**Utilizing Nighttime Light Data to Predict Money Laundering Activities**

Agung Andiojaya[1]; Riana Rizka[1]; Titi Kanti Lestari[2]; R.R. Nefriana[3]

*[1]Indonesia Financial Transaction Report and Analysis Center; [2]Atmajaya University, Jakarta, Indonesia; [3]BPS - Statistics Indonesia*

**Abstract:**

The relationship between money laundering and GDP has been extensively researched by several academics. While some of them indicated having a negative association, others claimed the contrary. Two years ago, when the pandemic hit the world, the environment of data creation and consumption underwent a significant transformation. Data supply has been driven by limited funding and human activity, but on the other hand, the amount of data needed to combat pandemic effects has increased[1]. This essay will examine the connections between nighttime light (NTL), economic expansion, and money laundering in Indonesian provinces. The examination of the usage of NTL in foretelling the potential rise in cases of money laundering is further expanded upon in this article. The research based on the panel data model's findings will indicate whether the NTL data is statistically significant or not significant for forecasting the rise in cases of money laundering brought on by the province's economic expansion.

**Keywords:** Nighttime-light; VIIRS; STRs; Money Laundering; Big Data

## Introduction

In recent years, there has been an increase in interest in a broad field of study known as the use of nighttime light data to anticipate socio-economic activities. The goal is to find patterns and relationships between the distribution and intensity of nighttime light and various socioeconomic indices, such as economic activity, population density, and income levels, using satellite imagery of nighttime light.

The availability of high-resolution satellite data, which enables the precise charting of nighttime light patterns at a fine geographic scale, has made study in this area viable[2], [3]. The distribution of light throughout a region is depicted on maps made by researchers using these images, and connections between these maps and other socio-economic variables are found using these maps.

One of the key benefits of collecting nighttime light data is that it offers a dependable and consistent approach to gauge the degree of economic activity in a certain location, independent of the local economy's unique features. As a result, it is now feasible to compare socioeconomic situations between different areas and nations, which may be helpful for decision-makers, development agencies, and other parties interested in learning about a certain region's economic situation[4]. Especially when the Covid-19 strikes in 2020 in which many activities were prohibited. This situation affected many collecting data activities that rely on traditional surveys[1].

In the recent situation, besides nighttime light data being used as a predictor to socio-economic data, it can also be applied to predicting the amount of money laundering activities in a certain area. Since people and organizations try to utilize the financial system to hide the results of criminal actions, money laundering frequently takes place in regions with significant levels of economic activity. It may be able to identify places with a higher risk of money laundering by analyzing nighttime light data to pinpoint regions with high levels of economic activity. Therefore, this study offers a special chance to learn more about the potential of nighttime light data to be utilized in forecasting these unlawful actions.

**Methodology**

The data used in this study include Indonesia's Gross Domestic Regional Product (GDRP) at Constant Price by Province sourced from BPS-Statistics Indonesia. For nighttime-light (NTL) satellite data employ average day-night-band (DNB) values of Monthly Cloud-free DNB Composite NPP-VIIRS data sourced from https://eogdata.mines.edu/products/vnl/. While money laundering is approached using suspicious transaction records (STR). To obtain spatial data for each province in Indonesia, this study uses BPS working map shape files. In generating NTL data by province, this study utilizes Google Earth Engine (GEE). GEE is a cloud-based platform for planetary-scale geospatial analysis, which allows the process of a variety of geographical data at scale and handles large geographical datasets. GEE also provides access to numerous remotely sensed datasets and derived products, including VIIRS DNB and Global Administrative Unit Layers.

Money laundering is a challenging economic problem to model as its monetary value cannot be easily quantified. Therefore, this paper uses Suspicious Transaction Reports (STRs) as a proxy for money laundering cases. STRs are notifications of transactions that seem suspicious or could be unlawful that financial institutions submit to the financial intelligence unit (FIU). However, it is important to note that STRs are not a perfect measure of money laundering activities. Some money laundering activities may not be detected by financial institutions and therefore may not be reported as STRs. Despite these limitations, STRs remain an important tool for detecting and measuring money laundering activities, and are widely used in Anti Money Laundering (AML) efforts. The STRs data sourced from Buletin Statistik published by Indonesia Financial Transaction Report and Analysis Center (Pusat Pelaporan dan Analisis Transaksi Keuangan - PPATK) [10] - [13].

