International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Multilevel Panel Data Modelling for Unemployment Rate in Indonesian

Rizky Mandasari[a], Indahwati[b]*, Anang Kurnia[c]

[a,b,c]*Department of Statistics, Faculty of Mathematics and Natural Science, Bogor Agriculture University, Bogor, Indonesia*

[a]*Email: rizkykymandasari11@gmail.com*

[b]*Email: indahwati@apps.ipb.ac.id*

[c]*Email: anangk@apps.ipb.ac.id*

Abstract

Indonesia as a developing country is often faced with the problem of unemployment. Low unemployment rates can reflect good economic growth and reflect the well-being of the population, thus each country will attemp to lower the unemployment rate, including Indonesia. The unemployment rate of provinces in Indonesia from 2005 to 2015 observed every 6 months can be viewed as panel data. This study aims to model the trend of unemployment rate in Indonesia and external factors that influence it through multilevel modeling approach, where unemployment rate value at any point of time (level 1) is nested within the province (level 2). Selected models are models with random intercept and random time slopes which indicate there are varieties of unemployment rate between provincies that occurs at initial values and the rate of change over time. The average decline is about 0.05% every 6 months. Provinces with a high initial unemployment rate value were more likely to decline the unemployment rate compared to provinces with low initial unemployment rate values. After being controlled by the effect of time, the fixed effect that significantly affects unemployment rate is investment and population density. In addition, unemployment rate in August tended to be higher than unemployment rate in February.

Keywords: Panel Data; Multilevel; Repeated Measurement; Unemployment Rate.

----

* Corresponding author.

1. Introduction

Unemployment is one of the problems that is often faced by developing countries, one of which is Indonesia. A low unemployment rate may reflect good economic growth, and may reflect improvements in life quality of the population and increase in income distribution so that the welfare of the population increases. Based on data from the Indonesian Central Bureau of Statistics, the unemployment rate in Indonesia generally shows a downward trend from 2005 to 2015 [1], with varying degrees of decline across provinces.

The unemployment rate of the provinces in Indonesia from 2005 to 2015 is panel data that can be viewed as repeated measurement data. In repeated measurement data with time as replications, it is possible to study changes in response between times along with the factors that influence these changes, both at the population level and individual level. Multilevel modelling is an approach that can be used to model repeated measurement data, where the value of observations at each time point (level 1) is nested in the research subject (level 2) [4]. With this approach, it is possible to capture the correlation between observations within the same individual. In addition, repeated measurement data modelling with a multilevel modelling approach can be used to capture the varieties of growth patterns between different individuals. More generally, multilevel models can be expressed as Linear Mixed Models [2]. In this study, repeated measurement data for unemployment rate in each province observed every 6 months were analysed using a multilevel modelling approach where the unemployment rate value at each time point (level 1) was nested in the province (level 2). Some studies that use panel data include Zulvia using the multilevel modelling approach [7]. The results show that multilevel modelling with schools nested in districts / cities and the effect of repeated time can capture the varieties of average National Examination scores in each school. Anggara [3] used Generalized Estimating Equations (GEE) and Generalized Linear Mixed Models (GLMM) in modelling poverty panel data in NTT Province. Some previous research on the unemployment rate include Muchlisoh using small area estimation approach for estimation unemployment rate based on unit level model with first order autoregressive time effects [8]. Noviyanti states that economic growth, inflation, investment and minimum wages affect unemployment rate [9] . Mariana using the General Spatial Model (GSM) states that the percentage of poor population and district minimum wages has a positive effect on unemployment rate, while the GRDP at constant prices and the percentage of population working in the agricultural sector negatively affect unemployment rate [5]. Prihatiningsih using Geographical Weighted Regression (GWR) states that population density, percentage of poor population, district/city minimum wages and percentage of large industrial business units per working age population have a positive effect on unemployment rate [6]. This study aims to model the pattern of unemployment rate trend in Indonesia and discover the external factors that influence it with the multilevel modelling approach. Based on the previous presentation and the completeness of the data, the external factors included in the modelling were population density, percentage of poor population, minimum wages, GDP and investment.

