

PENGELOMPOKAN KOMODITAS EKSPOR PADA PROVINSI JAWA BARAT BERDASARKAN NEGARA TUJUAN DENGAN MODEL *FINITE MIXTURE WEIBULL*

Nurvia Anggraini, Irwan Susanto, dan Hasih Pratiwi

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

nurvianggraini@student.uns.ac.id

ABSTRAK

Model mixture telah banyak digunakan dalam beberapa analisis perekonomian. Perkembangan komoditas ekspor berdasarkan negara tujuan pada Provinsi Jawa Barat tahun 2018 didominasi oleh sektor non migas. Peran nilai ekspor tertinggi terdapat pada Negara Amerika Serikat sebesar 16,57%. Selanjutnya diikuti oleh Negara Jepang, Thailand, Filipina, dan Tiongkok. Data komoditas ekspor dapat dimodelkan dengan menggunakan finite mixture karena adanya pola multimodal data. Pola multimodal diindikasikan dengan adanya beberapa puncak pada histogram data serta uji signifikansi pola unimodal yang tidak dipenuhi. Hal ini menunjukkan adanya kecenderungan sifat heterogenitas data. Model mixture telah diimplementasikan pada data berdistribusi heavy tailed yang salah satunya adalah distribusi Weibull. Metode estimasi parameter model mixture dilakukan dengan metode estimasi maksimum likelihood menggunakan algoritma expectation maximization (EM), karena fungsi ekspektasi maksimum likelihood memiliki bentuk yang tidak closed form. Konsep utama algoritma EM adalah mengasosiasikan model data lengkap ke struktur yang tidak lengkap pada data observasi dengan tujuan untuk memudahkan komputasi pada estimasi maksimum likelihood. Banyaknya cluster model finite mixture yang terbentuk dipilih berdasarkan ukuran Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC) terkecil. Berdasarkan hasil analisis, terbentuk dua cluster pada data komoditas ekspor pada Provinsi Jawa Barat berdasarkan negara tujuan pada tahun 2018. Pada cluster pertama terdapat 176 negara tujuan dengan proporsi cluster sebesar 0,7881, sedangkan pada cluster kedua terdapat 36 negara tujuan dengan proporsi cluster sebesar 0,2119.

Kata kunci: Finite Mixture, Komoditas Ekspor, Algoritma EM, Cluster.

ABSTRACT

A mixture model has been widely used in several economic analyses. The development of export commodities by destination country in West Java Province in 2018 was dominated by the non-oil sector. The role of the highest export value was found in the United States of America at 16.57%. Then followed by Japan, Thailand, Philippines, and China. Export commodity data can be modeled using finite mixture because of the multimodal data patterns. A multimodal pattern is indicated by the presence of several peaks in the data histogram and the significance test of the unimodal pattern which is not met. This shows the tendency of heterogeneity of data. The mixture model has been implemented on heavy tailed distribution data, one of which is the Weibull distribution. The method of estimating the mixture model parameters is performed by the maximum likelihood estimation method using the expectation maximization (EM) algorithm, because the maximum likelihood expectation function has a form that is not closed form. The main concept of the EM algorithm is to associate the complete data model to the incomplete structure of the observational data with the aim of making it easier to compute the maximum likelihood estimate. The number of finite mixture model clusters formed is selected based on the smallest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) sizes. Based on the analysis results, two clusters were formed in the export commodity data in West Java Province based on destination countries in 2018. In the first cluster there were 176 destination countries with cluster proportions of 0.7881, while in the second cluster there were 36 destination countries with cluster proportions of 0.2119.

Key words: Finite Mixture, Export Commodity, EM Algorithm, Cluster

PENDAHULUAN

Perdagangan internasional adalah salah satu upaya penting suatu negara untuk mencapai pembangunan nasional. Perdagangan internasional meliputi perdagangan barang dan jasa. Berkembangnya arus globalisasi yang semakin pesat menuntut kemajuan teknologi dan transportasi dalam melakukan perpindahan barang atau jasa oleh setiap negara di dunia menjadi lebih cepat, efektif, dan efisien (Sabaruddin, 2015).