Considering the data on cases of money laundering, which have different values and trends for each province, this study uses panel data analysis to obtain an overall picture regarding the interactions between variables. In general, the panel data analysis model can be expressed in equation (1) as follows:

$$STRs_{it} = \alpha + \beta NTL_{it} + \varepsilon_{it} \tag{1}$$

Note:

$STRs_{it}$ = STRs of $i^{th}$ province, $t^{th}$ quarter
α = join intercept
β = Coefficient of regression or slope
$NTL_{it}$ = Nighttime-light Data of $i^{th}$ province, $t^{th}$ quarter
$\varepsilon_{it}$ = error term of $i^{th}$ province, $t^{th}$ quarter
$i$ = 1 to N
$t$ = 1 to T

The random effect method might be appropriate for these data because the analysis of this research concentrates on the impact of the independent variable to the dependent variable by taking into account individual characteristics (province's value) and building the assumption of uncorrelated unobservable variable to the independent variable. Multiple tests were run to determine the panel data effect technique in order to provide strong support for this assumption. Breusch-Pagan Lagrange and Haussman To determine which panel data impact will best model the data, multipliers are used [5]–[9].

This study creates a binary dummy variable in the model to denote the quarter before and after the pandemic in addition to investigating the NTL and STRs model and the link between NTL data and STRs before and after a pandemic. The baseline for the dummy variable is the time period from the first quarter of 2018 to the fourth quarter of 2019 before the covid-19 pandemic. Whilst the pandemic period ranges from Q1-2020 to Q4-2021. As a result, the model developed in this paper becomes:

$$STRs_{it} = \alpha + \beta NTL_{it} + DummyPandemic_{t} + DummyPandemic_{t} * NTL_{it} + \varepsilon_{it} \quad (2)$$

**Panel data effect test**

The results of the Breusch-Pagan Lagrange Multiplier tests reported in Table 1, indicate that the random effect is better than the common effect. The Breusch-Pagan Lagrange Multiplier shows that the test exhibits enough statistical reasons to reject the null hypothesis that leads to opting for random effects as the panel data analysis method.

Table 1. Hypothesis and significance result of panel data effect test

| Test | Hypothesis | P-value |
|---|---|---|
| Breusch-Pagan Lagrange Multiplier | H0: using Common Effect<br>Ha: using Random Effect | 1.131e-12 |

Based on the results of the random effect model testing in table 2, it shows that NTL, DummyPandemic and the interaction of NTL and DummyPandemic are statistically significant. $R^2$ shows 0.59905 which means that this model can explain the variation of the dependent variable around 60 percent.

There are 3 important findings in the results of this test. First, we obtain a positive NTL estimator and model statistical significance. These findings suggest that STRs alterations can be estimated statistically using NTL. The estimator value is positive, meaning that when there is an increase in light intensity, the NTL will increase and vice versa.

Second, the DummyPandemic variable is statistically significant and has a negative estimator value. This shows that the STRs value during the pandemic tends to be lower than before the pandemic. This situation shows that there has been a decrease in the number of suspicious transaction reports during the pandemic.

The third finding is the most interesting finding this research invents by looking at the interaction between NTL and DummyPandemic variables. By its positive estimator, these interaction variables try to say that although in general the number of STRs is slightly declining during pandemics, the NTL still consistently has the same direction with the number of STRs. This finding strengthens the first finding.

With the overall significance of the model, it can be said that NTL can be used as a predictor variable for STRs.

Table 2. Result of Random Effect Model

| | Estimate | Std. Error | Pr (>\|z\|) |
|---|---|---|---|
| Intercept | 162.426 | 74.982 | 0.030295 * |
| NTL | 254.107 | 12.834 | <2.2e-16 *** |
| DummyPandemic | -184.346 | 66.117 | 0.005301 ** |
| NTL*DummyPandemic | 128.734 | 12.027 | <2.2e-16 *** |
| $R^2$ | | | 0.59905 |
| Adjusted $R^2$ | | | 0.59668 |
| Chi Square on 3 DoF | | | 759.161 |
| p-value | | | < 2.22e-16 |
| Note on Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | |

When the model was generated by using data during the pandemic, interesting findings were made (Q1-2020 to Q4-2021). From table 3 can be seen that the model as a whole is significant and the value of adjusted R2 is quite high, namely 0.73484, which means that this model can explain variations in the dependent variable of around 73 percent. The model presents that during a pandemic, NTL is good enough to be used as a proxy for measuring STR.