2. Methodelogy

2.1. Data Sources

The data used in this study is secondary data in the form of the unemployment rate of the provinces in Indonesia and several other variables namely Percentage of Poor Population (PPP), Population Density (PD), Provincial

Minimum Wage (PMW), Gross Regional Domestic Products (GRDP) and Investment (INV). The variables used are obtained from the Central Bureau of Statistics of the Republic of Indonesia (http://www.bps.go.id/).

2.2. Method of Analysis

The stages of analysis used in this study are as follows:

a. Data preparation: Data consists of 22 series from 2005 to 2015. Data from 2005 - 2014 (20 series) are used as training data and 2015 (2 series) data are used as testing data.
b. Data exploration. This stage is used to determine the structure of the model to be formed.
c. Stages of unemployment rate modelling :

1. Formation of the initial model used to observe the varieties of unemployment rate described by the province using intraclass correlation [7].

$$\mathrm{ICC} = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0}+\sigma^2_{u_e}} \quad (1)$$

The coefficients above are obtained from Level 1 and Level 2 models as follows:

Level 1 model:

$$Y_{it} = \beta_{0i} + e_{it} \quad (2)$$

Level 2 model:

$$\beta_{0i} = \gamma_{00} + u_{0i} \quad (3)$$

Where Y_{it} : the unemployment rate value in the province of i, i = 1, 2, ..., 34 in time (series) t, t = 1,2, ..., 20, β_{0i} : intercept for the province i,e_{it} : error at level 1, u_{0i}: the effect of random intercepts, $u_{0i} \sim N(0, \sigma^2_u)$.

2. Model formation with time explanatory variable at level 1:

Level 1 model :

$$Y_{it} = \beta_{0i} + \beta_1 T_{it} + e_{it} \quad (4)$$

Level 2 model :

$$\beta_{0i} = \gamma_{00} + u_{0i} \quad (5)$$

3. Selection of random effect.

The possibility of the model being formed is compared to observing the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) values and Likelihood Ratio Tests (LRTs), with the following LRTs [2]:

$$\mathrm{LRTs} = -2\log\left(\frac{\mathrm{L_{nested}}}{\mathrm{L_{reference}}}\right) \sim\lambda^2_{\mathrm{df}} \qquad (6)$$

4. The selection of level 1 explanatory variables using forward selection method.

Level 1 model:

$$\mathrm{Y_{it}} = \beta_{0i} + \beta_{1i}\mathrm{T_{it}} + \beta_2\mathrm{PPP_{it}} + \beta_3\mathrm{PMW_{it}} + \beta_4\mathrm{GRDP_{it}} + \beta_5\mathrm{INV_{it}} + \mathrm{e_{it}} \qquad (7)$$

Level 2 model:

$$\beta_{0i} = \gamma_{00} + u_{0i} \qquad (8)$$

$$\beta_{1i} = \gamma_{10} + u_{1i} \qquad (9)$$

with $\binom{u_{oi}}{u_{1i}} \sim N(\mathbf{0},\mathbf{\Sigma})$; $\mathbf{\Sigma} = \begin{pmatrix} \sigma^2_{u0} & \sigma_{u0u1} \\ \sigma_{u0u1} & \sigma^2_{u1} \end{pmatrix}$

5. Forming a model with the 2nd level explanatory variable.

6. Comparing the possible models that are formed using the goodness of fit model.

d. Conducting interpretation of the final model and calculate model accuracy level using Root Mean Squared Error (RMSE).

$$\mathrm{RMSE} = \sqrt{\frac{\Sigma_{i=1}^{34}\Sigma_{t=1}^{20}e_{it}^2}{780}} \qquad (10)$$

$\mathbf{e_{it}}$ is the difference of observation value with estimate value in province i and time t.

e. Performing model validation on 2015 data using Mean Absolute Percentage Error (MAPE), MAPE indicates how much error in forecasting is compared to the actual value.

$$\mathrm{MAPE} = \frac{\mathbf{\Sigma_{i=1}^{34}\Sigma_{t=1}^{2}\left|\frac{e_{it}}{y_i}\right|}}{\mathbf{78}} \qquad (11)$$

3. Result and Discussion

3.1. Data Exploration

Based on data from SAKERNAS published by Indonesian Central Bureau of Statistics BPS, the unemployment rate in Indonesia in 2005-2015 was published twice a year in February and August. Figure 1 shows that each province has diverse pattern of unemployment rate reduction that generally shows a down linear pattern to time.

