IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Prediction of CIF Components Proportion of Indonesian Import Value Using Multivariate Fractional Logit Model

[1] Mardiah; [2]Asep Saefuddin; [3] Indahwati

[1] Statistics Department, Bogor Agricultural University,
Bogor, 16680, Indonesia

[2] Statistics Department, Bogor Agricultural University,
Bogor, 16680,Indonesia

[3] Statistics Department, Bogor Agricultural University,
Bogor, 16680, Indonesia

Abstract - International Merchandise Trade Statistics (IMTS) recommends to use a free on board (FOB) valuation for exports and cost, insurance, and freight (CIF) valuation for imports. CIF is a sum of FOB, freight, and insurance value of imported goods. IMTS suggests countries that record import value on CIF to have an additional method to decompose CIF into FOB, freight, and insurance value. FOB, insurance, and freight fraction follow multivariate fractional model. The model is to give prediction value of three CIF components fraction. Based on MAPE and RMSEP value, mode of transport, transit status, and group of two digit HS code are three covariates that the best precision of the predicted value.

Keywords - *CIF value decomposition, Modeling Fractional Outcomes, Quasi Maximum Likelihood*

1. Introduction

International Merchandise Trade Statistics (IMTS) is an international guidelines for exports and imports documentation in trade between countries. It contains concept, definition, classification, and regulation on the trade of goods between countries. In IMTS 2010 second revision, countries almost universally apply a Free on Board (FOB-type) values for the valuation of exports and most countries use a Cost, Insurance, and Freight (CIF-type) values for the valuation of imports [3].

FOB-type values include the transaction of the goods and services performed to deliver goods to the border of exporting country. CIF-type values include FOB value and the value of services performed to deliver goods from the border of exporting country to the border of importing country. Indonesia and many other countries compile their import value in CIF-type because customs administrations usually determine the custom value of the goods in this basis.

There are two specific recommendations for countries about statistical value for imported goods mentioned in IMTS (concept and definition) section 4.8 and 4.9:

1. The statistic of imported goods be a CIF-type value; however countries are encouraged to compile FOB-type value of imported goods as supplementary information.
2. Countries that compile only CIF-type value for imported goods are encouraged to compile separate data for freight and insurance, at the most detailed commodity and partner level possible.

The research focuses on obtain decomposition of CIF imports value. FOB value of imports is important in order to compare the imports value of importing country and exports value of exporting country. Insurance and freight value separated could determine the services value share of the imports activities.

In order to decompose CIF into FOB, insurance, and freight value, that could be done easily if its component were completely available in the custom declarations of imports. Missing value of FOB, insurance, and freight are the main problem to satisfy IMTS recommendations. Custom declarations that contain missing value is about 90 percent every month. Based on data of Indonesia's import BC 2.0 softcopy document period of January 2016, 75,900 documents (86.62 percent) had missing values.

Statistics Indonesia (BPS) as the institution that officially released the Indonesia's import statistic has an approach to convert CIF into FOB value, CIF ratio estimation. That approach divided CIF become three fractions, 94.79 percent FOB, 4.74 percent freight, and 0.47 percent insurance. The three fractions were applied to all import transaction regardless of supplier country, mode of transportation, and other characteristics of the goods imported.

CIF components value in the custom declarations can be vary greatly, from one to millions USD so that the data has a very wide range and large deviation. Modeling CIF components value might also make the error very large, while a good model should have small error. Modeling fraction of FOB, insurance, and freight is one of the alternative solutions to obtain a better model. The three fractions is bounded to the range (0,1). Several methods to modeling fractional outcomes have been studied such as Tobit Model, Nonlinear Least Squares, Fractional Logit Model, Beta Model, and Simplex Model.

The fractional logit model has a flexibility that the dependent variables should not have any distributional assumption [3]. Its estimation method use quasi maximum likelihood that doesn't assume any distribution. This study aims to predict the fraction of FOB value using multivariate fractional regression approaches, that based on quasi parametric method (Fractional Logit Model). The three fractions of CIF component become dependent variables that variates on (0,1) range. From five covariates used, there would be a procedure of variable selection. Any covariates combinations that gives higher accuracy would be chosen as the best model to predict fractions of CIF component.

