Estimating Economic Activity in Zimbabwe Using Big Data

7TH IMF STATISTICAL FORUM

NOVEMBER 15, 2019

Gene Leon and Frederico Lima (IMF)

Motivation

- In recent decades, the production of accurate national account statistics in Zimbabwe has been hampered by low capacity and large disruptions including:

  - Land reform during the 2000s

  - Hyperinflation in the late 2000s

  - Electricity shortages (2015/6, and again in 2019)

  - De-dollarization starting in 2016

- These events have also contributed to a growing informal sector, which is estimated to be among the largest in the world (Medina and Schneider, 2018)

- Recent research shows that big data can improve GDP measurement, especially in countries with low statistical capacity (e.g. Chen and Nordhaus, Henderson et al., 2012)

Our contribution

- We combine traditional and nontraditional data sources, including night lights and vegetation health indices derived from satellite data, to create alternative estimates of economic activity in Zimbabwe

- We focus both on the <u>level</u> of economic activity and its <u>growth rate</u>

- Our estimates can help benchmark official data (e.g. the recent GDP rebase by Zimbabwean authorities), in particular where it fails to capture the informal or illegal sector

- In addition, our calculations are directly relevant for policymakers and for IMF engagement with the Zimbabwean authorities (e.g. Staff Monitored Program, Debt Sustainability Analysis)

Literature Review

- **Using Big Data to improve measurement in National Accounts**
  - e.g., Chen and Nordhaus, 2011; Henderson et al., 2012; Pinkovskiy and Sala-i-Martin, 2016; Galimberti, 2018; Hu and Yao, 2019; Martinez, 2019

- **Specific applications to LICs**
  - e.g., Bandhari et al. 2011; Bundervoet et al. 2015; Basihos 2016; Do et al. 2017; Skoufias et al. 2017; Debbich, 2019

- **Big data applications in economics**
  - e.g., Cavallo, 2013; Donaldson and Storeygard, 2016; Proville et al, 2017

- **Informal economy in Zimbabwe**
  - Pact (2015), Global Witness (2017), UNIDO (2017)

Our work combines traditional with non-traditional data

1. Non-traditional indicators, derived from satellite images:
   - ▶ Night lights
   - ▶ Vegetation & Drought indices

2. More traditional proxies for economic activity, which are at least partly collected by national authorities:
   - ▶ CO_2 emissions (European Commission / EDGAR)
   - ▶ Energy consumption (EIA)
   - ▶ Exports & Imports (WTO, UNCTAD)
   - ▶ Agricultural production (FAO)
   - ▶ Socioeconomic (urbanization, human capital, governance, … from WB)

Methodology

- The results we are presenting today are based on a standard methodology followed in the literature:

  - GDP level: cross-section regression

  - GDP growth: Henderson et al. (2012)

- To account for variation in economic structure, we limit the sample of comparable countries in two ways:

  - Only countries with high statistical capacity (as measured by the World Bank)

  - Only countries where agriculture > 10% of total value added

- In ongoing work, we examine the robustness of these results using alternative estimators (e.g., machine learning) and find similar results.