The varieties of this unemployment rate decreasing pattern will be captured using a random time effect structure that is a random linier time slope. In addition to the random linear time slope is also added the effect of random intercept because based on figure 1 it appears that the initial value of unemployment rate varies for each province. In Figure 1 it can also be seen that overall unemployment rate in Indonesia tends to be low in February and then increases in August, therefore in the modelling stage, a dummy variable is added to capture the phenomenon.

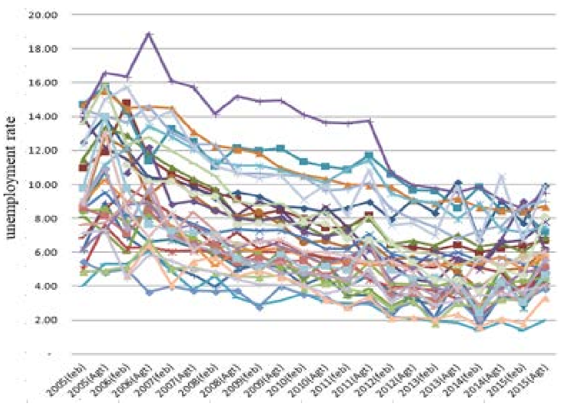

Figure 1: The trend of unemployment rate based on province in Indonesia in 2005-2015

Figure 2 shows that based on island geography, Java generally had the highest unemployment rate in 2005-2015, while the lowest unemployment rate occurred in the islands of BALI and LOMBOK (NTT and NTB Provinces). Figure 2 also shows that the six (group) islands in Indonesia have a different pattern of unemployment rate reduction in each year so that based on the exploration in Figure 2 the modelling stage is used for dummy variables for the six (groups) islands which is an explanatory level 2, since island geography is a fixed and unchanging variable over time.

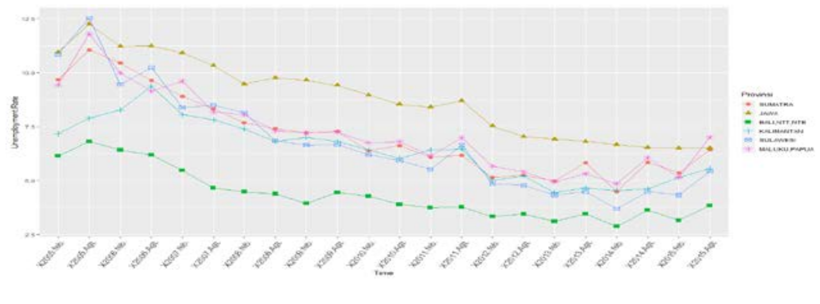

Figure 2: Average unemployment rate of six (groups) islands in Indonesia

Figure 3 shows that as a whole Banten province has the highest unemployment rate in 2005-2015 when compared to other provinces in Indonesia where the highest unemployment rate occurred in August 2006 in Banten province which was 18.91%. NTT and BALI provinces generally tend to have the lowest unemployment rate compared to other provinces in Indonesia from 2005-2015. Based on Figure 3 the average unemployment rate in Indonesia also shows decreasing pattern every year. This indicates that it is necessary to add fixed effect structure of time into the modelling.

