DEWAN EKONOMI NASIONAL
REPUBLIK INDONESIA

Forecasting Indonesian National and Provincial GDP using Nighttime Light Index

Krisna Gupta Timothy Kinmekita Ginting Meizahra Afidatie

15 April 2026

Why measure GDP at higher frequency?

- GDP is the anchor variable for most macroeconomic decisions: fiscal planning, monetary policy, private investment.
- Official quarterly numbers are released with **substantial lag** and occasional revisions.
- When the stakes are high, policy makers often need to **nowcast** — infer the most recent quarter from leading or coincident indicators.
- Because GDP is a political signal, there is an incentive to overestimate; *independent* validation is valuable.

Can nighttime lights serve as a credible proxy for Indonesian GDP growth, nationally and provincially?

Why nighttime lights?

- Satellite-based measure of light emitted from Earth at night: streetlights, buildings, vehicles, commercial activity.
- Literature: Henderson, Storeygard, and Weil (2012), Bickenbach et al. (2016), Martínez (2022) show NTL tracks economic activity globally, including for countries with weak statistical infrastructure.
- Data pipeline from NASA's Black Marble is reproducible and transparent.
- High spatial resolution (500 m) — allows both national and provincial analysis.
- But: is it genuinely useful for **Indonesia**?

Data: NASA Black Marble VIIRS

- Source: NASA Black Marble Standard Product Suite, via `blackmarblepy` (Stefanini Vicente and Marty 2023).
- Daily radiance measurements, cloud-corrected and calibrated.
- We aggregate to monthly (Jan 2012 – Dec 2024) then to quarterly year-on-year log growth.
- GDP data from BPS quarterly real GDP series (BPS 2025), matched by quarter.
- Provincial GDRP from BPS for 34 provinces.

Indonesia seen from space, 2023

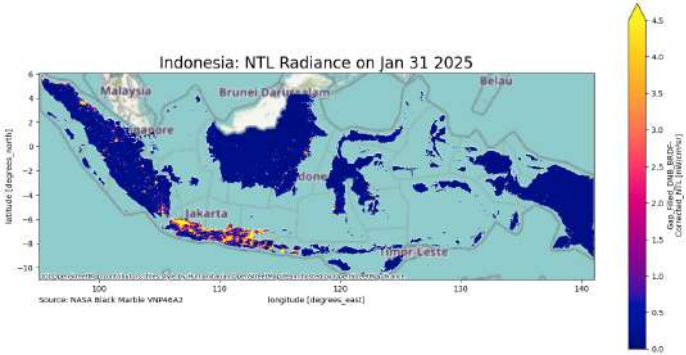


Figure 1: Annual NTL in Indonesia, 2023

- Stark contrast between Java and the rest of Indonesia — mirrors GDP & population distribution.
- Mining and palm-oil regions are visibly darker despite high economic output.

National series

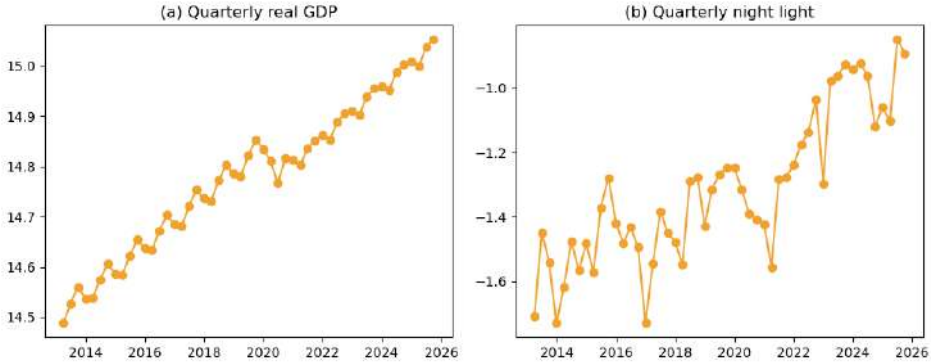


Figure 2: Quarterly real GDP (left) and NTL index (right), 2012–2024

Two layers of analysis

National level

- Bivariate time-series.
- OLS $\rightarrow$ ARDL with lag selection by AIC.
- Six ARDL specifications around COVID/scarring/quarterly dummies.
- Train/test forecast.