Night lights are strongly correlated with official GDP statistics in Zimbabwe and across countries
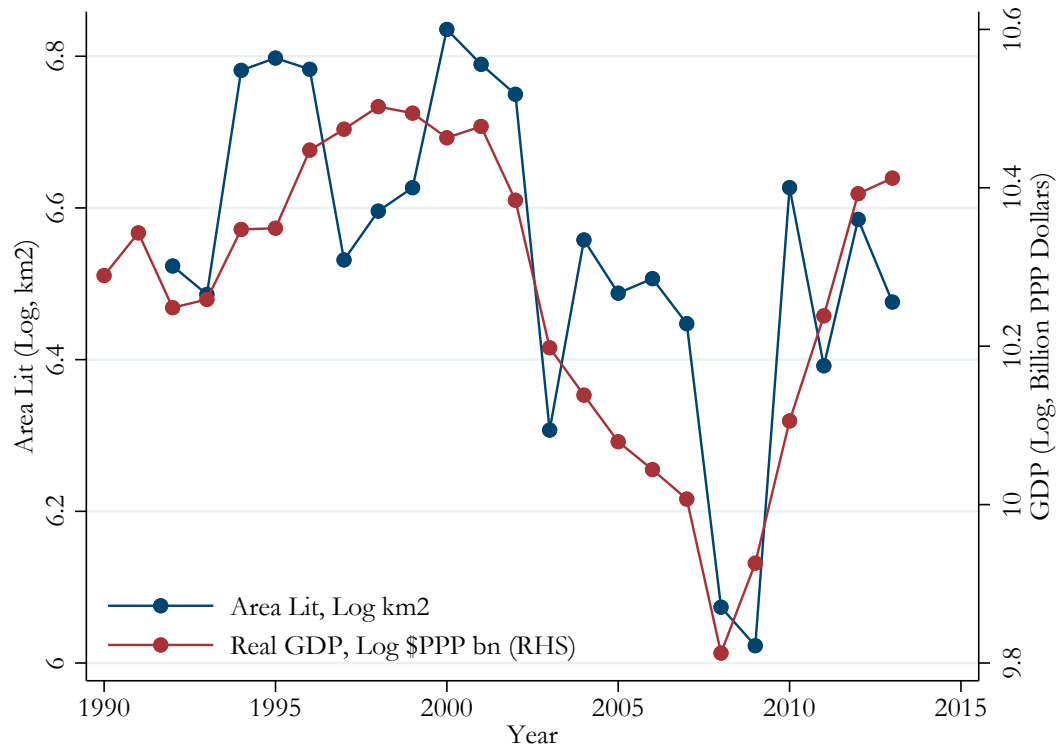
GDP and Night Lights across countries in 2012
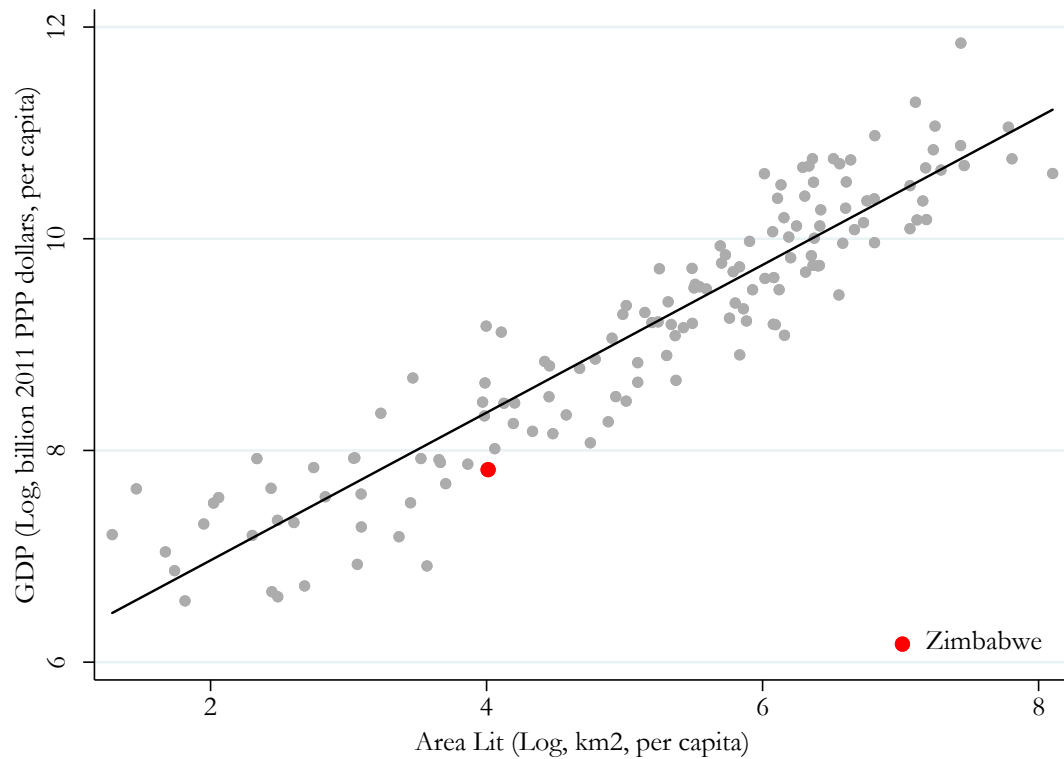
