Nonlinear Dynamics and Systems Theory, **22** (5) (2022) 550–560

The Analysis of Demand and Supply of Blood in Hospital in Surabaya City Using Panel Data Regression

A. Muhith [1], I. H. Susanto [2], D. Rahmalia [3], D. Adzkiya [4] and T. Herlambang [5*]

[1] *Nursing Science Department, Universitas Nahdlatul Ulama Surabaya, Indonesia.*
[2] *Department of Accounting Magister, University of Wijaya Kusuma Surabaya, Indonesia.*
[3] *Mathematics Department, University of Islam Darul Ulum Lamongan, Indonesia.*
[4] *Department of Mathematics, Institute Teknology Sepuluh Nopember, Indonesia.*
[5] *Information System Department, Universitas Nahdlatul Ulama Surabaya, Indonesia.*

Received: October 11, 2021; Revised: November 25, 2022

Abstract: Blood is a vital component in body health because it distributes oxygen, food, and hormones in the whole body. However, there are some cases such as the lack of blood, accidents, or other diseases when humans need blood transfusions, which depend on the demand and supply of blood in hospitals. In this research, panel data regression is used to analyse the demand and supply of blood in hospitals in Surabaya city. There are three models in panel data regression, namely, common effect (CE), fixed effect (FE), and random effect (RE). In this panel data regression, the number of demands of blood type O, A, B, and AB is the independent variable. In contrast, the blood supply is the dependent variable. First, we will determine the best model, common effect (CE), fixed effect (FE), or random effect (RE), through the Chow test, Hausman test, and Lagrange Multiplier test. From the result, the best model of the quantity of blood supply is fixed effect (FE). Then, the fixed effect (FE) model parameters are tested by using the F-test and T-test for testing the impact of independent variables on the dependent variable and R-squared for finding the proportion of effectiveness of independent variables. According to our simulation results, the R-squared is 0.998, which is very satisfactory.

Keywords: *panel data regression; demand and supply of blood; fixed effect model; statistics.*

Mathematics Subject Classification (2010): 62J02, 62J05, 62J07, 62M10.

* Corresponding author: `mailto:teguh@unusa.ac.id`

1 Introduction

Blood is a vital component in body health because it distributes oxygen, food, and hormones in the whole body. However, there are some cases such as the lack of blood, accidents, or other diseases when a human needs blood transfusion which depends on the demand and supply of blood in hospital [1, 2].

In this research, panel data regression is used to analyse the demand and supply of blood in hospitals in Surabaya city. Panel data is the data combining time-series data and cross-section data. Time-series data cover an object for an extended period. Cross-section data consist of many things such as a company, factory, restaurant, a place with some attributes. Thus, panel data regression is the regression using panel data. There are three models in panel data regression, namely, common effect (CE), fixed effect (FE), and random effect (RE).

From the previous research, the effects of independent variables and dependent variables have been applied by the correlation method in a Neural Network (NN) [3, 4] and Adaptive Neuro-Fuzzy Inference System (ANFIS) [5, 6]. There is a work on stability analysis of stochastic neural networks [7]. Let us also mention the results on control design using Sliding PID [8] and Linear Quadratic Regulator [9].

In this panel data regression, the number of demands of blood type O, A, B, AB is the independent variable, while blood supply is the dependent variable.

First, we determine the best model, common effect (CE), fixed effect (FE), or random effect (RE), through the Chow test for determining a better model between the common effect (CE) model and the fixed effect (FE) model in the panel data model, the Hausman test for determining a better model between the random effect (RE) model and the fixed effect (FE) model in the panel data model, and the Lagrange Multiplier test for determining a better model between the common effect (CE) model and the random effect (RE) model in the panel data model. Then, the simulation results are applied by EViews software.

From the result, the best model of the quantity of blood supply is fixed effect (FE). The F-test and T-test test the parameters of the fixed effect (FE) model for testing the impact of independent variables on the dependent variable, and R-squared is used for finding the proportion of effectiveness of independent variables.

