Amanda Pratama Putra

The Use of Mobile Positioning Data as Proxy of Economic Activity in Indonesia

Amanda Pratama Putra[1]; Heny Wulandari[2]

[1] BPS - Statistics Indonesia
[2] BPS - Statistics Indonesia

Abstract:

The Coronavirus Disease 2019 (COVID-19) pandemic has devastated the economic activity in many countries, including Indonesia. The alternative data sources to measure the economic activity especially in pandemic situation can be a valuable mine of information. In the present research, the method to measure economic activities based on mobile positioning data (MPD) has been developed. The sample dataset of mobile network subscribers activity was aggregated at the area of municipality level by the daily interval, where the activity is defined as the number of mobile transaction and detected location, and the number of the unique users. This paper shows that mobile positioning data (MPD) can be a proxy to estimate regional economic activity in Indonesia, especially during the pandemic condition.

Keywords: mobile positioning data, economic activity, proxy measurement

1. Introduction:

The first Coronavirus Disease 2019 (COVID-19) case was reported in late December 2019 in Wuhan, China, and since mid-January 2020, it has been spreading rapidly throughout China and other countries. Multiple nations have implemented mandatory quarantine or even national lockdown as the coronavirus outbreak becomes a global pandemic. In Indonesia, to combat the COVID-19 pandemic, the government issued Government Regulation Number 21 of 2020 about Large-Scale Social Restrictions/*Pembatasan Sosial Berskala Besar* (PSBB) in 31 March 2020. And this condition strongly disrupted economic activity.

In order to formulate immediate policies and strategic responses, the government requires economic situation in near real-time. However, there are several limitations to the indicators traditionally used to capture the economic situation. The indicators usually are available with substantial lags and often at the national level only (Beyer, Franco-Bedoya, & Galdo, 2020). The existing limitations encourage the exploration of other alternative data sources as a proxy for economic activities.

Recently, new approaches using big data to predict economic activity have been proposed to overcome data limitations and measurement issues. Kovacs (2019) states that the use of big data as an alternative to new data sources is important because big data can complement and enhance official statistics that already exist or are in the development stage. In addition, big data can help meet the availability of data or can be used to generate statistics that are available quickly and up to date, one of which is used in emergencies, such as natural disasters or, in this case during the pandemic.

In this paper, we focus on the exploration of mobile positioning data (MPD), a large-scale dataset of transaction records and locations of customers from cellular operators (Mobile Network Operators/MNO) which are processed and stored in a system. Mobile positioning data has been used in various contexts, such as tourism (Raun, Ahas, & Tiru, 2016), transport planning (Liu, et al., 2014), and traffic measurement (Dong, et al., 2015). Arhipova, et al. (2019) analyse the economic activity at the municipality level in Latvia based on mobile phone data. The data of person activity was aggregated at the area of each base station by the 15 minutes interval, where the activity is defined as the number of outgoing and incoming calls, sent and received short message service (SMS), as well as the number of the unique users. The study has concluded that the economic activity in municipalities can be estimated, and in particular cases, positive dynamics of regional development have been detected.

2. Data and Methodology:

A. Data

This research combines the use of mobile positioning data as alternative data sources with gross regional domestic product (GRDP) as official statistics of economic measurement.

1. Mobile Positioning Data (MPD)

Mobile positioning data (MPD) in this paper refers to the large-scale location data of subscribers of MNOs that are processed and stored in operator systems (Altin, Tiru, Saluveer, & Puura, 2015). The data for this study has been collected from Telkomsel, the biggest MNO in Indonesia, more specifically on sample dataset combination of CDR (Call Detail Record) data, signalling data, and UPCC (Unified Policy and Charging Controller) data. CDR data are only created when a mobile phone user makes a call or text message and typically consists of an anonymous (but consistent) ID of the caller and called person, the type of activity, such as a call or text message, the duration of the call, the timestamp of the call event, the identifiers of the cell tower to which the user was connected which can be linked to the exact geographical location of the cell tower. UPCC generate data from internet broadband activity of subscribers. MPD have been used to support the production of official statistics in Indonesia, one such example is a study on the use of MPD to estimate commuter flows (Putra, Setyadi, Lestari, & Esko, 2019).

For this research, we use the MPD sample dataset which consists of the number of detected subscribers aggregated by day from 1 January 2020 to 30 April 2020. It includes pre-pandemic and pandemic conditions of COVID-19 in each province in Indonesia. In this research, 5% of the data were randomly sampled with a proportional distribution following the distribution of the population. It is distributed evenly following the day and the administrative area. The number of unique users is counted during particular time providing an insight into the average activity in the area covered by a particular base station (Arhipova, Berzins, Brekis, Binde, & Opmanis, 2019).