Provincial level

- Panel of 34 provinces $\times$ 47 quarters.
- Static: pooled OLS, FE, TWFE.
- Dynamic: DFE, PMG, MG (Pesaran–Shin–Smith).

Are the series stationary?

Augmented Dickey–Fuller tests:

Series	ADF stat	p-value	5% critical	Verdict
log GDP, levels	-0.38	0.91	-2.93	I(1)
log GDP, Δ	-2.97	0.04	-2.93	stationary
log NTL, levels	$+1.14$	0.996	-2.94	I(1)
log NTL, Δ	-2.77	0.06	-2.94	borderline stationary

Both series look I(1); first differences stationary.

Are they cointegrated?

Two tests, two verdicts:

- **Engle–Granger residual ADF**: reject "no cointegration" at 1% (p $= 0.0003$)
- **Johansen trace test**: fail to reject rank $= 0$ ($9.45 < 15.49$ at 5%)

With $T = 51$ and two variables, the Johansen test has low power, and the evidence is **borderline** rather than decisive.

ARDL-in-levels (Pesaran–Shin–Smith bounds framework) remains valid whether the series are I(0) or I(1), so we proceed.

Does NTL *lead* GDP? Granger-causality

	lag 1	lag 2	lag 3	lag 4
NTL $\rightarrow$ GDP (p)	0.326	**0.007**	0.229	0.343
GDP $\rightarrow$ NTL (p)	**0.006**	**0.018**	**0.007**	**0.004**

- NTL Granger-causes GDP **only at lag 2**.
- GDP Granger-causes NTL at **every** lag, much more robustly.
- Pattern is more consistent with NTL *tracking* GDP contemporaneously than leading it.
- This qualifies the leading-indicator framing.

National: OLS is biased

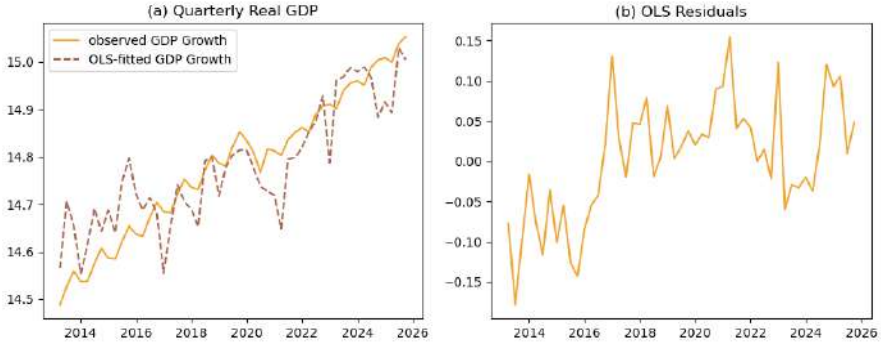
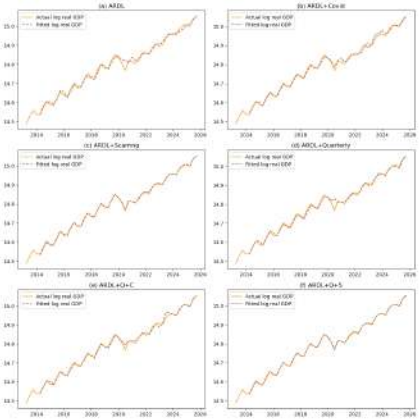


Figure 3: OLS fit and residuals

- OLS coefficient: 0.54*, **but residuals are systematically biased:** over-predicts pre-2017, under-predicts post-2017**.

National: six ARDL specifications



(a) baseline (b) +Covid (c) +Scarring (d) +Quarterly (e) +Q+C (f) +Q+S.

Specifications with the scarring dummy — panels (c) and (f) — produce the tightest fits.