Berdasarkan data dari Badan Pusat Statistik, komoditas ekspor pada Provinsi Jawa Barat didominasi oleh sektor non migas khususnya industri pengolahan yaitu sebesar 98,89%. Sedangkan sektor migas hanya menyumbang sekitar 0,80%. Adapun perkembangan negara tujuan ekspor akan disajikan dalam Tabel 1. Pada Tabel 1, dapat diamati bahwa Amerika Serikat tetap menjadi pangsa pasar utama meskipun mengalami penurunan nilai sebesar 1,18%. Selanjutnya disusul oleh Jepang yang mengalami kenaikan sebesar 8,13%, Thailand naik sebesar 7,64%, Tiongkok naik sebesar 3,23%, dan Filipina naik sebesar 1,06%.

Tabel 1. Negara Tujuan Utama Ekspor pada Provinsi Jawa Barat

Negara Tujuan	Nilai (US\$ Juta)					Pertumbuhan (%) YoY	Peran Terhadap Total (%)
	2014	2015	2016	2017	2018		
Amerika Serikat	4.924,5	4.957,1	5.020,7	5.090,4	5.030,4	-1,18	16,57
Jepang	3.343,1	2.840,5	2.718,1	2.877,4	3.111,4	8,13	10,25
Thailand	1.686,6	1.631,3	1.726,5	2.161,9	2.327,0	7,64	7,66
Filipina	980,8	933,7	1.217,6	2.061,8	2.083,6	1,06	6,86
Tiongkok	1.251,2	1.424,7	1.490,7	1.688,7	1.743,2	3,23	5,74
Lainnya	15.160,0	13.906,5	13.553,1	15.325,0	16.067,3	4,84	52,92
Total	27.346,1	25.693,7	25.726,7	29.205,2	30.362,9	3,96	100,00

Sumber: Badan Pusat Statistik, 2019

Dalam mengelompokkan distribusi data komoditas ekspor dengan pendekatan numerik, terdapat banyak alternatif distribusi probabilitas yang digunakan untuk mengestimasi model distribusi data komoditas ekspor. Estimasi model tersebut digunakan sebagai bagian dari proses analisis terhadap pengelompokan data komoditas ekspor berdasarkan negara tujuan pada Provinsi Jawa Barat. Analisis clustering dalam menentukan pengelompokan negara tujuan komoditas ekspor pada Provinsi Jawa Barat berperan penting untuk menetapkan kebijakan pemerintah. Analisis clustering yang memperhatikan model statistik disebut model based clustering (MBC).

Distribusi finite mixture telah banyak digunakan dalam penelitian model based clustering. Distribusi finite mixture relatif fleksibel dalam memodelkan distribusi pendapatan yang mempunyai sub-sub populasi berbeda, dimana sub-sub populasi tersebut dapat merefleksikan kelompok dengan homogenitas secara ekonomi (Pittau, 2016). Model finite mixture mampu merepresentasikan kelompok sub-sub populasi tersebut sebagai suatu komponen mixture. Misal x_1, x_2, \dots, x_n adalah sampel random berukuran n dimana x_i adalah vektor random berdimensi- p dengan fungsi densitas probabilitas $f(x_i)$ Suatu vektor variabel random $x =$

$[x_1, x_2, \dots, x_n]^T$ yang bertipe diskrit atau kontinu dikatakan berasal dari distribusi *finite mixture* jika memiliki fungsi kepadatan probabilitas $f(x_i)$ didefinisikan dengan (McLachlan and Peel, 1946):

$$f(x_i) = w_1 f_1(x_i) + \dots + w_k f_k(x_i) \quad (1)$$

dengan:

$f_k(x_i)$ = fungsi kepadatan probabilitas *mixture* untuk semua $k = 1, 2, \dots, K$;

w_k = proporsi *mixture* (*weight*);

$w = [w_1, w_2, \dots, w_n]^T$ = vektor parameter *weight* dari distribusi *finite mixture*.

Nilai-nilai dalam w harus memenuhi:

$$0 \leq w_k \leq 1$$

dan

$$\sum_{k=1}^K w_k = 1$$

Data komoditas ekspor pada Provinsi Jawa Barat memiliki karakteristik data kontinu, positif, dan *heavy-tailed*. Salah satu distribusi probabilitas yang berkarakteristik *heavy-tailed* adalah distribusi Weibull. Suatu variabel random X dikatakan berdistribusi Weibull dengan parameter $\beta > 0$ dan $\theta > 0$ jika memiliki fungsi kepadatan probabilitas:

$$f(x; \theta, \beta) = \frac{\beta}{\theta^\beta} x^{\beta-1} e^{-\left(\frac{x}{\theta}\right)^\beta} \quad (2)$$

dan bernilai 0 untuk yang lain. Notasi yang menunjukkan bahwa X memiliki bentuk densitas probabilitas pada (2) adalah $X \sim WEI(\theta, \beta)$ (Bain, 1992).