2. MATERIAL AND METHOD

2.1 Data

The raw data of the research was obtained from customs declaration of import BC 2.0 softcopy documents (Customs Office, Ministry of Finance), period of January-December 2016. The data was validated and filtered, to get full data containing value of FOB, insurance, freight, and CIF.

The data filtered contains 45,403 observations. That was splitted into two data sets, training and testing. 40,862 observation as training data for constructing the model and 4,541 observation as testing data for model evaluation.
There are five covariates (imported good characteristic) and three response variable in this research.

Table 1 Dependent and Independent Variables

Variables	Name	Type
Dependent	Y_1 = FOB fraction	Numeric
	Y_2 = Insurance fraction	Numeric
	Y_3 = Freight fraction	Numeric
Indepen-dent	X_1 = mode of trensport	Categoric
	X_2 = transit status	Categoric
	X_3 = region of partner country	Categoric
	X_4 = group of two digit HS code*	Categoric
	X_5 = group of netto	Categoric

*HS = Harmonized System (Commodity classification by World Classification Organizations)

Based on the data used in this research, there are two mode of transportations, sea and air transportation. Transit status is differentiated into "yes" (transit) and "no" (without transit). Commodity of imported goods and region of supplier country divided into eight classifications based on its freight proportion. Net weight of imported goods grouped into three groups, 1-30 kgs, 31-70 kgs, and more than 71 kgs.
For the independent categorical variables, the dummy variables was made.

2.2 Fractional Logit Model

One of regression method for fractional data is fractional logit model (FLM). FLM is a quasi-likelihood method that doesn't requires any distributional assumption, but only requires the conditional mean to be correctly specified for consistent parameter estimates [5]. The drawbacks of linear models for data in (0,1) range can be analogous to the drawbacks to the linear probability model for binary data [6]. The most common alternative solution to solve that problem is to model the log-odds (logit) of the response variable.

$$E\left(\log(\frac{y}{1-y}) \mid \mathbf{x}\right) = \mathbf{x\beta} \tag{1}$$

$$E(y \mid \mathbf{x}) = G(\mathbf{x\beta}) = \frac{\exp(\mathbf{x\beta})}{1+\exp(\mathbf{x\beta})} \tag{2}$$

Fractional logit model has an identical likelihood function:

$$l(\mathbf{b}) = G(\mathbf{xb})^{y}(1-G(\mathbf{xb}))^{1-y} \tag{3}$$

Parameters can be estimated by maximizing the log liklehood function:

$$\log(l(\mathbf{b})) = y \times \log(G(\mathbf{xb})) + (1-y) \times \log(1 - G(\mathbf{xb})) \quad (4)$$

Maximization of log-likelihood can be achieved using standard log-likelihood maximization techniques, to obtain the **b** parameter estimation.

2.3. Multivariate Fractional Logit Model

Multivariate fractional data is an extension of fractional data in which multiple response variables satisfy a sum constraint [2]. Multivariate fractional logit model is a method that corresponds to the data, with a logit link function. Its quasi likelihood function is a joint function of each its response variable:

$$L = \prod_{k=1}^{K} l(\mathbf{b}_k) = \prod_{k=1}^{K} G(\mathbf{xb}_k)^{y_k}(1 - G(\mathbf{xb}_k))^{1-y_k} \quad (5)$$

For K response variables.

Its quasi log likelihood function is following:

$$\log(L) = \sum_{k=1}^{K} y_k \times \log[G(\mathbf{xb}_k)] + (1-y_k) \times \log[1 - G(\mathbf{xb}_k)] \quad (6)$$

2.4. Model Evaluation

From several models obtained, one best model would be chosen. All models would be applied to testing data to measure the error. Error measurement statistics required to determine the predictive accuracy. Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error of Prediction (RMSEP) are some of the predictive accuracy measurements.

$$MAPE = \left(\frac{1}{n} \sum_{i-1}^{n} \frac{|actual - predicted|}{|actual|} \right) \times 100\% \quad (10)$$

$$RMSEP = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (actual - predicted)^2} \quad (11)$$

where n is the number of observation in testing data.

As described in the previous chapter, five covariates used to construct the model. In choosing the best model, there is a procedure of variable selection. From several models with any covariates combinations obtained, their MAPE and RMSEP value would be compared each other. Model chosen is the model with lowest MAPE and RMSEP value.