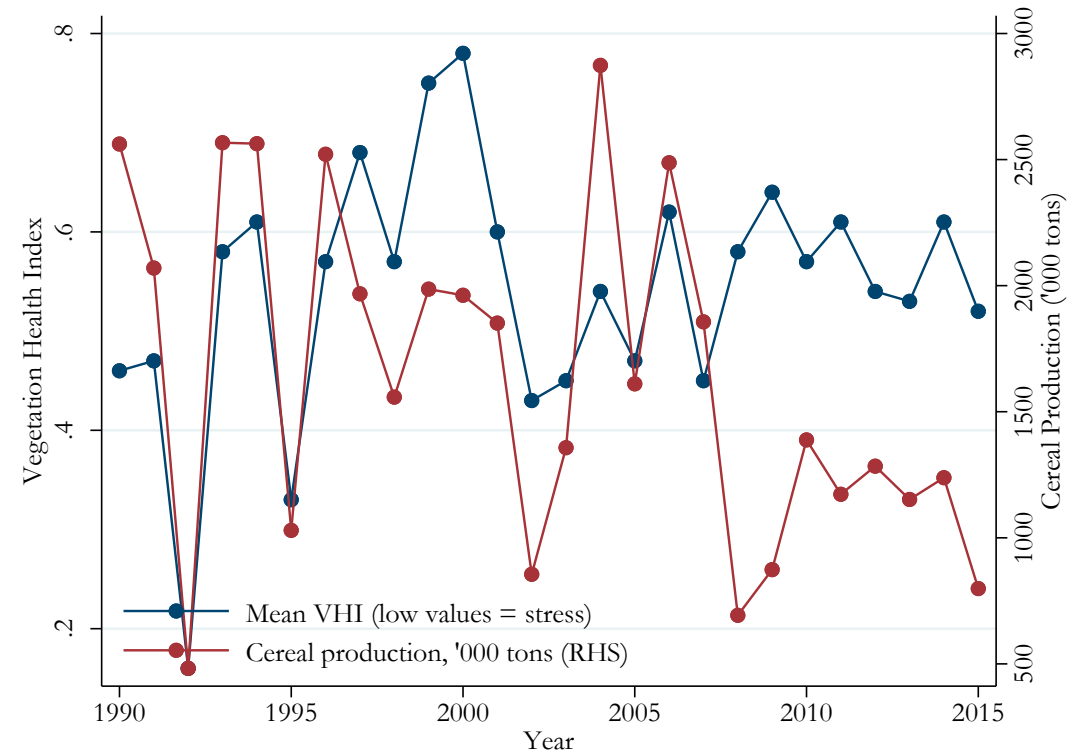**Controlling for countries where agriculture predominates**
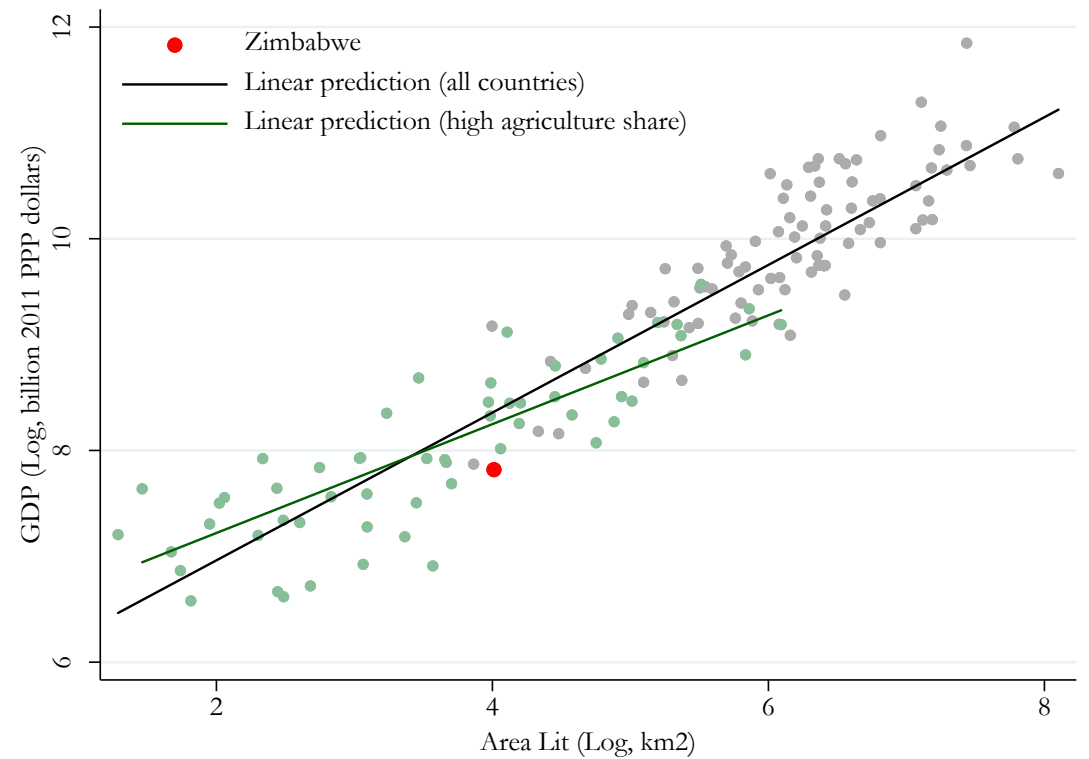
Satellite data also seem to be good proxies for the evolution of economic activity over time in Zimbabwe