Cabang utama inferensi statistik adalah estimasi. Terdapat beberapa metode dalam mengestimasi parameter. Metode yang digunakan untuk menaksir parameter apabila distribusi populasi diketahui adalah metode estimasi maksimum likelihood (MLE). Misal terdapat variabel independen $x_i, i = 1, \dots, n$ dan $\Psi = [w, \theta]^T$ maka fungsi likelihood dari model *finite mixture* (1) didefinisikan dengan

$$L(\Psi) = \prod_{i=1}^n \left[\sum_{k=1}^K w_k f_k(x_i; \theta_k) \right]$$

Beberapa penelitian telah menggunakan model *finite mixture* Weibull seperti analisis data reliabilitas menggunakan data waktu dengan estimasi parameter melalui algoritma EM (Elmahdy, 2017), model *mixture* Weibull untuk analisis laju kesembuhan bagi 100 pasien kanker payudara di Rumah Sakit Dr. Sardjito Yogyakarta (Dwidayati, 2013), serta model *finite mixture* Weibull untuk merepresentasikan kelompok negara kaya dan negara miskin yang diestimasi melalui metode estimasi maksimum likelihood (Paap dan Van Dijk, 1998). Algoritma *expectation maximization* (EM) adalah metode berulang untuk menemukan kemungkinan maksimum dalam model statistik untuk kasus dimana persamaan tidak dapat diselesaikan secara langsung. Pada algoritma EM memiliki dua tahap yaitu tahap *expectation* (E) dan tahap *maximization* (M). Misal diberikan nilai awal $\Psi^{(0)}$ algoritma EM untuk estimasi model *finite mixture* sebagai berikut:

1) Tahap *Expectation* (E)

Fungsi ekspektasi bersyarat x untuk data lengkap saat iterasi ke- s yaitu:

$$Q(\Psi; \Psi^{(s)}) = E(\ln L(\Psi) | x_i, \Psi^{(s)})$$

$$= \sum_{i=1}^n \sum_{k=1}^K z_{ik}^{(s)} \ln [w_k^{(s)} f_k(x_i | \theta_k^{(s)})]$$

$$= \sum_{i=1}^n \sum_{k=1}^K z_{ik}^{(s)} \{ \ln(w_k^{(s)}) + \ln(f_k(x_i | \theta_k^{(s)})) \}$$

Penduga $\hat{z}_{ik}^{(s)}$ diproses melalui

$$\hat{z}_{ik}^{(s)} = \frac{w_k^{(s)} f_k(x_i | \theta_k^{(s)})}{\sum_{h=1}^K w_h^{(s)} f_h(x_i | \theta_h^{(s)})}$$

dengan $i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, K$. Penduga $\hat{z}_{ik}^{(s)}$ dari tahap E selanjutnya digunakan untuk tahap M pada saat iterasi ke- $(s+1)$.

2) Tahap Maximization (M)

Pada tahap *maximization*, parameter estimasi maksimum likelihood dicari dengan cara memaksimalkan ekspektasi likelihood yang dihasilkan dari tahap *expectation* (E). Jika z_{ik} diketahui, maka MLE data lengkap pada w_k adalah

$$\hat{w}_k = \frac{\sum_{i=1}^n z_{ik}}{n}, \quad k = 1, \dots, K$$

Pada iterasi ke- $(s+1)$, parameter bobot w_k diduga dengan

$$\hat{w}_k^{(s+1)} = \frac{\sum_{i=1}^n \hat{z}_{ik}^{(s)}}{n}$$

Tahap E dan M dilakukan secara iteratif. Prosedur akan berhenti saat

$$L(\hat{\Psi}^{(s+1)}) \geq L(\hat{\Psi}^{(s)})$$

Untuk $s = 0, 1, 2, \dots$ dimana $\hat{\Psi} = [\hat{w}, \hat{\theta}]^T$ (Dempster *et al.*, 1977).