2 Panel Data Modeling

Panel data is the data combining time-series data and cross-section data. Time-series data cover an object for a long time. Cross-section data consist of many entities (for example, a company, factory, restaurant, place) with some attributes (for example, cost, benefit, the volume of production, the number of workers) in one period. Thus, panel data regression is the regression using panel data.

The regression model of time-series data is as follows:

$$Y_t = \alpha + \sum_{j=1}^{P} \beta^j x_t^j + \varepsilon_t, \qquad t = 1, 2, ..., T,$$

where T is the number of time-series data and P is the number of independent variables.

The regression model of cross-section data is as follows:

$$Y_i = \alpha + \sum_{j=1}^{P} \beta^j x_i^j + \varepsilon_i, \qquad i = 1, 2, ..., N,$$

where N is the number of cross-section data and P is the number of independent variables. So the regression model of panel data is as follows:

$$Y_{it} = \alpha + \sum_{j=1}^{P} \beta^j x_{it}^j + \varepsilon_{it}, \qquad t = 1, 2, ..., T, \qquad i = 1, 2, ..., N,$$

where N is the number of cross-section data, T is the number of time-series data, and P is the number of independent variables.

3 Estimation of Panel Data Regression

For estimating the parameters of the panel data model, there are three techniques.

3.1 Common Effect (CE) model (Pooled model)

In this model, time-series data and cross-section data are merged. By joining both of them, one can use the Ordinal Least Square (OLS) method or the least square technique to estimate the data panel model. It is assumed that the properties of data of the objects are similar in the interval of time [10].

However, this assumption deviates from the actual conditions because the characteristics of the objects are very different. Therefore, this model can be constructed as follows:

$$Y_{it} = \alpha + \sum_{j=1}^{P} \beta^j x_{it}^j + \varepsilon_{it}, \qquad t = 1, 2, ..., T, \qquad i = 1, 2, ..., N,$$

where N is the number of cross-section data, T is the number of time-series data, P is the number of independent variables, Y_{it} is the dependent variable of the i-th object in the t-th time, x_{it}^j is the j-th independent variables of the i-th object in the t-th time, β^j is the coefficient (parameter) of the j-th independent variables, α is the intercept, ε_{it} is the error component of the i-th object in the t-th time.

3.2 Fixed Effect (FE) model

This model estimates panel data by adding dummy variables. There are different effects among objects through the difference of their intercepts. In the fixed effect (FE) model, an object is an unknown parameter, and it will be estimated by dummy variables. This model can be constructed as follows [10]:

$$Y_{it} = \alpha + \sum_{j=1}^{P} \beta^j x_{it}^j + \sum_{k=2}^{n} \alpha_k D_k + \varepsilon_{it}, \qquad t = 1, 2, ..., T, \qquad i = 1, 2, ..., N,$$

where D_k is the dummy variable.

3.3 Random Effect (RE) model

In this method, the differences in object and time characteristics are formed by the error from the model. Because two components contribute to the error results, such as object and time, this method needs to be expanded to become the error from the object component, the error from the time component, and the combined error. The random effect (RE) model is as follows [11]:

$$Y_{it} = \alpha + \sum_{j=1}^{P} \beta^j x_{it}^j + \varepsilon_{it}, \qquad t = 1, 2, ..., T, \qquad i = 1, 2, ..., N,$$

where $\varepsilon_{it} = u_i + v_t + w_{it}$, u_i is the error from the object component, v_t is the error from the time component, w_{it} is the combined error.

4 The Selection of Best Model

For selecting the best model, the common effect (CE) model, fixed effect (FE) model or random effect (RE) model, there are some tests such as the Chow test, Hausman test, and Lagrange Multiplier test.

4.1 Chow test

The Chow test is used for determining a better model between the common effect (CE) model and the fixed effect (FE) model in the panel data model [12].

