

# opción

Revista de Antropología, Ciencias de la Comunicación y de la Información, Filosofía,  
Lingüística y Semiótica, Problemas del Desarrollo, la Ciencia y la Tecnología

Año 36, abril 2020 N°

91

Revista de Ciencias Humanas y Sociales

ISSN 1012-1537/ ISSNc: 2477-9385

Depósito Legal pp 198402ZU45



Universidad del Zulia  
Facultad Experimental de Ciencias  
Departamento de Ciencias Humanas  
Maracaibo - Venezuela

# Modeling of money supply using LASSO regression with Cross-Validation

**Trisha Magdalena Adelheid Januaviani<sup>1</sup>**

<sup>1</sup>Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran Jl. Raya Bandung-Sumedang Km 21, Jatinangor 45363, Sumedang, Jawa Barat, Indonesia.  
[trishadelheid@unpad.ac.id](mailto:trishadelheid@unpad.ac.id)

**Sukono<sup>2</sup>**

<sup>2</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran Jl. Raya Bandung-Sumedang Km 21, Jatinangor 45363, Sumedang, Jawa Barat, Indonesia.  
[sukono@unpad.ac.id](mailto:sukono@unpad.ac.id)

**Eman Lesmana<sup>3</sup>**

<sup>3</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran Jl. Raya Bandung-Sumedang Km 21, Jatinangor 45363, Sumedang, Jawa Barat, Indonesia.  
[eman.lesmana@unpad.ac.id](mailto:eman.lesmana@unpad.ac.id)

## Abstract

The study aims to investigate the modeling of money supply using LASSO regression with cross-validation. The money supply spent can be applied to multiple regression analysis because it has many influencing factors. Multiple linear regression analysis is a statistical technique to examine the relationship between independent variables and independent variables. As a result, the number of independent variables in the money supply asked is greater than multicollinearity. In conclusion, if net foreign assets consisting of bills and non-residents liabilities in Indonesia have increased by one billion, then the money supply in circulation in Indonesia will decrease by 0.6259048 and 1.451317.

**Keywords:** Money Supply, LASSO, LARS, Cross-validation.

# Modelado de la oferta monetaria utilizando la regresión LASSO con validación cruzada

## Resumen

El estudio tiene como objetivo investigar el modelado de la oferta monetaria utilizando la regresión LASSO con validación cruzada. La oferta monetaria gastada se puede aplicar al análisis de regresión múltiple porque tiene muchos factores influyentes. El análisis de regresión lineal múltiple es una técnica estadística para examinar la relación entre variables independientes y variables dependientes. Como resultado, el número de variables independientes en la oferta monetaria solicitada es mayor que la multicolinealidad. En conclusión, si los activos extranjeros netos que consisten en facturas y pasivos para no residentes en Indonesia han aumentado en mil millones, entonces la oferta monetaria en circulación en Indonesia disminuirá en 0.6259048 y 1.451317.

**Palabras clave:** Oferta de dinero, LASSO, LARS, Validación cruzada.

## 1. INTRODUCTION

Least Absolute Shrinkage and Selection Operator (LASSO) regression is a method of shrinkage where the parameter  $b$  approaches zero or to zero. Estimator generated from LASSO has smaller variants and more endings representative. The LASSO method is proposed to carry on the variable (model) selection and the parameter estimation simultaneously (ARSLAN, 2012; AHMAD & AHMAD; 2019). These two procedures have been combined in a single minimization problem. LASSO is used to handle data that have multicollinearity. Multicollinearity is a correlation between the independent variable which is a weakness in the analysis because it has a big error.

Multicollinearity may considerably decrease the efficiency of the ordinary least squares (OLS) method in the sense of leading to wide confidence intervals as well as wrong signs for the parameter  $\beta$  (ROOZBEH, BABAIE-KAFAKI, & SADIGH, 2017; AHMAD & AHMAD; 2018).