METODE PELAKSANAAN

Data yang digunakan dalam penelitian ini adalah data komoditas ekspor berdasarkan negara tujuan yang diekspor dari Provinsi Jawa Barat ke-212 negara tujuan tahun 2018. Data diperoleh dari jurnal publikasi Badan Pusat Statistik tahun 2019. Data dikumpulkan berdasarkan dokumen ekspor atau Pemberitahuan Eskpor Barang (PEB) melalui eksportir dan telah diberikan ijin oleh Kantor Pelayanan Bea dan Cukai, selanjutnya dikirim ke Badan Pusat Statistik (BPS) dan Bank Indonesia (BI). Nilai barang yang dicatat dalam statistik ekspor adalah *Free On Board* (FOB), yaitu nilai barang sampai di pelabuhan muat setelah barang dimuat ke kapal dalam satuan US\$ (Badan Pusat Statistik, 2019). Komoditas ekspor yang digunakan adalah semua barang yang keluar dari wilayah Indonesia baik yang bersifat komersial maupun non komersial seperti bantuan barang kredit lunak, hadiah, dan sebagainya. *Software* yang digunakan dalam penelitian ini adalah *software R* versi 3.5.2 dengan *package* yang digunakan adalah *mixR*. Penelitian yang dilakukan penulis adalah melalui pemodelan *finite mixture* dengan tahapan sebagai berikut:

- (1) Identifikasi pola distribusi pada data komoditas ekspor pada Provinsi Jawa Barat berdasarkan negara tujuan apakah memenuhi pola univariat *multimodal*. Jika data memenuhi *multimodal*, maka terdapat indikasi adanya model *finite mixture*. Lalu, dilakukan pemilihan distribusi data berdasarkan pola data *multimodal*.

(2) Estimasi model *finite mixture* Weibull dengan metode estimasi maksimum likelihood melalui algoritma *expectation-maximization*. Estimasi model dilakukan sampai iterasi ke- s dimana akan diperoleh nilai estimator yang konvergen.

(3) Seleksi banyaknya komponen *mixture* dalam model *finite mixture* berdasarkan ukuran AIC terkecil dan BIC terkecil (Celeux, *et al.*, 2018). Nilai AIC didefinisikan dengan

$$AIC = -2 \ln L(\Psi) + 2p \quad (3)$$

sedangkan nilai BIC didefinisikan dengan

$$BIC = 2 \ln L(\Psi) + p \ln(n) \quad (4)$$

dengan:

$L(\Psi)$ = fungsi likelihood dari penduga maksimum likelihood $\hat{\Psi}$

p = banyaknya parameter dalam model *finite mixture*

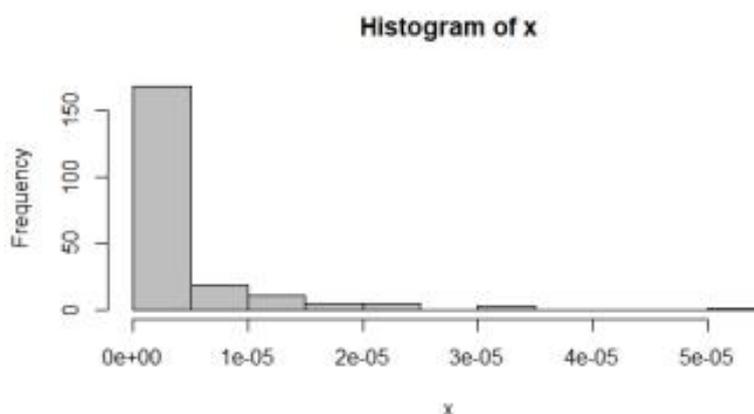
n = banyaknya data observasi

(4) Menentukan hasil *clustering* dari data serta merepresentasikan hasil *clustering* tersebut.

HASIL DAN PEMBAHASAN

A. Identifikasi Pola Data Komoditas Ekspor

Diberikan observasi x_i sebagai nilai komoditas ekspor pada Provinsi Jawa Barat berdasarkan negara tujuan pada tahun 2018. Berikutnya dilakukan identifikasi pola distribusi data apakah memenuhi adanya pola *multimodal*. Apabila memenuhi pola *multimodal*, maka dapat dilakukan analisis melalui pemodelan *finite mixture* karena terjadi pengelompokan data yang konvergen secara lokal.