Table 3. Result of Random Effect Model of STR and NTL during pandemic (Q1-2020 to Q4-2021)

| | Estimate | Std. Error | Pr (>\|z\|) |
|---|---|---|---|
| Intercept | -39.619 | 80.847 | 0.6241 |
| NTL | 254.107 | 15.136 | <2e-16 *** |
| $R^2$ | | | 0.73484 |
| Adjusted $R^2$ | | | 0.73377 |
| Chi Square on 3 DoF | | | 688.128 |
| p-value | | | < 2.22e-16 |
| Note on Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | |

Figure 2 presents the value of real STRs, while Figure 3 shows the predicted value of STRs calculated based on the model. In general, we can see that the trend of predicted value of STRs almost mimics the trend of real data especially for big provinces.

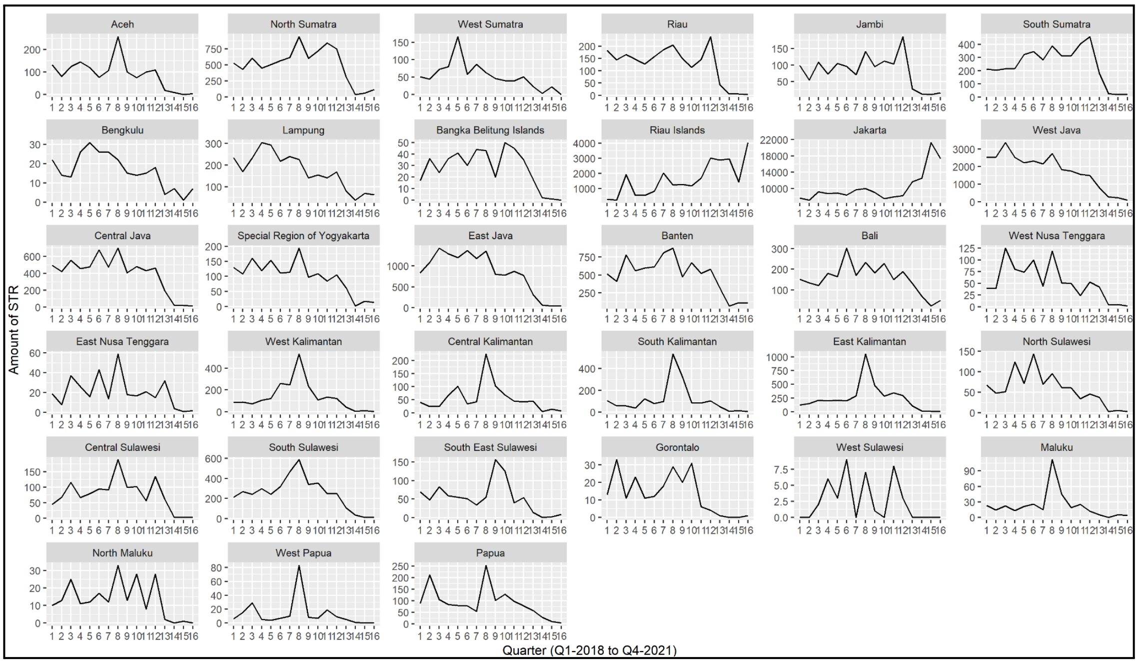


Figure 2. The trend of STR by Provinces during Q1-2018 to Q4-2021

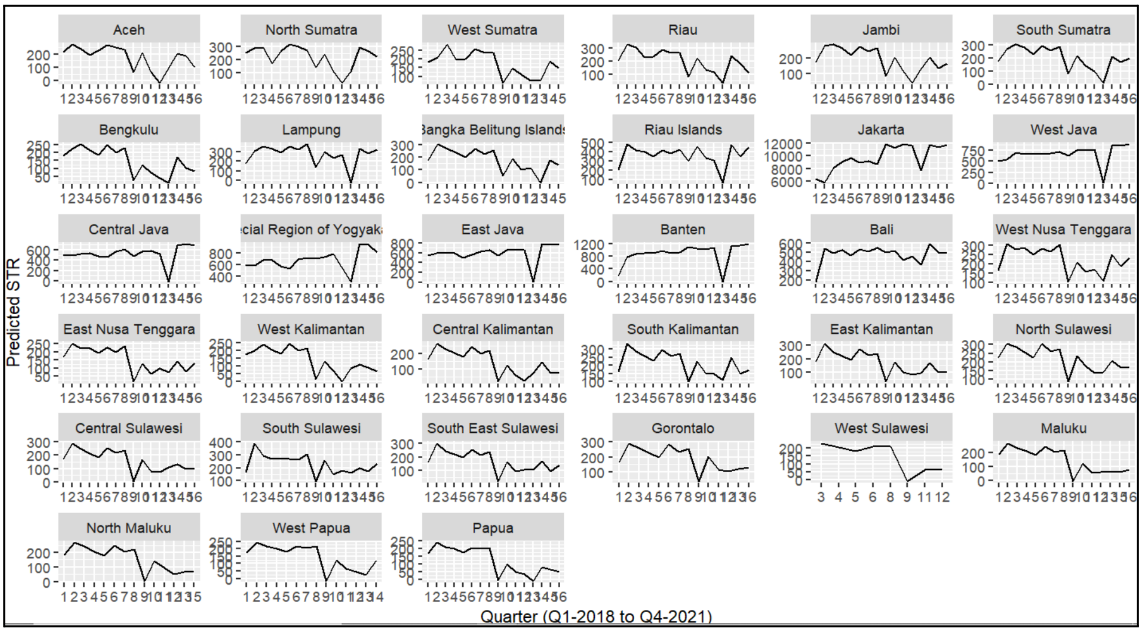


Figure 3. The trend line of predicted STR by Provinces from Q1-2018 to Q4-2021

## Conclusion

In spite of its drawbacks and cautions, using nighttime light data to estimate the volume of money laundering offenses is a viable strategy. The amount of economic activity in a specific location may be determined from the nighttime light statistics, and money laundering may be more likely to occur in economically active places. To effectively estimate the danger of money laundering in a specific location, it is crucial to combine nighttime light data with other information sources and to be aware of the methods' limits. The most thorough and precise evaluation of the risk of money laundering is likely to be obtained by using a multidisciplinary approach that takes into account several data sources. Despite these drawbacks, the application of nighttime light data as a tool for anticipating money laundering operations shows promise and has the ability to provide insightful information about this unlawful activity.

**Bibliography**


[1] OECD, "Strategic Foresight for the COVID-19 Crisis and Beyond: Using Futures Thinking to Design Better Public Policies." OECD Publishing Paris, 2020.