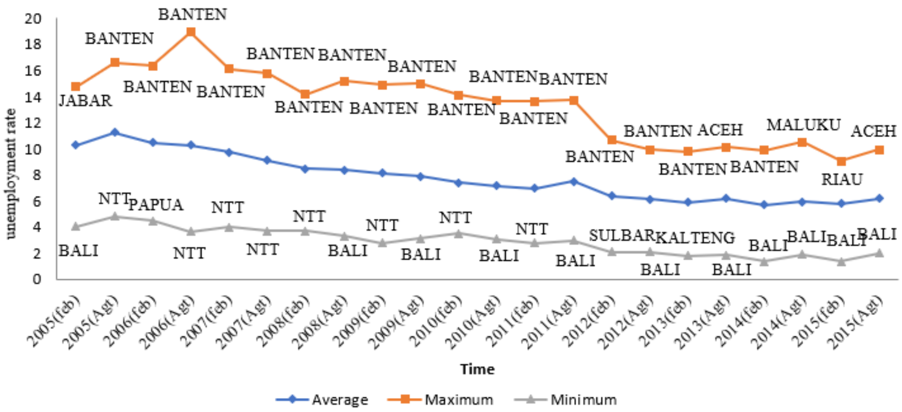

Figure 3: The maximum and minimum value of unemployment rate in Indonesia

3.2. Modelling of Unemployment Rate

Based on the modelling stages, 3 models were then compared using goodness of fit criteria presented in table 1.

Table 1: Comparison of possible models

Goodness of fit	Model 1	Model 2	Model 3
Log Likelihood	1981.1	1071.4	1065.9
AIC	1993.1	1089.9	1093.9
BIC	2002.3	1103.2	1115.3
RMSE	0.8915	0.7089	0.7097

Model 1 is a model with random intercept effect and random linear time slope with fixed effect of time, model 2 is model 1 added fixed effect on population density, investment and month dummy variable with value of 0 for February and 1 for August. The model 3 is model 2 plus the fixed effect of level 2 explanatory variables, namely the geographic code of the Indonesian region which is divided into 6 (groups) islands namely Java, Sumatra, Kalimantan, Sulawesi, (Maluku and Papua), (Bali and Lombok (NTT and NTB) . Based on Table I it can be seen that model 2 has smaller good value than models 1 and 3. Testing LRTs between model 1 vs model 2 and model 1 vs model 3 produces p-value that is equal to 0.000, thus rejecting H0 which mean it is needed to add explanatory variables as fixed effect found in model 2 and model 3. In the LRTs test between model 2 and model 3, the p-value is 0.358, the conclusion is H1 is not rejected, this means that there is no need to add an explanatory variable to model 3 or it can be said that there is no difference on unemployment rate in the inter-island region in Indonesia. Based on the goodness of fit model criteria, it can be concluded that the chosen model is model 2. Model 2 also has small RMSE value compared to models 1 and 3, meaning that the estimated value of the unemployment rate generated by model 2 is closer to the actual value than model 1 and model 3.

3.3. Final Model Interpretation of Unemployment Rate

Based on previous reviews, the final model of Unemployment Rate (UR) in Indonesia is as follows:

$$\widehat{\mathrm{UR}}_{\mathrm{it}} = 9.7119 - 0.05011\ \mathrm{T}_{\mathrm{it}} + 0.000047\ \mathrm{INV}_{\mathrm{it}} + 0.000207\ \mathrm{PD}_{\mathrm{it}} + 0.3191\mathrm{Month} + \hat{\mathrm{u}}_{1\mathrm{i}}\mathrm{T}_{\mathrm{it}} + \hat{\mathrm{u}}_{0\mathrm{i}} \qquad (12)$$

The random value $\hat{u}_{1i}$ is a random linier time slope and $\hat{u}_{0i}$ is the effect of different random intercept for each province.