Gambar 1. Histogram Data Komoditas Eskpor pada Provinsi Jawa Barat tahun 2018

Berdasarkan Gambar 1, terlihat bahwa grafik data menunjukkan adanya beberapa puncak distribusi. Hal ini menunjukkan bahwa data memenuhi pola *multimodal*. Selain itu, pola data diatas bersifat *heavy-tailed* yaitu distribusi dengan kemencengan yang ekstrim, sehingga distribusi probabilitas yang tepat untuk menggambarkan komponen *mixture* adalah distribusi probabilitas yang mempunyai karakteristik *heavy-tailed* yang salah satunya adalah distribusi Weibull. Selanjutnya, dilakukan uji signifikansi *individual distribution* untuk memastikan apakah pola *multimodal* dipenuhi. Pengujian distribusi dipilih berdasarkan distribusi yang memiliki karakteristik *heavy tailed*, positif, dan kontinu, dalam hal ini yaitu distribusi Weibull, Gamma, dan Lognormal. Selain itu, turut dilakukan pengujian pada Distribusi Normal.

Tabel 2. Uji Signifikansi *Individual Distribution*

Distribusi	P-value	Keputusan Uji
Normal	<0,005	H_0 ditolak
Weibull	<0,010	H_0 ditolak
Gamma	<0,005	H_0 ditolak
Lognormal	0,379	H_0 gagal ditolak

Dalam melakukan uji signifikansi, hipotesis *null* menyatakan bahwa data memenuhi pola *unimodal* yang berarti data mengikuti distribusi probabilitas tunggal. Kemudian hipotesis alternatif menyatakan bahwa data tidak memenuhi pola *unimodal* yang berarti data mengikuti distribusi *finite mixture*. Berdasarkan Tabel 2, dapat diamati bahwa distribusi Normal, Weibull, dan Gamma *unimodal* tidak mampu merepresentasikan data komoditas ekspor pada Provinsi Jawa Barat. Hal ini terjadi karena nilai *p-value* pada masing-masing distribusi kurang dari tingkat signifikansi (α) sebesar 0,05. Sedangkan distribusi Lognormal memenuhi pola *unimodal* karena memiliki nilai *p-value* lebih dari tingkat signifikansi (α) sebesar 0,05. Berikutnya, dilakukan proses estimasi model *finite mixture* sebagai analisis dari data komoditas ekspor pada Provinsi Jawa Barat dengan metode estimasi maksimum likelihood menggunakan algoritma *expectation maximization* (EM).

B. Seleksi Model

Dalam penelitian ini, kriteria untuk memilih komponen model *finite mixture* adalah ukuran *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC) dalam perhitungan (3) dan (4). Nilai AIC dan BIC dipilih yang terkecil. Banyaknya komponen yang dihitung maksimal hingga 2 komponen *mixture* karena pada komponen 3 dan komponen yang lebih besar dari 3, perhitungan ukuran AIC dan BIC tidak konvergen.

Tabel 3. Seleksi Model berdasarkan AIC dan BIC

Distribusi	Banyak Komponen	AIC	BIC	Banyak Iterasi EM
Gamma	2	7591,7295816	7608,512513	105
Weibull	2	7588,0316026	7604,814534	116
Lognormal	2	7593,4119471	7610,1948785	71
Normal	2	8066,0411353	8082,8240667	27

Berdasarkan Tabel 3, nilai AIC dan BIC terkecil terdapat pada model *finite mixture* Weibull dengan dua komponen yaitu sebesar 7588,0316026 untuk ukuran AIC dan 7604,814534 untuk ukuran BIC. Hasil komputasi estimasi model *mixture* Weibull diberikan oleh

$$f(x_i) = w_1 WEI(\theta_1, \beta_1) + w_2 WEI(\theta_2, \beta_2)$$

Fungsi $WEI(\theta, \beta)$ ditunjukkan oleh (2). Nilai parameter *weight* $\hat{w}_1 = 0,7881$ dan $\hat{w}_2 = 0,2118$. Sedangkan parameter bentuk $\hat{\beta}_1 = 0,4838$ pada komponen pertama dan $\hat{\beta}_2 = 0,6892$ pada komponen kedua; dan parameter skala $\hat{\theta}_1 = 5,2677 \times 10^{-6}$ pada komponen pertama dan $\hat{\theta}_2 = 4,9286 \times 10^{-8}$ pada komponen kedua. Parameter *weight* pada masing-masing komponen menginterpretasikan besarnya nilai besaran ekspor dalam populasi.

