

A REPEATED CROSS-SECTIONAL MODEL FOR ANALYZING UNEMPLOYMENT DATA IN BOGOR*

Ulfah Sulistyowati¹, Khairil Anwar Notodiputro^{2‡}, I Made Sumertajaya³

¹Department of Statistics, IPB University, Indonesia, ulfah_sulistyowati@apps.ipb.ac.id

²Department of Statistics, IPB University, Indonesia, khairil@apps.ipb.ac.id

³Department of Statistics, IPB University, Indonesia, imsjaya@apps.ipb.ac.id

‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 4 No 2 (2020), 254 - 260

Copyright © 2020 Ulfah Sulistyowati, Khairil Anwar Notodiputro, I Made Sumertajaya. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In general, the form of data encountered in statistical problems is panel data and cross-sectional data. There are times in certain conditions, the data formed in the form of a combination of panel data with cross-sectional data, which is commonly referred to as repeated cross-sectional data. Repeated cross-sectional data is often done in research with individual observations. In this study, a repeated cross-sectional analysis was carried out using a fixed influence model with observations in the form of an area (village) in Bogor, West Java to analyze unemployment factors. The results obtained are that ongoing village development affects the unemployment rate in Bogor.

Keywords: cross-sectional data, unemployment, repeated.

1. Introduction

The surveys conducted generally produce panel data, namely combination data in the form of cross-section and time series. So that makes it easier for researchers to look for the development of an object in a different year. Panel data is used if a researcher wants to know the analysis of differences between objects or between times. However, currently, there is a survey method that results in panel data becoming incomplete. So that formed cross-sectional data and panel data on the data collection.

* Received Jul 2019; Accepted Jul 2020; Published online on Jul 2020

Repeated Cross-Section analysis can be used on that data. By way of copying the data into several groups, then modeled based on the formed cohort. A cohort is a group of people in a certain population and geographical area, which is defined as having the same life event in a certain period. Usually, research that uses cohorts is a long-term study conducted every 5 years. The focus of cohort analysis is a particular cohort or category, not on specific individuals. Usually, the cohort used is all people born in the same year (called a birth cohort), all people employed at the same time, all people who retire in the span of one or two years, or people who graduate in the same year. Unlike pure panel studies, the sample of this study does not need to be the same person but the group that experiences the same daily life events (Neuman, 2015).

Rachmi et al. (2016) also conducted research using repeated cross-sectional on Stunting, Underweight, and Overweight in Children Age 2 - 4.9 years in Indonesia: Prevalence Trends and Associated Risk Factors. Research on repeated cross-sectional that has been carried out so far still uses individual objects, while the research conducted using object area (village) and no one has done the research.

The research conducted is repeated cross-sectional analysis that occurs in objects in the form of areas. The researcher wanted to apply the repeated cross-sectional analysis to the Bogor Sakernas data for 2015 - 2016. Individuals who were the object of research were villages in the Regency and City of Bogor. With the condition of the village being more complex than the individual, can repeated cross-sectional still be used?

2. Methodology

Research carried out using Bogor Sakernas data from 2015 – 2016 (BPS, 2015). Also, it uses supporting data from the 2014 Podes data and District data in figures 2016 - 2017. The research conducted examines unemployment factors in Bogor. The response variable is the unemployment rate in the village originating from Sakernas data.

The study consisted of 1 response variable and 5 explanatory variables. The response variable is the unemployment rate at the village level in Bogor. The explanatory variables (Table 1) are the number of industries in the village (X1), the number of groceries (X2), the number of shops (X3), village income (X4) and village area (X5). Besides, it was added to by weighting in the form of the population in each village. In this study using a repeated cross-sectional data analysis in which observations must be cohort. Cohorts that will be formed are based on 4 criteria, namely the location of the village in the Regency or City (D1), the village is located near or far from the capital (D2), ongoing village development or not (D3) and the availability of internet cafes in the village (D4). The data used is raw data consisting of February 2015, August 2015, February 2016, and August 2016. In August 2015, more data was collected compared to February 2015. With the condition of the census block samples in February 2015 it was taken back in August 2015. So when aggregated to the village level, the villages taken in February 2015 will be taken back in August 2015.