Reading the ARDL table

- **NTL coefficient shrinks** drastically once dynamics are added: 0.54 (OLS) $\rightarrow$ 0.01–0.08 (ARDL).
- Contemporaneous NTL is significant at 1% in the baseline, only 10% in the preferred ARDL+Scar.
- The scarring dummy is highly significant at all four lags.
- Bulk of the fit comes from **lagged GDP + the post-2020 dummy**, not from NTL.

NTL carries information about GDP, but not enough to be a stand-alone predictor.

The “scarring” coefficient — handle with care

- The post-2020 dummy coefficient at $t = 0$ is about -0.02 $\rightarrow$ naively, “GDP 2% lower persistently after the pandemic.”
- **Three reasons to discount the 2% number:**
  1. The +Scar spec flips AR coefficients’ signs and magnifies the trend — model-fit artefact as much as structural finding.
  2. The dummy absorbs *any* post-2020 shift: BPS methodology changes, commodity prices, reforms.
  3. Identification rests on only ~20 post-shock quarters.
- Read as **suggestive**, not a precise estimate of pandemic damage.

Out-of-sample forecast

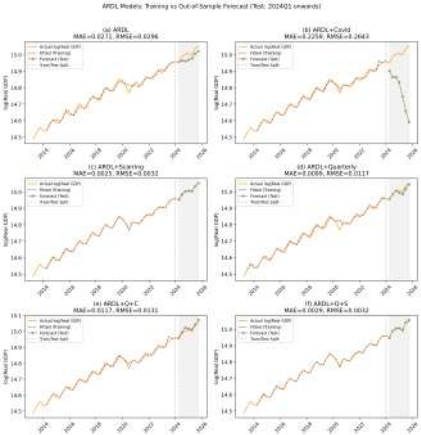


Figure 4: ARDL train/test split: hold-out 2024+

- +Scar and +Q+S produce the smallest MAE/RMSE.
- But: hold-out window is **only 4–6 quarters**. Too short to make strong forecast-quality claims.
- Rolling/recursive evaluation is future work.

Provincial: static models

	Pooled OLS	FE	TWFE
NTL coefficient	0.56***	0.28***	**0.055*** **
R^2	0.55	0.35	0.85

- Pooled OLS: driven by **between-province** variation (Java vs rest of Indonesia).
- Moving to within-province (FE) and time-controlled (TWFE) shrinks the coefficient by an order of magnitude.
- **TWFE = 0.055** is the most informative: within a province, controlling for common time shocks, NTL has little marginal predictive power.

Provincial cross-section

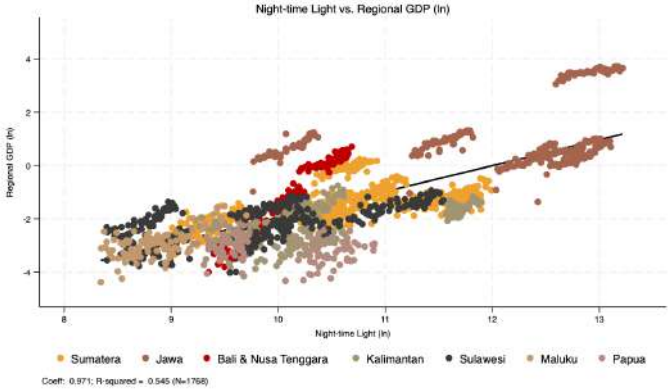


Figure 5: NTL vs GDRP across 34 provinces

- Java clusters high on both axes.
- Resource provinces generate lightless output, Agricultural provinces sit somewhere in between.
- The NTL–GDRP *slope* plausibly differs across island groups.

Provincial: dynamic estimators

	DFE	PMG	MG
Long-run NTL, baseline	0.332***	0.374	0.486***
Long-run NTL, +Scarring	0.589**	0.426	−0.713
Pooling: short run	common	common	free
Pooling: long run	common	common	free

- DFE (strongest pooling): significant throughout.
- PMG (long-run only pooled): insignificant in baseline.
- MG (fully free): large and positive in baseline, **sign-flipped** in +Scarring.

