Northern East regions. The places with negative change in nightlights, beyond the big urban agglomerations, are very spatially fragmented. Overall, population is agglomerating around some metropolitan areas as Milan, Verona, Padova, Bologna, Modena, Florence, Roma, Naples, Cagliari and Catania. Population changes are positively correlated to change in personal income in the most of North and Center, but is negatively correlated in the South. In numbers, the correlation between the changes in the period 2012-2019 of nightlights and income is negative and equal to about -0.6, between changes in nightlights and population negative and equal to about -0.8, while the correlation between changes in population and personal income is positive and about 0.8.

At country level Henderson et al. (2012) find radical different results. A possible explanation could be the wider units of observations (countries versus municipalities), which can cover a sort of spatial composition effect, given that also for Italy as country nightlights, personal income and population are positively correlated over time. However, the evidence at finer level should severely bias the analysis, or the interpretation of results (Henderson et al., 2018). Other contributions exploiting such relationship at more disaggregated level are unfortunately still more subject to such a possible bias (Donaldson and Storeygard, 2016).

4 The estimate of nightlights as proxy for economic activity

In this section, Section 4.1 contains the formulation of a theoretical model which can account for the main stylized facts discussed in Section 3 Then, Section 4.2 discusses a possible bias in the estimate from the geographical aggregation of units with spatial concentrated intensity. Finally, we separately estimate the model for personal income and population at different level of geographical aggregation in Sections 4.3 and 4.4 respectively.

4.1 From a data-driven economic model to the econometric model

A possible explanation for the spatial patterns observed in Figures 1 and 2 is that income has a positive long-run trend due to technological progress and the accumulation of other "material" factors,i.e:

$$y_{it} = k_{it} m_{it}^{\mu}, \tag{1}$$

where y_{it} is the income of region i at period t, k the regional technological progress, m is a measure of materials factors, such as population, plants, etc., and $\mu > 0$ the elasticity of income to material factors. Eq. (1) can account for both the positive correlation between income and population and for the higher growth rate of income with respect to population.

Instead, nightlights are growing for the material factors, but also present a tendency to decrease for environmental and economic motives, especially in the regions with high initial levels (Italian nightlights are very high compared to the rest of Western countries, see Cottarelli et al. 2022, Illuminazione pubblica: spendiamo troppo), i.e.:

$$n_{it} = n_{i0} \exp\left(g\left(n_{i0}\right) t\right) m_{it}^{\varphi}, \tag{2}$$

where n_{it} is the nightlights in region i at period t, $g\left(n_0\right)$ the exogenous (negative) growth rate of nightlights and is increasing in the initial level ($g' < 0$), and $\varphi > 0$ the elasticity of nightlights to material factors. Eq. (2) can account for both the correlation between nightlights and population and for the observed decrease in regions with the highest levels (see Figure 2).

Taking together Eqq. (1) and (2) we get the relationship between income and nightlights

$$y_{it} = n_{it}^{\mu/\varphi} k_{it}\, n_{i0}^{-\mu/\varphi} \exp\left(-\left(\mu/\varphi\right) g\left(n_{i0}\right) t\right), \tag{3}$$

which can account for the positive relationship between income and nightlights, but adds other factors, as the level of regional technological progress, the initial level of nightlights and the fact that such relationship is time varying.

The relationship between the change of nightlights and income is complex. From a dynamic perspective Eq. (3) leads to:

$$\frac{\dot{y}_{it}}{y_{it}} = \left(\frac{\mu}{\varphi}\right) \frac{\dot{n}_{it}}{n_{it}} + \frac{\dot{k}_{it}}{k_{it}} - \left(\frac{\mu}{\varphi}\right) g\left(n_{i0}\right) \tag{4}$$

Eq. (4) can account for three empirical facts: i) the positive relationship between the growth of nightlights and income (μ/φ measure the intensity); ii) income is growing at higher rate than nightlight thanks to exogenous technological progress ($\dot{k}_{it}/k_{it} > 0$); and iii) the big urban agglomerations are the regions where the decoupling between the income dynamics and nightlights dynamics is the most likely since they are the regions with the highest exogenous reduction in nightlights given their relative high initial level of nightlights (n_{i0} is higher in urban agglomerations and $g' < 0$).

The baseline econometric model is derived by (3), i.e.:

$$\log\left(y_{it}\right) = \alpha_t + \beta_t \log\left(n_{it}\right) + \gamma X_{it} + \epsilon_{it}, \tag{5}$$

where y_{it} is the variable to be explained, i.e. the income or population of region i at period t; n_{it} is the intensity of nightlights; X_{it} is a vector containing some regional characteristics, such as geographic and institutional specificity; β_t is the elasticity between income (population) and the intensity of nightlights at regional level, which is expected about 1; α_t is a time-varying parameter which is expected to decrease over time reflecting the downward trend in nightlights ; ϵ_{it} a random term, which represents a measurement error.