Table 1: List of explanatory variables.

variable	Information	satuan
X1	Number of industries	unit
X2	Number of groceries	unit
X3	Number of shops	unit
X4	Village income	rupiah
X5	Area of village	hectar

with weight in the form $X_i = \frac{x}{\text{number of villager}} \times 1000$

Before doing the modeling, the data formed will be copied in several cohorts. The cohort criteria used are classifications that have not changed for a long time. The researcher used 4 criteria for building a cohort. The Cohort formed are presented in Table 2.

Table 2: List of Cohorts in Bogor.

District / City	Position	Internet cafe	Village Development	Cohorts
District	Close	Available	Available	1
District	Close	Available	Unavailable	2
District	Distant	Available	Available	3
District	Distant	Available	Unavailable	4
District	Distant	Unavailable	Available	5
City	Close	Unavailable	Available	6
City	Close	Available	Unavailable	7
City	Close	Unavailable	Unavailable	8

The cohort is formed based on 4 criteria, namely the location of the village in the regency or city, the position of the village near or far from the city center, there is a village development or not and there are internet shop services or not. From the cohort formed the model is made as follows:

$$\bar{y}_{ct} = \bar{x}_{1ct}\beta_{1t} + \bar{x}_{2ct}\beta_{2t} + \bar{x}_{3ct}\beta_{3t} + \bar{x}_{4ct}\beta_{4t} + \bar{x}_{5ct}\beta_{5t} + \bar{\alpha}_{ct} + \bar{u}_{ct}$$

where:

\bar{y}_{ct} = average unemployment rate at c-cohort at t

\bar{x}_{1ct} = average number of industries in c-cohort at t-time

\bar{x}_{2ct} = average number of grocery in the c-cohort at t-time

\bar{x}_{3ct} = average number of stores in c-cohort at t-time

\bar{x}_{4ct} = average village income in c-cohort at t

\bar{x}_{5ct} = average village area in c-cohort at t

$c = 1, 2, 3, \dots, 8; t = 1, 2, 3, 4;$

$$\bar{\alpha}_{ct} = \delta_1 d_{c1} + \delta_2 d_{c2} + \delta_3 d_{c3} + \delta_4 d_{c4} + \delta_5 d_{c1} t_1 + \delta_6 d_{c1} t_2 + \delta_7 d_{c1} t_3 + \delta_8 d_{c2} t_1 + \delta_9 d_{c2} t_2 + \delta_{10} d_{c2} t_3 + \delta_{11} d_{c2} d_3 + \delta_{12} t_1 + \delta_{13} t_2 + \delta_{14} t_3$$

where:

d_1 = district / city

d_2 = the position of the village is near or far from the city center

d_3 = village development or not

d_4 = availability of internet cafe in the village

With assumption $E\{x_{it}u_{it}\} = 0$

In addition to the above model, a model is formed for every time

$$\bar{y}_c = \bar{x}'_{1c}\beta_1 + \bar{x}'_{2c}\beta_2 + \bar{x}'_{3c}\beta_3 + \bar{x}'_{4c}\beta_4 + \bar{x}'_{5c}\beta_5 + \bar{\alpha}_c + \bar{u}_c$$

with $c = 1, 2, 3, \dots, 8$

$$\bar{\alpha}_{c1} = \delta_1 d_{c1} + \delta_2 d_{c2} + \delta_3 d_{c3} + \delta_4 d_{c4}$$

If the individual effect α_i does not correlate with the explanatory variable (x'_{it}), the model can be easily estimated consistently from cross-sectional repeated by collecting all observational data and estimating using ordinary least squares with $\alpha_i + u_{it}$ as an error value.

The main problem when estimating β of the equation is that $\bar{\alpha}_{ct}$ depends on unobserved t and must correlate with \bar{x}'_{ct} (if α_i correlates with x'_{it}). Therefore, treating $\bar{\alpha}_{ct}$ as part of term random error tends to cause inconsistent estimates. One alternative that can be done is to consider $\bar{\alpha}_{ct}$ as a fixed unknown parameter assuming that the range of periods is ignored. If the average cohort consists of very many individual observations, then this assumption looks more plausible and the natural estimator for β is the estimator in the pseudo panel (Verbeek & Vella, 2005), as follows :

$$\hat{\beta}_w = \left(\sum_{c=1}^C \sum_{t=1}^T (\bar{x}_{ct} - \bar{x}_c)(\bar{x}_{ct} - \bar{x}_c)' \right)^{-1} \sum_{c=1}^C \sum_{t=1}^T (\bar{x}_{ct} - \bar{x}_c)(\bar{y}_{ct} - \bar{y}_c)$$

In this case, $\hat{\beta}_w$ is an estimator in the pseudo panel.