What this disagreement means

- FE absorbs *intercept* heterogeneity across provinces — not *slope* heterogeneity.
- DFE/PMG impose a common NTL **slope** across all 34 provinces. MG does not.
- When the true elasticity differs across Java / Kalimantan / Sulawesi / Papua, DFE gives a precise average that doesn't represent any particular province.
- MG's sign-flip in +Scarring is not "the true elasticity is negative" — it's the signature of heterogeneity showing up once the pooling restriction is relaxed. *The cross-estimator disagreement is itself the finding.*

What about the ECT slowdown?

- Baseline DFE: ECT $= -0.0374 \rightarrow$ ~6.75 years to adjust.
- +Scarring DFE: ECT $= -0.0169 \rightarrow$ ~14.75 years.
- Tempting interpretation: **hysteresis** — the pandemic permanently slowed recovery.
- But: adding a scarring dummy mechanically *redefines* equilibrium, so part of the ECT drop is a specification artefact.
- We flag this as consistent with hysteresis, not a causal identification.
- Separating observation costs number of observations.

What we find

- NTL is **positively and generally significantly** correlated with Indonesian GDP.
- Most raw-OLS association is **between-province** rather than time-series: TWFE coefficient is ~0.05, not 0.56.
- There is a clear structural break in GDP around 2020, and accounting for it materially improves forecast fit.
- Pattern of provincial estimator disagreement points to **genuine slope heterogeneity** across islands.

What we do not find

- Evidence that NTL is a strong **stand-alone** predictor — the ARDL forecast is driven by lagged GDP and the post-2020 dummy.
- Robust Granger-causal evidence that NTL **leads** GDP (only lag 2 works).
- A clean causal reading of the **2% scarring** estimate (dummy absorbs any post-2020 shock).
- Convincing evidence that the **adjustment speed** has changed (ECT slowdown is partly a specification artefact).
- Cross-estimator agreement in the provincial panel — pooled estimates likely paper over heterogeneity.

Takeaway

NTL for Indonesia is best framed as:

1. A **complement** to official GDP in nowcasting pipelines, not a substitute.
2. A **between-province** indicator of structural variation.
3. An **imperfect** crisis signal — NTL diverged from GDP precisely during the 2020 shock, where independent validation is most valuable.

Useful, but not a magic wand.

Thank you

Paper and reproducibility repo:

- Main paper: github.com/den-econ/nitelite
- Python pipeline (national): `appendix.ipynb`
- Stata pipeline (provincial): `stata/Do/Nighttime Light_Reg.do`
- Data: NASA Black Marble (Stefanini Vicente and Marty 2023), BPS (BPS 2025)

Contact: krisna@dewanekonomi.go.id

References

Bickenbach, Frank, Eckhardt Bode, Peter Nunnenkamp, and Mareike Söder. 2016. "Night Lights and Regional GDP." Journal Article. *Review of World Economics* 152 (2): 425–447. https://doi.org/https://doi.org/10.1007/s10290-016-0246-0.

BPS. 2025. *[Seri 2010] 4. Laju Pertumbuhan PDB Menurut Pengeluaran, 2025. Diakses Pada 14 September 2025*. Https://www.bps.go.id/id/statistics-table/2/MTA4IzI=/-seri-2010–4–laju-pertumbuhan-pdb-menurut-pengeluaran–persen-.html.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *The American Economic Review* 102 (2): 994–1028. http://www.jstor.org/stable/23245442.

Martínez, Luis R. 2022. "How Much Should We Trust the Dictator's GDP Growth Estimates?" Journal Article. *Journal of Political Economy* 130 (10): 2731–2769. https://doi.org/10.1086/720458.

Stefanini Vicente, Gabriel, and Robert Marty. 2023. *BlackMarblePy: Georeferenced Rasters and*

DEWAN EKONOMI NASIONAL
REPUBLIK INDONESIA

THANK YOU