Table 2: The Estimate of fixed effect parameter of final model

Fixed Effect	Dummy	Estimate	P-Value
Intercept		9.7119	<.0001*
Time		-0.05011	<.0001*
INV		0.000047	0.0039*
PD		0.000207	0.0039*
Month Dummy	1	0.3191	<.0001*

* significant at $\alpha = 0.05$

In table 2 it can be seen that the estimated coefficient of time effect remains negative with -0.05011, it can be interpreted that every 6 months there is a decrease of unemployment rate around 0.05% if other variables remain. The predicted value of positive investment parameters indicates that provinces with high investment values tend to have higher unemployment rate than low-investment provinces. This is in line with the box-plot presentation of investment variables in figure 4, where high value investments are generally owned by provinces in Java. In figure 2 it is also seen that high unemployment rate is also owned by Java. Similarly, the estimate of parameters for positive population density variables indicate that provinces with high population density tend to have higher unemployment rate than provinces with lower population densities. Provinces with high population density are generally provinces located in Java Island. The month dummy variable is significant at the 5% level, this reinforces the tendency that appears descriptively in figure 1 where the unemployment rate in August tends to be higher than the unemployment rate in February. The estimated value of the month dummy variable parameter of 0.3191 shows that the average difference in unemployment rate in August and February is 0.3191% if the other variables remain.

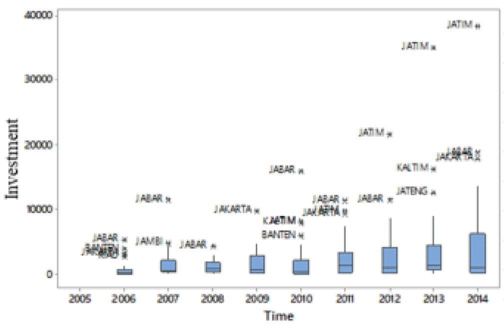

Figure 4: Box-plot of investment variables by year

In addition to the estimation of fixed effects, multilevel modelling also obtained the estimated value of covariance parameters of random effects as can be seen in Table 3.

Table 3: The Estimate of random effects

	Estimate	P-value
$\sigma^2_{\text{intercept}}$	10.3006	0.0002*
$\sigma_{\text{intercept,time}}$	-0.04094	0.0025*
σ^2_{time}	0.000213	0.0026*

* significant at $\alpha = 0.05$

Based on Table 3, it can be seen that entire estimated covariance parameters values of random effects significant at the 5% significance level. A significant range of random intercepts shows that there is varieties of initial unemployment rate values between provinces in Indonesia, with an estimated intercepts value of 10.3. Significant time slope variants indicate a variation in the decline in unemployment rate per unit of time between provinces in Indonesia, with a slope range of 0.000213. Significantly negative estimation of variants' intercepts and time slope indicate negative relationship between intercept and slope (the correlation value is -0.874), indicating that provinces that have high initial unemployment rate have lower rate of unemployment rate compared to provinces with low initial unemployment rate values. Visually this can be seen in Figure 5.

In Figure 5, it appears that Banten province has the highest intercept and the biggest negative time slope compared to other provinces, this means that Banten is the province that has the highest initial unemployment rate and a fairly large decline in unemployment rate over time. This is inversely proportional to the provinces of Papua and NTT where the initial unemployment rate of the two provinces was quite low and tended to experience low decline in unemployment rate over time.

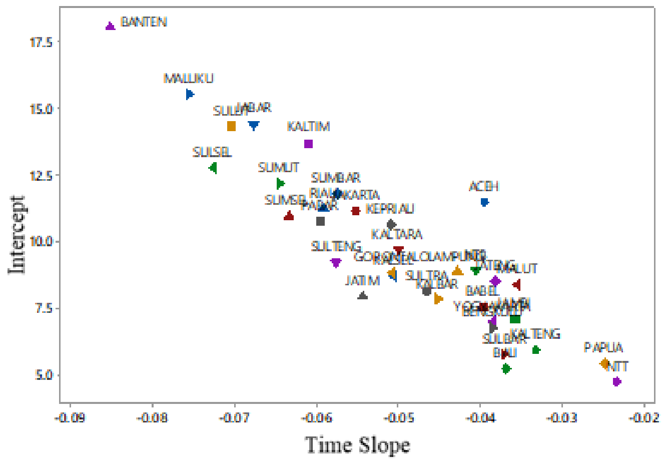

Figure 5: Scatter plot between time slope and intercept for each province

Thus it can be said that provinces with high initial unemployment rate tend to experience significant decrease in unemployment rate at each time point, this reflects that provinces with high unemployment rate tend to have government's attention to reduce unemployment rates.