Where \bar{x}_c = is the time average of the average cohort observation in the cohort c . Deaton (1985), based on the asymptotic estimator when the number of individuals (N) tends to be infinity with constant cohort size. To find out the effect occurs when then the equation formed will be broken down for each time. This equation will be compared with the model that has been formed by the researcher. The steps of this study will be explained as follows:

1. Form a cohort on data that has been formed.
2. Data exploration, see the development of the unemployment rate every year.
3. Estimating a random model for data that has been formed
4. Form the equation for each cohort assuming $E\{x_{it}u_{it}\} = 0$
5. Determine at what time, the factors that have been determined affect the unemployment rate
6. Check the value of RMSE and R2 on each model, the smallest value is the best model that can be used to estimate the unemployment rate in each cohort.

3. Results and Discussion

Research conducted before we go into modeling looks at the comparison of the actual unemployment rate published from BPS with the estimated value of the specified model.

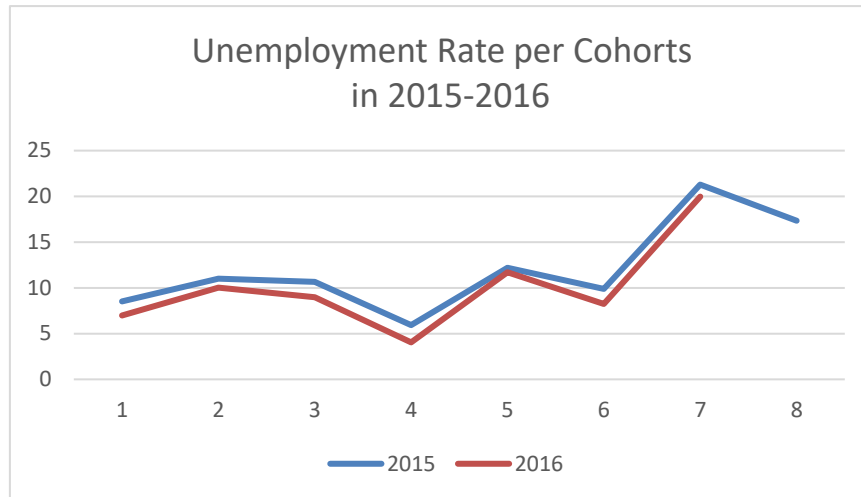


Figure 1: Unemployment rate per cohorts in 2015 – 2016.

In the graph above, it can be seen that the unemployment rate in 2016 has decreased. The highest unemployment rate occurred in the 7th cohort, where the village was in Bogor City, close to the capital city, there was an internet shop service available in the village and there was no ongoing village development. The next highest unemployment rate is in cohort 5, the village is in Bogor Regency, far from the capital, there is no internet shop available and ongoing village development in the village. With the same pattern between 2015 and 2016, the unemployment rate in 2016 could be the result of policies implemented in 2015. It can be said that the policies carried out in 2015 gave positive results on the unemployment rate in Bogor. In 2016, the 8th cohort did not have a point, because the data in 2016 was less, so the data in cohort 8 were not taken in 2016.

Estimates were carried out using a Repeated Cross Section Data Analysis, assuming that data is normal for each time. The models obtained are as follows:

$$\begin{aligned}
 Y_{ct} = & -1,44t_1 - 3,51t_2 + 0.75t_3 + 7.00d_1^* + 0.76d_2 + 0.42d_3 + 0.924d_4 - 4.02d_1t_1 \\
 & - 4.36d_1t_2 + 1.22d_1t_3 + 1.18d_2t_1 + 0.52d_2t_2 - 2.16d_2t_3 + 2.43d_1d_2 \\
 & - 0.160X_1 + 0.114X_2 + 0.066X_3 + 0.046X_4 + 1.88X_5
 \end{aligned}$$

d_1^* = significant to α is 5%

In the model, only village development and area affect the unemployment rate. Whereas villages with the development of village development compared to villages that are not under construction of villages, the unemployment rate is 7% smaller compared to villages that are not under construction.

3.1 Fixed Effect Model

Apart from being seen from the random model above, modeling for the Fixed Model is done with time. The time used in this study is 4, namely February 2015, August 2015,

February 2016, and August 2016. The models formed for each year are different are presented in Table 3.

Table 3: The fixed model formed for each year.