2. Gross Regional Domestic Product (GRDP)

Data on quarterly GRDP at the provincial level is used as an indicator of regional economic activity measurement. In this study, we use quarterly GRDP data, 2020 Q1 (Januari – March) and 2020 Q2 (April – June) based on the constant price approach for each province in Indonesia. Indonesia consists of 34 provinces with various contributions of economic sectors to its GRDP. GRDP data is target data that will be used as a benchmark to evaluate MPD's ability as a proxy that describes economic conditions. For the modelling process, the GRDP value is transformed into log natural form (ln).

B. Methodology

MPD is passive positioning data that can provide real-time subscribers mobility so that it is appropriate to be used as a proxy for economic activity. The number of unique customers detected through MPD can also indirectly describe population density which is an inseparable part of economic activity. This study consists of several stages:

1. Calculate the total number of subscribers aggregated by day and province.
2. Cluster the provinces using k-means method into two clusters based on the daily pattern of unique subscribers. The median number of subscribers is normalized with Z-score normalization for every province respectively.
3. Summarise the statistics of unique subscribers from step 1 into median subscribers aggregated by quarter and province. As the result, there are 68 records consist of 34 median unique subscribers per day which represent the conditions on 2020-Q1 and another 34 records for 2020-Q2. Before it is used in the modelling stage, the value is normalized using min-max normalization method to mask the actual value.
4. Estimate simple linear ordinary least square (OLS) regressions to predict GRDP using MPD for each province clusters by following this model:

$$\ln grdp \quad = \beta_0 + \beta_1 \times cluster_2 + \beta_2 \times median\ subs + \beta_3 \times median\ subs \times cluster_2 + \varepsilon$$

5. Evaluate the regression model to check the model significance and variable significance.

3. Result and Analysis:

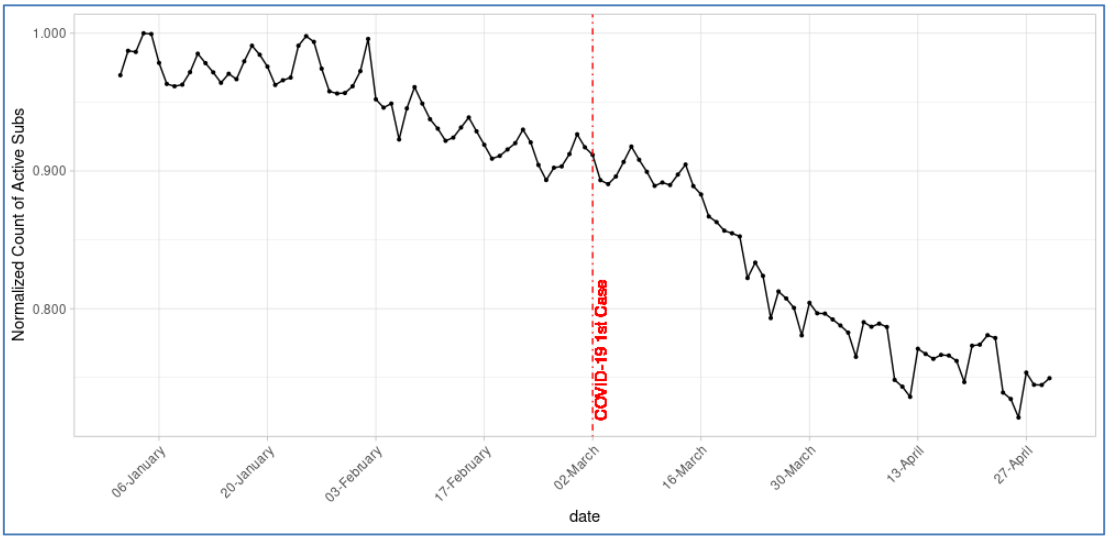

Figure 1. Normalized count of unique users

Based on the MPD, the number of unique users declined gradually after the first case of coronavirus was detected in Indonesia in early March 2020. In a short time, it dropped drastically when the social restriction policy was implemented in mid-March 2020 as can be seen in figure 1. Subscribers activity detected fell by nearly 20 percent in April 2020.