3. Result and Discussion

3.1 Data Explorations

The value of three CIF components (FOB, freight, and insurance) has a very wide range and deviation. Its value might be from 1 to thousands even million USD. Modeling value of CIF component might be misleading. FOB, freight, and insurance fraction of CIF are the best solution to obtain predictive model without any transformation.
Comparison of two data sets are following:

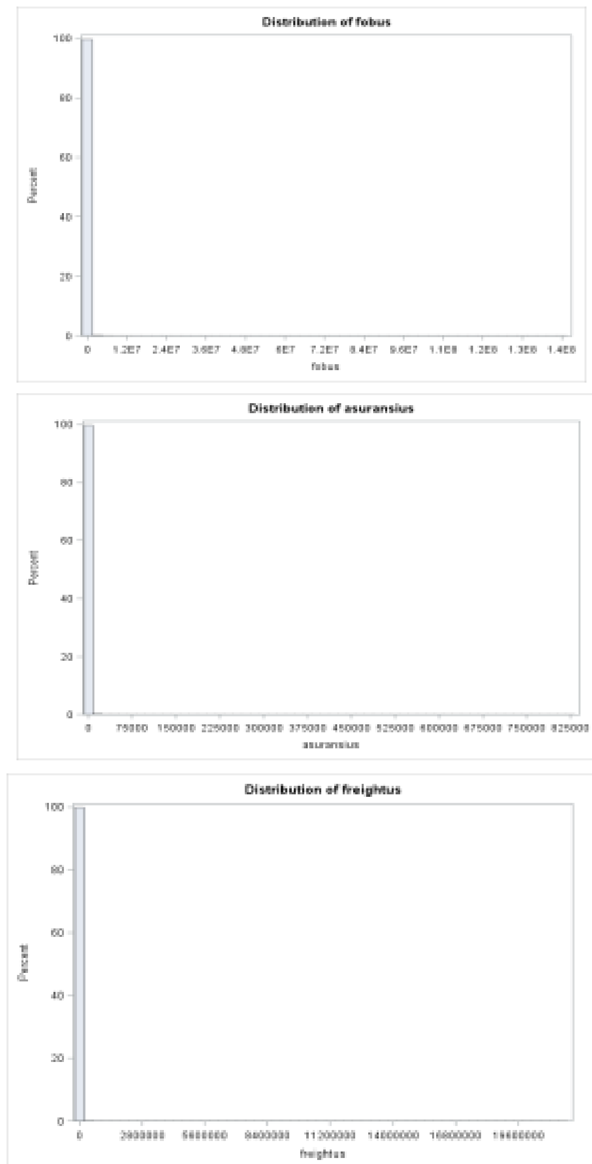

Fig. 1 Histograms of value of CIF component

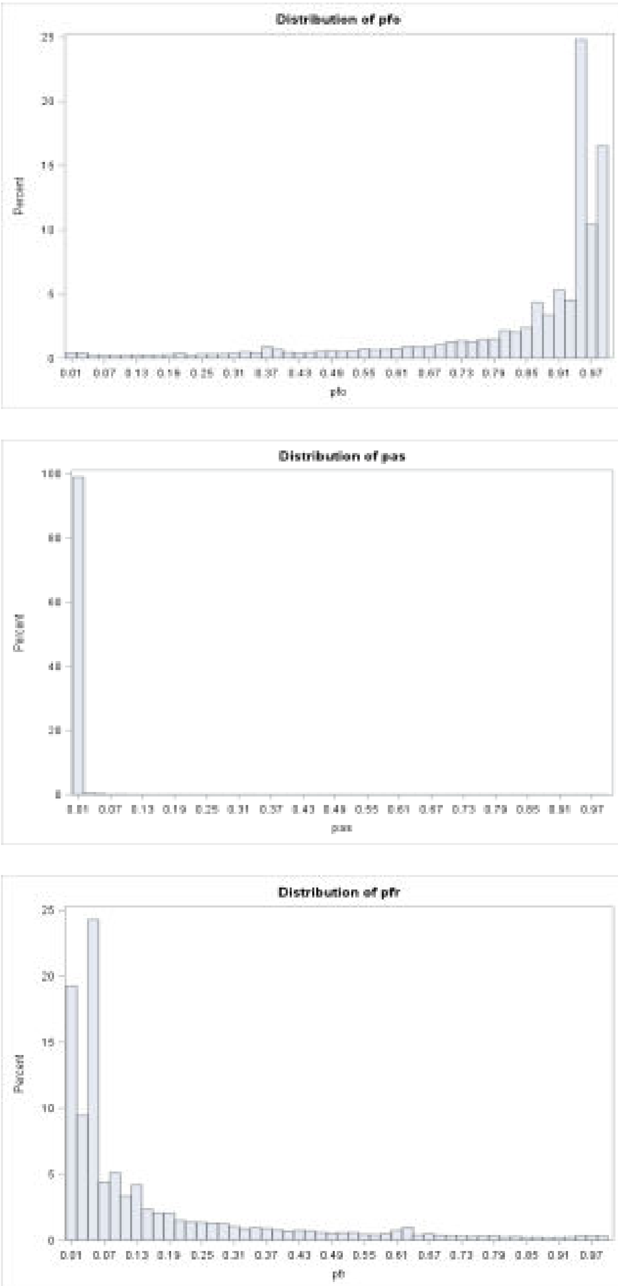

Fig.2 Histograms of Fraction of CIF component

Table 2 Statistics of Value of CIF Components (USD)

Statistic	FOB	Insurance	Freight
Minimum	0.01	0.01	0.01
Q1	616.0	3.9	67.3
Median	3,248.3	17.5	296.8
Q3	16,764.9	67.7	929.2
Maximum	144,696,236	832,003	21,704,435
Standard ddeviation	762,629.7	5,674.3	102,738.3

Table 3 Statistics of Fraction of CIF Components

Statistic	Fraction of CIF components (USD)		
	FOB	Insurance	Freight
Minimum	0.0001	0.0001	0.0001
Q1	0.8085	0.0047	0.0311
Median	0.9454	0.0050	0.0498
Q3	0.9647	0.0050	0.1852
Maximum	0.9997	0.9914	0.9956
Standard ddeviation	0.2129	0.0169	0.2119

3.2 Descriptive Analysis