GDP and Night Lights

Crop output and Vegetation Health

Our results indicate that the level of official GDP may be underestimated

TABLE 1. Estimating GDP in Zimbabwe in 2012 (in 2011 $PPP)

	Baseline				Interactions with Agriculture share > 10			
	Only Night Lights	+ Socio-economic	+ Agriculture Indicators	+ Trade Indicators	Only Night Lights	+ Socio-economic	+ Agriculture Indicators	+ Trade Indicators
Estimated GDP	57.2	42.9	48.1	48.3	49.0	44.0	36.5	37.0
	(2.8)	(4.3)	(5.2)	(6.5)	(3.1)	(7.8)	(6.8)	(9.9)
	[50.9 , 62]	[34.3 , 51.4]	[37.8 , 58.4]	[35.5 , 61.1]	[42.9 , 55.1]	[28.6 , 59.4]	[23.1 , 49.9]	[17.5 , 56.4]
Deviation from Official GDP (%)	73%	31%	47%	48%	50%	35%	12%	13%
N	108	108	108	108	108	108	108	108
Adjusted R^2	0.85	0.90	0.91	0.94	0.87	0.91	0.92	0.95

Notes: This table displays alternative estimates of Zimbabwe 2012 GDP. Official GDP was 32.6 billion 2011 $PPP. All regressions exclude countries scoring in the bottom quartile of statistical capacity. Robust standard errors in round brackets, and 95 percent confidence intervals in square brackets.

- Including agriculture and trade indicators brings our GDP estimate more in line with the official data… but Zimbabwe official data is itself subject to mismeasurement (e.g., mining exports).