[2] A. Andiojaya, R. R. Nefriana, R. Rizka, and T. K. Lestari, "GDP forecasting through nighttime-light data: Evidence from Indonesia," *J. Inov. Ekon.*, vol. 7, no. 4, Sep. 2022, doi: https://doi.org/10.22219/jiko.v7i04.20239.

[3] J. Gibson, S. Olivia, G. Boe-Gibson, and C. Li, "Which night lights data should we use in economics, and where?," *J. Dev. Econ.*, vol. 149, p. 102602, Mar. 2021, doi: 10.1016/j.jdeveco.2020.102602.

[4] B. Adhikari and S. Dhital, "Decentralization and regional convergence: Evidence from night‐time lights data," *Econ. Inq.*, vol. 59, no. 3, pp. 1066–1088, Jul. 2021, doi: 10.1111/ecin.12967.

[5] M. Ekananda, *Analisis Ekonometrika Data Panel Edisi 2: Teori Lengkap Dan Pembahasan Menyeluruh Bagi Peneliti Ekonomi, Bisnis, Dan Sosial*, 2nd ed. Jakarta: Mitra Wacana Media, 2018.

[6] N. D. Nachrowi and H. Usman, *Pendekatan Populer dan Praktis Ekonometrika untuk Analisis Ekonomi dan Keuangan*, 1st ed. Jakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia, 2006.

[7] B. H. Baltagi and Q. Li, "A joint test for serial correlation and random individual effects," *Stat. Probab. Lett.*, vol. 11, no. 3, pp. 277–280, Mar. 1991, doi: 10.1016/0167-7152(91)90156-L.

[8] W. H. Greene, *Econometric analysis*, 7th ed. Boston: Prentice Hall, 2012.

[9] D. N. Gujarati and D. C. Porter, *Basic econometrics*, 5th ed. Boston: McGraw-Hill Irwin, 2009.

[10] PPATK. (2018). *Buletin Statistik Anti Pencucian Uang & Pencegahan Pendanaan Terorisme (APUPPT) Edisi Desember 2018.*

[11] PPATK. (2019). *Buletin Statistik Anti Pencucian Uang & Pencegahan Pendanaan Terorisme (APUPPT) Edisi Desember 2019.*

[12] PPATK. (2020). *Buletin Statistik Anti Pencucian Uang & Pencegahan Pendanaan Terorisme (APUPPT) Edisi Desember 2020.*

[13] PPATK. (2021). *Buletin Statistik Anti Pencucian Uang & Pencegahan Pendanaan Terorisme (APUPPT) Edisi Desember 2021.*