From the data used in this research, customs declaration of import BC 2.0 documents period of January-December 2017 validated and filtered, total CIF value is 1,993.04 million USD. Its components, FOB, freight, and insurance value are each 1,888.05; 95.88; and 9.11 million USD, so that its fractions each are 83.89; 15.57; and 0.54 percent on average.

Table 4 Fractions of CIF component on Average by Imported Goods Characteristics (percent)

Charac-teristic	Categories	FOB	Insurance	Freight
Mode of transport	Air	77.84	0.57	21.59
	Transit	74.77	0.53	24.70
	No Transit	78.19	0.57	21.23
	Sea	93.89	0.49	5.62
	Transit	93.14	0.48	6.37
	No Transit	93.94	0.49	5.57
Region of partner country	Region 1	96.38	0.17	3.45
	Region 2	93.68	0.58	5.73
	Region 3	91.94	0.53	7.53
	Region 4	83.98	0.51	15.51
	Region 5	76.99	0.59	22.42
	Region 6	55.35	0.63	44.02
	Region 7	23.28	0.43	76.29
	Region 8	6.57	0.52	92.91
Group of	Group 1	95.96	0.42	3.62

IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

2 digit HS code	Group 2	91.42	0.39	8.18
	Group 3	87.39	0.54	12.07
	Group 4	81.62	0.54	17.84
	Group 5	74.46	0.61	24.93
	Group 6	65.91	0.85	33.23
	Group 7	51.23	0.57	48.21
	Group 8	25.20	0.20	74.60
Group of netto	1-30 kg	79.63	0.59	19.77
	31-70 kg	80.94	0.63	18.43
	> 71 kg	86.88	0.49	12.63

If BPS approach were applied in the data, the fractions of CIF component each are 94.77; 4.74; and 0.47 percent on average. Deviation of freight fraction is quiet large, while deviation of insurance fraction is quite small.