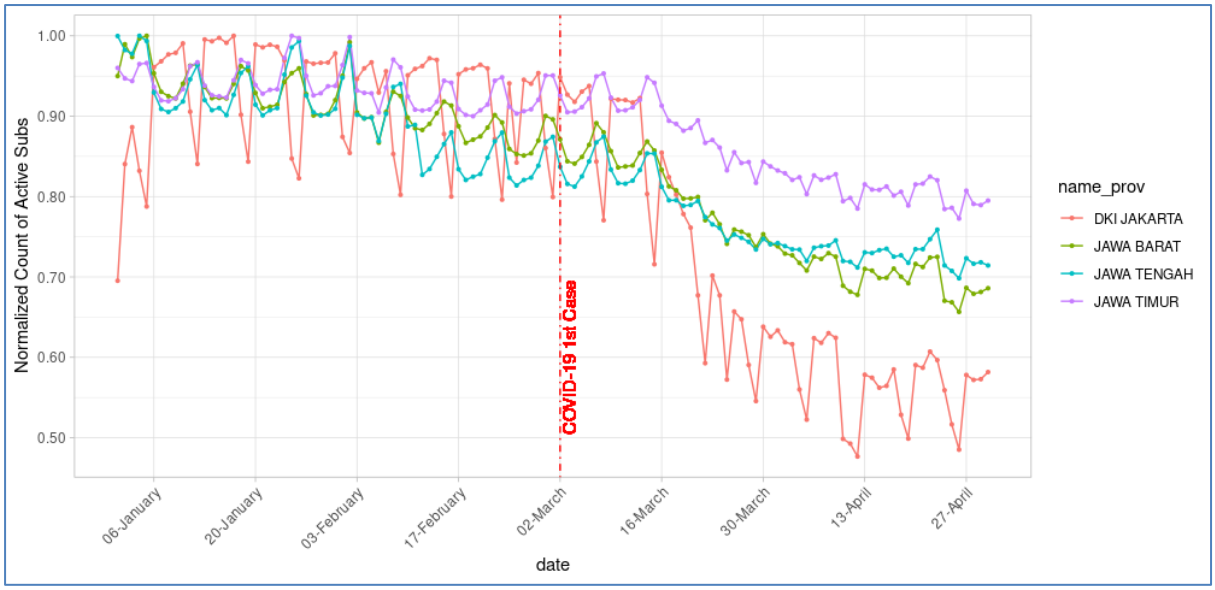

Figure 2. Normalized count of unique users in several provinces

If we look into the pattern in several provinces presented in Figure 2, it is clear that the decline in the number of unique users in DKI Jakarta Province as the provincial capital was steeper compared to other large provinces such as Jawa Barat, Jawa Tengah, and Jawa Timur. This indicates that the COVID-19 pandemic may have a greater economic impact on the centre of economic activity. From this chart, we also found the indication of seasonality based on weekday. For example, the average number of unique subscribers in DKI Jakarta at the weekends was lower than weekdays, but the opposite pattern occured in Jawa Barat, where the number of unique subscribers was higher on weekends than on weekdays.