The money supply in Indonesia is influenced by factors such as net foreign assets (NFA) and domestic assets net (Net Domestic Assets/NDA). Assets net foreign income is the difference from claims to non-Residents and obligations to non-residents. While the assets domestic net is Bank Indonesia's net bill to the population which consists of, among others, net bills to the central government and bills to other sectors (MALYSHEVA, 2019).

Multiple regression analysis can be applied to the money supply because it has many influencing factors. Regression analysis is a statistical analysis system for determining the effect of independent variables on a dependent variable (AMOOZAD-KHALILI, ROSTAMIAN, & ESMAEILPOUR-TROUJENI, 2019). The money supply as an independent variable Y and the influencing factor is the independent variable X. Because independent variables in the money supply more than two likely the data have multicollinearity.

Analyzing data on the money supply in Indonesia using the LASSO method decides the best model with two methods, namely cross-validation. Cross-validation (CV) is a model evaluation technique that utilizes data separation.

## 2. METHODS AND MATERIALS

The money supply is the total value of money in the hands of the community (CHUNG & ARIFF, 2016). Two types of money supply are narrow money supply (M1) and broad money supply (M2) (SHIRVANI, 2013). The narrow money supply consists of currency and demand deposits:

$$M1 = C + D \quad (1)$$

*M1*: Money supply in the narrow sense

*C*: Currency

*D*: Demand deposit

Currency is banknotes and coins used by the public as a means of payment while demand deposits are money used by certain circles as a means of payment such as checks (HE, 2017). The broad money supply (M2) is the sum of currency, demand deposits and quasi money as follows:

$$M2 = M1 + TD \quad (2)$$

*M1*: The money supply in a broad sense

*TD*: Quasi money or term savings

The money supply is the obligation of the monetary system (Central Bank, Commercial Bank, and Rural Credit Bank or BPR) to the domestic private sector (not including the central government and non-residents). Obligations that are a component of the money supply consist of currency held by the public (excluding commercial banks and rural banks), demand deposits, quasi money owned by the domestic private sector, and securities other than shares issued by the monetary system owned by the domestic private sector with the remaining period is up to one year.

Several factors affect the money supply, namely Net Foreign Assets) and Net Domestic Assets consist of Net Claims to the Central Government and Claims to other sectors (private sector, regional government, financial institutions, and non-financial companies) especially in the form of loans. Dependent and independent variables in money supply data are as follows:

$Y$ : Money Supply

$X_1$ : Bill on Non-Residents

$X_2$  : Obligations to Non-Residents

$X_3$  : Net Bill on the Central Government

$X_4$  : Claims on Other Financial Institutions

$X_5$ : Bill on Regional Government

$X_6$  : Claims on SOE Non-Financial Companies

$X_7$  : Private Sector Bill

### *2.1. Ordinary Least Squares*

The Ordinary Least Squares (OLS) method is one of the parameter estimators in the regression model. An OLS operator is used to determine the coefficients (ALOISIO, ALAGGIO, & FRAGIACOMO, 2019). The model used is the sample model because the population model chooses errors that cannot be directly observed, while the sample model is as follows (MAHABOOB, B. & BHUMIREDDY, 2018):

$$y = X\beta + e \quad (3)$$

The estimated model of equation (3) is stated as follows:

$$\hat{y} = X\hat{\beta} \quad (4)$$

so we get the parameter  $\hat{\beta}$  estimator

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (5)$$

### *.2.2. Multicollinearity*

The correlation between the independent variables is called multicollinearity. Multicollinearity meant that there was a perfect or definite linear relationship between some or all variables that explained the regression model. Data use of fewer variables makes the model weak while the use of numerous variables can be computationally intensive and may lead to problems associated with multicollinearity (VU, MUTTAQI, & AGALGAONKAR, 2015). Ragnar Frisch in 1934 was the first person to introduce multicollinearity.