3.4. Validation of Final Model Unemployment Rate

Based on figure 6, it can be seen that the estimated unemployment rate obtained from 2005 to 2014 tends to approach the actual value.

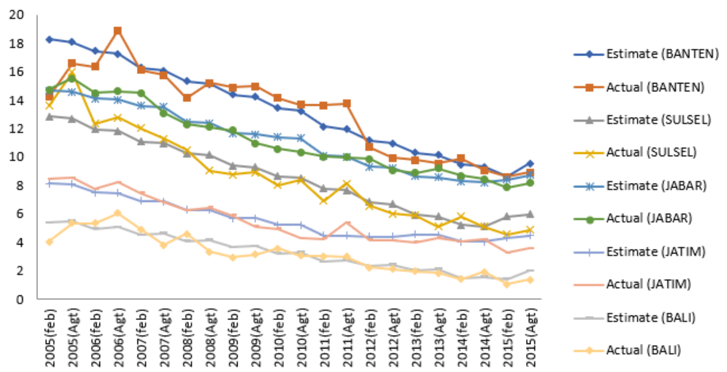

Figure 6: Actual data plot and estimated UR modelling

Model validation is carried out using data in the last year of 2015 consisting of data in August and February. In the model validation stage, the MAPE value is 25.54% which means that the magnitude of the final model error in predicting the unemployment rate is 25.54%.

4. Conclusion

Multilevel modelling of panel data for the unemployment rate in Indonesia can effectively capture the varieties that occurs in the pattern of unemployment rate trend between provinces in Indonesia. The varieties of unemployment rate between provinces in Indonesia occurs at the initial value and rate of decline over time. Provinces with a high initial value of unemployment rate tend to experience a greater decrease in unemployment rate per unit time. The final model discribe that external factors that significantly affect unemployment rate in Indonesia after being corrected by the effect of time are investment and population density, besides that unemployment rate in August tends to be higher than in February.

References

[1] [BPS] Badan Pusat Statistik. Keadaan Agkatan Kerja di Indonesia Februari 2015. Jakarta: Badan Pusat Statistik, 2015.

[2] B.T. West, K.B. Welch, A.T. Galecki. Linier Mixed Model A Practical Guide Using Statistical Software. United States: Taylor & Francis Group, 2007.

[3] D. Anggara, Indahwati, A. Kurnia. "Generalized Linear Mixed Models Approaches To Modeling Panel Data: Application To Poverty In East Nusa Tenggara." Global JURnal of Pure and Applied Mathematics, Research India Publication. Vol. 11, pp. 2867-2875, 2015.

[4] J.J. Hox. Multilevel Analysis Techniques and Applications. New Jersey (US): Lawrance Erlbaum

Associates, Inc, 2002.

[5] Mariana. “Spatial regression approach in modeling of Unemployment Rate.” Jurnal Matematika dan Pembelajarannya. Vol. 1, 2013.

[6] O. Prihatiningsih. 2012. “Menentukan Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka (TPT) Provinsi Jawa Barat dengan Regresi Terboboti Geografis (RTG).” Skripsi, Institut Pertanian Bogor, Bogor, 2012.

[7] P. Zulvia, A. Kurnia, A.M. Soleh. “Multilevel modeling and panel data analysis in educational research (Case study: National examination data senior high school in West Java).” AIP Conference Proceedings, 2017.

[8] S. Muchlisoh, A. Kurnia, K.A. Notodiputro, I.W. Mangku. “ Small Area Estimation of Unemployment Rate Based on Unit Level Model with First Order Autoregressive Time Effects.” JURnal of Applied Probability and Statistic. Vol. 12, pp. 65-77, 2017.

[9] V. Noviyanti. “Analisis Faktor-faktor yang Mempengaruhi Tingkat Pengangguran di Jawa Tengah Tahun 1991 sampai 2011.” Skripsi. Universitas Muhammadiyah Surakarta, Surakarta 2014.