The hypothesis used in the Chow test is as follows. The null hypothesis (H_0) represents the common effect (CE) model, whereas the alternative hypothesis (H_1) represents the fixed effect (FE) model. The Chow statistics is given by

$$\frac{(ESS1 - ESS2)/(N-1)}{(ESS2)/(NT - N - P)},$$

where

$ESS1$: Residual Sum Square of the fixed effect (FE) model,
$ESS2$: Residual Sum Square of the common effect (CE) model,
N : the number of cross-section data,
T : the number of time-series data,
P : the number of independent variables.

The Chow statistics follows the F-statistics distribution with the degree of freedom $(N - 1, NT - N - P)$. If the Chow statistics is larger than the critical value of the F-statistics distribution or the p-value is less than the significance level α, then H_1 is accepted and H_0 is rejected so that the selected model is the fixed effect (FE) model.

4.2 Hausman test

The Hausman test is used for determining a better model between the random effect (RE) model and the fixed effect (FE) model in the panel data model [12].

The hypothesis used in the Hausman test is as follows. The null hypothesis (H_0) represents the random effect (RE) model, whereas the alternative hypothesis (H_1) represents the fixed effect (FE) model. The m statistics is given by

$$m = (\beta - b)(M0 - M1)^{-1}(\beta - b),$$

where:

β : Statistics vector of fixed effect (FE) variables,
b : Statistics vector of random effect (RE) variables,
$M0$: covariance matrix for the fixed effect (FE) model,
$M1$: covariance matrix for the random effect (RE) model.

The m statistics follows the chi-square distribution with the degree of freedom equal to P. If m statistics is larger than the critical value of the chi-square distribution or the p-value is less than the significance level α, then H_1 is accepted and H_0 is rejected so that the selected model is the fixed effect (FE) model.

4.3 Lagrange Multiplier test

The Lagrange Multiplier test is used to determine a better model between the common effect (CE) model and the random effect (RE) model in the panel data model [13].

The hypothesis used in the Lagrange Multiplier test is as follows. The null hypothesis (H_0) represents the common effect (CE) model, whereas the alternative hypothesis (H_1) represents the random effect (RE) model. The LM statistics is given by

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\bar{e}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}^2} - 1\right],$$

where

N : the number of cross-section data,
T : the number of time-series data,
e_{it} : residual of the common effect (CE) model.

The LM statistics follows the chi-square distribution with the degree of freedom equal to P.

If the LM statistics is larger than the critical value of the chi-square distribution or the p-value is less than the significance level α, then H_1 is accepted and H_0 is rejected so that the selected model is the random effect (RE) model.

The Lagrange Multiplier test is not applied when the Chow test and the Hausman test show a better model is the fixed effect (FE) model [10].

5 Significance Test

After the best model is obtained, it is required to apply the significance test as follows.

5.1 F-test

The F-test is applied for testing the estimation results on whether the independent variables have effects on the dependent variable globally [14].

The hypothesis used in the F-test is as follows. The null hypothesis (H_0) represents "independent variables do not affect the dependent variable", whereas the alternative hypothesis (H_1) represents "independent variables affect the dependent variable".

The F-statistics is given by

$$F_{test} = \frac{MS(y)}{MS(e)},$$

where

$MS(e)$: mean square of regression,
$MS(y)$: mean square of residual.

The F_{test} statistics follows the F-statistics distribution with the degree of freedom equal to $(N + P - 1, NT - N - P)$. If the F_{test} statistics is larger than the critical value of the F-statistics distribution or the p-value is less than the significance level α, then H_1 is accepted and H_0 is rejected so that the conclusion is that there is the effect of the independent variables on the dependent variable.

5.2 T-test

The T-test is applied for testing the estimation results on whether the independent variables have effects on the dependent variable partially [14].

The hypothesis used in the T-test is as follows. The null hypothesis (H_0) represents "independent variables do not affect the dependent variable", whereas the alternative hypothesis (H_1) represents "independent variables affect the dependent variable".