Detect multicollinearity with the value of Variance Inflation Factor (VIF) where if the VIF value is more then 10 there is multicollinearity (VU ET AL, 2015). VIF equation can be written as follows:

$$VIF_j = \frac{1}{TOL_j}; \tag{6}$$

$$. TOL_j = 1 - R_j^2. \tag{7}$$

### *2.3. Cross-Validation*

Cross-validation is a measurement of assessing the performance of a predictive model and statistical analysis will generalize to an

independent dataset. Cross-validation divides the data into two parts, namely training data and test data. Training data is used to determine the value of  $\beta$ , while the test data used to test the predictability of  $X\beta$ . The cross-validation value obtained is an estimator for prediction errors (IZENMAN, 2008). In K-fold cross-validation, all observations are randomly partitioned into

K sub-samples. K-fold cross-validation, which repeatedly retains a percentage of the data set for prediction, maybe appropriate to balance the variance-bias tradeoff. Each sub-example is used as test data and the rest is used as training data. The cross-validation process is repeated up to K times, and every one sub-example is used only once in the test data. This method has an advantage when the amount of data observed is only small. The mathematical form of the CV K-fold is as follows (JIANG & CHEN, 2016):

$$\begin{aligned} \text{CV MSE} &= \frac{1}{k} \sum_{h=1}^k \text{MSE}_h \\ &= \frac{1}{k} \sum_{h=1}^k \sum (x_i, y_i) (y_i - \hat{y}_{-k}(x_i))^2 \end{aligned} \quad (8)$$

#### 2.4. LASSO

The Least Absolute Shrinkage and Selection Operator (LASSO) regression method is one of the independent variable shrinkage techniques that was introduced by Tibshirani in 1996 for the first time. Lasso is a technique to fit the linear regression model which minimizes

the residual squared error with a constraint on the sum of the absolute value of the regression coefficients (ZHENG & LIU, 2011). Shrinking the coefficient parameter  $\beta$  which correlates close to zero or becomes zero, so that the variance obtained is smaller than the final model which is more representative. LASSO addresses data that have multicollinearity. The general form of LASSO is as follows:

$$\min \|Y - X\beta\|_2^2 + \lambda\|\beta\|_1 \tag{9}$$

written as follows

$$\min_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j )^2 \tag{10}$$

With obstacles

$$D = \sum_{j=1}^p |\beta_j| \leq t \tag{11}$$

### 2.5. Least Absolute Regression and Shrinkage (LARS)

LASSO estimator solutions cannot be obtained in a closed-form but must use sacred programming. LASSO gained attention in 2002 after the LAR (Least Angle Regression) algorithm although it was first published in 1996. LAR was modified for LASSO which results in algorithmic efficiency in estimating LASSO coefficients using computation so that it is faster and more efficient than using quadratic

programming. Least Absolute Regression and Shrinkage (LARS) is a LAR modification for LASSO.

The LARS stages compute the LASSO parameters as follows. Defined  $i = 1, \hat{\mu}^1 = \mathbf{0}$  with dimensions  $n \times 1$ ,  $n$  is a lot of data. For example  $\hat{\mu}_A^i$  is an estimated value with an active variable, value  $\hat{\mu}$  will change according to stages. Defined  $\hat{\beta} = \mathbf{0}$  with dimensions  $n \times p$  where  $p$  the number of independent variables.

The data used is the data that is transformed.  $\mathbf{X}^*, \mathbf{Y}^*$  with

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{S_{X_j} \sqrt{n-1}}; S_{X_j} = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1}}; j = 1, 2, \dots, p \tag{12}$$

$$Y^* = Y - \bar{Y} \tag{13}$$

The initial stage looks for the largest correlation vector and absolute correlation value with the following equation in a row:

$$\hat{c}^i = \mathbf{X}^{*T} (\mathbf{Y}^* - \hat{\mu}_A^i) \tag{14}$$

$$\hat{C}^i = \max\{|\hat{c}_j^i|\} \tag{15}$$

Where  $A$  will change by following the following equation:

$$A = \{j \mid |\hat{c}_j^i| = \hat{C}^i\} \tag{16}$$

Next count equiangular vector *vector* ( $\mathbf{u}_A^i$ ) which is defined  $\mathbf{u}_A^i = X_A^i \omega_A^i$  (17)

with

$$X_A^i = [\dots s_j X_j^* \dots]_{j \in A} \tag{18}$$

where

$$s_j = \text{sign}\{\hat{c}_j^i\}, j \in A \tag{19}$$

equation (14) is described as

$$\omega_A^i = P_A G_A^{i-1} \mathbf{1}_A \tag{20}$$

$$G_A^i = X_A^{i T} X_A^i ; P_A^i = \left( \mathbf{1}_A^T G_A^{i-1} \mathbf{1}_A \right)^{\frac{1}{2}} \tag{21}$$

The next step to finding the inner product vector is defined as:

$$a^i \equiv X^{*T} \mathbf{u}_A^i \tag{22}$$

So

$$\hat{\gamma}^i = \min_{j \in A^c}^+ \left\{ \frac{\hat{c}_j^i - c_j^i}{p_A^i - a_j^i}, \frac{c_j^i + \hat{c}_j^i}{p_A^i + a_j^i} \right\} \tag{23}$$

Where used in LARS restrictions at a later stage. The LARS restriction according to Hestie begins with counting  $\gamma_j^i$  with the following equation:

$$\gamma_j^i = -\frac{\hat{\beta}_j^i}{s_j \omega_A^i} \tag{24}$$

Limiting the LAR algorithm for LASSO the following conditions:

$$\varphi^i = \min_{\gamma_j^i > 0} \{ \gamma_j^i \} \tag{25}$$

1. If  $\varphi^i \geq \hat{\gamma}^i$  where is the value  $\hat{\gamma}^i$  in equation (23), it does not violate the LARS constraints can proceed to the next stage to find the value  $\hat{\beta}^{i+1}$  and  $\hat{\mu}_A^{i+1}$

2. If  $\varphi^i$  has no value, then  $\varphi^i = \hat{\gamma}^i$  so that it does not violate the LARS barrier.

3. If  $\varphi^i < \hat{\gamma}^i$  then violating the LARS barrier. Stop the LARS process at this stage, remove the variables  $j$  from the calculation  $\hat{\beta}^{i+1}$  and  $A_+ = A - \{j\}$ , , If all independent variables have been entered, ignore this step.

4. After the LARS restriction stage, the next step is to update the value  $\hat{\beta}^{i+1}$  and  $\hat{\mu}_A^{i+1}$  with

$$\hat{\beta}^{i+1} = \hat{\beta}_j^i + \hat{\gamma}^i \omega_{Aj}^i S_j \tag{26}$$

$$\hat{\mu}_A^{i+1} = \hat{\mu}_A^i + \hat{\gamma}^i u_A^i \tag{27}$$

Because there is the standardization of data, then the value  $\hat{\beta}_{LASSO}$  will be returned to the actual data with the following equation:

$$\hat{\beta}_{LASSO(i+1)j} = \frac{\hat{\beta}_{LASSOj}^{i+1}}{Scale_j}, \tag{28}$$

$$Scale_j = \sqrt{\sum_{i=1}^n \sum_{j=1}^p (X_{ij} - \bar{X}_j)^2} \tag{29}$$

If  $\hat{\beta}_{LASSO}^i = \hat{\beta}_{MKT}^*$  then stop at this stage, but if  $\hat{\beta}_{LASSO}^i \neq \hat{\beta}_{MKT}^*$  then repeat from the initial stages of LARS with an update  $i$  become  $i = i + 1$ . Repetition occurs as much as the maximum amount of data.

### 3. RESULTS AND DISCUSSION

#### 3.1. Result

Multicollinearity in the initial data using VIF in Table 1.