The largest difference of the fractions of CIF components is the freight fractions between air and sea transportation, 15.97 percent. Freight fraction of imported goods delivered by air transportation on average is quite high (21.59%), it has large difference with BPS approach (4.74%).

Insurance fraction is relatively the same among the characteristics of imported goods. Its range 0.4 to 0.6 percent, except three characteristics category: region 1, group 6, and group 8 of 2 digit HS code.

3.3 Multivariate Fractional Logit Model

The computation processes in this research using SAS/STAT 9.4 application with NLMIXED Procedure. Based on the best model chosen, three covariates are used in the MFLM: mode of transportations, transit status, and group of 2 digit HS code. The linear predictor are follows:

$$\eta_{ki} = b_{0ki} + (b_{mki} \times d_{mki}) + (b_{tki} \times d_{tki}) + (b_{h1ki} \times d_{h1ki}) + (b_{h2ki} \times d_{h2ki}) + (b_{h3ki} \times d_{h3ki}) + (b_{h4ki} \times d_{h4ki}) + (b_{h5ki} \times d_{h5ki}) + (b_{h6ki} \times d_{h16i}) + (b_{h7ki} \times d_{h7ki})$$

where:
$k = k^{th}$ response variable
$i = i^{th}$ observation in testing data
d_m = dummy variable for mode of transportation, 0 for sea and 1 for air
d_t = dummy variable for transit status, 0 for transit and 1 for no transit
d_{h1}-d_{h7} = dummy variable for group of 2 digit HS code

The **b** coefficients in the linear predictor is as shown in table 5 below:

Table 5 Parameter Estimation **b** in the MFLM

Parameter Estimation	FOB fraction	Insurance fraction	Freight fraction
b_0	-1.2843	-6.2503	1.2813
b_m	1.4159**	-0.1269	-1.4722**
b_t	0.1964**	0.0590	-0.2040**
b_{h1}	3.7608	0.7824	-3.8539*
b_{h2}	2.6188	0.7403	-2.6369
b_{h3}	2.7022	1.0176	-2.7357
b_{h4}	2.1156	1.052	-2.1322
b_{h5}	1.8199	1.1364	-1.8347
b_{h6}	1.5351	1.4566	-1.56
b_{h7}	1.0123	1.0483	-1.0216

*Significant at the alpha 10%
**Significant at the alpha 5%

Coefficient interpretations:
b_m is a dummy variable for mode of transportation, where its value was 0 for sea transport and 1 for air transport. For FOB fraction model, b_m = 1.4159 meant the linear predictor for observation with mode of transportation = "Air" was 1.4159 higher than observation with mode of transportation = "Sea" (other covariates was assumed fixed). Related to the predicted value , prediction of FOB fraction value for observation with mode of transportation = "Air" was 0.3047 higher than observation with mode of transportation = "Sea" (other covariates was assumed fixed), in the level of significance alpha 0.05.
MAPE and RMSEP value of the model are given in the table 6.

Table 6. MAPE and RMSEP value

Model Evaluation	FOB fraction	Insurance fraction	Freight fraction
MAPE (%)	64.15	75.24	1,142.9
RMSEP	0.1755	0.0162	0.1745

The MAPE and RMSEP value above is the smallest value obtained, from many combinations of covariates. This result showed that multivariate fractional logit model for three fractions of CIF component which had the highest accuracy in predicted value is model that consists of mode of transportation, transit status, and group of two digit HS code as covariates.

4. Conclusions and Suggestions

Based on the model there are three factors (covariates) which provides the lowest MAPE and RMSEP. These factors are mode of transportation, transit status, and group of two digit HS code.

For next studies, there are some suggestions as follow:
1. Use additional factors such as port of importation and shipping company (domestic or foreign).