The T-statistics is given by

$$T_{test} = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)},$$

where

$\hat{\beta}_k$: the k-th parameter,
$SE(\hat{\beta}_k)$: standard deviation of the k-th parameter.

The T_{test} statistics follows the T-statistics distribution with the degree of freedom equal to $(NT - N - P)$. If the T_{test} statistics is larger than the critical value of the T-statistics distribution or the p-value is less than the significance level $\alpha/2$, then H_1 is accepted and H_0 is rejected so that the conclusion is that there is the effect of the independent variables on the dependent variable.

5.3 R-squared

The determination coefficient (R^2) is used for measuring the fitness rate of panel data regression. It is a proportion of the contribution of independent variables and dummy variables to that of the dependent variable [15].

The coefficient (R^2) is determined using

$$R^2 = \frac{ESS}{TSS},$$

where

ESS : sum of square of regression,
TSS : total of sum of square.

The value of R^2 is between 0 and 1. If R^2 approaches 1, then in this model, the effect of independent variables is stronger.

6 Results

This research shows the effects of the demand for blood types O, A, B, AB on the quantity of blood supply in five hospitals in Surabaya city from January 2015 until December 2017. In creating panel data regression, the data used are as follows. Cross-section or object data:

1. Angkatan Laut hospital,
2. Unair hospital,
3. Haji hospital,
4. Adi Husada hospital,
5. Darmo hospital.

The data used are monthly data from January 2015 until December 2017.

In this research, the analysis of demand and supply of blood in some hospitals in Surabaya city is done by panel data regression using EViews software. There are three models in panel data regression, such as common effect (CE), fixed effect (FE), and random effect (RE). In this panel data regression, the number of demands of blood type O, A, B, AB is an independent variable, while blood supply is the dependent variable.

6.1 The selection of best model

First, we use the Chow test to determine a better model between the common effect (CE) model and the fixed effect (FE) model in the panel data model.

The Chow test results can be seen in Figure 1. Figure 1 shows the p-value of cross-section F is $0.0039 < 0.05$. Therefore, H_1 is accepted and H_0 is rejected so that the selected model is the fixed effect (FE) model.

Second, we use the Hausman test to determine a better model between the random effect (RE) model and the fixed effect (FE) model in the panel data model.

The Hausman test results can be seen in Figure 2. Figure 2 shows the p-value of cross-section random is $0.003 < 0.05$. Therefore, H_1 is accepted and H_0 is rejected so that the selected model is the fixed effect (FE) model.

Because both the Chow and the Hausman test show that the fixed effect (FE) model is the best model, the Lagrange Multiplier test is not required.

Redundant Fixed Effects Tests
Equation: Untitled
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	4.003197	(4,171)	0.0039
Cross-section Chi-square	16.112418	4	0.0029

Figure 1: Chow test result.

Correlated Random Effects - Hausman Test
Equation: Untitled
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	16.012786	4	0.0030

Figure 2: Hausman test result.

The fixed effect (FE) panel data regression model for the quantity of blood supply used is shown in Figure 3. The coefficient of the quantity of blood type O is 0.982566, the coefficient of the quantity of blood type A is 1.003406, the coefficient of the quantity of blood type B is 0.976337, the coefficient of the quantity of blood type AB is 0.990383, the intercept is 1.376641.

Dependent Variable: S_BLOOD
Method: Panel Least Squares
Date: 06/19/20 Time: 11:36
Sample: 2015M01 2017M12
Periods included: 36
Cross-sections included: 5
Total panel (balanced) observations: 180

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BLOOD_O	0.982566	0.009145	107.4399	0.0000
BLOOD_A	1.003406	0.013104	76.57403	0.0000
BLOOD_B	0.976337	0.008611	113.3822	0.0000
BLOOD_AB	0.990383	0.027919	35.47346	0.0000
C	1.376641	0.951421	1.446931	0.1497

Figure 3: Fixed effect (FE) model panel data regression.