- Using agricultural countries as our main comparator sample tends to lower the GDP point estimate, but also increases parameter uncertainty

Our estimates also suggest a smaller contraction (and recovery) during the 2000s hyperinflation period

Night Lights-only Model
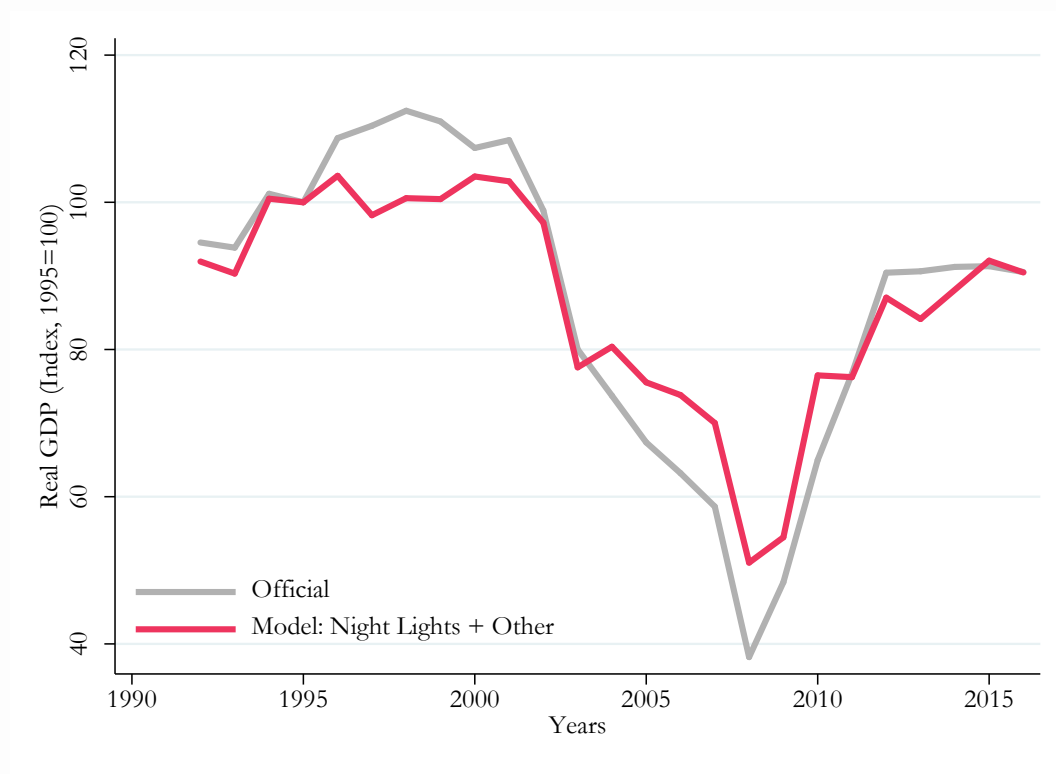
Night Lights + Other Predictors
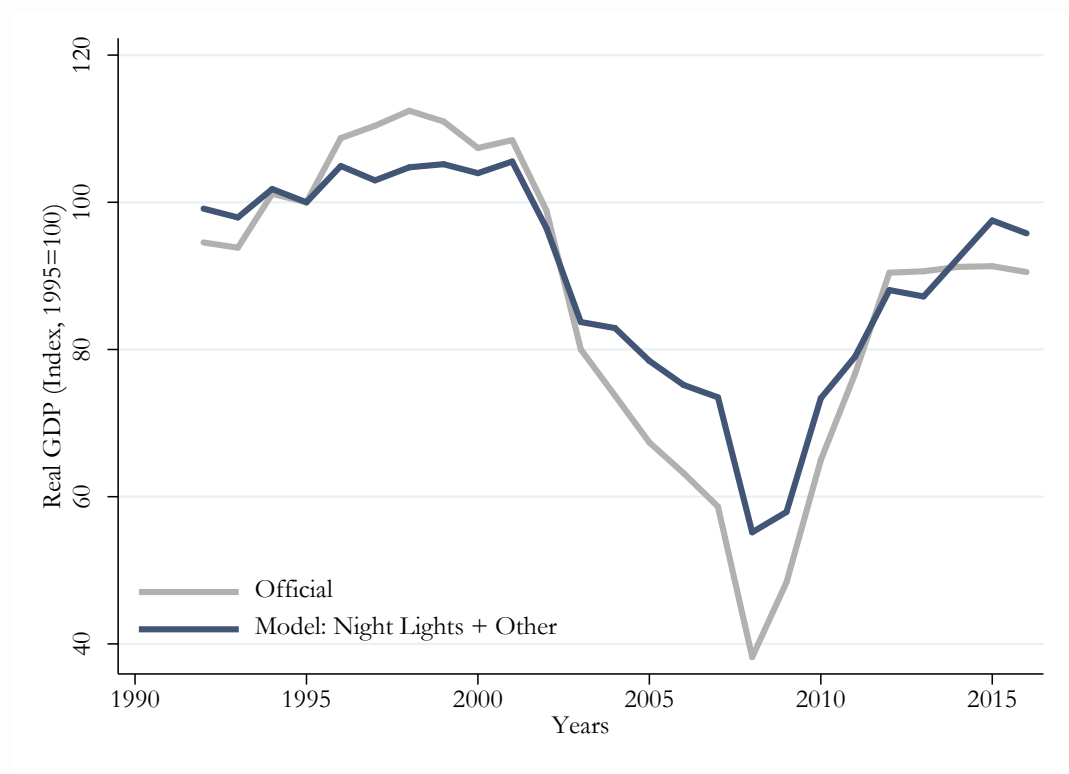
- The smaller contraction during the 2000s compared to official data may be explained by a shift towards informal activities in response to a high inflation environment