We will estimate the model for different years to test the time stability of estimated coefficients and for different geographical levels, i.e. NUTS 2, 3 and municipalities, to test the scale stability of estimated coefficients. In the estimate we do not consider any X_{it} because we are interested to the "pure" elasticity of income/population to nightlights.

4.2 The spatial aggregation of baseline econometric model

Since we provide estimation at different level of geographical aggregation is interesting to explore theoretically the implications for the estimates of possible bias induced by specifi spatial pattern.

Given the baseline econometric model in Eq. (5) and ignoring the contribution of X_{it}, i.e.:

$$y_{it} = \exp(\alpha_t) n_{it}^{\beta_t} \exp(\epsilon_{it}) . \tag{6}$$

Suppose that the total number of regions can be partitioned in M macro regions, where m_j is the number of regions in macro region j and the set of region in j. Therefore, the income of macro region j Y_{jt} is defined as:

$$Y_{jt} = \exp(\alpha_t) \sum_{i \in m_j} n_{it}^{\beta_t} \exp(\epsilon_{it}) , \tag{7}$$

i.e.

$$Y_{jt} = \exp(\alpha_t) N_{jt}^{\beta_t} \sum_{i \in m_j} a_{ijt}^{\beta_t} \exp(\epsilon_{it}) , \tag{8}$$

where $n_{it} \equiv a_{ijt} N_{jt}$, $\sum_{i \in m_j} a_{ijt} = 1$ and $\sum_{i \in m_j} n_{it} = N_{jt}$. Taking:

$$Y_{jt} = \exp(\alpha_t) N_{jt}^{\beta_t} m_j \left(\frac{1}{m_j} \right) \sum_{i \in m_j} a_{ijt}^{\beta_t} \exp(\epsilon_{it}) , \tag{9}$$

in the limit of a large m_j and the independence between a_{ijt} and $\exp(\epsilon_{it})$ we can approximate Y_{jt} as:

$$Y_{jt} \approx \exp(\alpha_t) N_{jt}^{\beta_t} m_j \left(\frac{1}{m_j} \right) \sum_{i \in m_j} a_{ijt}^{\beta_t} \left(\frac{1}{m_j} \right) \sum_{i \in m_j} \exp(\epsilon_{it}) , \tag{10}$$

since $E[XY] = E[X] E[Y]$ and $E[X] \approx (1/N) \sum_{i=1}^{N} X_i$.

Taking the log of both sides:

$$\log Y_{jt} \approx \alpha_t + \beta_t \log N_{jt} + \log \left(\sum_{i \in m_j} a_{ijt}^{\beta_t} \right) + \log \left(\frac{\sum_{i \in m_j} \exp(\epsilon_{it})}{m_j} \right) \tag{11}$$

If we estimated the following model at macro-region level via OLS:

$$\log Y_{jt} = \alpha_t + \beta_t \log N_{jt} + e_{jt}, \tag{12}$$

then:

$$\hat{\beta}_t \xrightarrow{p} \beta_t + \mathrm{Corr}\left(\log(N_{jt}), e_{jt}\right) \left[\frac{\sigma_{e_{jt}}}{\sigma_{\log N_{jt}}} \right] \tag{13}$$

On the base of Eq. (11) $\mathrm{Corr}\left(\log(N_{jt}), e_{jt}\right)$ could be non-zero for the relationship between N_{jt} (the total intensity of nightlights) and m_j (its cardinality is the number of regions inside the macroregion j).

The two polar cases are when the within-macroregion distribution is uniform, i.e. $n_{it} = N_{jt}/m_j$ and $a_{ijt} = 1/m_j$, and when all the intensity of nightlights for each macroregion is concentrated in only region, i.e $n_{it} = 0$ for all $i \neq q$ and $n_{qt} = N_{jt}$ and $a_{ijt} = 0$ for all $i \neq q$ and $a_{qjt} = a$.

In the uniform distribution, Eq. (11) becomes:

$$\log Y_{jt} = \beta_t \log N_{jt} + (1 - \beta_t) \log m_j + \log \left(\frac{\sum_{i \in m_j} \exp(\epsilon_{it})}{m_j} \right) , \tag{14}$$

and therefore:[1]

$$\text{Corr}(\log N_{jt}, e_{jt}) = \text{Sign}(1-\beta_t)\text{Corr}(\log N_{jt}, \log m_j) \tag{15}$$

under the assumption that

$$\text{Corr}\left(\log(N_{jt}), \log\left(\frac{\sum_{i\in m_j}\exp(\epsilon_{it})}{m_j}\right)\right) \to 0 \text{ as } m_j \to \infty, \tag{16}$$