Table 1. VIF value of data

Model	Collinearity Statistics	
	Tolerance	VIF
$X_1$	0.028	35.798
$X_2$	0.048	20.733
$X_3$	0.267	3.746
$X_4$	0.014	72.375
$X_5$	0.519	1.928
$X_6$	0.048	20.949
$X_7$	0.005	195.009

Based on Table 1, it can be concluded that there is multicollinearity in the data because the VIF is more than 10. Then the data is standardized based on equations (11) and (12) and find the value using the ordinary least squares method in table 2.

Table 2. Coefficient of OLS

Model	$\hat{\beta}_{OLS}$
$X_0$	854.608.641.238
$X_1$	30.428
$X_2$	-19.667
$X_3$	14.088
$X_4$	-4.343
$X_5$	-9.330
$X_6$	-3.496
$X_7$	19.768

In standardization data, VIF value is sought to find out whether or not there is multicollinearity in the independent variables which can be seen in Table 3.

Table 3. VIF value on standardized data

Model	Collinearity Statistics	
	Tolerance	VIF
$X_1$	0.067	14.973
$X_2$	0.123	8.121
$X_3$	0.445	2.246
$X_4$	0.143	6.996
$X_5$	0.724	1.381
$X_6$	0.182	5.480
$X_7$	0.057	17.660

There is multicollinearity because there is still a VIF value of more than 10 in the independent variable so that the data can be applied to the LASSO method.

### 3.2. Analysis

Calculation of the LASSO method in money supply data uses R Studio software. Table 4, shows the value of the LASSO coefficient estimator using Rstudio as follows:

Table 4. Koefisien Parameter LASSO

<i>i</i>	$\widehat{\beta}_{LASSO}$						
	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$	$\widehat{\beta}_7$
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0.4 4
2	0.75	0	0	0	0	0	0.7 6
3	0.75	0	0.04	0	0	0	0.7 8
4	0.82	0	0.25	0	0	0.34	0.7 1
5	0.81	0	0.29	0	-1.49	0.38	0.7 1
6	0.72	0	0.83	1.6	-17.31	0.64	0.5 8
7	0.63	-1.45	0.83	0.61	-8.69	0.13	0.9 5

Iteration stops when the value  $i = 7$  because  $\widehat{\beta}_{LASSO} = \widehat{\beta}_{OLS}$ . Next is to choose a model from LASSO using a cross-validation equation (7) which can be seen in Figure 1.

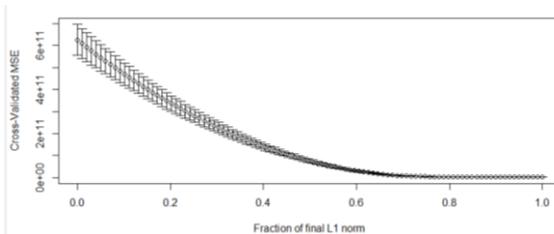


Figure 1: Cross-Validation

The smallest MSE value of cross-validation is 1.354.902.248 at the lowest point. The best model according to cross-validation is the iteration model  $i = 7$ .

#### **4. CONCLUSION**

Based on data from the Indonesian money supply from 2013 to 2018, the best model of LASSO regression is obtained using cross-validation as follows:

$$\begin{aligned} Y = & -0.6259048X_1 - 1.451317X_2 + 0.83394348X_3 \\ & + 0.6116285X_4 - 8.69414X_5 + 0.1332662X_6 \\ & + 0.9487676X_7 \end{aligned}$$

If net foreign assets consisting of bills and non-residents liabilities in Indonesia have increased by one billion, then the money supply in circulation in Indonesia will decrease by 0.6259048 and 1.451317. Whereas the net assets in the country, which consist of Net Claims to the Central Government, Other Financial Institutions, Non-Financial SOEs and Private Sector, if it increases one billion, then increases respectively by 0.83394348, 0.6116285, 0.1332662 and 0.9487676 on the money supply circulate. The money supply will decrease 8.694140 when bills to local governments increase by one billion.