Conclusion and Next Steps

- We combine traditional indicators of economic activity with Big Data to produce alternative estimates of economic activity in Zimbabwe

- We find that official GDP statistics somewhat underestimate the level of economic activity, and overestimate the contraction/recovery cycle during the 2000s
  - The difference could be due to unrecorded (informal) activity, but also methodology and model uncertainty

Ongoing work:

- Capturing informal / illegal mining production (gold, diamonds)
  - Some evidence that official statistics don't capture all mining production. Accounting for smuggling / mismeasurement of mining exports could increase official GDP by up to 10% in some years.

- Using payments data (e.g., mobile money transactions)

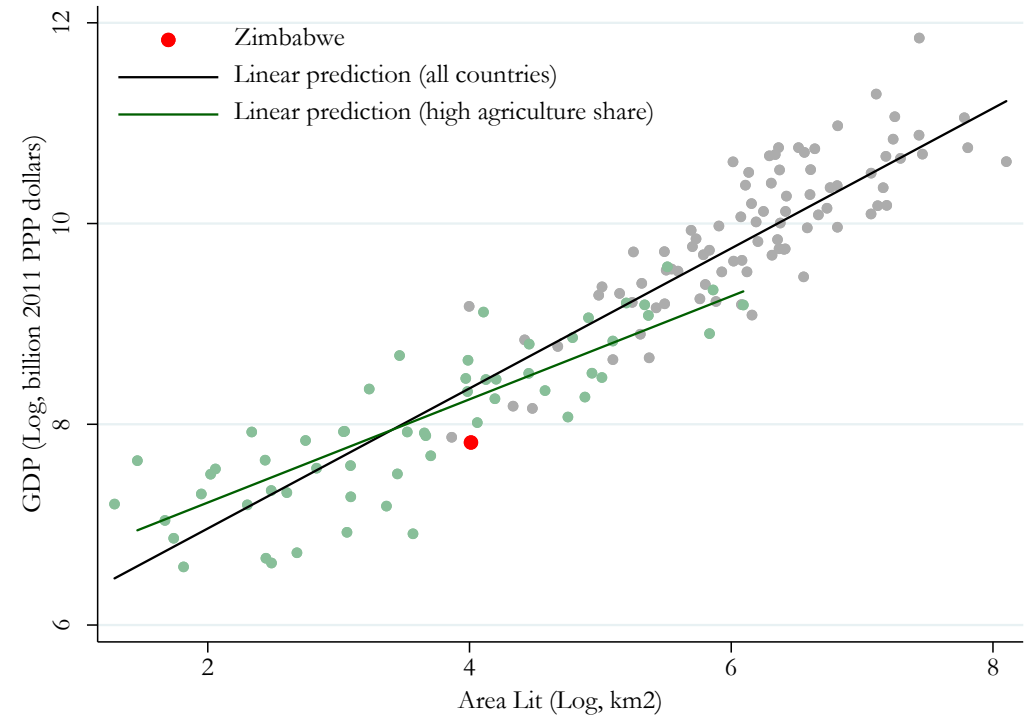- Incorporating high-frequency indicators into GDP growth estimates