where $\text{Sign}(x)$ is equal to 1 if $x > 0$, equal to 0 if $x = 0$ and -1 otherwise. Instead, in the extremely concentrated distribution Eq. (11) becomes:

$$\log Y_{jt} = \beta_t \log N_{jt} + \log\left(\frac{\sum_{i\in m_j}\exp(\epsilon_{it})}{m_j}\right); \tag{17}$$

and therefore:

$$\text{Corr}(\log(N_{jt}), e_{jt}) = 0 \tag{18}$$

Therefore, from Eq. (13)

$$\hat{\beta}_t = \begin{cases} \beta_t & \text{when the distribution is extremely concentrated;} \\ \beta_t + \text{Sign}(1-\beta_t)\text{Corr}(\log N_{jt}, \log m_j)\left[\dfrac{\sigma_{e_{jt}}}{\sigma_{\log N_{jt}}}\right] & \text{when the distribution is uniform.} \end{cases} \tag{19}$$

We conjecture that the intermediate cases between uniform and extremely concentrated distribution have a covariance in absolute value between zero and $|\text{Corr}(\log N_{jt}, \log m_j)|$. Hence, in the intermediate cases $\hat{\beta}_t$ should belong to the interval defined in Eq. (19).

Below, we find evidence that NUTS 3 and NUTS 2 regions presents such concentrated spatial distributions, generating therefore a bias in the estimated coefficient.

4.3 The nightlights as proxy for personal income

Figure 3 reports the estimated α and β of Eq. (5) and the explained variance (R^2) of different estimated models for years . Overall, nightlights appear a good proxy for personal income at any level of aggregation and year in the sample The explained variance is not significantly changing over time and never below 78%. It is decreasing from municipalities to NUTS 3, and strongly decreasing to NUTS2. Aggregate income was grown of 11.3%, while nightlights of 10.6% in the period. If we take municipal level on average income was growing at 8% while nightlights at 15%. This means that a heterogeneity is present at municipal level, with the municipalities with the highest income growing less. Taking as given the level of aggregation, we observe that the estimated intercept in Eq. (5) are slightly decreasing over time, that is the scale parameter is changing over time in favour of income. The estimated elasticity is slightly increasing over time and always greater than one for municipalities and NUTS 3, while at NUTS 2 is lower than 1. This implies that income is growing at the faster pace than nightlights and this excess of growth is increasing over time. The estimate of β_t coefficient is almost equal for municipalities and NUTS 3 regions. According to Eq. (19) this

[1] It holds $\text{Corr}(\alpha X, Y) = \text{Sign}(\alpha)\text{Corr}(X, Y)$.

result can be explained by the extreme inequality of municipal nightlights within the same NUTS 3 regions (in 2012 about 95% ot total variance at NUTS 3 level is due to within NUTS3 regions variance). On the contrary, the estimate of β_t coefficient for NUTS 2 regions is lower than the one of NUTS 3 regions; according to Eq. (19) this result can be explained by the more limited inequality of NUTS 3 nightlights within the same NUTS 2 regions (in 2012 about 78% of total variance at NUTS 2 level is due to within NUTS 2 regions variance) together with the fact that true β_t is higher than 1 (aggregation bias is therefore negative).

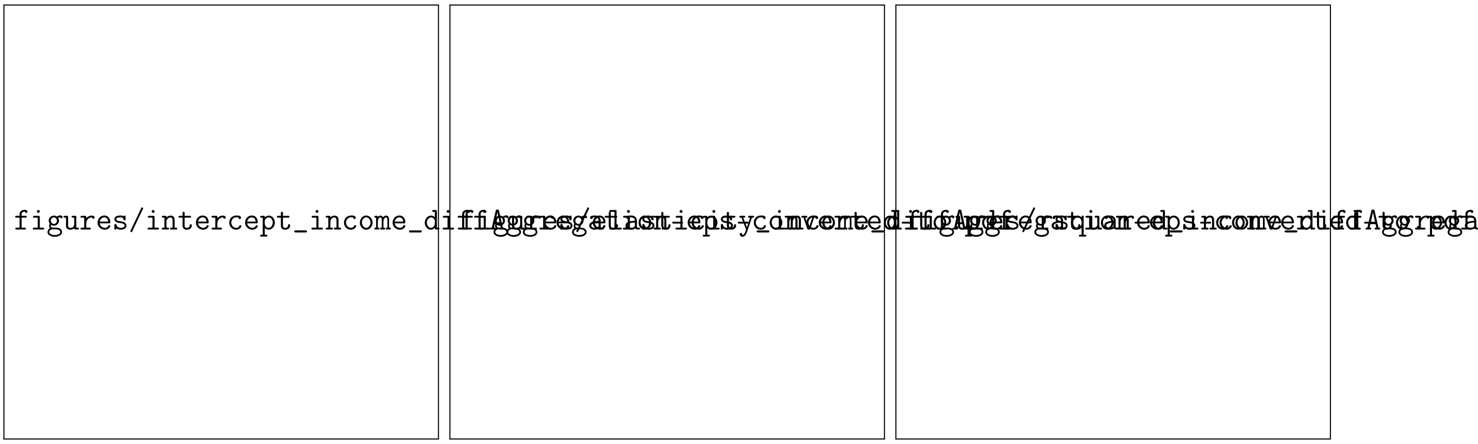

Figure 3: Nightlights of Italian NUTS 2 regions, NUTS 3 regions and municipalities versus regional income. The estimated α and β of Eq. (5) and the explained variance (R^2) of the estimated models.