Variable	Feb-15		Agu-15		Feb-16		Agu-16	
	Coef	p-value	Coef	p-value	Coef	p-value	Coef	p-value
Constant	0,86	0,877	9,38	0	10,3	0,008	5,36	0,093
X1	-2,06	0,159	-0,1	0,467	-0,12	0,904	1,001	0,253
X2	0,558	0,259	0,0364	0,663	-0,032	0,924	-0,176	0,541
X3	1,17	0,419	-0,15	0,414	-0,219	0,819	2,34 *)	0,009
X4	0,013	0,975	0,0091	0,874	0,037	0,901	-0,273	0,302
X5	11,85	0,093	-2,38	0,213	3,76	0,403	0,04	0,992
cluster								
1	3,55	0,556	2,36	0,26	-7,48	0,068	-4,02	0,246
2	-3,1	0,82	1,37	0,801	-12,27	0,167	0,08	0,992
3	-3,88	0,597	0,47	0,846	-2,81	0,551	1,42	0,727
4	3,62	0,673	-0,14	0,958	18,31	0,003	10,19	0,042
5	-5,3	0,612	-5,04	0,295	4,09	0,549	-5,79	0,33
6	-	-	0,87	0,689	-	-		

The estimation of the model above shows that in 2016 the unemployment rate in Bogor has decreased. Judging from the constant value in August 2016, the unemployment rate without the influence of any variable shows 5.36%. It could be that the number is the impact or effect of government policies in the form of village funds that have been implemented since 2014. Besides, there is one variable that affects the unemployment rate in the village, namely the number of shops in the village. The number of shops in the village has a positive effect on the unemployment rate, increasing one shop in the village area, then the unemployment rate will increase by 2.34%. So it can be concluded that the number of shops in the village is not one solution that can reduce unemployment. The more shops, the competition between the stores are also getting higher, so the store with a concept that is not too attractive will soon close and generate new unemployment.

Table 4: List the value of RMSE for all models.

Model	MSE	RMSE
Random	79.907	8.939
Fixed Model on February 2015	137.620	11.731
Fixed Model on August 2015	76.687	8.757
Fixed Model on February 2016	73.457	8.571
Fixed Model on August 2016	39.918	6.318

3.2 Best Model

To find out the best comparison of models, it can be seen from the value of R² or the value of RMSE. In Table 4, it can be explained that in the random model compared to the fixed model, it is still good for the fixed model when in August 2016. The possibility is that there is an improvement in unemployment in that year.

3.3 Unemployment Rate in Bogor

The highest unemployment rate in Bogor is in the 7th cohort, where the village is located in the city of Bogor, close to the government system, there is an internet cafe and there is no ongoing village development. Some of the villages included in the cohort are Bondong Village and Batutulis Village, both of which are in South Bogor. The second is in cohort 5, where villages are in Bogor Regency, far from the government, there are no internet cafes and village development is underway. The villages included in the group are Mekarjaya Village, Bunar Village, and Parakan Village. If analyzed by the model above, at that time the villages were probably experiencing the exploitation of taverns that were not too much enjoyed, resulting in a lot of unemployment in the village. Besides, in several articles on government activities in Bogor, said that in the South Bogor area there were 200 unemployed people due to layoffs due to the bankruptcy of the companies they occupied. So that at the time of the South Bogor Subdistrict was at a high unemployment rate.

4. Conclusion

That the Repeated Cross-Sectional method can be used to predict the level of unemployment at the area level, not just the individual level. With some provisions that must be followed, one of them forms a cohort for the individuals studied, besides that the time spent researching as much as possible is a period of research, such as once a year, or once every two years.

References

- [BPS] Badan Pusat Statistik Provinsi Jawa Barat. (2015). *Keadaan Angkatan Kerja di Jawa Barat*. Jawa Barat (ID): BPS.
- Deaton, A. (1985). Panel Data from Time Series of Cross Section. *Journal of Econometrics*, 30(1–2): 109–126.
- Neuman, A. (2015). *Social Research Methods: Qualitative and Quantitative Approaches*. US: Allyn and Bacon.
- Rachmi, C. N., Agho, K. E., Li, M., & Baur, L. A. (2016). Stunting, underweight and overweight in children aged 2.0–4.9 years in Indonesia: prevalence trends and associated risk factors. *PloS One*, 11(5): e0154756.
- Verbeek, M., & Vella, F. (2005). Estimating Dynamic Models from Repeated Cross-Sections. *Journal of Econometrics*, 127(1): 83–102.