Pengelompokan data menggunakan model *finite mixture* dapat diperoleh berdasarkan parameter pada distribusi Weibull. Berdasarkan hasil analisis menggunakan *software* R, diperoleh dua *cluster* negara tujuan eksportir dengan algoritma EM konvergen sampai iterasi ke-116. Pada *cluster* pertama terdapat 176 negara tujuan dengan proporsi komponen sebesar 0,7881. Pada *cluster* kedua terdapat 36 negara tujuan dengan proporsi komponen

sebesar 0,2119. *Cluster 1* menginterpretasikan negara-negara tujuan dengan nilai eksportir rendah yang memiliki nilai *Free on Board* (FOB) maksimal sebesar US\$ 79 juta dimana sebagian besar terdapat pada wilayah Afrika, sebagian wilayah Eropa Utara, dan Amerika Selatan. *Cluster 2* menginterpretasikan negara-negara tujuan dengan nilai eksportir tinggi yang memiliki nilai *Free on Board* (FOB) lebih dari US\$ 79 juta dimana sebagian besar terdapat pada wilayah Asia Tenggara, Amerika Utara, dan Eropa Barat.

SIMPULAN

Data komoditas ekspor pada Provinsi Jawa Barat memiliki pola multimodal sehingga dapat dianalisis dengan model finite mixture Weibull dua komponen. Dua komponen tersebut menggambarkan dua cluster negara tujuan eksportir. Cluster pertama terdapat 176 negara tujuan nilai eksportir rendah dimana nilai *Free on Board* (FOB) maksimal sebesar US\$ 79 juta dan sebagian besar terdapat pada wilayah Afrika, sebagian wilayah Eropa Utara, dan Amerika Selatan. Cluster kedua terdapat 36 negara tujuan nilai eksportir tinggi dimana nilai *Free on Board* (FOB) lebih dari US\$ 79 juta dan sebagian besar terdapat pada wilayah Asia Tenggara, Amerika Utara, dan Eropa Barat. Pada penelitian berikutnya, penulis merekomendasikan agar analisis mixture dapat dilakukan menggunakan distribusi probabilitas lain yang lebih cocok untuk data ekonomi.

REFERENSI

- Badan Pusat Statistik. (2019). *Ekspor Menurut Provinsi Asal Barang*. Diakses dari <https://bps.go.id/publication>.
- Bain, L.J. and Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics* 2 ed. California: Duxbury Press.
- Celex, G., Fruhwirth-Schnatter, S., and Roberts, C.P. (2018). Model Selection for Mixture Model – Perspective and Strategies in, *Handbook of Mixture Analysis*. NW: CRC Press.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: series B (Methodological)*. 39(1), 1-22.
- Dwidayati, N., Kartiko S.H., dan Subanar. (2013). Estimation of the Parameters of a Mixture Weibull Model for Analyze Cure Rate. *Applied Mathematical Sciences*. 7(116), 5767-5778.
- Elmahdy, E.E. (2017). Modelling Reliability Data with Finite Weibull or Lognormal Mixture Distributions. *Applied Mathematics and Information Sciences*. 11(4), 1081-1089.
- McLachlan, G. and Peel, D. (1996). *Finite Mixture Models*. United States of America: A Wiley-Interscience Publication.
- Paap, R. dan Van Dijk, H.K. (1998). Distribution and Mobility of Wealth of Nations. *European Economic Review*. 42(7), 1269 – 1293.
- Pittau, M.G., Zelli, R. and Massari, R. (2016). Evidence of Convergence Clubs Using Mixture Models. *Econometric Reviews*. 35(7), 1317 – 1342.
- Sabaruddin, S.S. (2015). *Dampak Perdagangan Internasional Indonesia Terhadap Kesejahteraan Masyarakat: Aplikasi Structural Path Analysis*. Buletin Ekonomi Moneter dan Perbankan. 17(4).