4.4 The nightlights as proxy for the population

Figure 4 reports the estimated α and β of Eq. (5) and the explained variance (R^2) of the estimated models. Taking as given the level of aggregation, we observe that the estimated intercept in Eq. (5) are slightly decreasing over time, pointing out to a different trend in the two variables (aggregate income is grown of 0.2%, while nightlights of 10.6% in the period). Increasing the level of aggregation the income is still growing faster than nightlights, but lower and lower (the estimated intercepts are closer and closer). This suggests that the municipalities/NUTS 3 regions with the highest intensity of nightlights are the one with the lowest increase. The estimated elasticity is slightly increasing over time and always greater than one for municipalities. Taking as given the level of aggregation, as expected the explained variance is not significantly changing over time. The expGlained variance instead is decreasing from municipalities to NUTS 3, and strongly decreasing to NUTS2, but never below 78%. Overall, nightlights appear a good proxy for population at any level of aggregation and year in the sample. The estimate of β_t coefficient for population at different levels of aggregation shows the same pattern of the estimates for income, but the magnitude of the change between different levels is lower as well as the estimate of β_t is closer to one. This agrees with Eq. (19), where the bias is proportional to the sign of $1 - \beta_t$.

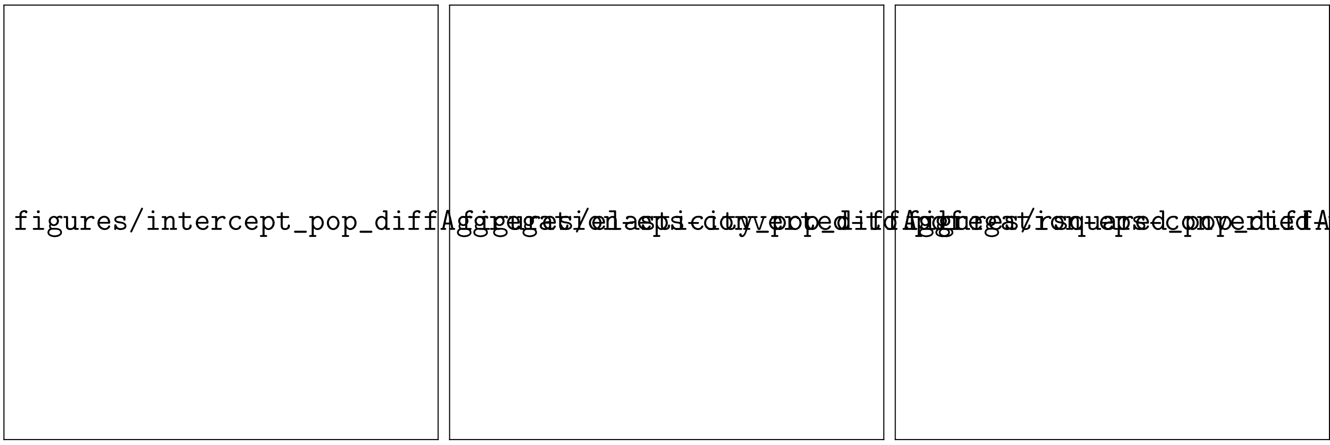

Figure 4: Nightlights of Italian NUTS 2 regions, NUTS 3 regions and municipalities versus regional population. The estimated α and β of Eq. (5) and the explained variance (R^2) of the estimated models.

5 Concluding remarks

TO BE WRITTEN

References

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

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

Gibson, J., S. Olivia, and G. Boe-Gibson (2020). Night lights in economics: Sources and uses 1. *Journal of Economic Surveys 34*(5), 955–980.

Henderson, J. V., T. Squires, A. Storeygard, and D. Weil (2018). The global distribution of economic activity: nature, history, and the role of trade. *The Quarterly Journal of Economics 133*(1), 357–406.

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

Martinez, L. R. (2022). How much should we trust the dictator's gdp growth estimates? *Journal of Political Economy 130*(10), 2731–2769.

Michalopoulos, S. and E. Papaioannou (2018). Spatial patterns of development: A meso approach. *Annual Review of Economics 10*, 383–410.

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences 103*(10), 3510–3517.