2. Apply other method (parametric method) such as Beta model.

Appendix

Group of 2 digit HS code and regions of partner country are two covariates that used to construct the model. Categories of those two covariates are as follow:

Group of 2 digit HS code

Group	Code of 2 digit HS*
Group 1	02, 05, 11, 14, 16, 18, 30, 50, 66, 71, 79, 92, 93, 98
Group 2	04, 19, 21, 22, 23, 42, 47, 52, 63, 72, 75, 78, 86, 91
Group 3	09, 17, 20, 25, 27, 28, 31, 42, 43, 48, 54, 57, 58, 59, 61, 62, 64, 68, 69, 73, 74, 76, 80, 82, 83, 84, 85, 88, 90, 96
Group 4	13, 15, 38, 39, 40, 53, 56, 65, 70, 81, 89, 94, 95, 97
Group 5	01, 03, 29, 32, 33, 34, 35, 36, 44, 45, 55, 60, 87
Group 6	24, 26, 37
Group 7	06, 07, 08, 10, 12, 49, 67
Group 8	51

*2 digit HS code called ***Chapter*** in World Classification Organization (WCO) commodity classifications (Harmonized system)

Regions of partner countries

Regions	Countries
Region 1	Myanmar, Ethiopia, Mauritius, Fiji, East Timor, Luxembourg, Iceland, Liechtenstein
Region 2	Brunei Darussalam, Morocco, Algeria, Equatorial Guinea, Nigeria, Cote D'ivoire, Zambia, Malawi, Panama, Uruguay, Puerto Rico, Malta
Region 3	Singapore, Malaysia, Lao People's Dem. Rep., Tunisia, Cameroon, Seychelles, Uganda, Swaziland, Bahamas, Dominican Rep., Bulgaria, Serbia
Region 4	Hong Kong, Taiwan, China, Vietnam, India, Bangladesh, Sri Lanka, Iran, Saudi Arabia, Egypt, Libyan, South Africa, Canada, Guatemala, Colombia, Virgin Island (British), United Kingdom, Netherlands, France, Germany, Belgium, Switzerland, Denmark, Norway, Sweden, Finland, Ireland, Italy, Hungary, Latvia, Slovenia
Region 5	Japan, Rep. of Korea, Macau, Thailand, Philippines, Cambodia, Pakistan, Israel, Turkey, United Arab Emirates, Tanzania, Australia, New Zealand, United States, Mexico, Honduras, Chile, Brazil, Austria, Spain, Portugal, Greece, Poland, Romania, Slovakia, Czech Rep., Russia Fed.
Region 6	Kuwait, Congo, Micronesia, Argentina, Ecuador, Peru, Ukraine, Estonia, Croatia
Region 7	Mongolia, Sudan, Tokelau
Region 8	Papua New Guinea, Jordan

References

[1] [DJBC] Direktorat Jenderal Bea dan Cukai Kementerian Keuangan Republik Indonesia, Buku Tarif Kepabeanan Indonesia, Jakarta (ID): Direktorat Jenderal Bea dan Cukai Kementerian Keuangan Republik Indonesia, 2012.

[2] Horas RM, Basnou C, and Pino J "A multivariate fractional regression approach to modeling land use and cover dynamics in a Mediterranean landscpae", Computers, Environment, and Urban Systems, Vol. 54, 2015, pp. 47-55.

[3] Kieschnick R, and McCullough BD, "Regression Analysis of Variates Observed on (0,1): Percentage, Proportions, and Fractions" [SMIJ] Statistical Modelling: An International Journal, vol. 3, 2003, pp. 193-213.

[4] Liu W, and Xin J, "Modelling Fractional Outcomes with SAS", paper 1304, 2014.

[5] Papke LE, and Woolridge JM, "Econometric Methods For Fractional Response Variables WithAn Application To 401 (K) Plan Participation Rates", [JAE] Journal of Applied Econometrics, Vol.11, 1996, pp. 619-632.

[6] Swearingen CJ, Castro MSM, amd Bursac Z, "Modelling Percentage Outcomes: The Beta Regression Macro", paper 335, 2011.

[7] [UNSD] United Nations Statistics Division (USA), *Manual on Statistics of Internaional Trade of Services 2010*, New York (USA): United Nations Publications, 2011.

Mardiah currently pursuing masters degree program in Applied Statistics in Bogor Agricultural University, Indonesia. Attained bachelor degree from STIS Jakarta in 2008.

Asep Saefuddin is lecturer at Department of Statistics , Bogor Agricul-tural University, Indonesia. His main interest is in Geoinformatics, Spatial Analysis, Statistical Modeling.

Indahwati is lecturer at Department of Statistics , Bogor Agricul-tural University, Indonesia. Her main interest is in Statistical Modelling, Sampling Design and Methodology.