## REFERENCES

- AHMAD, I., & AHMAD, S. 2018. “Multiple Skills and Medium Enterprises’ Performance in Punjab Pakistan: A Pilot Study”. **Journal of Social Sciences Research**, Vol. 7: 44-49.
- AHMAD, I., & AHMAD, S. 2019. “The Mediation Effect of Strategic Planning on The Relationship between Business Skills and Firm’s Performance: Evidence from Medium Enterprises in Punjab, Pakistan”. **Opcion**, Vol. 35, N° 24: 746-778.
- ALOISIO, A., ALAGGIO, R., & FRAGIACOMO, M. 2019. “Dynamic Identification of a Masonry Façade from Feismic Response Data Based on an Elementary Ordinary Least Squares approach”. **Engineering Structures**. Vol. 197, N° 2: 109-415. Netherlands.
- AMOOZAD-KHALILI, M., ROSTAMIAN, R., & ESMAEILPOUR-TROUJENI, M. 2019. **Economic Modeling of Mechanized and Semi-Mechanized Rainfed Wheat Production Systems using Multiple Linear Regression Model**. Information Processing in Agriculture. China.
- ARSLAN, O. 2012. “Weighted LAD-LASSO method for robust parameter estimation and variable selection in regression”. **Computational Statistics and Data Analysis**. Vol. 56, N° 6: 1952–1965. Netherlands.
- CHUNG, T. & ARIFF, M. 2016. “A test of the linkage among money supply, liquidity and share prices in Asia”. **Japan and The World Economy**. Vol. 39: 48–61. Netherlands.
- HE, Y. 2017. “A Study on the Relationship between Money Supply and Macroeconomic Variables in China”. **Mediterranean Journal of Social Sciences**. Vol. 8, N° 6: 99-107. Italy.
- IZENMAN, A. 2008. **Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning**. New York: Springer. USA.
- JIANG, P. & CHEN, J. 2016. “Neurocomputing Displacement Prediction of Landslide Based on Generalized Regression Neural Networks with K-Fold Cross-Validation”. **Neurocomputing**. Vol. 74, N° 16: 2502–2510. Netherlands.

- MAHABOOB, B. & BHUMIREDDY, V. 2018. “A Treatise on Ordinary Least Square Estimation of Parameters of Linear Model”. **International Journal of Engineering & Technology**. Vol. 7, N<sup>o</sup> 4: 518-522. India.
- MALYSHEVA, S. 2019. **The rehabilitation of idleness: The production of new values and meanings for leisure in the late 19th and early 20th centuries.**
- ROOZBEH, M., BABAIE-KAFAKI, S., & SADIGH, A. 2017. “A heuristic approach to combat multicollinearity in the least trimmed squares regression analysis”. **Applied Mathematical Modelling**. Vol. 57: 105-120. Netherlands.
- SHIRVANI, H. 2013. “The Relative Importance of the Determinants of the US Money Supply”. **Research in Business and Economics Journal**. Vol. 8, N<sup>o</sup> 1: 1–16. USA.
- VU, D., MUTTAQI, K., & AGALGAONKAR, A. 2015. “A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables”. **Applied Energy**. Vol. 140: 385–394. Netherlands.
- ZHENG, S. & LIU, W. 2011. “An experimental comparison of gene selection by Lasso and Dantzig selector for cancer classification”. **Computers in Biology and Medicine**. Vol. 41, N<sup>o</sup> 11: 1033–1040. Netherlands.



**UNIVERSIDAD  
DEL ZULIA**

---

**opción**

Revista de Ciencias Humanas y Sociales

Año 36, N° 91 (2020)

Esta revista fue editada en formato digital por el personal de la Oficina de Publicaciones Científicas de la Facultad Experimental de Ciencias, Universidad del Zulia.  
Maracaibo - Venezuela

[www.luz.edu.ve](http://www.luz.edu.ve)

[www.serbi.luz.edu.ve](http://www.serbi.luz.edu.ve)

[produccioncientifica.luz.edu.ve](http://produccioncientifica.luz.edu.ve)