6.2 Significance test

In fixed effect (FE) model panel data regression, we will test the impact of independent variables on the dependent variables using the F-test and T-test. Furthermore, we also measure the proportion of the independent and dummy variables' contribution and that of dependent variables using R-squared.

The F-test is applied for testing the estimation results on whether the independent variables have effects on the dependent variables globally. The F-test results can be seen in Figure 4. Figure 4 shows the p-value of the F-test is 0.000 $<$ 0.05. Therefore, H_1 is accepted and H_0 is rejected, i.e., there is an effect of the independent variables on the dependent variables.

Effects Specification

Cross-section fixed (dummy variables)			
Root MSE	3.490123	R-squared	0.998008
Mean dependent var	193.3889	Adjusted R-squared	0.997915
S.D. dependent var	78.41534	S.E. of regression	3.580790
Akaike info criterion	5.437751	Sum squared resid	2192.572
Schwarz criterion	5.597399	Log likelihood	-480.3976
Hannan-Quinn criter.	5.502481	F-statistic	10708.81
Durbin-Watson stat	1.469845	Prob(F-statistic)	0.000000

Figure 4: F-test results.

The T-test is applied for testing the estimation results on whether the independent variables have effects on the dependent variables partially. The T-test results can be seen in Figure 5. Figure 5 shows the p-value of the T-test on the number of demands of blood type O is 0.000 $<$ 0.05, the number of requests of blood type A is 0.000 $<$ 0.05, the number of requests of blood type B is 0.000 $<$ 0.05, the number of demands of blood type AB is 0.000 $<$ 0.05. Therefore, H_1 is accepted and H_0 is rejected, i.e., there is an effect of the independent variables on the dependent variables.

Dependent Variable: S_BLOOD
Method: Panel Least Squares
Date: 06/19/20 Time: 11:36
Sample: 2015M01 2017M12
Periods included: 36
Cross-sections included: 5
Total panel (balanced) observations: 180

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BLOOD_O	0.982566	0.009145	107.4399	0.0000
BLOOD_A	1.003406	0.013104	76.57403	0.0000
BLOOD_B	0.976337	0.008611	113.3822	0.0000
BLOOD_AB	0.990383	0.027919	35.47346	0.0000
C	1.376641	0.951421	1.446931	0.1497

Figure 5: T-test results.

R-squared is used to measure the proportion of the contribution of independent and dummy variables and that of the dependent variables. The R-squared test can be seen in Figure 6. Figure 6 shows R-squared is 0.998. It means that the effects of the independent variables on the dependent variables are 99.8%.

7 Conclusions

There are three models in panel data regression, namely, common effect (CE), fixed effect (FE), and random effect (RE). In this panel data regression, the number of demands of

Effects Specification

Cross-section fixed (dummy variables)			
Root MSE	3.490123	R-squared	0.998008
Mean dependent var	193.3889	Adjusted R-squared	0.997915
S.D. dependent var	78.41534	S.E. of regression	3.580790
Akaike info criterion	5.437751	Sum squared resid	2192.572
Schwarz criterion	5.597399	Log likelihood	-480.3976
Hannan-Quinn criter.	5.502481	F-statistic	10708.81
Durbin-Watson stat	1.469845	Prob(F-statistic)	0.000000

Figure 6: T-test results.

blood type O, A, B, AB is an independent variable, while blood supply is the dependent variable. First, we determine the best model, common effect (CE), fixed effect (FE), or random effect (RE), through the Chow test, Hausman test, and Lagrange Multiplier test. From the result, the best model of the quantity of blood supply is fixed effect (FE). Then, the parameters of the fixed effect (FE) model are tested by the F-test and T-test for testing the impact of the independent variables on the dependent variable and R-squared is used for finding the proportion of effectiveness of the independent variables. In our simulation, the R-squared is 0.998, which is a very good result. As a future work, we are planning to employ some machine learning techniques to analyze the demand and supply of blood. Furthermore, by combining mathematical science and business management, we strive to provide a feedback for stakeholders before making any decision.