Thank you!

Robustness: relation of GDP and Night Lights as a function of statistical capacity and agriculture share

Controlling for statistical capacity
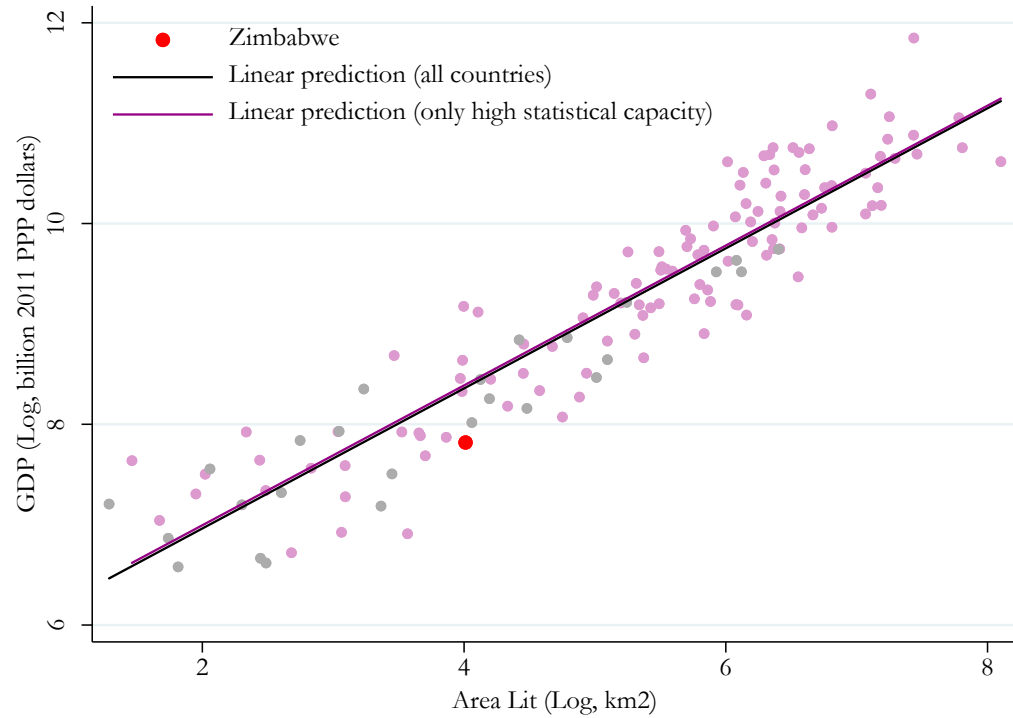
Controlling for agriculture share