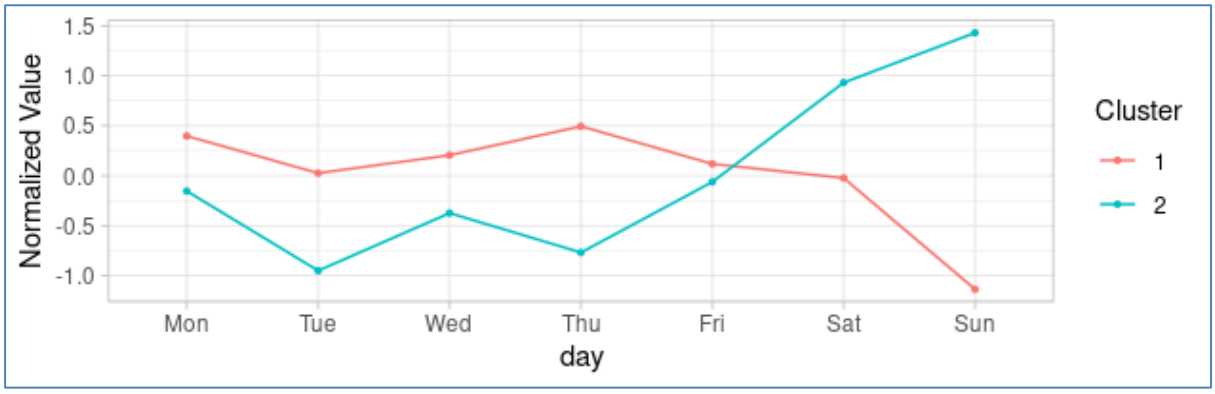

Figure 3. Clustering province by similar subscriber activity

Cluster analysis is conducted to group the province with similar MPD pattern based on daily aggregate. The cluster analysis uses K-means method (MacQueen, 1967) to find each province's similarity in terms of daily pattern. Based on silhouette index (Rousseeuw, 1987) to find optimal number of clusters, 2 clusters are established in which each group consist of 20 provinces and 14 provinces respectively. The comparison of daily pattern in each cluster is presented in figure 3. The cluster 1 groups provinces with higher subscriber activities in weekday meanwhile the cluster 2 groups provinces with higher subscriber activities in weekend. It shows that the cluster 1 is a group of provinces with high work activities so that the mobility pattern tends to be more active on weekdays.

In this study, the regression model explains the relationship between MPD and the constant price of GRDP with consideration of province cluster. The results show that the number of subscribers from MPD is significantly associated with the constant price of GRDP with coefficient determination of 0.7976. It means the model explains 79,76 percent of GRDP variability. Moreover, the residuals of the model meet the normality assumption confirmed by Shapiro-wilk test (Shapiro & Wilk, 1965).

Regression Parameter	Estimate	p-value
Intercept (β_0)	16.4097	<.0001***
Cluster$_2$ (β_1)	0.4797	.0058**
Median subs (β_2)	12.0977	<.0001***
Median subs * Cluster$_2$ (β_3)	-8.5714	<.0001***

Table 1. Regression model result

From the regression model summary, the number of subscribers is significantly associated with constant price of GRDP (p-value < 0.05). The province cluster which represents the different patterns of daily activity also affects the relationship between the number of subscribers and GRDP. The estimated GRDP for cluster 2 is smaller than cluster 1 which shows the negative interaction coefficient (β_3) between the cluster variable and the number of subscribers as shown in the table 1.

4. Discussion and Conclusion:

The COVID-19 pandemic has caused a decline in economic activity in Indonesia, which is reflected by a decrease in the number of subscribers activity captured in MPD. This study finds a significant correlation between the number of MNO subscribers and the value of constant price of GRDP in 34 provinces in Indonesia during the Q1 and Q2 periods of 2020. In addition, cluster analysis to group the 34 provinces into two clusters based on their daily MPD pattern has an important issue in the regression model which increase the coefficient determination of the regression model. Based on this study, it can be concluded that MPD can be a proxy in predicting the pattern of regional economic activity in Indonesia, especially when there are economic changes due to changes in human mobility patterns. Nevertheless, more research relying on larger samples and on longer observation periods will be useful.

References:

Altin, L., Tiru, M., Saluveer, E., & Puura, A. (2015). Using Passive Mobile Positioning Data in Tourism and Population Statistics. *the New Techniques and Technologies for Statistics 2015 proceedings.* Brussel: Eurostat.

Arhipova, I., Berzins, G., Brekis, E., Binde, J., & Opmanis, M. (2019). Mobile Phone Data Statistics as Proxy Indicator for Regional Economic Activity Assessment. *International Conference on Finance, Economics, Management and IT Business (FEMIB)*, 27-36.

Beyer, R., Franco-Bedoya, S., & Galdo, V. (2020). Examining the economic impact of COVID-19 in India through daily electricity consumption and nighttime light intensity. *World Development*.

Dong, H., Wu, M., Ding, X., Chu, L., Jia, L., Qin, Y., & Zhou, X. (2015). Traffic zone division based on big data from mobile phone cellular base stations. *Transportation Research,58*, 278–291.

Kovacs, K. (2019). Big data for official statistics : The UN perspective. *Regional workshop on the use of mobile phone data.*

Liu, F., Janssens, D., Cui, J. X., Wang, Y. P., Wets, G., & Cools, M. (2014). Building a validation measure for activity-based transportation models based on mobile phone data. *Expert Systems with Applications, 41(14)*, 6174–6189.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297). University of California Press.

Putra, A. P., Setyadi, I. A., Lestari, T. K., & Esko, S. (2019). Measuring Commuting Statistics in Indonesia Using Mobile Positioning Data. *Asia Pacific Economic Statistik Week.* Bangkok: UNESCAP.

Raun, J., Ahas, R., & Tiru, M. (2016). Measuring tourism destinations using mobil tracking data. *Tourism Management, 57*, 202-212.

Rousseeuw, P. J. (1987). Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics*, 53–65.

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 591-611.