Acknowledgment

This research was supported by Kemendikbud with contract number 313/E4.1/AK.04.PT/2021,027/AMD-SP2H/LT-MULTI-TERAPAN/LL7/2021, 895/UNUSA/Adm-LPPM/VII/2021 year 2021 and thanks to LPPM Universitas Nahdlatul Ulama Surabaya.

References

[1] A. Muhith, T. Herlambang, M. Y. Anshori, R. Rizqina, D. Rahmalia and Hermanto. Estimation of Whole Blood (WB) and Anti-Hemophiliate Factor using Extended Kalman Filter in PMI Surabaya. *Journal of Physics: Conference Series* **1538** (1) (2020) 012034.

[2] A. Muhith, T. Herlambang, A. Haris and D. Rahmalia. Estimation of Whole Blood (WB) using Unscented Kalman Filter in Surabaya. *IOP Conference Series: Materials Science and Engineering* **874** (1) (2020) 012028.

[3] D. Rahmalia and T. Herlambang. Prediksi Cuaca Menggunakan Algoritma Particle Swarm Optimization-Neural Network (PSONN). In: *Proc. Seminar Nasional Matematika dan Aplikasinya* 2017, 41–48.

[4] D. Rahmalia and N. Aini. Pengaruh Korelasi Data pada Peramalan Suhu Udara Menggunakan Backpropagation Neural Network. *Zeta-Math Journal* **4** (1) (2018) 1–6.

[5] D. Rahmalia and A. Rohmatullah. Pengaruh Korelasi Data pada Peramalan Kelembaban Udara Menggunakan Adaptive Neuro Fuzzy Inference System (ANFIS). *Applied Technology and Computing Science Journal* **2** (1) (2019) 10–24.

[6] D. Rahmalia, T. Herlambang, A. S. Kamil, R. A. Rasyid, F. Yudianto, L. Muzdalifah and E. F. Kurniawati. Comparison between Neural Network (NN) and Adaptive Neuro Fuzzy Inference System (ANFIS) on sunlight intensity prediction based on air temperature and humidity. *Journal of Physics: Conference Series* **1538** (1) (2020) 012044.

[7] B. Benaid, H. Bouzahir, C. Imzegouan and F. El Guezar. Stability Analysis for Stochastic Neural Networks with Markovian Switching and Infinite Delay in a Phase Space. *Nonlinear Dynamics and Systems Theory* **19** (3) (2019) 372–385.

[8] T. Herlambang, S. Subchan, H. Nurhadi and D. Adzkiya. Motion Control Design of UN-USAITS AUV Using Sliding PID. *Nonlinear Dynamics and Systems Theory* **20** (1) (2020) 51–60.

[9] T. Herlambang, D. Rahmalia, H. Nurhadi, D. Adzkiya and S. Subchan. Optimization of Linear Quadratic Regulator with Tracking Applied to Autonomous Underwater Vehicle (AUV) Using Cuckoo Search. *Nonlinear Dynamics and Systems Theory* **20** (3) (2020) 282–298.

[10] B. H. Baltagi. *Econometric Analysis of Panel Data.* Springer, 2008.

[11] J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* MIT press, Cambridge, MA, 2002.

[12] W. H. Greene. *Econometric Analysis.* Pearson Education India, 2003.

[13] C. Hsiao. *Analysis of Panel Data.* Cambridge university press, 2014.

[14] M. Srihardianti, M. Mustafid and A. Prahutama. Metode Regresi Data Panel Untuk Peramalan Konsumsi Energi di Indonesia. *Jurnal Gaussian* **5** (3) (2016) 475–485.

[15] A. Y. D. Kristina. Pengaruh Pendapatan Asli Daerah, Indeks Pembangunan Manusia dan Tenaga Kerja Terhadap Produk Domestik Regional Bruto (38 Kabupaten/Kota di Provinsi Jawa Timur Tahun 2011–2016). *Jurnal Ilmu Ekonomi JIE* **1** (2) (2017) 176–188.