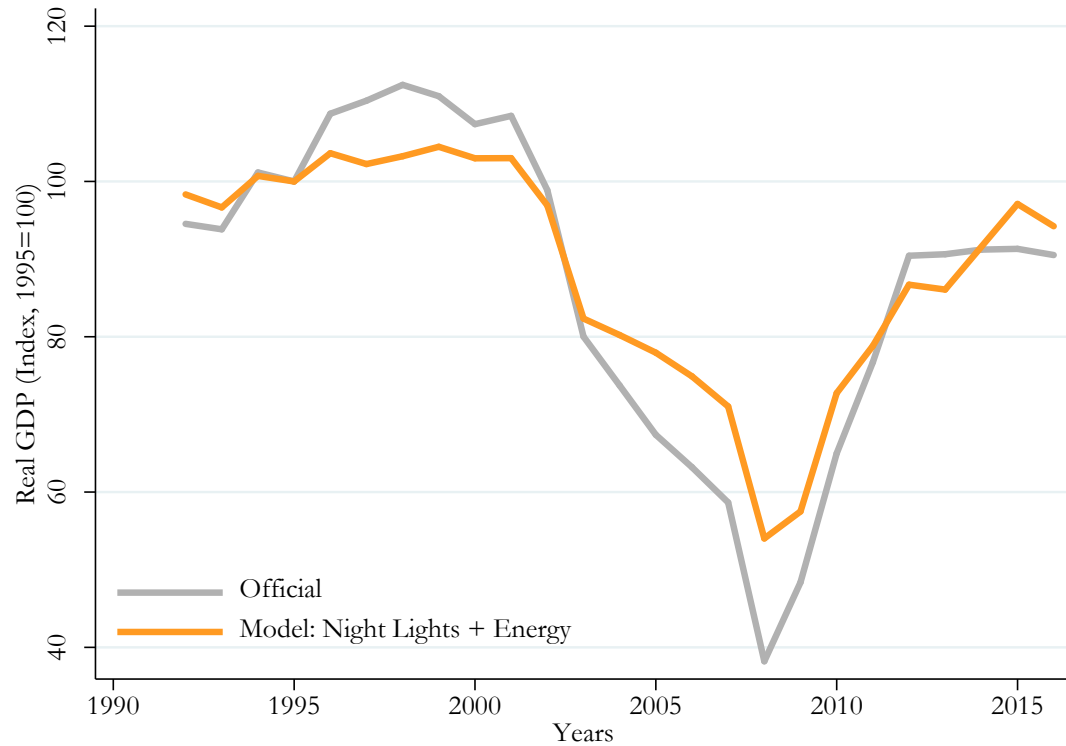# Robustness: Estimating GDP in levels instead of per capita

	TABLE 1. Estimating GDP in Zimbabwe in 2012 (in 2011 $PPP)							
	Baseline				Interactions with Agriculture share > 10			
	Only Night Lights	+ Socio-economic	+ Agriculture Indicators	+ Trade Indicators	Only Night Lights	+ Socio-economic	+ Agriculture Indicators	+ Trade Indicators
Estimated GDP	46.2	46.1	36.5	35.6	53.6	53.7	33.4	29.8
	(3.0)	(6.4)	(4.5)	(5.5)	(5.4)	(12.7)	(5.9)	(6.7)
	[41.9 , 53.6]	[33.6 , 58.6]	[27.8 , 45.3]	[24.8 , 46.5]	[43 , 64.3]	[28.8 , 78.6]	[21.8 , 45]	[16.7 , 42.8]
Deviation from Official GDP (%)	46%	41%	12%	9%	64%	65%	2%	-9%
N	108	108	108	108	108	108	108	108
Adjusted R^2	0.85	0.90	0.91	0.94	0.87	0.91	0.92	0.95

Notes: This table displays alternative estimates of Zimbabwe 2012 GDP. Official GDP was 32.6 billion 2011 $PPP. All regressions exclude countries scoring in the bottom quartile of statistical capacity. Robust standard errors in round brackets, and 95 percent confidence intervals in square brackets.

Robustness: Estimated GDP growth adding alternative predictors

Adding controls for energy

Adding controls for trade